The Impact of the Ocean Observing System on Estimates of the California Current Circulation Spanning Three Decades

Andrew M. Moore[1*], Michael G. Jacox[2,3], William J. Crawford[1], Bruce Laughlin[1], Christopher A. Edwards[1] and Jérôme Fiechter[1]

1 Department of Ocean Sciences, 1156 High Street

University of California, Santa Cruz CA 95062.

2 Institute or Marine Sciences, 1156 High Street

University of California, Santa Cruz CA 95062.

3 Environmental Research Division, Southwest Fisheries Science Center

NOAA, Monterey, California.

* Corresponding author: ammoore@ucsc.edu

April 18, 2017

Abstract

Data assimilation is now used routinely in oceanography on both regional and global scales for computing ocean circulation estimates and for making ocean forecasts. Regional ocean observing systems are also expanding rapidly, and observations from a wide array of different platforms and sensor types are now available. Evaluation of the impact of the observing system on ocean circulation estimates (and forecasts) is therefore of considerable interest to the oceanographic community. In this paper, we quantify the impact of different observing platforms on estimates of the California Current System (CCS) spanning a three decade period (1980-2010). Specifically, we focus attention on several dynamically related aspects of the circulation (coastal upwelling, the transport of the California Current and the California Undercurrent, thermocline depth and eddy kinetic energy) which in many ways describe defining characteristics of the CCS. The circulation estimates were computed using a 4-dimensional variational (4D-Var) data assimilation system, and our analyses also focus on the impact of the different elements of the control vector (*i.e.* the initial conditions, surface forcing, and open boundary conditions) on the circulation. While the influence of each component of the control vector varies between different metrics of the circulation, the impact of each observing system across metrics is very robust. In addition, the mean amplitude of the circulation increments (*i.e.* the difference between the analysis and background) remains relatively stable throughout the three decade period despite the addition of new observing platforms whose impact is redistributed according to the relative uncertainty of observations from each platform. We also consider the impact of each observing platform on CCS circulation variability associated with low-frequency climate variability. The low-frequency nature of the dominant climate modes in this region allows us to track through

time the impact of each observation on the circulation, and illustrates how observations from some platforms can influence the circulation upto a decade into the future.

Keywords: California Current; Observation impacts; Adjoints; 4D-Var.

1 Introduction

The California Current System (CCS) along the west coast of North America forms the equatorward branch of the North Pacific subtropical gyre (see Fig. 1). It comprises a dynamicaly rich and highly variable circulation which has been the subject of many previous studies (see Hickey (1998) and Checkley and Barth (2009) for some excellent in-depth reviews). A dominant feature of the CCS circulation is the presence of a pronounced seasonal cycle in coastal upwelling. During the spring and summer, the winds between Washington and Baja California are equatorward and upwelling favorable. This results in cold, nutrient rich waters at the ocean surface which in turn enhance ocean primary productivity. The associated offshore Ekman transport also sets up an offshore pressure gradient that drives an equatorward coastal jet. During the fall and winter, the North Pacific high pressure system gives way to the Aleutian low and the winds north of $\sim$ 40N become poleward which promotes downwelling along the coast of Northern California, Oregon and Washington. A poleward flow at depth is also often present, the so-called California Undercurrent (CUC), located over the continental shelf between 100 m and 300 m, with a velocity $\sim 0.1 - 0.3$ m s^{-1} (Hickey, 1998). The CUC is relatively poorly observed, although it has been observed along the entire west coast of the U.S. (Pierce et al., 2000). While a poleward current is to be expected on the grounds of mass conservation in the presence of coastal upwelling (e.g. Gill, 1982), the dynamics of the CUC are not well understood, although a recent study by Connolly et al. (2014) suggests that the CUC may be associated with an alongshore pressure gradient. At some times of the year, a surface poleward return flow is also observed along the coast, the so-called Davidson Current, which some have attributed to a surfacing of the CUC, although

there is no census of opinion on this (Hickey, 1979).

The circulation is dominated by the first baroclinic mode, with the result that as the sea surface goes down in response to offshore transport during the upwelling seasons, the pycnocline shoals making nutrient rich sub-thermocline waters more accessible. Poleward of Cape Mendocino, the CC and coastal jet form fairly coherent circulation features (Fig. 1). At Cape Mendocino, inertia carries the CC farther offshore where it becomes baroclinically unstable contributing to a field of energetic mesoscale and sub-mesoscale eddies, leading to a region of elevated eddy kinetic energy offshore (Kelly et al., 1998).

The CCS is also influenced by several known modes of climate variability that include the El Niño Southern Oscillation (ENSO), the North Pacific Gyre Oscillation (NPGO), and the Pacific Decadal Oscillation (PDO). Through a combination of changes in the local atmospheric circulation and remotely generated coastally trapped waves, ENSO exerts a significant influence on the physical and biogeochemical conditions in the CCS (Jacox et al., 2015; Frischknecht et al., 2015; Jacox et al, 2016). For example, thermocline depth, upwelling intensity, and the depth of the upwelling source waters can be dramatically different during El Niño years, leading to warmer than normal ocean temperatures and depleted nutrients along much of the California coast. Similarly, changes in the large scale atmospheric circulation over the NE Pacific associated with the PDO and NPGO have been linked to low frequency variability in the CCS (e.g. Di Lorenzo et al., 2008; 2009; Johnson and Mantua, 2015).

Recently, Neveu et al. (2016; hereafter N16) have computed a sequence of historical circulation estimates of the CCS spanning the period 1980-2010. Using the Regional Ocean Modeling System (ROMS) and a state-of-the-art 4-dimensional variational (4D-Var) data assimilation system, these analyses combine model circulation estimates with all available

quality controlled ocean observations in the region to yield analyses of the ocean circulation environment that are more reliable than either the model or the observations considered in isolation. The focus of the present study is to explore the extent to which the different observing platforms constrain different aspects of the circulation that characterize the CCS (e.g. upwelling, alongshore transport, the CUC, eddy kinetic energy, etc), and changes in the circulation associated with the dominant low frequency modes of climate variability identified above.

Observation impact studies are now routine at many numerical weather centers (e.g. Langland and Baker, 2004; Errico, 2007; Zhu and Gelaro, 2008; Gelaro and Zhu, 2009; Lupu et al., 2011, 2012; Jung et al., 2013; Tyndall and Horel, 2013; Lorenc and Marriott, 2014). There have been some efforts in oceanography also to quantify the impact of the observing system on ocean analyses using Observing System Experiments (OSEs; e.g. Balmaseda at al., 2007; Oke and Schiller, 2007; Smith and Haines, 2009), spectral analysis of the representer matrix (Le Hénaff et al., 2009), quantification of the degrees of freedom of the observing system (Moore et al., 2011a), assessment of observation footprints (Oke and Sakov, 2012), and ensemble methods (Storto et al., 2013). A more extensive review of these efforts can be found in Oke et al. (2014a,b). In the present study we use an adjoint-based method developed by Langland and Baker (2004), that is commonly used by the meteorological community to quantify the impact of individual observations on different aspect of an analysis-forecast system.

This study is unique in that it quantifies the impact of an ocean observing system during a period spanning three decades starting from an initial period served only by in situ hydrographic data through to the present day where multiple satellite observing systems and

in situ assets are in place. Section 2 outlines the ocean model, data assimilation system, and circulation analyses that form the foundation of this study. The methodology used to quantify the observation impacts is described in section 3, along with several metrics that quantify important aspects of the CCS circulation. The impact of the control vector components and observations on each metric are presented in sections 4 and 5 respectively. These calculations are extended in section 6 to include climate variability. A summary and conclusions are presented in section 7.

2 The ROMS CCS 4D-Var Circulation Analyses

N16 describe a sequence of circulation analyses for the CCS that were computed using ROMS 4D-Var, spanning the 31 year period 1980-2010 (hereafter referred to as WCRA31). A full description of the ROMS 4D-Var configuration and observation data sets used can be found in N16, so only a brief description of the salient features that are important for the observation impact calculations will be presented here. Readers should consult N16 for more details.

The ROMS model domain and bathymetry used in this study are shown in Fig. 1 and span the region 30°N-48°N, 134°W-115.5°W. The horizontal resolution is 1/10°, and in the vertical 42 terrain-following σ-levels were used that vary in thickness between 0.3 m and 8 m over the continental shelf and between 7 m and 100 m in the deep ocean. Veneziani et al. (2009) have demonstrated that this configuration of ROMS captures very well many aspects of the observed CCS circulation.

In keeping with the usual notation (Moore et al., 2011b) we will denote by $\mathbf{x}$ the ROMS

state-vector of grid-point prognostic variables that comprises potential temperature T, salinity S, horizontal velocities (u, v), and sea surface displacement (ζ), so that $\mathbf{x} = (T, S, \zeta, u, v)^T$ and superscript T denotes the vector transpose. The non-linear ROMS will be denoted as $M(t_i, t_{i-1})$, so that:

$$\mathbf{x}(t_i) = M(t_i, t_{i-1})(\mathbf{x}(t_{i-1}), \mathbf{f}(t_{i-1}, t_i), \mathbf{b}(t_{i-1}, t_i)) \tag{1}$$

where $\mathbf{x}(t_i)$ represents propagation forward in time of the state-vector by ROMS subject to surface forcing, $\mathbf{f}(t_{i-1}, t_i)$, and lateral open boundary conditions, $\mathbf{b}(t_{i-1}, t_i)$, during the time interval $[t_{i-1}, t_i]$. In the ROMS 4D-Var data assimilation system, the control vector comprises the model initial conditions, surface forcing, and open boundary conditions, and the background vectors for each are denoted $\mathbf{x}^b$, $\mathbf{f}^b$ and $\mathbf{b}^b$ respectively.

During WCRA31, the wind stress, heat flux and freshwater flux components of $\mathbf{f}^b$ were derived using a combination of daily averaged atmospheric products. These include ECMWF Reanalyses: ERA40 (2.5° resolution; Källberg et al., 2004), ERA Interim (0.7° resolution; Dee et al., 2011), and the Cross-Calibrated, Multi-Platform ocean wind product of Atlas et al. (2011; 25 km resolution). The ocean surface fluxes were computed using the bulk formulations of Liu et al. (1979) and Fairall et al. (1996a,b).

The prior time evolving open boundary conditions $\mathbf{b}^b$ were taken from the global Simple Ocean Data Assimilation product (version SODA POP 2.2.4) of Carton and Giese (2008). The Chapman (1985) boundary conditions were applied to sea surface height, while the vertically integrated flow is subject to the Flather (1976) condition. To damp spurious boundary waves, a 100 km wide sponge layer was also used adjacent to the open boundaries, in which

viscosity increased linearly from 4 m^2s^{-1} in the interior to 100 m^2s^{-1} at the boundary.

A variety of in situ and satellite observations were assimilated into ROMS. A summary of the different observation platforms and the nominal instrument error standard deviations assigned to each platform are shown in Table 1 of N16. All of the satellite platforms measure sea surface temperature (SST) at a spatial resolution that is higher than that of the model, so all observations were combined into super observations where appropriate before being assimilated into the model using the standard formula for a linear, unbiased, minimum variance estimate (Daley, 1991). A time series of the total number of observations available from each observing platform during each data assimilation cycle is shown in Fig. 2, and illustrates how the ocean observing system grew rapidly during the 2000-2010 decade. Particularly noteworthy is the relatively stable number of in situ measurements due to an increase in the number of hydrographic data from the Argo drifting float program (*i.e.* note the reduction in the variance of the number of in situ observations in Fig. 2 during the 2001-2010 decade).

The observations of SST from each of the satellite instruments were assimilated from individual swaths. However, in the case of satellite-derived sea surface height (SSH), the decision was made to assimilate the 1 day gridded composites of the mean dynamic topography from Aviso DUACS version DT-2010 (Dibarboure et al. 2011). The gridded SSH data were used because the current version of the ROMS 4D-Var system does not allow for prior errors or observation errors that are correlated in time. As such, information from individual along track observations can be quickly lost due to geostrophic adjustment and can become ineffective for constraining the model unless it is persisted through time. Assimilating the gridded products alleviates this issue by allowing the large-scale gyre circulation and eddy

158 field to geostrophically adjust to the entire SSH field. This is not an ideal solution, however,
159 because of the limitations of the objective mapping technique used to map the altimeter
160 observations onto a regular grid. Also, since satellite SSH observations near the coast are
161 known to be unreliable (Saraceno et al., 2008), only observations that are more than 50
162 km from the coast were assimilated. In addition, errors in SSH and SST observations are
163 assumed to be spatially uncorrelated in time as discussed in N16.

164 The 1980-2010 time interval spanned by WCRA31 was divided into a 8 day windows with
165 each window over-lapping neighboring windows by 4 days. All of the observations available
166 during each 8 day window were then assimilated into the model using the dual formulation
167 of the ROMS 4D-Var system (Gürol et al., 2014). The 4D-Var control vector $\mathbf{z}$ comprises all
168 of the parameters that will be adjusted, which in general will include the initial conditions,
169 surface forcing and open boundary conditions, and corrections for model error in the case of
170 weak constraint 4D-Var. The aim of 4D-Var is to find the control vector, $\mathbf{z}$, that minimizes
171 the cost function:

$$J_{NL} = (\mathbf{z} - \mathbf{z}^b)^T \mathbf{D}^{-1} (\mathbf{z} - \mathbf{z}^b) + (\mathbf{y}^o - H(\mathbf{z}))^T \mathbf{R}^{-1} (\mathbf{y}^o - H(\mathbf{z})) \tag{2}$$

172 where $\mathbf{y}^o$ is the vector of observations; $\mathbf{D}$ is the background error covariance matrix; $\mathbf{R}$ is
173 the observation error covariance matrix; and H is the observation operator that maps $\mathbf{z}$ to
174 the observation points. In the case of 4D-Var, H also includes the nonlinear ROMS M from
175 (1). The cost function J_{NL} was minimized indirectly using the incremental formulation of
176 Courtier et al. (1994) which takes advantage of a truncated Gauss-Newton method (Lawless
177 et al. 2005). In this case, the problem is linearized about $\mathbf{z}_k = \mathbf{z}^b + \sum_{l=1}^{k-1} \delta \mathbf{z}_l$, where k

refers to the k^{th} outer-loop and $\delta\mathbf{z}_l$ are the increments from all previous outer-loops. The increments $\delta\mathbf{z}_l$ are the solutions of a sequence of linear minimizations of a quadratic cost function J:

$$J_k = \delta\mathbf{z}_k^T\mathbf{D}^{-1}\delta\mathbf{z}_k + (\mathbf{d}_{k-1} - \mathbf{G}_{k-1}\delta\mathbf{z}_k)^T\mathbf{R}^{-1}(\mathbf{d}_{k-1} - \mathbf{G}_{k-1}\delta\mathbf{z}_k). \tag{3}$$

In (3), $\mathbf{d}_{k-1} = \mathbf{y}^o - H(\mathbf{z}_{k-1})$ is the so-called innovation vector; and $\mathbf{G}_{k-1}$ is the generalized observation operator which is the tangent linearization of H linearized about $\mathbf{z}_{k-1}$. Generally, minimization of (3) proceeds iteratively in 4D-Var via a sequence of so-called inner-loops, after which the state $\mathbf{z}_k$ about which $\mathbf{G}_k$ is linearized is updated using the newly identified increment (the outer-loop), and a new sequence of inner-loops is performed. During WCRA31, a single outer-loop and 15 inner-loops were employed so the subscripts l and k will be dropped in sequel.

Following Weaver and Courtier (2001), the background error covariance matrix $\mathbf{D}$ was modeled as the solution of a diffusion equation. Complete details of the parameters used in the background error covariance model can be found in N16. Furthermore, it was assumed that the observation error covariance matrix $\mathbf{R}$ is diagonal (*i.e.* the observation errors are assumed to be uncorrelated in time and space).

3 The Observation Impact Methodology

The method adopted here for quantifying the impact of the observations on the 4D-Var circulation estimates is that described by Langland and Baker (2004; hereafter LB04). As described in section 2, the dual form of the ROMS 4D-Var algorithm with one outer-loop was

employed in computing the circulation estimates, in which case the optimal control vector $\mathbf{z}^a$ can be expressed as:

$$\mathbf{z}^a = \mathbf{z}^b + \mathbf{D}\mathbf{G}^T(\mathbf{G}\mathbf{D}\mathbf{G}^T + \mathbf{R})^{-1}\mathbf{d} = \mathbf{z}^b + \mathbf{K}\mathbf{d} \tag{4}$$

where $\mathbf{K} = \mathbf{D}\mathbf{G}^T(\mathbf{G}\mathbf{D}\mathbf{G}^T + \mathbf{R})^{-1}$ is the Kalman gain matrix. If each 8 day assimilation window is numbered using the index j, and the starting time of each window in days is represented as t_j, then the interval spanned by each assimilation window is $[t_j, t_j + 8]$. The resulting circulation estimate $\mathbf{x}^a(t)$ (*i.e.* the analysis) during assimlation window j is then given by $\mathbf{x}^a(t) = M(t, t_j)(\mathbf{x}^a(t_j), \mathbf{f}^a(t_j, t), \mathbf{b}^a(t_j, t)) = M(t, t_j)(\mathbf{z}^a(t_j))$.

Following LB04, consider a scalar function $\mathcal{I}(\mathbf{x})$ of the state-vector $\mathbf{x}$. LB04 considered the impact of each observation on $\mathcal{I}(\mathbf{x})$ computed from forecasts of $\mathbf{x}$ initialized from a 4D-Var analysis using a numerical weather prediction model. Here, we adopt a variant of the same approach and consider the impact of each observation on $\mathcal{I}(\mathbf{x})$ during each 4D-Var analysis cycle following Trémolet (2008). The forecast problem was considered by Moore et el. (2011b). With this in mind, consider the change in $\mathcal{I}(\mathbf{x})$ that results at some time t during an analysis cycle from assimilating observations into the model, namely $\Delta\mathcal{I} = \mathcal{I}(\mathbf{x}^a(t)) - \mathcal{I}(\mathbf{x}^b(t))$. Following LB04, the function increment $\Delta\mathcal{I}$ can be expressed as:

$$
\begin{aligned}
\Delta\mathcal{I} &= \mathcal{I}(M(t,t_j)(\mathbf{z}^a(t_j))) - \mathcal{I}(M(t,t_j)(\mathbf{z}^b(t_j))) \\
&\simeq \mathcal{I}(M(t,t_j)(\mathbf{z}^b(t_j)) + \mathbf{M_b}(t)\delta\mathbf{z}) - \mathcal{I}(M(t,t_j)(\mathbf{z}^b(t_j))) \\
&\simeq (\partial\mathcal{I}/\partial\mathbf{z}|_{\mathbf{z}^b})^T\mathbf{M_b}(t)\delta\mathbf{z} \\
&= (\partial\mathcal{I}/\partial\mathbf{z}|_{\mathbf{z}^b})^T\mathbf{M_b}(t)\mathbf{K}\mathbf{d} = \mathbf{d}^T\mathbf{K}^T\mathbf{M_b}^T(t)(\partial\mathcal{I}/\partial\mathbf{z}|_{\mathbf{z}^b})
\end{aligned} \tag{5}
$$

where $\delta\mathbf{z} = \mathbf{z}^a(t_j) - \mathbf{z}^b(t_j)$, and a first-order Taylor expansion has been used to linearize $M(t,t_j)(\mathbf{z}^a(t_j))$ and $\mathcal{I}(\mathbf{x}^a(t))$. In (5), $\mathbf{M_b}(t)$ is the tangent linearization of ROMS linearized about the background circulation $\mathbf{x}^b(t) = M(t,t_j)(\mathbf{z}^b(t_j))$.

The last equality in (5) shows that the change $\Delta\mathcal{I}$ in $\mathcal{I}$ due to assimilating the observations can be computed as the dot-product of the innovation vector $\mathbf{d}$ and the vector $\mathbf{g} = \mathbf{K}^T\mathbf{M_b}^T(t)(\partial\mathcal{I}/\partial\mathbf{z}|_{\mathbf{z}^b})$. The vector $\mathbf{g}$ represents an integration of the derivative of $\mathcal{I}$ by the adjoint model $\mathbf{M_b}^T(t)$ all multiplied by the transpose of the Kalman gain matrix. Since each element of the vectors $\mathbf{d}$ and $\mathbf{g}$ is uniquely associated with an individual observation, the contribution (or "impact") of each observation on $\Delta\mathcal{I}$ can be quantified.

As discussed in N16, the cost function (3) of ROMS dual 4D-Var is minimized using the Lanczos formulation of the Restricted $\mathbf{D}$-Preconditioned Conjugate Gradient algorithm (RPCG) of Gratton and Tshimanga (2009) as described by Gürol et al. (2014). Specifically, the Kalman gain matrix is decomposed as $\mathbf{K} = \mathbf{D}\mathbf{G}^T\mathbf{V}_m\mathbf{T}_m^{-1}\mathbf{V}_m^T\mathbf{G}\mathbf{D}\mathbf{G}^T$, where $\mathbf{V}_m$ is the matrix of Lanczos vectors resulting from m inner-loops, and $\mathbf{T}_m$ is the tridiagonal matrix of associated Lanczos vector coefficients. The Lanczos vectors form an orthonormal basis where $\mathbf{V}_m^T\mathbf{G}\mathbf{D}\mathbf{G}^T\mathbf{V}_m = \mathbf{I}_m$. Using the Lanczos formulation, the action of the transpose Kalman

gain $\mathbf{K}^T = \mathbf{G}\mathbf{D}\mathbf{G}^T\mathbf{V}_m\mathbf{T}_m^{-1}\mathbf{V}_m^T\mathbf{G}\mathbf{D}$ on any vector can be conveniently computed for any data assimilation cycle. In practice, $\mathbf{V}_m$ and $\mathbf{G}\mathbf{D}\mathbf{G}^T\mathbf{V}_m$ are routinely archived during 4D-Var so evaluation of $\mathbf{K}^T$ requires only one additional integration of the tangent linear model, $\mathbf{G}$, sampled at the observations points. In principle, $\mathbf{K}^T$ could be computed for any variational data assimilation system if the Lanczos vectors or conjugate gradient descent directions are archived.

In all of the observation impact calculations described in section 4, all but one of the scalar functions employed take the form $\mathcal{I} = (1/N)\sum_{n=1}^{N}\mathbf{h}_n^T\mathbf{x}_n$, where $\mathbf{x}_n \equiv \mathbf{x}(t_j + n\Delta t)$ are the individual time step values of $\mathbf{x}$ during the analysis cycle, Δt is the model time step, and N is the number of time steps per analysis cycle. Thus each $\mathcal{I}$, hereafter referred to as metrics, will represent a time average, and unless otherwise specified the averaging period is each 8 day assimilation cycle. The elements of the vector $\mathbf{h}_n$ are weights that define the chosen metric of the circulation. In this case, the change in $\mathcal{I}$ given by (5) due to assimilating the observations can be written as:

$$\Delta\mathcal{I} \simeq \mathbf{d}^T\mathbf{K}^T\frac{1}{N}\sum_{n=1}^{N}\mathbf{M}_{\mathbf{b}}^T(t_n)\mathbf{h}_n = \frac{1}{N}\mathbf{d}^T\mathbf{G}\mathbf{D}\mathbf{G}^T\mathbf{V}_m\mathbf{T}_m^{-1}\mathbf{V}_m^T\mathbf{G}\mathbf{D}\sum_{n=1}^{N}\mathbf{M}_{\mathbf{b}}^T(t_n)\mathbf{h}_n \qquad (6)$$

where $t_n = t_j + n\Delta t$, and the Lanczos vector expansion for $\mathbf{K}^T$ has been used in the second equality.

One of the metrics used is based on a quadratic measure, so that $\mathcal{I} = (1/N)\sum_{n=1}^{N}(\mathbf{z}_n - \overline{\mathbf{z}})^T\mathbf{E}(\mathbf{z}_n - \overline{\mathbf{z}})$ where $\overline{\mathbf{z}}$ is the time mean of $\mathbf{z}$, and $\mathbf{E}$ is a weight matrix. In this case:

$$\Delta\mathcal{I} \simeq \frac{2}{N}\mathbf{d}^T\mathbf{G}\mathbf{D}\mathbf{G}^T\mathbf{V}_m\mathbf{T}_m^{-1}\mathbf{V}_m^T\mathbf{G}\mathbf{D}\sum_{n=1}^{N}\mathbf{M}_{\mathbf{b}}^T(t_n)\mathbf{E}(\mathbf{z}^b{}_n - \overline{\mathbf{z}}^b). \qquad (7)$$

The control vector $\mathbf{z}$ appears in (7) rather than $\mathbf{x}$ because formally the tangent linear model $\mathbf{M}_b$ operates directly on the control vector. Since $\Delta\mathcal{I}$ in (5) involves only elements of the state-vector, the weight matrix $\mathbf{E}$ will be a sparse matrix.

3.1 The Chosen Metric, $\mathcal{I}(\mathbf{x})$

Five different metrics $\mathcal{I}(\mathbf{x})$ were used in this study to quantify the impact of the observing system on the different aspects of the CCS circulation highlighted in section 1. These include the transport of the California Undercurrent (CUC), coastal upwelling, the mean transport of the CCS across a mid-California section (37°N), offshore eddy kinetic energy, and the depth of the $\sigma = 26$ kg m^{-3} isopycnal surface, which is a good proxy for the depth of the pycnocline over much of the model domain.

3.1.1 CUC transport

The transport of the CUC was identified as a circulation metric because of the desire to understand how the observing network is able to constrain this important, but poorly observed, feature of the circulation. The 30 year average meridional velocity component on the model σ-level 10 (which passes through the depth range occupied by the CUC on the shelf) computed from WCRA31 is shown in Fig. 3a, and clearly indicates the presence of a coherent and persistent poleward flowing undercurrent feature along the continental slope. In this study, the CUC was identified by the transport across the region spanning the continental shelf in the depth range 100 m - 500 m as illustrated in Fig. 3b. Specifically, the metric $\mathcal{I}_{cuc}(\mathbf{x})$ was defined as the 8-day average transport carried by the CUC where the elements of the vector $\mathbf{h}_n$ are the area of each grid cell in the $x - z$ plane that falls within the target

area indicated in Fig. 3b. Two metrics of CUC transport were considered here: $\mathcal{I}^N_{cuc}(\mathbf{x})$ the total transport integrated over the region 40.5°N-47°N between Cape Mendocino and the model northern open boundary, and $\mathcal{I}^C_{cuc}(\mathbf{x})$ the total transport integrated over the region 34.5°N-40.5°N between Cape Mendocino and Point Conception. These regions capture the dynamics of the northern and central portions of the CCS respectively, which are delineated by different wind regimes (Dorman and Winant, 1995), and are illustrated in Fig. 1a.

A time series of $\mathcal{I}^C_{cuc}(\mathbf{x}^a)$ computed from WCRA31 is shown in Fig. 4a and indicates that in the central CCS region the CUC transport exhibits a pronounced seasonal cycle. The monthly mean seasonal cycle of central coast CUC transport in Fig. 4a indicates poleward transport year-round with maximum transport during June and transport close to zero in January. Also shown in Fig. 4a is a time series of the CUC transport computed from a run of the model without data assimilation (hereafter referred to as the "forward run," not to be confused with the background), but using the same background surface forcing and boundary conditions. The influence of data assimilation on the CUC transport estimates, as evidenced by the difference between the assimilative and forward runs in Fig. 4a, is clearly significant at times, particularly in the seasonal cycle. The mean seasonal cycle of central coast CUC transport of the forward run is also poleward year-round, except in March when it is slightly equatorward but very close to zero (Fig. 4a). Times series of $\mathcal{I}^N_{cuc}(\mathbf{x})$ (not shown) exhibit a peak poleward transport in August in both the analyses and the forward run. In the forward run, the mean seasonal cycle of $\mathcal{I}^N_{cuc}(\mathbf{x})$ exhibits equatorward transport during Feb-May, while in the analyses the equatorward transport is diminished by $\sim 50\%$ (not shown).

As noted in the introduction, while CUC has been observed along the entire west coast

of North America (Pierce et al., 2000), there are no long term records of CUC transport to compare with the model. However, the model transport is within the generally accepted range of 0.5-1.5 Sv (Hickey, 1979; Pierce et al., 2000). While the metrics $\mathcal{I}^C_{cuc}$ and $\mathcal{I}^N_{cuc}$ used here are based on the transport through the fixed volume indicated in Fig. 3b, observations suggest that the position of the CUC varies with latitude along the coast and with time (Pierce et al., 2000; Connolly et al., 2014). The fixed nature of the CUC target volume in the model may account for negative excursions of the $\mathcal{I}_{cuc}$ at some times of the year in Fig. 4a (which are not documented in the literature) when the CUC may fall partly outside the target or when other aspects of the flow impinge on the target. However, Fig. 4a shows the seasonal mean transport in the model is poleward year-round along the central California coast which is in agreement with observed seasonal mean estimates. Given that the seasonal movements of the CUC may not be captured by the fixed volume target considered here, a more appropriate interpretation of $\mathcal{I}_{cuc}$ would be as an indicator of the CUC transport rather than a rigid definition.

Figure 5a shows a time series of the CUC transport increments $\Delta\mathcal{I}^C_{cuc} = \mathcal{I}^C_{cuc}(\mathbf{x}^a) - \mathcal{I}^C_{cuc}(\mathbf{x}^b)$, and represents the influence of data assimilation on the background transport during each 8 day cycle. The transport increments are generally small compared to the analysis (cf Fig. 4a), indicating that corrections by the data assimilation system to the background CUC circulation during each 8 day cycle are small. This is a desirable feature of any data assimilation system because it indicates that the system is close to equilibrium. Large transport increments from one cycle to the next would be an indication of large circulation adjusts resulting from potentially detrimental dynamical imbalances. Figure 5a also shows the CUC transport increments $\Delta\mathcal{I}^C_{cuc}$ computed after invoking the tangent linear approximation ac-

cording to the last equality in (5). The tangent linear approximation inherent in (5) clearly provides a very good estimate of the actual difference in transport between the analysis and background, lending confidence to the observation impact calculations presented in sections 4 and 5 that are based on this approximation.

3.1.2 Upwelling transport

As discussed in section 1, seasonal wind-driven upwelling is an important feature of the circulation along much of the west coast of North America. The degree to which this is constrained by the observing system is therefore also of considerable interest. To this end, the metrics $\mathcal{I}^N_{up}(\mathbf{x})$ and $\mathcal{I}^C_{up}(\mathbf{x})$ were defined as the 8-day mean upwelling transport at 40 m depth averaged over the two regions shown in Fig. 1a spanning the northern and central portions of the CCS. In this case the elements of the vector $\mathbf{h}_n$ in (5) are the areas of each grid cell element in the latitude-longitude plane spanning each region.

There are no direct observations of upwelling along the coast, and generally it is inferred from the surface winds using Ekman theory (Bakun, 1973). However, the generally coarse resolution of atmosphere forecast models leads to uncertainties in the upwelling transport estimates, and onshore geostrophic flows can suppress upwelling expected from Ekman divergence alone (Marchesiello and Estrada, 2010). As a result, there can be large discrepancies between the expected rate of upwelling based on wind estimates, and upwelling derived from models (Jacox et al., 2014).

A time series of the analysis upwelling transport $\mathcal{I}^C_{up}(\mathbf{x}^a)$ is shown in Fig. 4b from WCRA31 and the forward run, and exhibits a pronounced seasonal cycle, with maximum upwelling occurring during May. The time series of $\mathcal{I}^N_{up}(\mathbf{x}^a)$ is quantitatively similar (not

shown), with peak upwelling (downwelling) occuring during June (December). The upwelling transport and seasonal cycle in both WCRA31 and the forward run is quite similar, although at times there are some sizeable differences during some cycles.

Figure 5b shows time series of the upwelling transport increments $\Delta\mathcal{I}^C_{up} = \mathcal{I}^C_{up}(\mathbf{x}^a) - \mathcal{I}^C_{up}(\mathbf{x}^b)$ and indicates that corrections made to the background upwelling by data assimilation are not large ($\sim$ 10%) when compared to the analysis (cf Fig. 4b). Time series of the transport increments computed according to the tangent linear approximation in the last equality of (5) are also shown in Fig. 5b and agree well very with those computed directly from $\mathcal{I}(\mathbf{x}^a) - \mathcal{I}(\mathbf{x}^b)$, confirming the validity of the tangent linear approximation for this metric also.

3.1.3 $\boldsymbol{\sigma} = 26$ kg m^{-3} isopycnal depth

Individual isopycnal surfaces within the CCS are important for characterizing the physical and biogeochemical aspects of the circulation. Therefore, it is of interest to understand the degree to which the depth and structure of individual isopycnals are constrained by the observations during data assimilation. We focus our attention here on the depth of the $\sigma = 26$ kg m^{-3} isopycnal, which is a convenient proxy for the pycnocline depth over much of the model domain. The function used in this case, $\mathcal{I}_{26}(\mathbf{x})$, is not an explicit function of the prognostic variables that comprise the state-vector $\mathbf{x}$, so an additional tangent linear approximation was required which can further limit the accuracy of the observation impact calculations. Specifically, when estimating $\mathcal{I}_{26}(\mathbf{x})$ from the ROMS fields we assume a linear equation of state for density $\rho(z) = \rho_0 - \alpha T(z) + \beta S(z)$ where α and β are the thermal expansion and salinity contraction coefficients of sea water respectively, and z denotes the

vertical coordinate (not to be confused with the 4D-Var control vector $\mathbf{z}$). However, it is important to note that the full non-linear equation of state was used during 4D-Var. If we denote the depth of the $\sigma = 26$ kg m^{-3} isopycnal surface in the background as z^b_{26}, then to first-order we can approximate the density in the analysis at the same depth as $\rho^a(z^b_{26}) \simeq 1026 + \delta z(\partial\rho^b/\partial z)|_{z^b_{26}}$ where $\rho^b(z)$ is the density profile of the background based on the linear equation of state, and δz is the change in depth of the isopycnal due to assimilating the observations. Therefore, to first-order $\delta z \simeq (\rho^a(z^b_{26}) - 1026)/(\partial\rho_b/\partial z)|_{z^b_{26}}$.

The metric $\mathcal{I}_{26}(\mathbf{x})$ was defined as the 8-day mean of δz averaged over the two regions shown in Fig. 1a spanning the northern and central CCS. Time series of $\mathcal{I}^C_{26}(\mathbf{x})$ are shown in Fig. 4c from WCRA31 and the forward run. In this case, z_{26} can be computed directly and the above approximation in isopycnal depth is not required. In both cases the pycnocline depth undergoes pronounced seasonal variations being shallowest during the peak of the upwelling season. In general, the pycnocline is systematically shallower in WCRA31 compared to the forward run by ~ 11 m, indicating that data assimilation is correcting a bias in the forward model.

Time series of $\Delta\mathcal{I}^C_{26}(\mathbf{x})$ computed directly from $\mathcal{I}(\mathbf{x}_a) - \mathcal{I}(\mathbf{x}_b)$ and using the tangent linear approximation (5) in conjunction with isopycnal depth approximation for δz above are shown in Fig. 5c. Despite the additional linear approximation required to estimate δz, the two time series generally agree well, although of the five metrics considered the correlation between the two curves in Fig. 5c is the lowest.

3.1.4 Eddy kinetic energy

A region of elevated eddy kinetic energy (EKE) exists downstream and offshore of Cape Mendocino associated with the baroclinically unstable nature of the CC (cf. Fig. 1). As shown by N16, data assimilation is able to qualitatively capture this region of elevated EKE very well, including the observed offshore propagation. Since this region is thought to considerably influence the horizontal and vertical transport of momentum and biogeochemical tracers (Gruber et al., 2012), it is of interest to quantify which elements of the observing system constrain most this important aspect of the circulation. To this end, a fourth metric was computed according to $\mathcal{I}_{EKE}(\mathbf{x}) = (1/N)\sum_{n=1}^{N}(\mathbf{x}_n - \overline{\mathbf{x}})^T\mathbf{E}(\mathbf{x}_n - \overline{\mathbf{x}})$ that defines the time mean volume mean EKE within the target area identified in Fig. 1 down to 500 m for each 4D-Var assimilation cycle. The matrix $\mathbf{E}$ is diagonal with elements given by $\Delta V_i/V$, where ΔV_i is the volume of each u and v grid cell within the target volume V, and zero elsewhere, and $\overline{\mathbf{x}}$ is the seasonal cycle of $\mathbf{x}$.

Figure 4d shows time series of $\mathcal{I}_{EKE}(\mathbf{x})$ from WCRA31 and the forward run. As discussed in N16 the 4D-Var analyses are significantly more energetic than the forward run, showing that data assimilation is effective at energizing the circulation in this important dynamical region. EKE in WCRA31 shows a pronounced seasonal cycle with a minimum in Spring, compared to a fairly uniform seasonal mean in the forward run (Fig. 4d). The variability of the EKE in the analyses is also higher than in the forward model and is most likely associated with a larger number of eddies in the former as quantified by N16. Time series of the EKE increments $\Delta\mathcal{I}_{EKE}(\mathbf{x})$ computed directly from $\mathcal{I}(\mathbf{x}_a) - \mathcal{I}(\mathbf{x}_b)$ are shown in Fig. 5d. Also shown are the EKE increments computed using the tangent linear approximation (7)

for a quadratic metric. Figure 5d shows that the tangent linear approximation of $\Delta\mathcal{I}_{EKE}(\mathbf{x})$ consistently under estimates the actual $\Delta\mathcal{I}_{EKE}(\mathbf{x})$, and has a negative time mean value. This is to be expected since the linearization for this metric in (7) neglects the second-order term $\delta\mathbf{x}^T\mathbf{E}\delta\mathbf{x}$ which is always positive. Nevertheless, equation (7) faithfully captures the time variations in $\Delta\mathcal{I}_{EKE}(\mathbf{x})$ as indicated by the relatively high correlation between the two EKE increment time series.

3.1.5 37°N transport

As a measure of the alongshore circulation variations associated with the CCS, we also considered a metric of the cycle average total transport across a section at 37°N from the coast to 127°W over the upper 500 m of the water column, denoted by $\mathcal{I}_{37N}(\mathbf{x})$. As illustrated in Fig. 1, $\mathcal{I}_{37N}(\mathbf{x})$ is a measure of the net transport associated with the equatorward flowing CC and coastal jet, and the poleward flowing CUC. The impact of the observing system on this particular metric has been studied in detail by Moore et al. (2015), but is also included here for completeness.

Figure 4e shows time series of $\mathcal{I}_{37N}(\mathbf{x})$ from WCRA31 and the forward run, and indicates that the data assimilation leads to a significantly different mean alongshore circulation off central California. The analysis transport generally has a peak poleward (equatorward) value during July (April), while in the forward run peak transport occurs during October. The 37°N transport increments computed from $\mathcal{I}(\mathbf{x}_a) - \mathcal{I}(\mathbf{x}_b)$ and from the tangent linear approximation (5) are shown in Fig 5e confirming that in this case also the tangent linear approximation yields a reliable estimate of the alongshore circulation increments.

4 Control Vector Impacts

The function increment $\Delta\mathcal{I}$ in (6) and (7) can be further decomposed into contributions from the individual control vector elements, namely $\Delta\mathcal{I}_{\mathbf{x}}$ due to the increments in the initial conditions, $\Delta\mathcal{I}_{\mathbf{f}}$ due to the increments in the surface forcing, and $\Delta\mathcal{I}_{\mathbf{b}}$ due to the increments in the open boundary conditions, so that $\Delta\mathcal{I} = \Delta\mathcal{I}_{\mathbf{x}} + \Delta\mathcal{I}_{\mathbf{f}} + \Delta\mathcal{I}_{\mathbf{b}}$. This is achieved by running the tangent linear model described by $\mathbf{G}$ separately subject to the separate control vector increment components of $\sum_{i=1}^{N} \mathbf{M}_{\mathbf{b}}(t_n)\mathbf{h}_n$ in the case of $\mathcal{I}_{up}$, $\mathcal{I}_{CUC}$, $\mathcal{I}_{26}$ and $\mathcal{I}_{37N}$ or $\sum_{i=1}^{N} \mathbf{M}_{\mathbf{b}}(t_n)\mathbf{E}(\mathbf{z}_n - \overline{\mathbf{z}})$ in the case of $\mathcal{I}_{EKE}$. The resulting control vector impacts can be very useful for monitoring the performance of the 4D-Var data assimilation system throughout the analysis period.

4.1 $\mathcal{I}_{cuc}$

To illustrate the influence of the control vector increments on the CUC transport, Fig. 6a shows time series of the rms annual averaged control vector contributions to $\Delta\mathcal{I}_{cuc}^{C}$ during WCRA31. During much of the 31 year period, adjustments to the initial conditions generally contribute most to $\Delta\mathcal{I}_{cuc}^{C}$, although there are periods when the surface forcing increments are also of almost equal importance. Prior to the assimilation of SSH into the analyses starting in 1992, the contribution of the open boundary condition increments to $\Delta\mathcal{I}_{cuc}^{C}$ is relatively small. However, during the period 1992-1998 following the initial introduction of SSH into the analyses, the contribution of the boundary condition increments to $\Delta\mathcal{I}_{cuc}^{C}$ is very significant. As shown later, much of this contribution comes from the western and southern open boundary, and is believed to be due to a mismatch between the mean dynamic topography

of the SODA boundary conditions and that of the Aviso data that were assimilated into the model (Moore et al., 2015). Figure 6a shows that the contribution of the boundary condition increments to $\Delta\mathcal{I}^C_{cuc}$ gradually diminishes over time after 1998, and becomes relatively small again during the 2000-2010 decade.

Time series of the contribution of the control vector increments during each 8-day 4D-Var cycle of a single representive year, 2005 (chosen because it spans a period when observations are present from all platforms), for $\Delta\mathcal{I}^C_{cuc}$ are shown in Fig. 7a to illustrate the competing influence of individual control vector elements. While the initial conditions typically exert the largest control on the CUC transport, they are often opposed by the increments in the surface forcing. The control vector increments have a quantitatively similar impact on $\mathcal{I}^N_{cuc}$ (not shown).

4.2 $\mathcal{I}_{up}$

Time series of the rms annual averaged control vector contributions to $\Delta\mathcal{I}^C_{up}$ during WCRA31 are shown in Fig. 6b. In this case, increments to the surface forcing contribute most to $\Delta\mathcal{I}^C_{up}$. Additional calculations (not shown) reveal that it is the alongshore wind stress increments that exert the greatest control on $\Delta\mathcal{I}^C_{up}$, which is consistent with the strong control exerted by alongshore wind stress on coastal upwelling and downwelling. The contribution of the open boundary condition increments to $\Delta\mathcal{I}^C_{up}$ is significant only during 1992-1996 following the introduction of SSH observations into the 4D-Var analyses. The control vector contributions to $\Delta\mathcal{I}^C_{up}$ during each 8-day 4D-Var cycle of 2005 are shown in Fig. 7b and show that while the influence of the initial conditions on $\Delta\mathcal{I}^C_{up}$ is small they often oppose the influence of the surface forcing. The control vector impacts on $\Delta\mathcal{I}^N_{up}$ (not shown) are quantitatively similar.

4.3 $\mathcal{I}_{26}$

The rms annual averaged control vector contributions to $\Delta\mathcal{I}_{26}^{C}$ are shown in Fig. 6c and reveal that the pycnocline depth increments are controlled primarily by the initial conditions, with a small contribution from the surface forcing. This is also the case for $\Delta\mathcal{I}_{26}^{N}$ (not shown), although the isopycnal depth increments are larger in the central CCS. The control vector impacts during each individual cycle for 2005 in Fig. 7c show that the increments in the initial conditions and surface forcing generally reinforce each other for this circulation metric.

4.4 $\mathcal{I}_{EKE}$

Like the isopycnal depth, increments in the EKE are controlled primarily by the initial conditions (Fig. 6d), with a small contribution from the surface forcing, the open boundary condition impacts being negligible. Despite being much more energetic than the forward run, Fig. 7d for the representative year 2005 gives the impression that the initial condition increments act to lower the EKE of the background circulation during each cycle of the representative year 2005. However, this is not really the case since as noted in section 3.1.4, the tangent linear approximation of $\mathcal{I}_{EKE}$ is an underestimate of the true increment in EKE because, by necessity, the positive definite 2nd-order contribution $\delta\mathbf{x}^T\mathbf{E}\delta\mathbf{x}$ is neglected (cf Fig. 5d). Nonetheless, Figs. 6d and 7d illustrate the important control that the initial conditions exert on EKE.

4.5 $\mathcal{I}_{37N}$

Similar to the CUC transport, the average transport across 37°N has significant contributions from all three components of the control vector (Fig. 6e). The influence of the open boundary conditions beginning with the introduction of SSH observations into the assimilation system in 1992 is particularly evident, and this particular metric has been discussed in detail by Moore et al. (2015). In contrast to CUC transport, the opposing influences of the initial condition and surface forcing increments are less evident (Fig. 7e).

5 Observation Impacts

5.1 Innovation vs impact

As described in section 3, the contribution and impact of each individual observation on each chosen metric $\mathcal{I}$ can be quantified according to the individual elements of the dot-products in (6) and (7). A general sense of the impact of each observation on a particular $\mathcal{I}$ can be obtained by examining the relationship between each element of the innovation vector $\mathbf{d} = (\mathbf{y} - H(\mathbf{z}_b))$ and the vector $\mathbf{g} = (1/N)\mathbf{G}\mathbf{D}\mathbf{G}^T\mathbf{V}_m\mathbf{T}_m^{-1}\mathbf{V}_m^T\mathbf{G}\mathbf{D}\sum_{n=1}^{N}\mathbf{M}_\mathbf{b}^T(t_n)\mathbf{h}_n$ in the case of linear $\mathcal{I}$ or $\mathbf{g} = (2/N)\mathbf{G}\mathbf{D}\mathbf{G}^T\mathbf{V}_m\mathbf{T}_m^{-1}\mathbf{V}_m^T\mathbf{G}\mathbf{D}\sum_{n=1}^{N}\mathbf{M}_\mathbf{b}^T(t_n)\mathbf{E}(\mathbf{z}^b{}_n - \overline{\mathbf{z}}^b)$ in the case of quadratic $\mathcal{I}$. For example, Fig. 8 shows a scatter plot of the product of each element of $\mathbf{g}$ and $\mathbf{d}$ (*i.e.* the observation impact for each observation) versus each element of $\mathbf{d}$ for $\Delta\mathcal{I}_{up}^C$ for all assimilation cycles spanning the period 1980-2010. Scatter plots are shown for observations of temperature, salinity and SSH separately. The overall "butterfly wing" shape of each scatter plot indicates that observations associated with small innovations

(*i.e.* those that do not depart significantly from the background) have the smallest impact on $\mathcal{I}_{up}^{C}$. Conversely, observations associated with large innovations (*i.e.* those that depart significantly from the background) exhibit a broad range of impacts on $\mathcal{I}_{up}^{C}$, although in all cases there is a generally tapering off of the impact range as the innovation increases. In the case of temperature and salinity observations, it is reassuring to note that the very largest innovations (upto ± 10°C and -2 to 1 psu), while infrequent, have almost no impact on $\mathcal{I}_{up}^{C}$ at all. These are observations that have escaped rejection by the background quality control check (see Moore et al., 2014 and N16 for more details) but that clearly do not have an overall detrimental influence on the circulation. Scatter plots for the other metrics $\mathcal{I}$ are qualitatively similar to those in Fig. 8 (not shown).

5.2 RMS annual mean impacts for each platform

In this section we consider the impact of observations from each different observing platform. The rms observation impacts for each platform averaged over each year are qualitatively and remarkably similar for each of the metrics considered here for both the central and northern CCS regions, so we will present only the case for $\Delta\mathcal{I}_{up}^{C}$. Time series of the rms impact of each observing platform on $\Delta\mathcal{I}_{up}^{C}$ averaged over each year are shown in Fig. 9a. During 1980, only *in situ* hydrographic observations were available, during which time the upwelling transport increments are due solely to the assimilation of these data. In 1981, SST observations from the AVHRR Pathfinder satellite became available and starting in 1982 they dominate $\Delta\mathcal{I}_{up}^{C}$. For the next decade AVHRR and *in situ* observations were the only data available until the introduction of Aviso SSH in 1992. Between 1992 and 2000 Aviso SSH and AVHRR SST collectively contribute almost all of the information about the upwelling transport

increments. During this period, the impact of SSH on $\mathcal{I}_{up}$ is $\sim 25-50\%$ that of AVHRR SST, although recall that SSH observations were not assimilated within 50 km of the coast. In 2000, two additional satellite platforms came online, AMSR and MODIS. Therefore, during the last decade of WCRA31 SST from three satellite platforms is generally available and must be combined to form multi-platform super observations when appropriate. Figure 9a shows that collectively SST has the largest impact on $\mathcal{I}_{up}^{C}$, and with the introduction of MODIS and AMSR the impact of Aviso SSH declines. The SST super observations typically exert the largest influence on $\Delta\mathcal{I}_{up}$, although the impact of MODIS SST observations is comparable when these data are present and do not need to be combined with other data to form multi-platform super observations. The impact of MODIS SST and SST super observations surpasses that of AVHRR because of the relative size of the nominal *a priori* observation error assumed for each platform: 0.6°C for AVHRR, 0.3°C for MODIS, 0.7°C for AMSR, and the standard errors for super observations depending on the platforms that contribute. As a result, during the 2001-2010 decade AVHRR and AMSR SST contribute least to $\Delta\mathcal{I}_{up}^{C}$ when available and not in the form of multi-platform super observations. The partition of the impact between each SST platform, SSH and *in situ* observations is also a reflection of the relative volume of each data type as shown in Fig. 2.

The impact of each observing platform on the contribution of each of the control vector components to $\Delta\mathcal{I}_{up}^{C}$ was also computed and mirrors that of the total increment in Fig. 9a (not shown).

The dominant impact of SST observations on 4D-Var upwelling transport estimates can be understood in terms of the influence of upwelling on surface temperatures. Any mismatch between the observed and modeled SST (quantified by the innovations $\mathbf{d}$) will be corrected

by 4D-Var based on Green's function information provided by the adjoint model. Figure 6b indicates that much of the circulation change required to better fit the model to the SST observations comes about through 4D-Var corrections to the wind, specifically alongshore wind stress as demonstrated by Broquet et al. (2011).

It is note worthy that the quantum jumps in the number of available observations in 1992 and 2000 (cf Fig. 2) are not reflected in $\Delta\mathcal{I}^C_{up}$. As each new observing platform comes online, the magnitude of the circulation increments remains largely unchanged, and is simply repartitioned between the pre-existing and new observing platforms. This is further evidence of the stability of the 4D-Var analysis system, and that the statistical equilibrium of the increments that is established early on the analyses (around 1982) is largely unaffected by the introduction of each new observing platform.

For all of the metrics, the in situ hydrographic observations appear to exert a negligible impact on the circulation. However, this can be misleading as illustrated in Fig. 9b which shows the rms annual average impact *per datum* for $\mathcal{I}^C_{up}$. When viewed in this way, the impact of each in situ observation on upwelling transport is clearly much larger than it is for satellite observations, particularly during the early 1980s when only in situ observations and AVHRR SST were available. Over time, however, the impact per datum undergoes a rapid decline as the total number of observations from all platforms increases. Figures 9a and 9b are qualitatively representative of the behavior of all the metrics in both regions.

5.3 Impacts for a representative year

Time series of the observation impacts for each platform during each 4D-Var cycle are shown in Fig. 10 for the representative year 2005 for each of the metrics considered in the central

CCS region. Those from the northern CCS region are quantitatively very similar and are not shown. Figure 10 shows that SST observations dominate the increments in all of the metrics $\mathcal{I}$. The impact of SSH observations appears to be largest for the horizontal transport metrics $\mathcal{I}_{cuc}$ (Fig. 10a) and $\mathcal{I}_{37N}$ (Fig. 10e). SST super observations typically have the largest influence for all $\mathcal{I}$. In addition, AMSR SST generally has a very small impact due to a combination of higher observation error and lower horizontal resolution ($\sim$ 30 km) compared to AVHRR and MODIS. Perhaps somewhat counter-intuitively SSH observations have relatively little impact on the EKE (Fig. 10d), indicating that 4D-Var is able to effectively utilize the signatures of eddies and filaments present in the SST observations to reconstruct consistent horizontal circulation fields for the eddies. N16 have shown that the number of eddies present during WCRA31 is $\sim 50\%$ higher than in the forward run. This highlights the power of the adjoint model to dynamically interpolate information from the observations between the observed and unobserved elements of the state vector.

By-and-large, Fig. 10 also shows that contributions of observations from each platform to each $\Delta\mathcal{I}$ tend to reinforce each other. Only in a few instances is there evidence for any opposing influence of observations from a platform, and the effect is usually quite small, which can be taken as an indication of dynamical consistency of observational information across all platforms.

5.4 Cumulative impacts

Another illuminating way to view the impact of observations from each platform is in terms of the cumulative change in each metric per unit time, which is defined here as:

$$\mathcal{I}_\tau = \frac{1}{\tau}\int_0^\tau |\Delta\mathcal{I}(t)|dt \tag{8}$$

where $t = 0$ denotes 9 Jan 1980 (the start time of the first 4D-Var cycle), and τ is the elapsed time since 9 Jan 1980. A time series of $\mathcal{I}_\tau$ for coastal upwelling in the central CCS is shown in Fig. 9c, and indicates that $\mathcal{I}_\tau$ very quickly reaches a near constant value. Therefore, the rate of accumulated upwelling transport associated with each data assimilation cycle remains relatively constant. Also shown in Fig. 9c is the contribution to $\mathcal{I}_\tau$ from each individual observation platform. The cumulative upwelling transport increment due to the in situ observations is initially high, but decreases slowly over time due to the influence first of AVHRR, and then to the successive introduction of each new observing platform. Nevertheless, the rate of accumulated transport associated with in situ observations is comparable to that of Aviso, MODIS and SST super observations during the 2000-2010 decade. Clearly AVHRR SST observations have the largest accumulated rate of upwelling transport because of the long succession of AVHRR/Pathfinder missions. The decline in the influence of AVHRR around 1999/2000 when MODIS came online is noticeable showing how the impact of the AVHRR radiometer is down-graded by the more accurate instrument aboard MODIS (cf section 5.2). Time series of $\mathcal{I}_\tau$ for each of the other metrics are qualitatively similar to Fig. 9c (not shown).

5.5 Impact per datum versus location

It is also instructive to compute the impact per datum of the observations based on their geographic location. For example, Fig. 11a shows the rms average impact per datum of

SST observations from AVHRR on central CCS upwelling ($\Delta\mathcal{I}^C_{up}$) within each grid cell of the model domain. Clearly SST observations in the central CCS region have the largest impact. The same is generally true of SST observations from MODIS (Fig. 11c), although in this case observations from farther afield also have a significant influence. In both cases, SST observations upstream of the target region near the coast also exert a significant influence. The corresponding rms impacts for AVHRR and MODIS on northern CCS upwelling ($\Delta\mathcal{I}^N_{up}$) are shown in Figs. 11b and 11c. While MODIS observations typically have a larger impact on $\Delta\mathcal{I}^C_{up}$ than AVHRR, their impacts are comparable in the case of $\Delta\mathcal{I}^N_{up}$.

The geographical distribution of the SST observation impacts for $\Delta\mathcal{I}_{cuc}$, $\Delta\mathcal{I}_{37N}$ and $\Delta\mathcal{I}_{EKE}$ have remarkably qualititively similar patterns to those shown in Fig. 11. For example, although not shown here, the SST impacts for $\Delta\mathcal{I}_{cuc}$ in the central (northern) CCS are qualitatively very similar to those shown in Figs. 11a and 11b (11c and 11d). Similarly, the SST impacts on $\Delta\mathcal{I}_{37N}$ and $\Delta\mathcal{I}_{EKE}$ qualitatively resemble those of central CCS upwelling in Figs. 11a-11d (not shown), indicating that observations outside and upstream of these target regions (cf Fig. 1) have a significant impact on these aspects of the circulation.

Figures 11e and 11f show the rms impact of the Aviso SSH observations on $\Delta\mathcal{I}^C_{up}$ and $\Delta\mathcal{I}^N_{up}$ respectively, and the geographical distribution of the impacts is very similar in both cases. Two areas along the western and southern open boundaries show up as particularly significant. These are associated with the SSH open boundary condition issues discussed in section 4.1, and are discussed in detail by Moore at al. (2015). In addition, there is a broad area of moderately high impact between 130°W and 125°W, 33°N and 39°N. In fact, these features are common to all of the metrics considered here (not shown) and most likely associated with pressure gradient influences on the circulation as hypothesized by Moore et

al. (2015). In general, Aviso observations have less impact than SST observations per grid cell.

The rms impacts for all in situ temperature and salinity observations regardless of depth are shown in Figs. 12a and 12b for $\Delta\mathcal{I}^C_{up}$ and Figs. 12c and 12d for $\Delta\mathcal{I}^N_{up}$. The geographical distributions of the temperature impacts mirror those of SST. For salinity observations, however, the salinity observations along the central California coast have the largest impact on upwelling in both regions. As in the case of SST and SSH, the geographic distributions of the in situ observations are qualitatively similar for the other metrics considered (not shown). Consistent with Fig. 9b, the impact of the in situ observations per datum per grid cell is larger than that of either SST or SSH by about an order of magnitude.

The average impact of each in situ temperature and salinity observation as a function of depth (averaged over the whole model domain) is shown in Fig. 13 for $\Delta\mathcal{I}^C_{up}$ and $\Delta\mathcal{I}^C_{cuc}$. In both cases, the impact per observation decreases with depth, although in the case of $\Delta\mathcal{I}^C_{cuc}$, the impact of subsurface temperature observations peaks around 100 m, and there is the hint of a secondary peak near 500 m.

The remarkable degree of similarity between the geographical distributions of the impacts for each platform across all five metrics is associated with the fact that the chosen $\mathcal{I}$ are not mutually exclusive, and all aspects of the circulation described by the chosen suite of $\mathcal{I}$ are intimately linked by the circulation dynamics.

6 Observation Impacts on Climate Variability

As discussed in section 1, the CCS is significantly influenced by climate variability on seasonal-to-decadal time scales associated with the dominant modes of climate variability in the tropical and North Pacific. Crawford et al. (2016; hereafter C16) have performed an analysis of climate-induced variability in the CCS circulation based on WCRA31. Using linear inverse modeling, C16 identified three dominant Principal Oscillation Patterns (POPs; Hasselmann, 1988) that collectively account for $\sim 55\%$ of the low-frequency variability in the CCS circulation. The POPs are identified by modeling the departures of the state-vector from the seasonal cycle $\delta\mathbf{x} = (\mathbf{x} - \overline{\mathbf{x}})$ as a linear stochastic system $d\delta\mathbf{x}/dt = \mathbf{A}\delta\mathbf{x} + \boldsymbol{\xi}(t)$, where $\mathbf{A}$ is the dynamical matrix that advances $\delta\mathbf{x}$ forward in time, and $\boldsymbol{\xi}(t)$ is white noise stochastic forcing. As shown by von Storch et al. (1994), an empirical approximation $\tilde{\mathbf{A}}$ of $\mathbf{A}$ can be derived according to $\tilde{\mathbf{A}} = \tau^{-1} ln(\mathbf{C}(\tau)\mathbf{C}^{-1}(0))$ where $\mathbf{C}(\tau) = \langle \delta\mathbf{x}(\tau)\delta\mathbf{x}(0)^T \rangle$ is the lag-τ covariance matrix of $\delta\mathbf{x}$ and $\mathbf{C}(0) = \langle \delta\mathbf{x}(0)\delta\mathbf{x}(0)^T \rangle$ is the zero lag covariance matrix. The POPs are identified as the leading eigenvectors $\boldsymbol{\phi}_j$ of $\tilde{\mathbf{A}}$, and represent in empirical approximations of the normal modes of the system (Pedlosky, 1976). The focus of C16 is low frequency variability, so monthly mean state-vector fields of $\mathbf{x}$, smoothed spatially using a Shapiro filter, were used to compute $\tilde{\mathbf{A}}$. The combined spatio-temporal averaging and smoothing isolates the low-frequency, low-wavenumber component of the circulation. In addition, to facilitate computation of $\tilde{\mathbf{A}}$, the dimension of the problem was reduced by projecting $\delta\mathbf{x}$ onto the leading 50 EOFs of $\mathbf{C}(0)$.

As described by C16, three dominant POPs contribute to climate variability in WCRA31. POP1 with a period of 3.6 years and e-folding decay time of 2.2 years captures much of the

variability in the CCS that is associated with ENSO. POP2 (5.2 year period, 2.7 year e-folding decay time) and POP3 (9.7 year period, 9.6 year e-folding decay time) capture CCS variability associated with the PDO and NPGO combined, which in the case of POP3 takes the form of a resonant response. These same POPs are not present in the forward run, indicating that they are being recovered by data assimilation from the observational data. It is therefore of considerable interest to quantify the impact of each observing platform on the low-frequency CCS variability that is described by the POPs. With this in mind we consider three additional metrics that quantify the amplitude of the projection of the 4D-Var circulation increments $\delta\mathbf{z}^a$ on each POP during each 8-day assimilation window. Since the POPs are defined in terms of the covariance of the state-vector $\mathbf{x}$ about the mean seasonal cycle $\overline{\mathbf{x}}$, we consider a metric of the form:

$$\mathcal{I}_{pop_j}(\mathbf{z}) = (1/N)\sum_{n=1}^{N} \boldsymbol{\phi}_j^{\dagger T}\mathbf{P}(\mathbf{z}_n - \overline{\mathbf{z}}) \tag{9}$$

where $\mathbf{P}$ is a smoothing operator[1] that isolates the low wavenumber components of the circulation as mentioned above. The POPs satisfy the biorthogonality relation $\boldsymbol{\phi}_j^H\boldsymbol{\phi}_n^\dagger = \delta_{jn}$, where $\boldsymbol{\phi}_n^\dagger$ are the eigenvectors of $\tilde{\mathbf{A}}^T$, and superscript H denotes the Hermitian transpose. The biorthogonality relation was used in the derivation of (9) and isolates the component of $\mathcal{I}$ associated with POP j. The increments in the POP amplitudes that arise from data assimilation can then be computed according to:

$$\Delta\mathcal{I}_{pop_j} = \frac{1}{N}\mathbf{d}^T\mathbf{G}\mathbf{D}\mathbf{G}^T\mathbf{V}_m\mathbf{T}_m^{-1}\mathbf{V}_m^T\mathbf{G}\mathbf{D}\sum_{n=1}^{N}\mathbf{M}_{\mathbf{b}}^T(t_n)\mathbf{P}^T\boldsymbol{\phi}_j^\dagger. \tag{10}$$

[1]Twenty applications of a 2nd-order Shapiro filter were used as described by C16.

Since in general the eigenvectors $\boldsymbol{\phi}_j$ and $\boldsymbol{\phi}_j^\dagger$ are complex, the observation impacts for the resulting complex metric $\mathcal{I}_{pop_j}$ must be computed for the real and imaginary components separately.

Time series of the modulus of the leading POP amplitudes, $|\mathcal{I}_{pop_j}(\mathbf{x}^a)|$, during WCRA31 are shown in Fig. 14. Also shown in Fig. 14 are time series of $|\Delta\mathcal{I}_{pop_j}|$ and indicate that the changes in POP amplitude during each 4D-Var cycle are small.

In all cases it is the initial conditions that have the largest impact on the POP amplitude increments $|\Delta\mathcal{I}_{pop_j}|$ (not shown). The observation impacts for $|\Delta\mathcal{I}_{pop_j}|$ (not shown) are qualitatively similar to those shown in Figs. 9a, 9b and 9c for central CCS upwelling, suggesting that each observing platform influences the week-to-week variations and climate-induced variability of the CCS to a similar degree. The rms impact per datum in each model grid cell of AVHRR SST observations on the real and imaginary components of each POP are shown in Fig. 15, and the geographical distribution of the SST impacts mirror the SST structure of the real and imaginary component of each POP presented in C16 (not shown). For example, the SST impact on $Im(\mathcal{I}_{pop_1})$ in Fig. 15b has a structure similar to that of the SST associated with response of the U.S. west coast to large amplitude ENSO events.

The average impact of in situ temperature and salinity observations on the real and imaginary amplitude components of POP3 are shown in Fig. 16. While the impact typically decreases with depth, there is a pronounced peak associated with temperature observations at a depth of $\sim$ 100 m. The same is true for POP1 and POP2 also (not shown). This is most likely because variations in the upper ocean heat content and thermocline depth play an important role for the low-frequency variations in the circulation associated with each POP as discussed by C16.

Each POP evolves in time according to $a_j e^{\omega_j t}\boldsymbol{\phi}_j$, where a_j is the complex POP amplitude and ω_j are the eigenvalues of $\tilde{\mathbf{A}}$ and represent the complex frequencies (*i.e.* $Re(\omega_j)$ is the POP decay rate and $Im(\omega_j)$ is the POP oscillation frequency). Based on the deterministic evolution of the POPs, the impact of each observation on $\Delta\mathcal{I}_{pop_j}$ can therefore be traced through time for the entire duration of WCRA31. The POPs generally occur in complex conjugate pairs, and the time evolution of the state-vector anomaly $\delta\mathbf{x}_j(t,\tau)$ at time t associated with the increment in POP j introduced at observation time τ is a combination of the POP and its complex conjugate according to $\delta\mathbf{x}_j(t,\tau) = \Delta\mathcal{I}_{pop_j} e^{\omega_j(t-\tau)}\boldsymbol{\phi}_j + \Delta\mathcal{I}^*_{pop_j} e^{\omega^*_j(t-\tau)}\boldsymbol{\phi}^*_j$ where the superscript $*$ denotes the complex conjugate. The contribution of $\delta\mathbf{x}_j(t,\tau)$ to each of the circulation metrics defined in section 3.1 is given by $\mathbf{h}^T\delta\mathbf{x}_j(t,\tau)$ or $(\mathbf{z}^b(t)-\overline{\mathbf{z}}^b)^T\mathbf{E}\delta\mathbf{x}_j(t,\tau)$ in the case EKE. Therefore, the time evolution of the contribution of the observations to all linear functionals will be the same regardless of the choice of $\mathbf{h}$, except for a difference in phase and amplitude (including the sign) depending on $\mathbf{h}$. This is therefore a very powerful diagnostic for tracing the influence of each observation or observing platform through time associated with the climate variability described by each POP.

Another useful diagnostic is $c_j(t) = \int_0^t \mathbf{h}^T\delta\mathbf{x}_j(t,\tau)d\tau$ or $c_j(t) = \int_0^t (\mathbf{z}^b(\tau)-\overline{\mathbf{z}}^b)^T\mathbf{E}\delta\mathbf{x}_j(t,\tau)d\tau$ which represents the contribution of *all* observations over the interval $[0,t]$ to the chosen circulation metric at time t associated with POP j. For example, Fig. 17 shows the contribution of each observing platform to central CCS upwelling through time. Satellite SST contributes most over time, although in situ temperature measurements are an important contributor to POP1 upwelling during the 1980s (Fig. 17a). However during this time the in situ observations generally oppose the contribution from SST. During the 2000s, Aviso SSH is important for POP3 upwelling (Fig. 17e) and its contribution is almost 90° out of

phase with that of SST. Coastal upwelling associated with POP2 (Fig. 17c) and POP3 (Fig. 17e) also have largest amplitude during this period, which as shown by Jacox et al. (2014) and C16 is due to the predominant in-phase relationship between the PDO and NPGO at this time. Time series of $c_j(t)$ for CUC transport, isopycnal depth and 37°N transport are qualitatively similar to those of upwelling apart from a phase shift and/or a change in sign.

Time series of $c_j(t)$ for EKE display quite a different character, however, which are also shown in Fig. 17. For POP1, the 1980s and 2000s are characterized by highest amplitude EKE associated with ENSO variability (Fig. 17b). For POP3, on the other hand, the amplitude of EKE is only marginally higher during the 2000s (Fig. 17f). During this time the tendency of satellite SST observations to increase the amplitude of EKE is largely offset by SSH observations. POP2 EKE (Fig 17d) exhibits similar behavior to POP1. Figure 17 is a further illustration of the competition between the contribution of observations from different observing platforms to different aspects of the low-frequency circulation.

The integrand $f_j(t,\tau) = \mathbf{h}^T \delta\mathbf{x}_j(t,\tau)$ or $f_j(t,\tau) = (\mathbf{z}^b(\tau) - \overline{\mathbf{z}}^b)^T \mathbf{E}\delta\mathbf{x}_j(t,\tau)$ of $c_j(t)$ provides a detailed view of the observations that contribute to the individual events that are apparent in Fig. 17. For example, Fig. 18 shows $f_3(t,\tau)$ for POP3 central CA upwelling associated with AVHRR SST (Fig. 18a), MODIS SST (Fig. 18b) and Aviso SSH (Fig. 18c) over the periods when these observations were assimilated into the model. In each figure, the abscissa represents t and the ordinate is the observation time τ. The extended period of influence of the observations due to the low frequency of POP3 ($2\pi/Im(\sigma_3) = 9.7$ yrs, $1/Re(\sigma_3) = 9.6$ yrs) is remarkable, and observations can significantly influence the upwelling for decade or more. Notice, for example, how MODIS SST observations during the summer of 2003 have a large influence of upwelling that extends through to mid-2010. Similarly, the influence on

upwelling of Aviso SSH observations made the 1997/1998 El Niño and 1999/2000 La Niña lasts for a decade. The opposing influences of SST and SSH on POP3 upwelling during the period 2000-2010 noted earlier is also apparent in Fig. 18b when assimilation of MODIS SST generally decreases upwelling, while Fig. 18c shows that assimilation of Aviso SSH generally increases upwelling. Notice also how the influence of AVHRR on upwelling decreases around 2000 when the more accurate MODIS observations come online. The extended impact of each observing platform over time associated with the POPs is apparent in all of the circulation metrics considered here (not shown).

7 Summary and Conclusions

In this study, we have have explored the impact of different satellite and in situ observing platforms on circulation estimates of the CCS computed using 4D-Var data assimilation, and spanning the three decade long period 1980-2010. Specifically, we have examined the impact of the existing observing network on estimates of several important defining aspects of the CCS that include coastal upwelling, the transport near the coast of the CUC, the depth of the thermocline, the alongshore transport, and eddy kinetic energy. In addition, we have also quantified the impact of the observations on estimates of three leading modes of circulation variability associated with ENSO, the NPGO and the PDO. The method used was that of Langland and Baker (2004), employed routinely in NWP, and based on a tangent linear approximation, and was found to perform well for both the linear and quadratic circulation metrics considered here. The impact of different components of the 4D-Var control vector was also quantified using the same approach. The relative impact of the control vector

components was found to be strongly metric dependent in that transport-based metrics are influenced by all components, while isopycnal depth and EKE depend most strongly on the initial conditions. The amplitude of different modes of climate variability of the circulation is also primarily controlled by the initial conditions. In addition, different components of the control vector can have opposing impacts on the circulation, highlighting the some times complex interplay between the control vector elements during the 4D-Var procedure.

Perhaps surprisingly, observations associated with the largest innovations do not necessarily exert the largest impact on the circulation estimates. In fact, the observations that depart most from the model background typically have the least impact on the circulation. This indicates that the 4D-Var procedure is able to down-weight the influence of such observations when information from other neighboring observations indicates that the large innovations are unreliable.

Overall, the observation impacts were found to be qualitatively similar for all of the circulation metrics considered. In hindsight, this is perhaps not too surprising because all of the circulation metrics considered are linked by the CCS dynamics. For example, CUC transport is a subset of 37N transport at that latitude; upwelling transport, pycnocline depth and alongshore transport of the CC will be strongly correlated; EKE will be related to alongshore transport through modulations of baroclinic instability; and EKE and pycnocline depth will be correlated due to offshore propagation of the eddies. All of these dynamical connections are manifested in the observation impact calculations via the adjoint model and the Kalman gain matrix. The observation impact calculations also illustrate the stability of the 4D-Var system through time, in that as different observing platforms come online through time, the amplitude of the circulation increments remains relatively unchanged,

and the impact of the observations on each circulation metric is redistributed between the different platforms.

SST observations typically have the largest overall impact because of the shear volume of data that is assimilated. The relative ranking of the individual SST observation platforms (*i.e.* AVHRR vs MODIS) is controlled by the a priori observation error. When multiple SST platforms provide observations at the same time, super observations typically have the largest impact since they are assigned the lowest standard error.

However, while relatively few in number compared to satellite data, in situ hydrographic observations have by far the largest impact per datum, and a single hydrographic profile can have the same impact as many more satellite observations. This suggests that the 4D-Var analyses may derive further benefit from the in situ observations if the satellite SST observations are decimated. In fact, satellite observations are routinely decimated in NWP for this very reason. In situ observations are also important of course because they provide the only source of information about the subsurface conditions. Nonetheless, our calculations show that near surface hydrographic data have the largest impact, although for some metrics obsevations at around 100 m depth have the largest impact. This is especially true for the impact of in situ observations on the low-frequency climate variability captured by the analyses because of the important role played by upper ocean heat content and thermocline dynamics in this case.

A novel aspect of the current work is quantification of the observation impacts on climate variability that is captured by the 4D-Var analyses. In particular, we are able to trace the impact of each observation through time via the time evolution of the Principal Oscillation Patterns (POPs). A remarkable finding is that observations that project onto POPs that

describe slowly damped, low frequency variability can exert a signicant influence on the circulation upto a decade into the future. Throughout we also see very clear evidence for how the impact of observations from different platforms can reinforce or oppose each other depending on the mode of climate variability under consideration (eg ENSO vs PDO and NPGO).

This study also highlights the value of quantifying the observation impacts as a means of monitoring both the effectiveness of the data assimilation system and observing array. In the system considered here, issues were encountered due to an inconsistency between the mean sea surface height of the Aviso data and that of the open boundary conditions used to constrain the model. Had we been monitoring the observation impacts in conjunction with the 4D-Var analyses, we could have taken mitigating steps to correct this inconsistency. Routine monitoring of the observation impacts will be our modus operandi in the future. While other methods, such as OSSEs, can be used to monitor the influence of different observing platforms on the circulation analyses (and subsequent forecasts), the method used here of LB04 is much more cost effective than OSSEs and is easy to implement within the existing ROMS system.

The non-local influence of observations on different aspects of the circulation was also noted here. For example, observations both upstream and downstream of a particular target area can have a significant impact on the target metric. Similarly, observations can impact ocean circulation estimates many years into the future. This is direct evidence of the influence of ocean dynamics on the propagation of observational information during the 4D-Var process. Obvious processes include advection, wave dynamics and large-scale balances such as geostrophy, and we are currently working on ways to further unravel the relative impact

of each of these different processes.

Another interesting aspect of this study is that it demonstrates how information about observed features of the circulation, such as SST and SSH, can influence very poorly observed (or in some cases indirectly unobserved) aspects of the circulation, such as coastal upwelling, undercurrent transport, and eddy kinetic energy. Again this underscores the power of the adjoint operators in equations (6) and (7) which affects the information exchange between the observed and unobserved elements of the state vector (or any function $\mathcal{I}$ of them). This in turn provides valuable quantitative information about the value of observations from the different elements of the observing network. Some of the dynamical connections between the observation impacts and the associated circulation metrics are relatively straightforward to understand. For example, assimilation of satellite SST observations can strongly influence 4D-Var estimates of coastal upwelling because of the signature of the upwelling process in surface temperature. As demonstrated here, the SST observations mainly control upwelling transport estimates via changes in the alongshore wind stress, information that is derived from the circulation Green's functions, courtesy of the adjoint model. Since the chosen metrics $\mathcal{I}$ are not mutually exclusive, this information will influence other aspects of the circulation as well, such as CUC transport and EKE. As such, the impact of the observations on some aspects of the circulation can be surprising, and more effort is warranted to identify ways to isolate the particular dynamical pathways of each observation through the 4D-Var system.

Acknowledgements

The authors wish to acknowledge the support of NOAA through a grant from the South-

west Universities Research Association (SURA) as part of the Coastal Ocean Modeling Testbed (COMT; NA13NOS0120139) and the National Science Foundation (OCE 1061434).

References

Atlas, R., R.N. Hoffman, J. Ardizzone, S.M. Leidner, J.C. Jusem, D.K. Smith, D. Gombos, 2011: A cross-calibrated, multiplatform ocean surface wind velocity product for meteorological and oceanographic applications. *Bull. Amer. Meteor. Soc.*, **92**, 157-174, doi: 10.1175/2010BAMS2946.1.

Bakun, A., 1973: Coastal upwelling indices, west coast of North America, 1946?71, NOAA Tech. Rep., NMFS SSRF-671, 103 pp., U.S. Dept. of Commerce.

Balmaseda, M.A., D.L.T. Anderson and A. Vidard, 2007: Impact of Argo on analyses of the global ocean. *Geophys. Res. Lett.*,**34**, L16605. http://dx.doi.org/10.1029/2007GL0304452.

Broquet, G., A.M. Moore, H.G. Arango, and C.A. Edwards, 2011: Corrections to ocean surface forcing in the California Current System using 4D-variational data assimilation. *Ocean Modelling*, **36**, 116-132.

Carton, J.A. and B.S. Giese, 2008: A reanalysis of ocean climate using simple ocean data assimilation (SODA). *Mon. Wea. Rev.*, **136**, 2999-3017.

Chapman, D.C., 1985: Numerical treatment of cross-shelf open boundaries in a barotropic coastal ocean model. *J. Phys. Oceanogr.*, **15**, 1060-1075.

Checkley, D.M. and J.A. Barth, 2009: Patterns and process in the California Current system. *Prog. Oceanogr.*, **83**, 49-64.

Connolly, T.P., B.M. Hickey, I. Schulman and R.E. Thomson, 2014: Coastal trapped waves, alongshore pressure gradients, and the California Undercurrent. *J. Phys. Oceanogr.*, **44**, 319-342.

Courtier, P., J.-N. Thépaut and A. Hollingsworth, 1994: A strategy for operational imple-

894 menation of 4D-Var using an incremental approach. *Q. J. R. Meteorol. Soc.*, **120**,
895 1367-1388.
896 Crawford, W.J., A.M. Moore, M.G. Jacox, J. Fiechter, E. Neveu and C.A. Edwards, 2016: A
897 Resonant Response of the California Current Circulation to Forcing by Low Frequency
898 Climate Variability. *Deep Sea Research II*, In press.
899 Daley, R., 1991: Atmospheric Data Analysis. Cambridge University Press, 457pp.
900 Dee, D.P. et al, 2011: The ERA-Interim reanalysis: configuration and performance of the
901 data assimilation system, *Q.J.R. Meteorol. Soc.*, **137**, 553-597, doi: 10.1002/qj.828.
902 Dibarboure, G., M.-I. Pujol, F. Briol, P. Y. Le Traon, G. Larnicol, N. Picot, F. Mertz, and
903 M. Ablain, 2011: Jason-2 in DUACS: Updated system description, first tandem results
904 and impact on processing and products. *Mar. Geod.*, **34**, (3-4), 214-241.
905 Di Lorenzo, E.,N. Schneider, K.M. Cobb, P.J.S. Franks, K. Chhak, A.J. Miller, J.C. McWilliams,
906 S.J. Bograd, H. Arango, E. Curchitser, T.M. Powell and P. Rivíere, 2008: North Pacific
907 Gyre Oscillation links ocean climate and ecosystem change. *Geophys. Res. Letts.*, **35**,
908 L08607, doi:10.1029/2007GL032838.
909 Di Lorenzo, E., J. Fiechter, N. Schneider, A. Bracco, A.J. Miller, P.J.S. Franks, S.J. Bograd,
910 A.M. Moore, A.C. Thomas, W. Crawford, A. Peña and A.J. Hermann, 2009: Nutrient
911 and salinity decadal variations in the central and eastern North Pacific. *Geophys. Res.*
912 *Letts.*, **36**, L14601, doi:10.1029/2009GL038262.
913 Dorman, C.E. and C.D. Winant, 1995: Buoy observations of the atmosphere along the west
914 coast of the United States, 1981-1990. *J. Geophys. Res.*, **100**, 16,029-16,044.
915 Errico, R.M., 2007: Interpretations of an adjoint-derived observational impact measure.
916 *Tellus*, **59A**, 273-276.

Fairall, C.W., E.F. Bradley, J.S. Godfrey, G.A. Wick, J.B. Ebson and G.S. Young, 1996a: Cool-skin and warm layer effects on the sea surface temperature. *J. Geophys. Res.*, **101**, 1295-1308.

Fairall, C.W., E.F. Bradley, D.P. Rogers, J.B. Ebson and G.S. Young, 1996b: Bulk parameterization of air-sea fluxes for tropical ocean global atmosphere Coupled-Ocean Atmosphere Response Experiment. *J. Geophys. Res.*, **101**, 3747-3764.

Flather, R.A., 1976: A tidal model of the northwest European continental shelf. Memoires de la Societe Royale des Sciences de Liege 6 (10), 141-164.

Frischknecht, M., M. Münnich and N. Gruber, 2015: Remote versus local influence of ENSO on the California Current System. *J. Geophys. Res.*, **120**, 1353-1374, doi:10.1002/2014JC010531.

Gelaro, R. and Y. Zhu, 2009: Examination of observation impacts derives from Observing System Experiments (OSEs) and adjoint models. *Tellus*, **61A**, 179-193.

Gill, A.E., 1982: Atmosphere-Ocean Dynamics, Academic Press, 662pp.

Gratton, S., and J.Tshimanga, 2009: An observation-space formulation of variational assimilation using a restricted preconditioned conjugate gradient algorithm. *Q. J. R. Meteorol. Soc.*, **135**, 1573-1585.

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

Hasslemann, K., 1988: PIPs and POPs - a general formalism for the reduction of dynamical systems in terms of Principal Interaction Patterns and Principal Oscillation Patterns. *J. Geophys. Res.*, **93**, 11015-11020.

Hickey, B.M., 1979: The California Current System: hypotheses and facts. *Progress in*

940 *Oceanography*, **8**, 191-279.

941 Hickey, B.M., 1998: Coastal oceanography of western North America from the tip of Baja,
942 California to Vancouver Island. *The Sea*, **11**, 345-393, Eds. A.R. Robinson and K.H.
943 Brink, John Wiley and Sons.

944 Jacox, M.G., A.M. Moore, C.A Edwards and J. Fiechter, 2014: Spatially resolved upwelling
945 in the California Current System and its connections to climate variability *Geophysical*
946 *Research Letters*, **41**, 3189-3196.

947 Jacox, M.G., J. Fiechter, A.M. Moore and C.A. Edwards, 2015: El Niño and the California
948 Current Coastal Upwelling Response. *J. Geophys. Res.*, **120**., doi:10.1002/2014JC010650.

949 Jacox, M.G., E.L. Hazen, C.A. Edwards, P.T. Drake, A.M. Moore and S.J. Bograd, 2016: Im-
950 pacts of the 2015-2016 El Nino on the California Current System: An early assessment
951 and comparison to past events. *Geophys. Res. Letts.*, doi:10.1002/2016GL069716.

952 Johnstone, J.A., and N.J. Mantua, 2014: Atmospheric controls on northeast Pacific tem-
953 perature variability and change, 1900-2012. *Proceedings of the National Academy of*
954 *Sciences*, **111**, 14360-14365.

955 Jung, B.-J., H.K. Kim, T. Auligné, X. Zhang and X.-Y. Huang, 2013: Adjoint-derived
956 observation impact using WRF in the western North Pacific. *Mon. Wea. Rev.*, **141**,
957 4080-4097.

958 Källberg, P., A. Simmons, S. Uppala and M. Fuentest, 2004: The ERA-40 Archive. ERA-40
959 Project Report Series No. 17.

960 Kelly, K.A., R.C. Beardsley, R. Limeburner, K.H. Brink, J.D Paduan and T.K. Chereskin,
961 1998: Variability of the near-surface eddy kinetic energy in California Current based
962 on altimetric, drifter, and moored current data. *J. Geophys. Res.*, **103**, 13,067-13,083.

963 Langland, R.H. and N.L. Baker, 2004: Estimation of observation impact using the NRL
964 atmospheric variational data assimilation adjoint system. *Tellus*, **56A**, 189-201.
965 Lawless, A.S., S. Gratton, and N.K. Nichols, 2005: Approximate iterative methods for vari-
966 ational data assimilation. Int. *J. Numer. Meth. Fl.*, **1**, 1-6.
967 Le Hénaff, M., P. De Mey and P. Marsaleix, 2009: Assessment of observational networks
968 with the representer matrix spectra method-application to a 3D coastal model of the
969 Bay of Biscay. *Ocean Dyn.*, **59**, 3-20.
970 Liu, W.T., K.B. Katsaros and J.A. Businger, 1979: Bulk parameterization of the air-sea
971 exchange of heat and water vapor including the molecular constraints at the interface.
972 *J. Atmos. Sci.*, **36**, 1722-1735.
973 Lorenc, A.C. and R.T. Marriott, 2014: Forecast senistivity to observations in the Met Office
974 global numerical weather prediction system. *Q. J. Roy. Meteorol. Soc.*, **140**, 209-224.
975 Moore, A.M., H.G. Arango, G. Broquet, C. Edwards, M. Veneziani, B. Powell. D. Fo-
976 ley, J. Doyle, D. Costa and P. Robinson, 2011a: The Regional Ocean Modeling Sys-
977 tem (ROMS) 4-dimensional variational data assimilation systems. Part III: Observa-
978 tion impact and observation sensitivity in the California Current System. *Progress in*
979 *Oceanography*, **91**, 74-94.
980 Moore, A.M., H.G. Arango, G. Broquet, B.S. Powell, J. Zavala-Garay and A.T. Weaver,
981 2011b: The Regional Ocean Modeling System (ROMS) 4-dimensional variational data
982 assimilation systems. Part I: System overview and formulation. *Progress in Oceanog-*
983 *raphy*, **91**, 34-49.
984 Moore, A.M., H.G. Arango, G. Broquet, C. Edwards, M. Veneziani, B. Powell. D. Foley,
985 J. Doyle, D. Costa and P. Robinson, 2011c: The Regional Ocean Modeling System

(ROMS) 4-dimensional variational data assimilation systems. Part II: Performance and application to the California Current System. *Progress in Oceanography*, **91**, 50-73.

Moore, A.M., C.A. Edwards, J. Fiechter and M.G. Jacox, 2015: Observing System Impacts on Estimates of California Current Transport. In "*Coastal Ocean Observing Systems*," Yonggang Liu, Heather Kekering and Robert Weisberg, Eds., Elsevier, Chapter 19, 351-369.

Neveu, E., A.M. Moore, C.A. Edwards, J. Fiechter, P. Drake, M.G. Jacox and E. Nuss, 2016: An historical analysis of the California Current using ROMS 4D-Var. Part I: System configuration and diagnostics. *Ocean Modelling*, **99**, 131-151.

Oke, P.R. and A. Schiller, 2007: Impact of Argo, SST, and altimeter data on an eddy resolving ocean reanalysis. *Geophys. Res. Lett*, **34**, L19601. http://dx.doi.org/10.1029/2007GL031549.

Oke, P.R. and P. Sakov, 2012: Assessing the footprint of a regional ocean observing system. *J. Mar. Syst.*, **105**, 30-51. http://dx.doi.org/10.1016/j.jmarsys.2012.05.009.

Oke, P.R., G. Larnicol, Y. Fujii, G.C. Smith, D.J. Lea, S. Guinehut, et al., 2015: Assessing the impact of observations on ocean forecasts and reanalyses: part 1, global studies. *J. Oper. Oceanogr.*, **8**, s49-s62.

Oke, P.R., G. Larnicol, E.M. Jones, V. Kourafalou, A.K. Sperrevik, F. Carse, et al., 2015: Assessing the impact of observations on ocean forecasts and reanalyses: part 2, regional applications. *J. Oper. Oceanogr.*, **8**, s63-s79.

Pedlosky, J., 1979: Geophysical Fluid Dynamics. Springer-Verlag, pp710.

Pierce, S.D., R.L. Smith, P.M. Kosro, J.A. Barth and C.D. Wilson, 2000: Continuity of the poleward undercurrent along the eastern boundary of the mid-latitude north Pacific.

Deep Sea Res. II, **47**, 811-829.

Saraceno, M., P.T. Strub and P.M. Kosro, 2008. Estimates of sea surface height and near surface alongshore coastal currents from combinations of altimeters and tide gauges. *J. Geophys. Res.*, **113**, C11013, doi:10.1029/2008JC004756.

Smith, G.C. and K. Haines, 2009: Evaluation of the S(T) assimilation method with the Argo dataset. *Q. J. R. Meteorol. Soc.*, **135**, 739-756.

Storto, A., S. Masina and S. Dobricic, 2013: Ensemble spread-based assessment of observation impact: application to a global ocean analysis. *Q. J. R. Meteorol. Soc.*, **139**, 1842-1862.

Trémolet, Y., 2008: Computation of observation sensitivity and observation impact in incremental variational data assimilation. *Tellus*, **60A**, 964-978.

Tyndall, D.P. and J.D. Horel, 2013: Impacts of Mesonet observations on meteorological surface analyses. *Wea. and Forecasting*, **28**, 254-269.

von Storch, H., G. Bürger, R. Schnur and J.-S. von Storch, 1995: Principal oscillation patterns: A review. *J. Climate*, **8**, 377-400.

Weaver, A.T. and P. Courtier, 2001: Correlation modelling on the sphere using a generalized diffusion equation. *Q. J. R. Meteorol. Soc.*, **127**, 1815-1846.

Veneziani, M., C.A. Edwards, J.D. Doyle and D. Foley, 2009a: A central California coastal ocean modeling study: 1. Forward model and the influence of realistic versus climatological forcing. *J. Geophys. Res.*, **114**, C04015, doi:10.1029/2008JC004774.

Figure Captions

Figure 1: (a) The ROMS CCS model domain and bathymetry. Also shown is a schematic representation of some important dynamical features of the circulation in the region. The central and northern CCS target regions for upwelling and thermocline depth referred to in the main text are shown (black lines), as well as the target region for eddy kinetic energy (white dashed line) and the 37N section. (b) A zoom of the northern CCS region showing the 100m and 500m isobaths used to define the target region for the undercurrent transport.

Figure 2: Time series of $\log_{10}$ of the number of observations from each observing platform during each 8 day data assimilation cycle of WCRA31. AVHRR (dark blue), Aviso (red), MODIS (light blue), AMSR (purple), SST super observations (green) and in situ hydrographic observations (black).

Figure 3: (a) The annual mean meridional velocity component on σ-level 10 computed from WCRA31. (b) A vertical section of the Mar-May average meridional velocity along 36°N showing the mean structure of the CUC and the target depths (dashed gray lines) used to define $\mathcal{I}_{cuc}$. The thin black line denotes s-level 10.

Figure 4: Time series of each circulation metric from WCRA31 (blue) and from the forward run without data assimilation (red; not to be confused with the background) for (a) $\mathcal{I}^C_{cuc}$, (b) $\mathcal{I}^C_{up}$, (c) $\mathcal{I}^C_{26}$, (d) $\mathcal{I}_{EKE}$ and (e) $\mathcal{I}_{37N}$, where superscrip C denotes the metrics for the central CCS region. Each sub-panel to the right shows the mean seasonal cycle for each metric from WCRA31 (blue) and from the forward run (red).

Figure 5: Time series of the increments of each circulation metric computed using the tangent

linear assumption (blue) and directly from xa-xb (red) for (a) $\Delta\mathcal{I}^C_{cuc}$, (b) $\Delta\mathcal{I}^C_{up}$, (c) $\Delta\mathcal{I}^C_{26}$, (d) $\Delta\mathcal{I}_{EKE}$ and (e) $\Delta\mathcal{I}_{37N}$. The correlation r between the two time series is indicated in each case.

Figure 6: The rms annual mean impact of each component of the control vector on (a) $\Delta\mathcal{I}^C_{cuc}$, (b) $\Delta\mathcal{I}^C_{up}$, (c) $\Delta\mathcal{I}^C_{26}$, (d) $\Delta\mathcal{I}_{EKE}$, and (e) $\Delta\mathcal{I}_{37N}$. Initial conditions (red), surface forcing (green) and open boundary conditions (blue).

Figure 7: The impact of each component of the control vector during each 4D-Var cycle of 2005 for (a) $\Delta\mathcal{I}^C_{cuc}$, (b) $\Delta\mathcal{I}^C_{up}$, (c) $\Delta\mathcal{I}^C_{26}$, (d) $\Delta\mathcal{I}_{EKE}$, and (e) $\Delta\mathcal{I}_{37N}$. Initial conditions (red), surface forcing (green) and open boundary conditions (blue).

Figure 8: A scatter plot of the innovation vs impact on $\Delta\mathcal{I}^C_{up}$, for observations of (a) SSH, (b) temperature, and (c) salinity.

Figure 9: (a) Time series of the rms annual mean impact of each observing platform on $\Delta\mathcal{I}^C_{up}$. (b) Time series of $\log_{10}$ of the rms annual average impact per datum of each observing platform on . $\Delta\mathcal{I}^C_{up}$. (c) Time series of the absolute value of the cumulative increment per year in $\Delta\mathcal{I}^C_{up}$ associated with each observing platform. The black curve shows the total.

Figure 10: The impact of each observing platform during each 4D-Var cycle of 2005 for (a) $\Delta\mathcal{I}^C_{cuc}$, (b) $\Delta\mathcal{I}^C_{up}$, (c) $\Delta\mathcal{I}^C_{26}$, (d) $\Delta\mathcal{I}_{EKE}$ and (e) $\Delta\mathcal{I}_{37N}$.

Figure 11: Log_{10} of the rms impact per grid cell of AVHRR on (a) $\Delta\mathcal{I}^C_{up}$ and (b) $\Delta\mathcal{I}^N_{up}$, MODIS on (c) $\Delta\mathcal{I}^C_{up}$ and (d) $\Delta\mathcal{I}^N_{up}$, and Aviso on (e) $\Delta\mathcal{I}^C_{up}$ and (f) $\Delta\mathcal{I}^N_{up}$. The units are $\log_{10}$(Sv). The central and northern CCS target areas are also shown.

Figure 12: Log_{10} of the rms impact per grid cell of in situ temperature observations on (a) $\Delta\mathcal{I}^C_{up}$ and (b) $\Delta\mathcal{I}^N_{up}$, and in situ salinity observations on (c) $\Delta\mathcal{I}^C_{up}$ and (d) $\Delta\mathcal{I}^N_{up}$. The units

are $\log_{10}$(Sv). The central and northern CCS target areas are also shown.

Figure 13: Vertical profiles of the average impact of in situ observations of temperature (black) and salinity (red) for (a) $\Delta\mathcal{I}^C_{up}$ and (b) $\Delta\mathcal{I}^C_{cuc}$.

Figure 14: Time series of $|\mathcal{I}_{pop}|$ (blue) and $|\Delta\mathcal{I}_{pop}|$ (red) for (a) POP1, (b) POP2 and (c) POP3.

Figure 15: Log_{10} of the rms impact per grid cell of AVHRR on (a) $\text{Re}(\Delta\mathcal{I}_{pop_1})$, (b) $\text{IM}(\Delta\mathcal{I}_{pop_1})$, (c) $\text{Re}(\Delta\mathcal{I}_{pop_2})$, (d) $\text{IM}(\Delta\mathcal{I}_{pop_2})$, (e) $\text{Re}(\Delta\mathcal{I}_{pop_3})$, and (f) $\text{IM}(\Delta\mathcal{I}_{pop_3})$. The units are $\log_{10}$(Sv).

Figure 16: Vertical profiles of the average impact of in situ observations of temperature (black) and salinity (red) for (a) $\text{Re}(\Delta\mathcal{I}_{pop_3})$ and (b) $\text{Im}(\Delta\mathcal{I}_{pop_3})$.

Figure 17: Time series of $c_1(t)$ for (a) coastal upwelling and (b) EKE, and for $c_2(t)$ (c) and (d), and $c_3(t)$ (e) and (f) for each observing platform: satellite SST (blue), Aviso (red), in situ temperature (black), and in situ salinity (green). The total for all platforms is also shown (dashed).

Figure 18: Hovmoller plots of $f_3(t,\tau)$ for central CA upwelling where t is time and τ is the observation time for (a) AVHRR SST, (b) MODIS SST and (b) Aviso SSH.

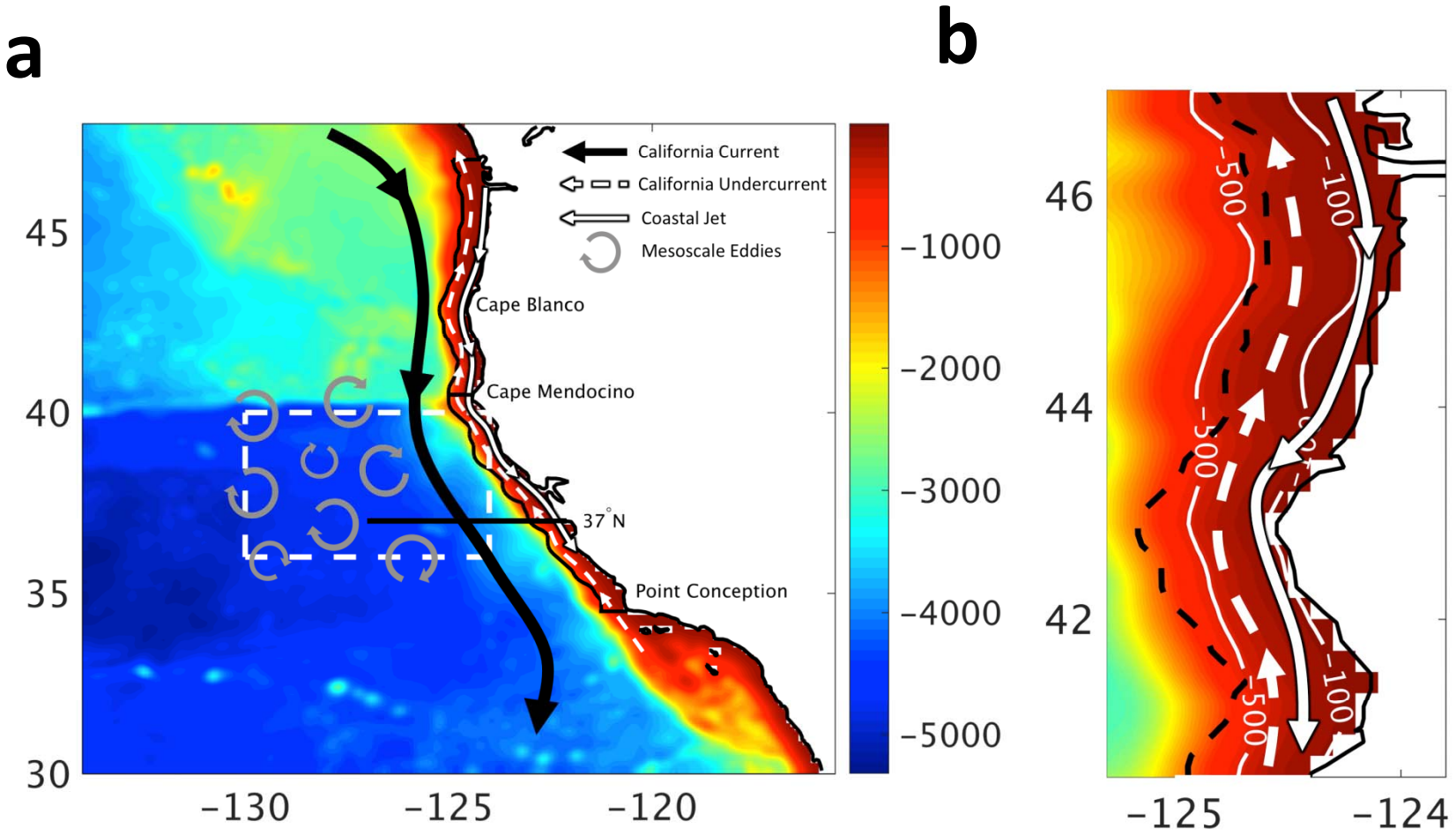

Figure 1: (a) The ROMS CCS model domain and bathymetry. Also shown is a schematic representation of some important dynamical features of the circulation in the region. The central and northern CCS target regions for upwelling and thermocline depth referred to in the main text are shown (black lines), as well as the target region for eddy kinetic energy (white dashed line) and the 37°N section. (b) A zoom of the northern CCS region showing the 100m and 500m isobaths used to define the target region for the undercurrent transport.

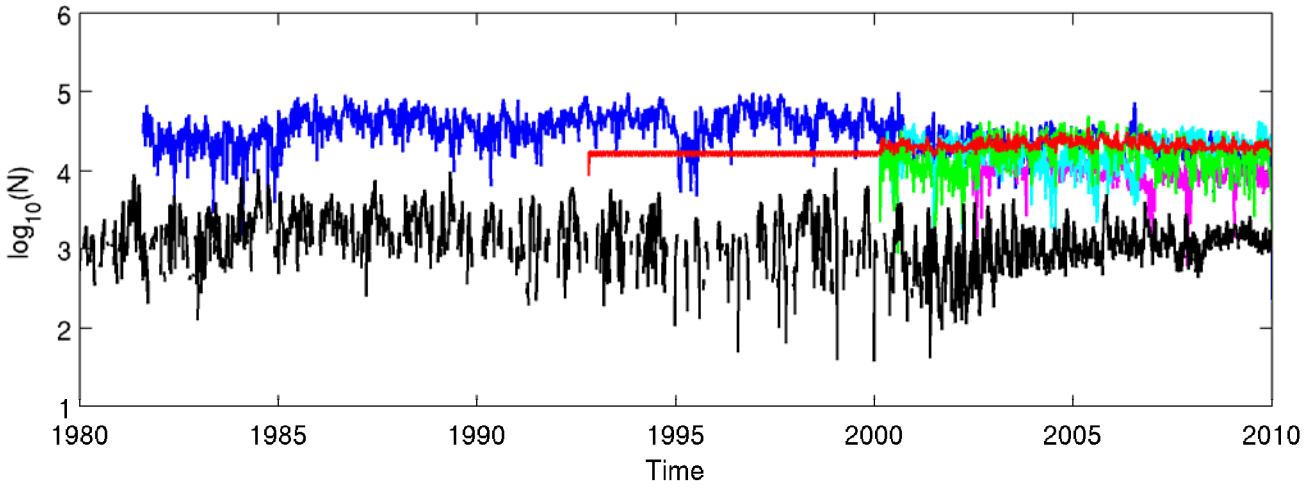

Figure 2: Time series of $\log_{10}$ of the number of observations from each observing platform during each 8 day data assimilation cycle of WCRA31. AVHRR (dark blue), Aviso (red), MODIS (light blue), AMSR (purple), SST super observations (green) and in situ hydrographic observations (black).

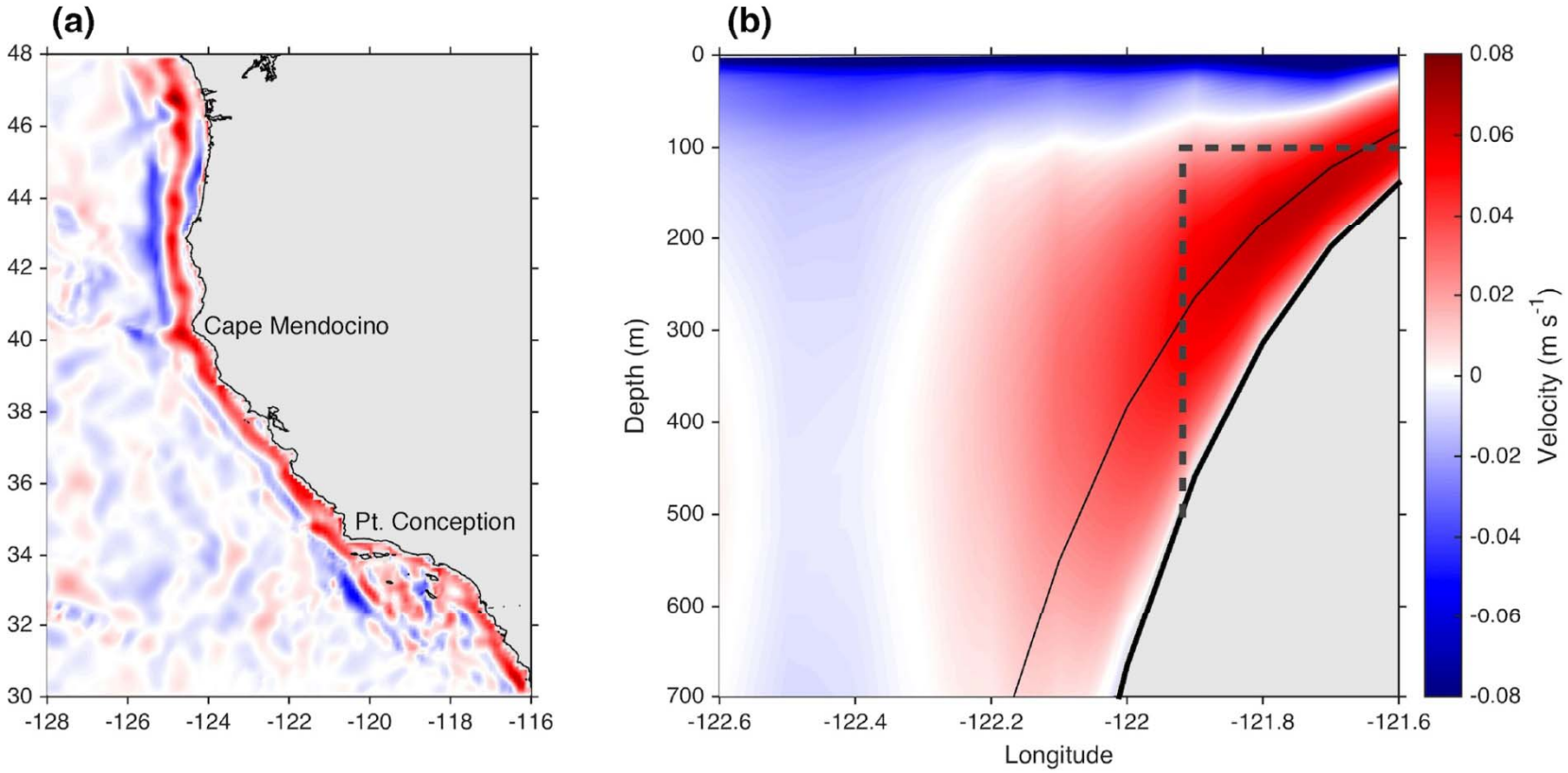

Figure 3: (a) The annual mean meridional velocity component on s-level 10 computed from WCRA31. (b) A vertical section of the Mar-May average meridional velocity along 36°N showing the mean structure of the CUC and the target depths (dashed gray lines) used to define $\mathcal{I}_{cuc}$. The thin black line denotes s-level 10.

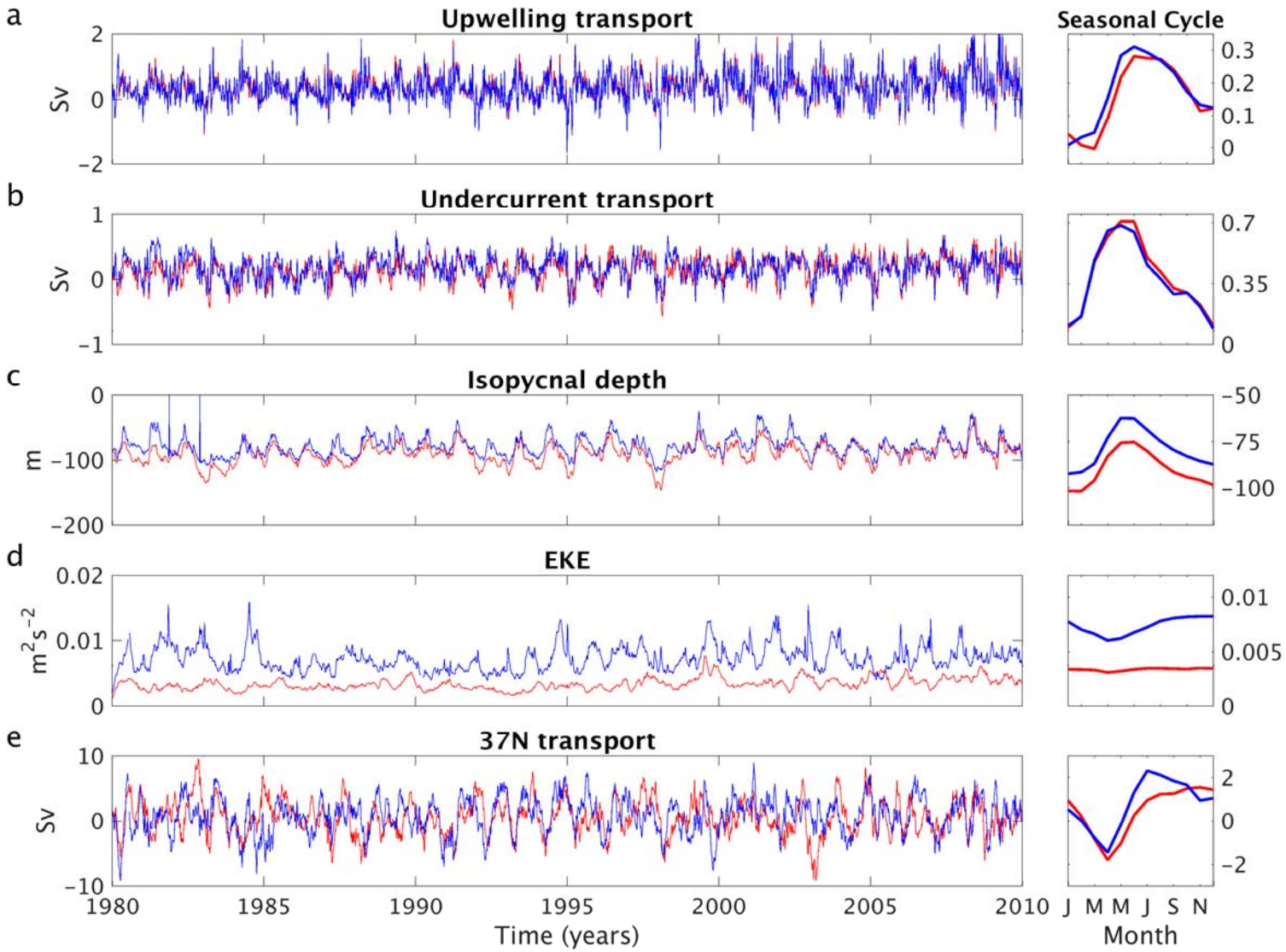

Figure 4: Time series of each circulation metric from WCRA31 (blue) and from the forward run without data assimilation (red) for (a) $\mathcal{I}^{C}_{cuc}$, (b) $\mathcal{I}^{C}_{up}$, (c) $\mathcal{I}^{C}_{26}$, (d) $\mathcal{I}_{EKE}$ and (e) $\mathcal{I}_{37N}$. Each sub-panel to the right shows the mean seasonal cycle of each metric from WCRA31 (blue) and from the forward run (red).

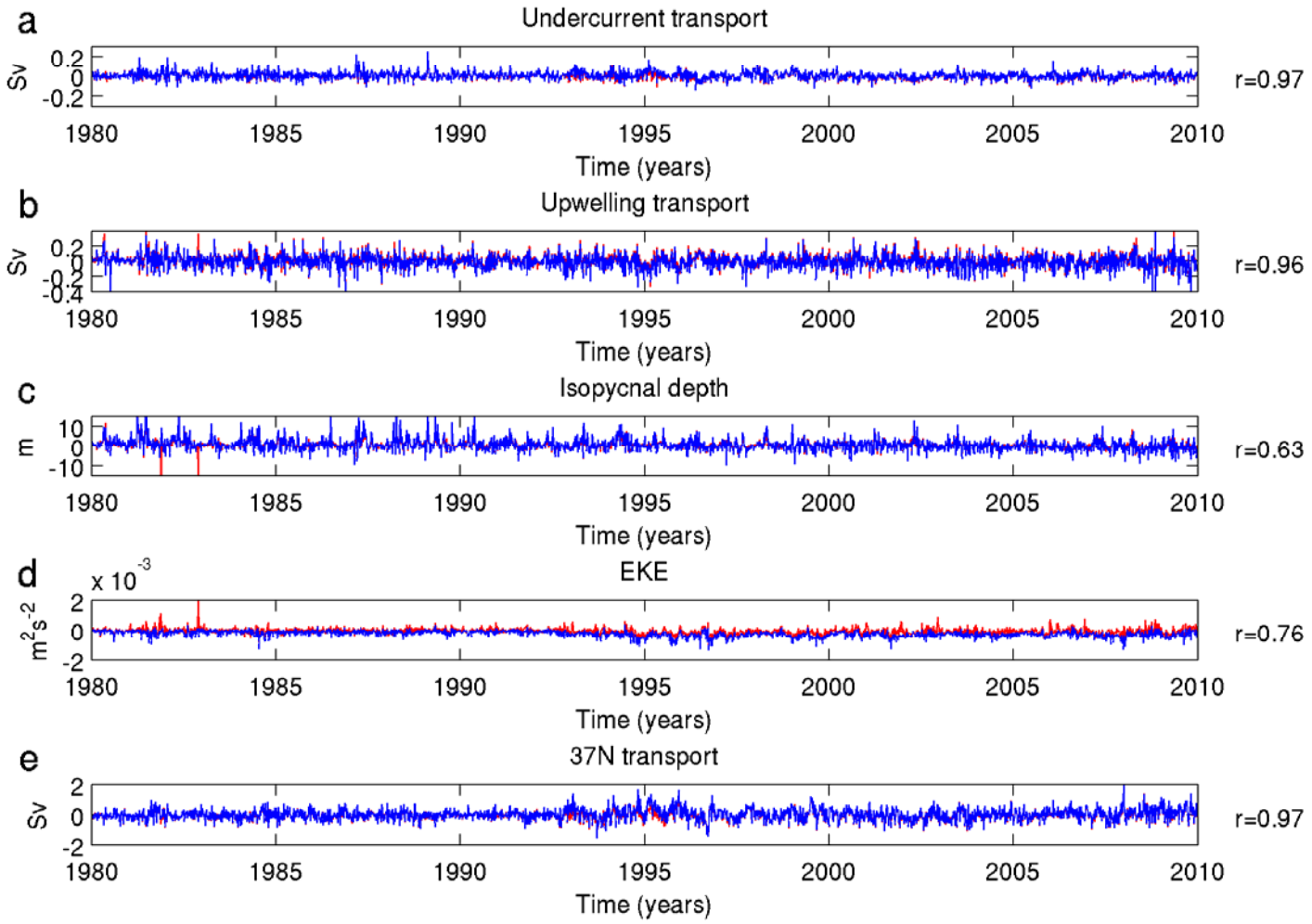

Figure 5: Time series of the increments of each circulation metric computed using the tangent linear assumption (blue) and directly from $\boldsymbol{x}^a$-$\boldsymbol{x}^b$ (red) for (a) $\Delta\mathcal{I}^C_{cuc}$, (b) $\Delta\mathcal{I}^C_{up}$, (c) $\Delta\mathcal{I}^C_{26}$, (d) $\Delta\mathcal{I}^C_{EKE}$ and (e) $\Delta\mathcal{I}^C_{37N}$. The correlation, r, between the two time series is indicated in each case.

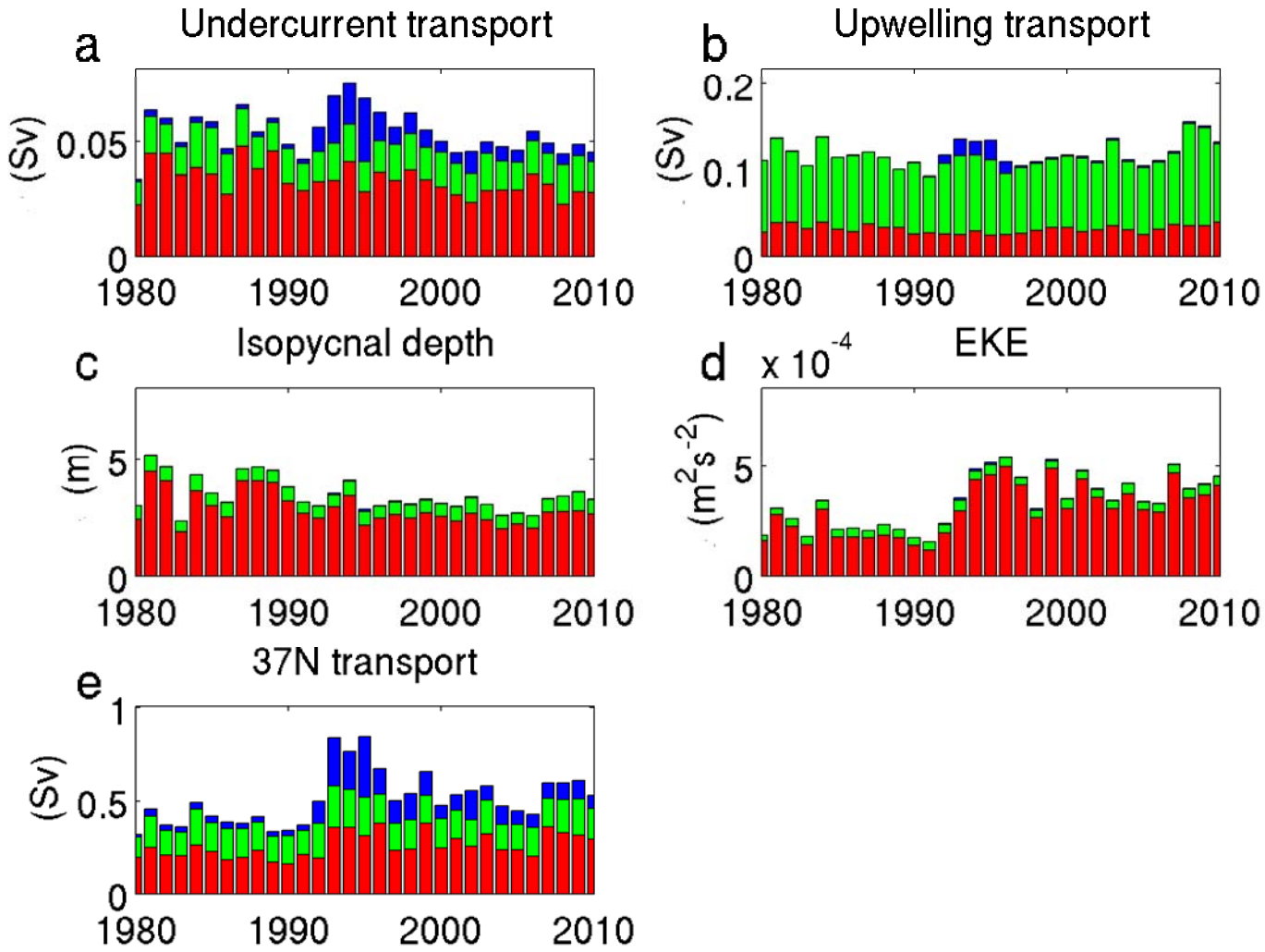

Figure 6: The rms annual mean impact of each component of the control vector on (a) $\Delta\mathcal{I}^{C}_{cuc}$, (b) $\Delta\mathcal{I}^{C}_{up}$, (c) $\Delta\mathcal{I}^{C}_{26}$, (d) $\Delta\mathcal{I}_{EKE}$, and (e) $\Delta\mathcal{I}_{37N}$. Initial conditions (red), surface forcing (green) and open boundary conditions (blue).

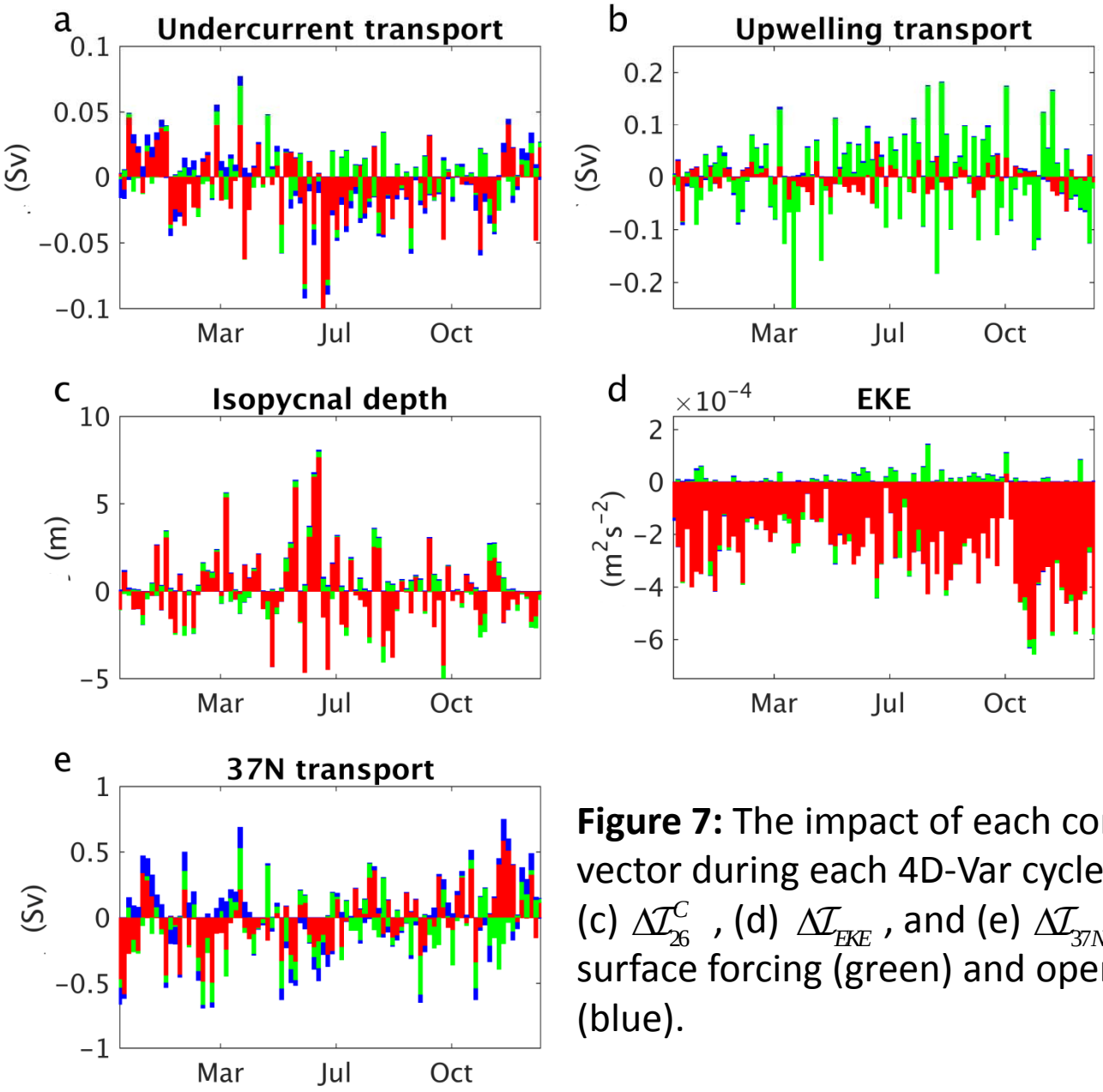

Figure 7: The impact of each component of the control vector during each 4D-Var cycle of 2005 for (a) $\Delta\mathcal{I}^{C}_{alc}$, (b) $\Delta\mathcal{I}^{C}_{up}$ (c) $\Delta\mathcal{I}^{C}_{26}$, (d) $\Delta\mathcal{I}_{EKE}$, and (e) $\Delta\mathcal{I}_{37N}$. Initial conditions (red), surface forcing (green) and open boundary conditions (blue).

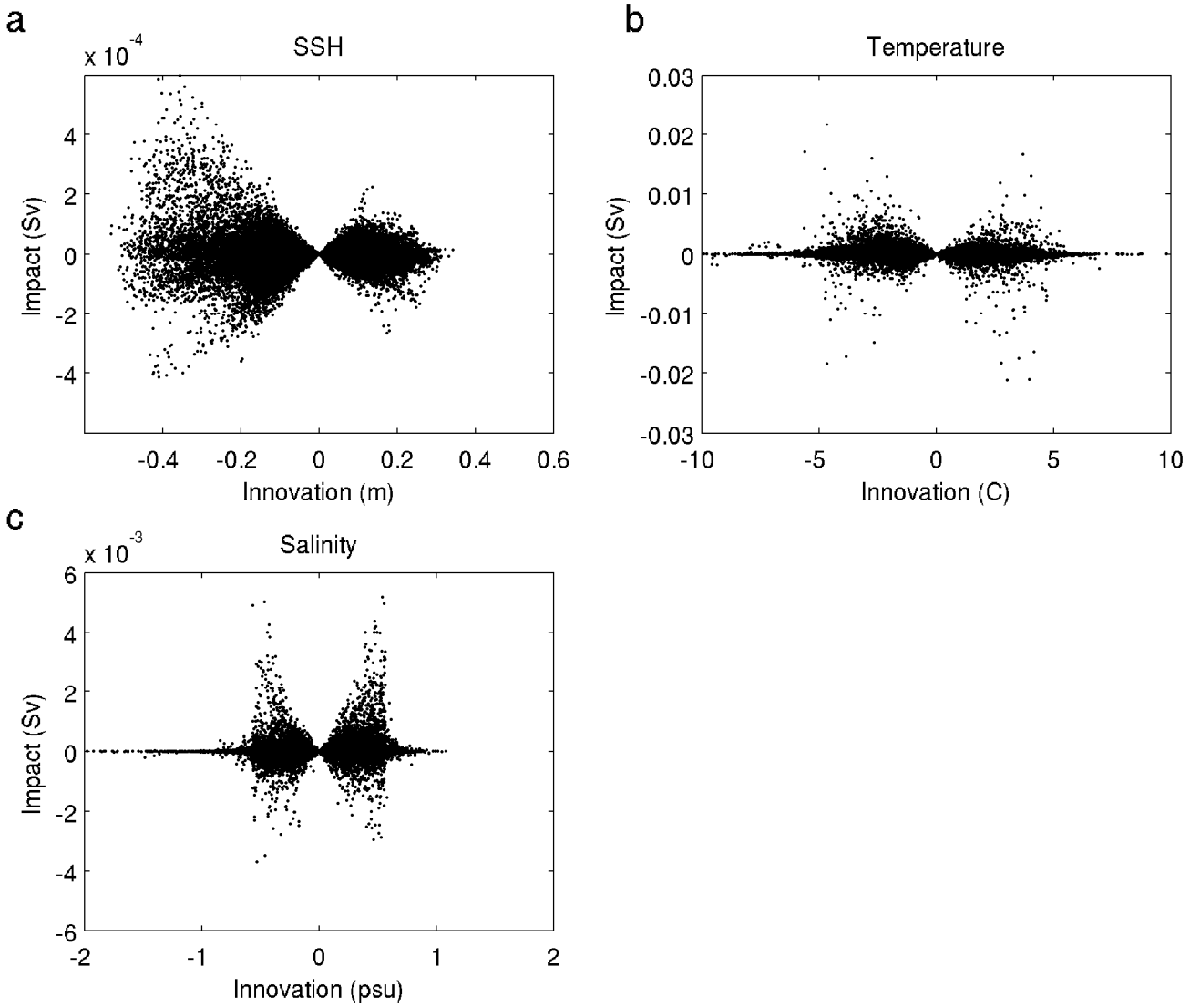

Figure 8: A scatter plot of the innovation vs impact on $\Delta\mathcal{I}_{up}^{C}$, for observations of (a) SSH, (b) temperature, and (c) salinity.

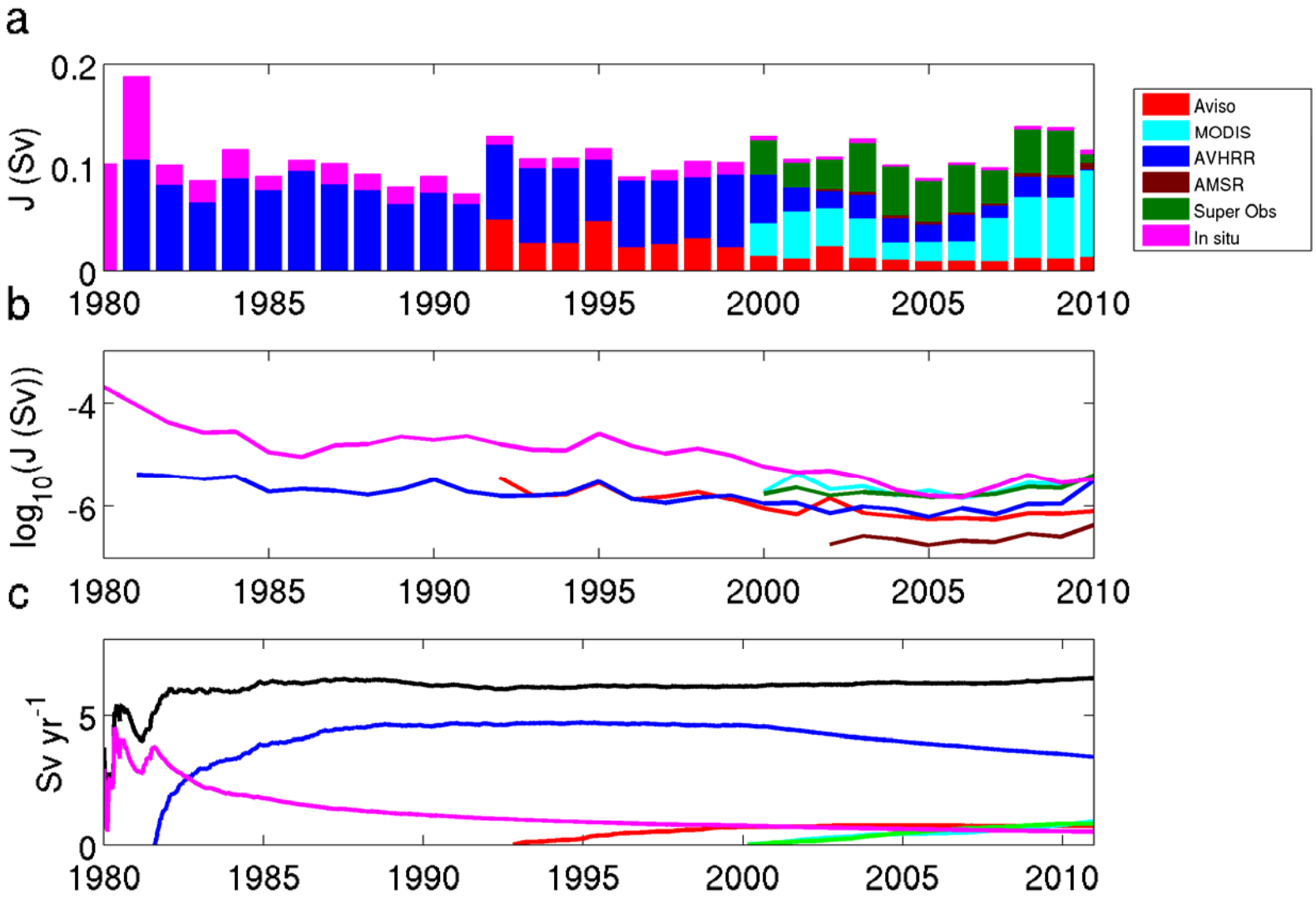

Figure 9: (a) Time series of the rms annual mean impact of each observing platform on $\Delta\mathcal{I}^C_{up}$. (b) Time series of $\log_{10}$ of the rms annual average impact per datum of each observing platform on $\Delta\mathcal{I}^C_{up}$. (c) Time series of the absolute value of the cumulative increment per year in $\Delta\mathcal{I}^C_{up}$ associated with each observing platform. The black line is the total.

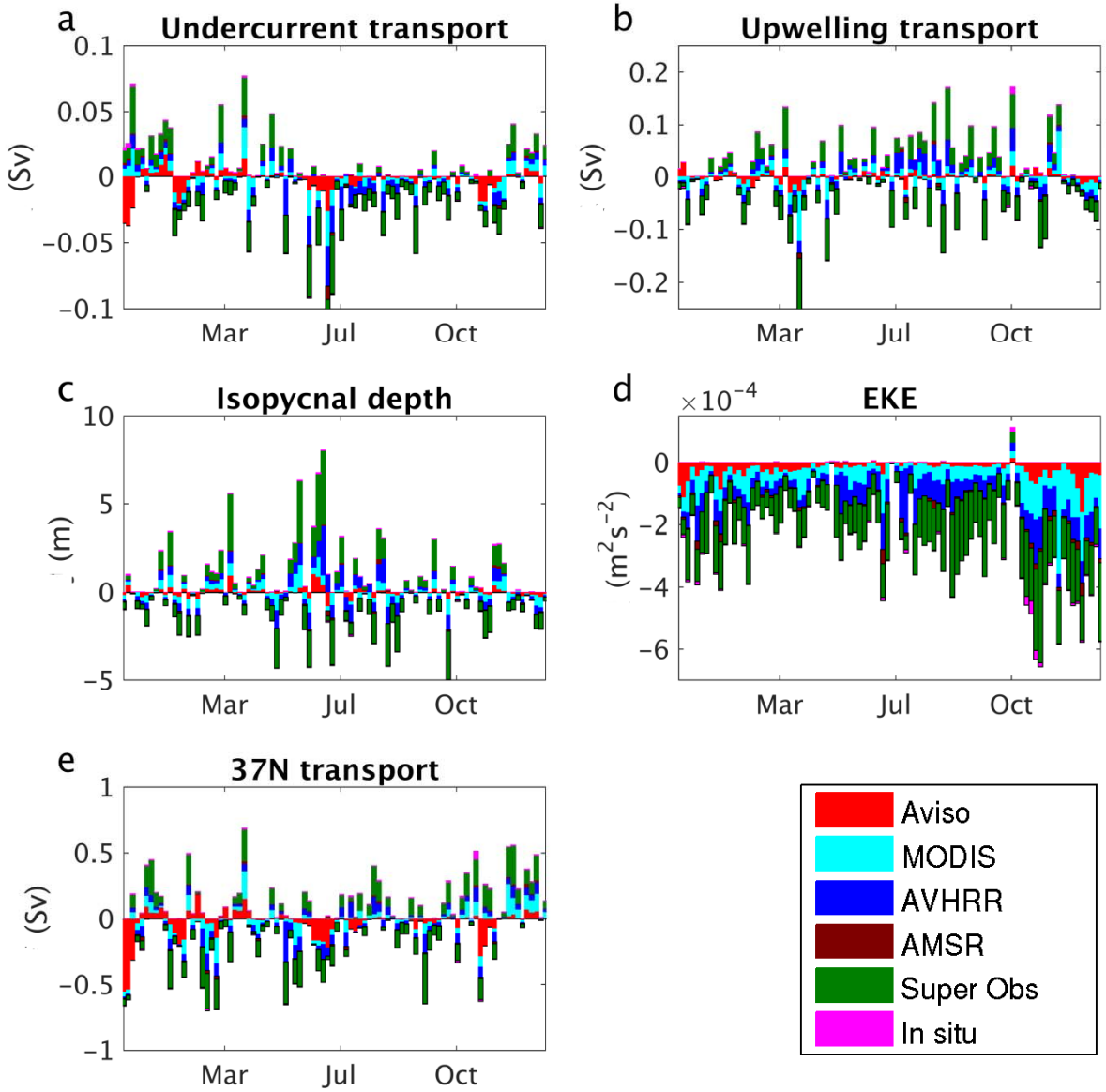

Figure 10: The impact of each observing platform during each 4D-Var cycle of 2005 for (a) $\Delta\mathcal{I}^{C}_{cuc}$, (b) $\Delta\mathcal{I}^{C}_{up}$, (c) $\Delta\mathcal{I}^{C}_{26}$, (d) $\Delta\mathcal{I}_{EKE}$ and (e) $\Delta\mathcal{I}_{37N}$.

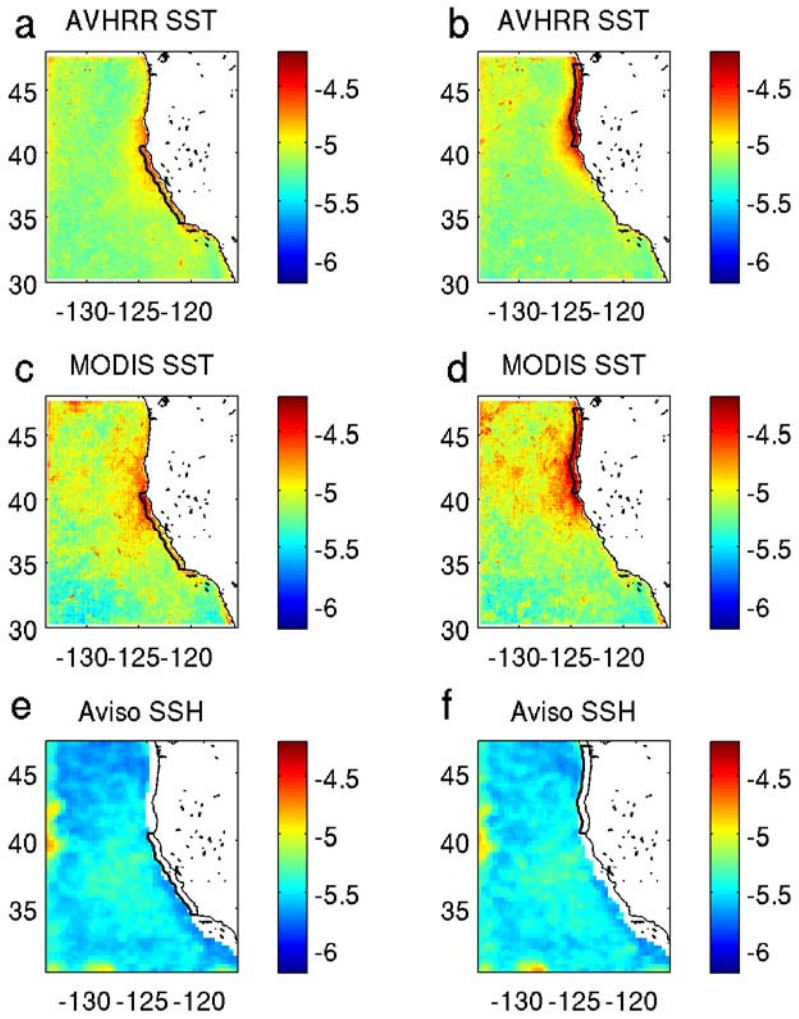

Figure 11: Log_{10} of the rms impact per grid cell of AVHRR on (a) $\Delta\mathcal{I}_{up}^{C}$ and (b) $\Delta\mathcal{I}_{up}^{N}$, MODIS on (c) $\Delta\mathcal{I}_{up}^{C}$ and (d) $\Delta\mathcal{I}_{up}^{N}$, and Aviso on (e) $\Delta\mathcal{I}_{up}^{C}$ and (f) $\Delta\mathcal{I}_{up}^{N}$. The units are $\log_{10}$(Sv). The central and northern CCS target areas are also shown.

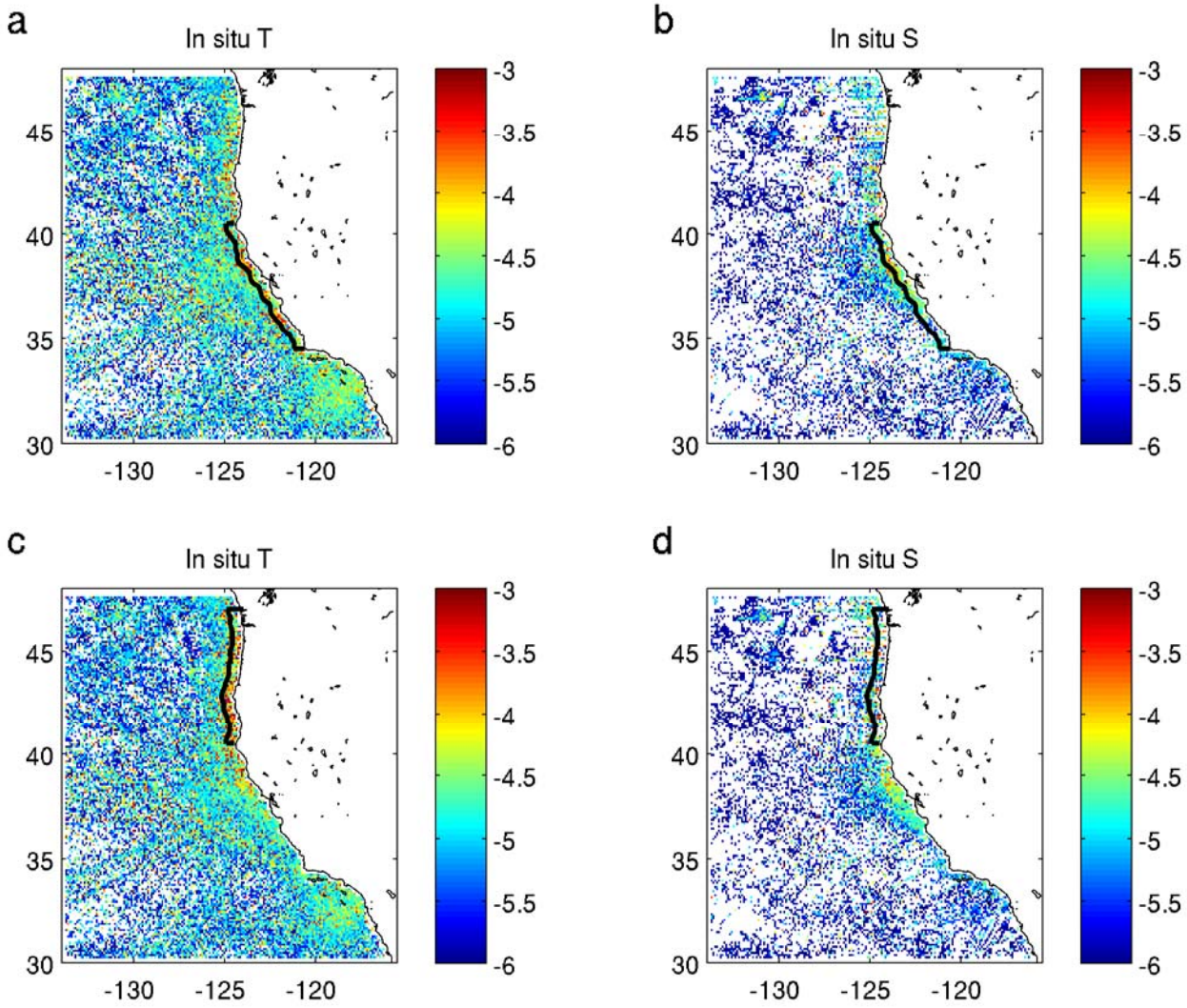

Figure 12: Log_{10} of the rms impact per grid cell of in situ temperature observations on (a) $\Delta\mathcal{I}_{up}^{C}$ and (b) $\Delta\mathcal{I}_{up}^{N}$, and in situ salinity observations on (c) $\Delta\mathcal{I}_{up}^{C}$ and (d) $\Delta\mathcal{I}_{up}^{N}$. The units are $\log_{10}$(Sv). The central and northern CCS target areas are also shown.

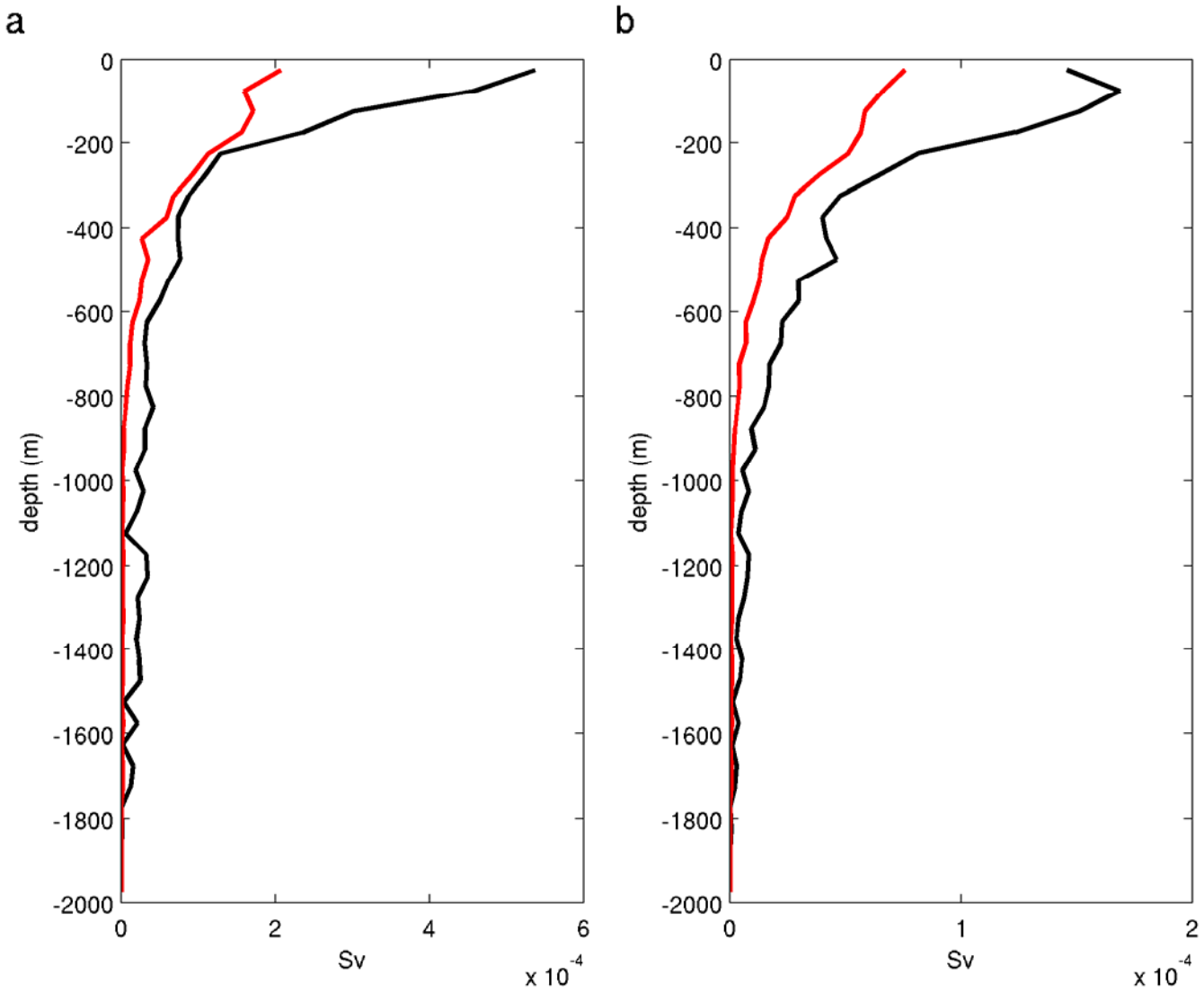

Figure 13: Vertical profiles of the average impact of in situ observations of temperature (black) and salinity (red) for (a) $\Delta\mathcal{I}^{C}_{up}$ and (b) $\Delta\mathcal{I}^{C}_{cuc}$.

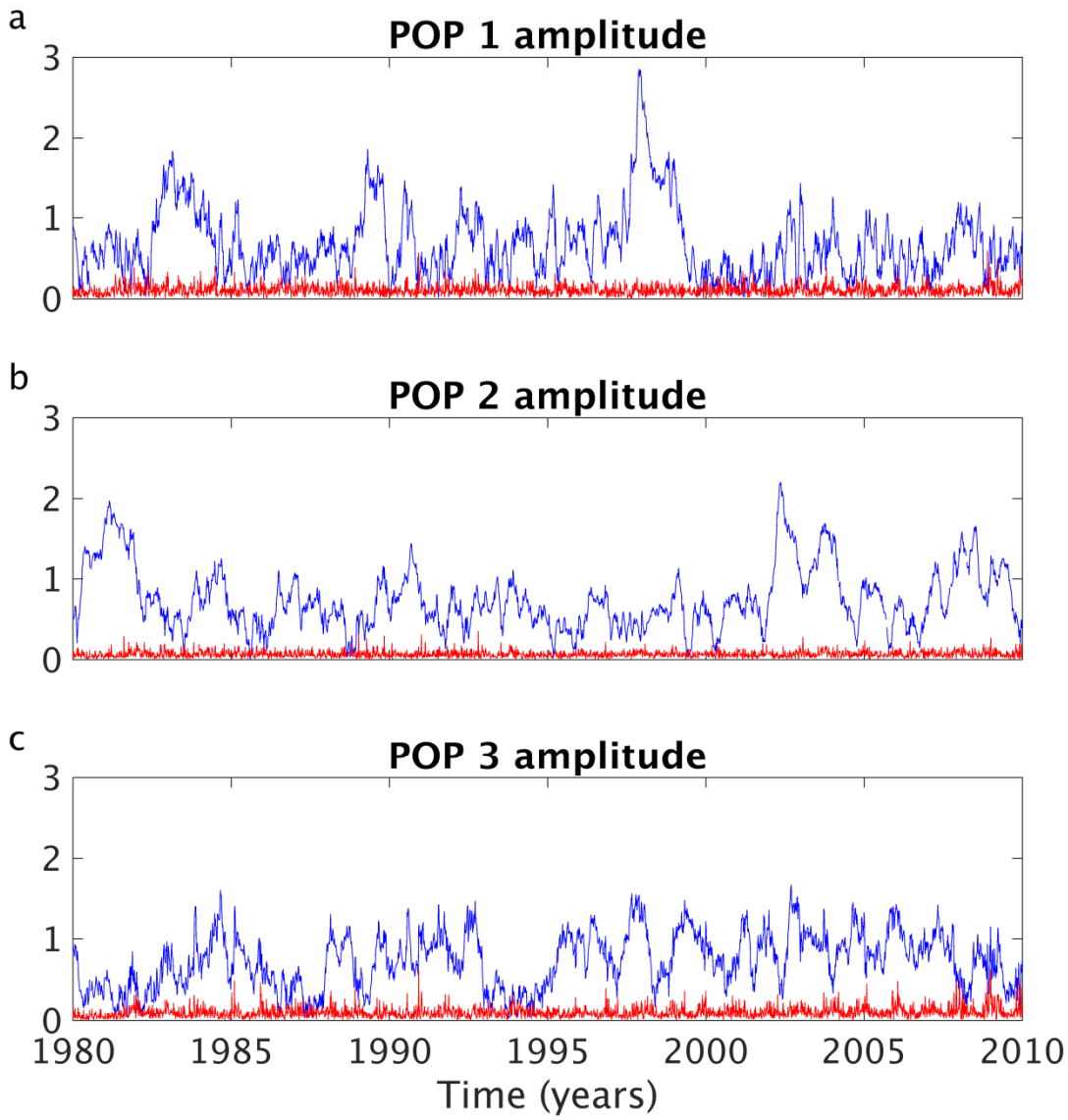

Figure 14: Time series of $\left|\mathcal{I}_{pop}\right|$ (blue) and $\left|\Delta\mathcal{I}_{pop}\right|$ (red) for (a) POP1, (b) POP2 and (c) POP3.

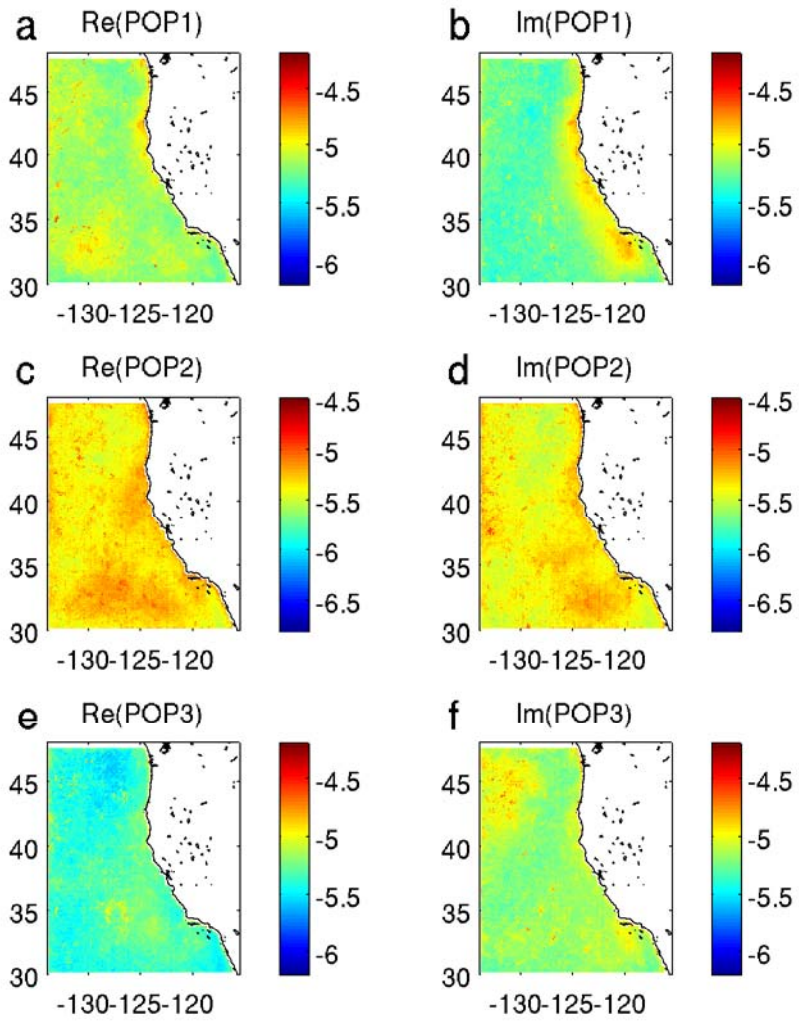

Figure 15: Log_{10} of the rms impact per grid cell of AVHRR on (a) $\mathrm{Re}\left(\Delta\mathcal{I}_{pop_1}\right)$, (b) $\mathrm{Im}\left(\Delta\mathcal{I}_{pop_1}\right)$, (c) $\mathrm{Re}\left(\Delta\mathcal{I}_{pop_2}\right)$, (d) $\mathrm{Im}\left(\Delta\mathcal{I}_{pop_2}\right)$, (e) $\mathrm{Re}\left(\Delta\mathcal{I}_{pop_3}\right)$, and (f) $\mathrm{Im}\left(\Delta\mathcal{I}_{pop_3}\right)$. The units are $\log_{10}$(Sv).

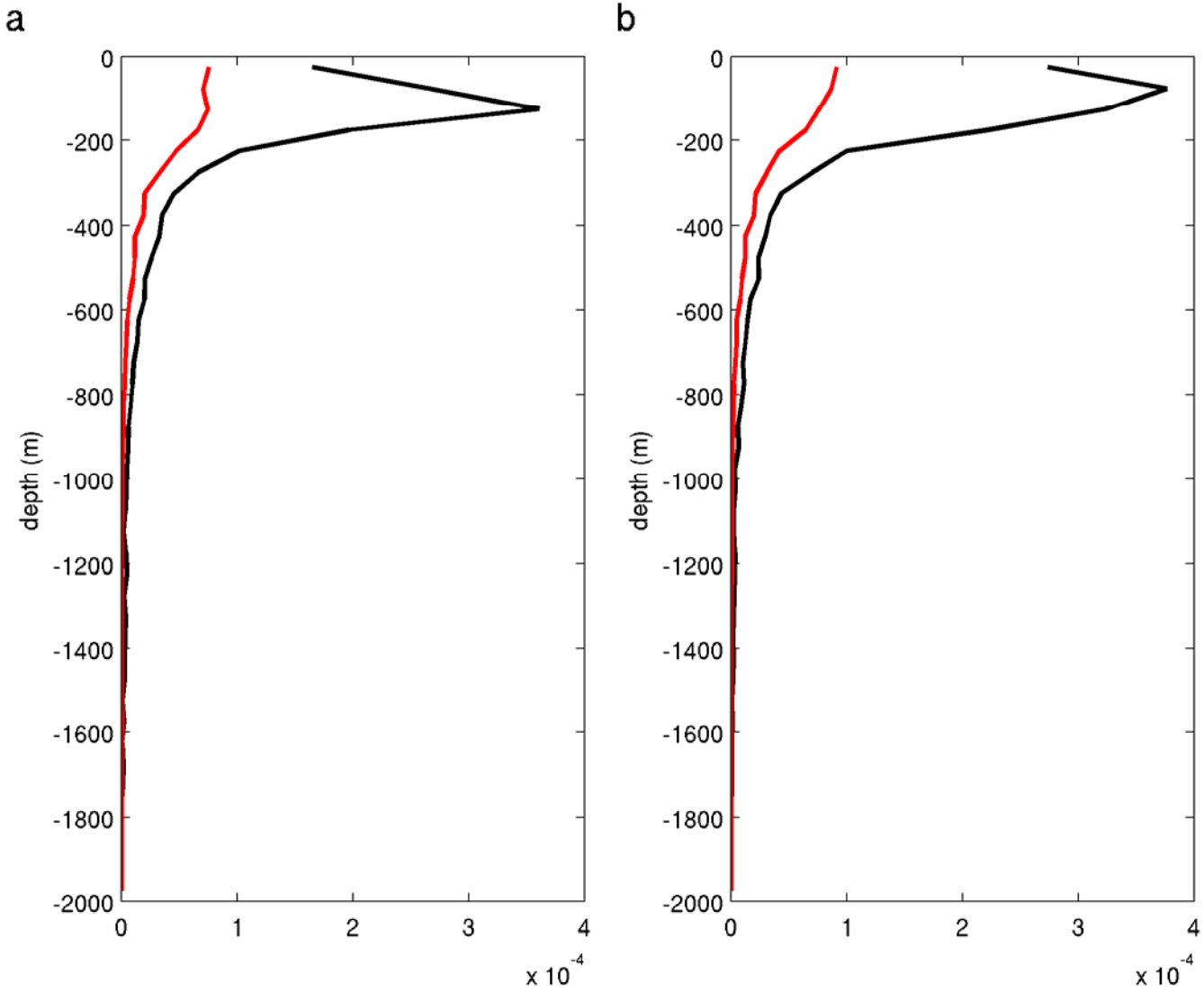

Figure 16: Vertical profiles of the average impact of in situ observations of temperature (black) and salinity (red) for (a) $\mathrm{Re}\left(\Delta\mathcal{I}_{pop_3}\right)$ and (b) $\mathrm{Im}\left(\Delta\mathcal{I}_{pop_3}\right)$.

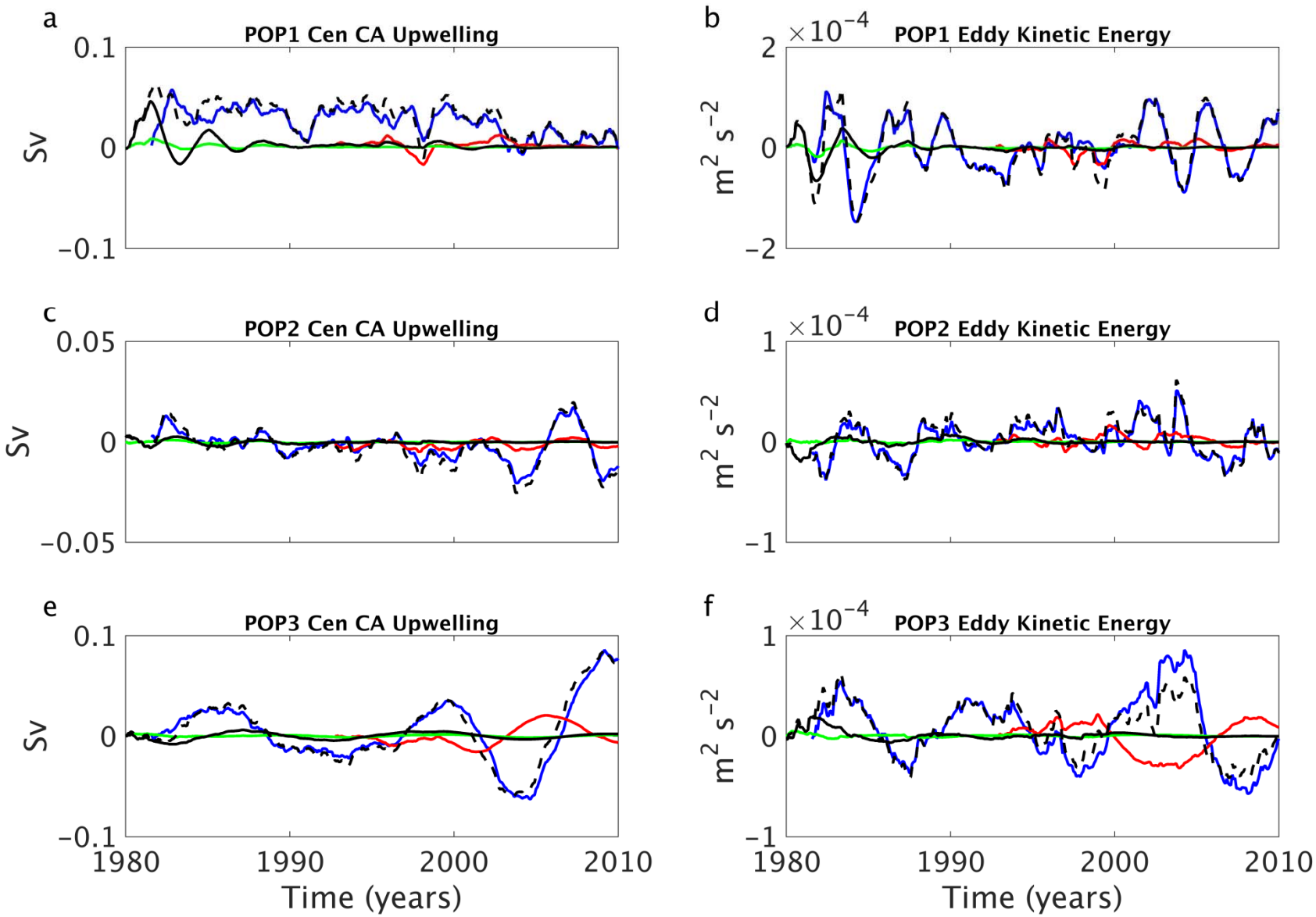

Figure 17: Time series of $c_1(t)$ for (a) coastal upwelling and (b) EKE, and for $c_2(t)$ (c) and (d), and $c_3(t)$ (e) and (f) for each observing platform: satellite SST (blue), Aviso (red), in situ temperature (black), and in situ salinity (green). The total for all platforms is also shown (dashed).

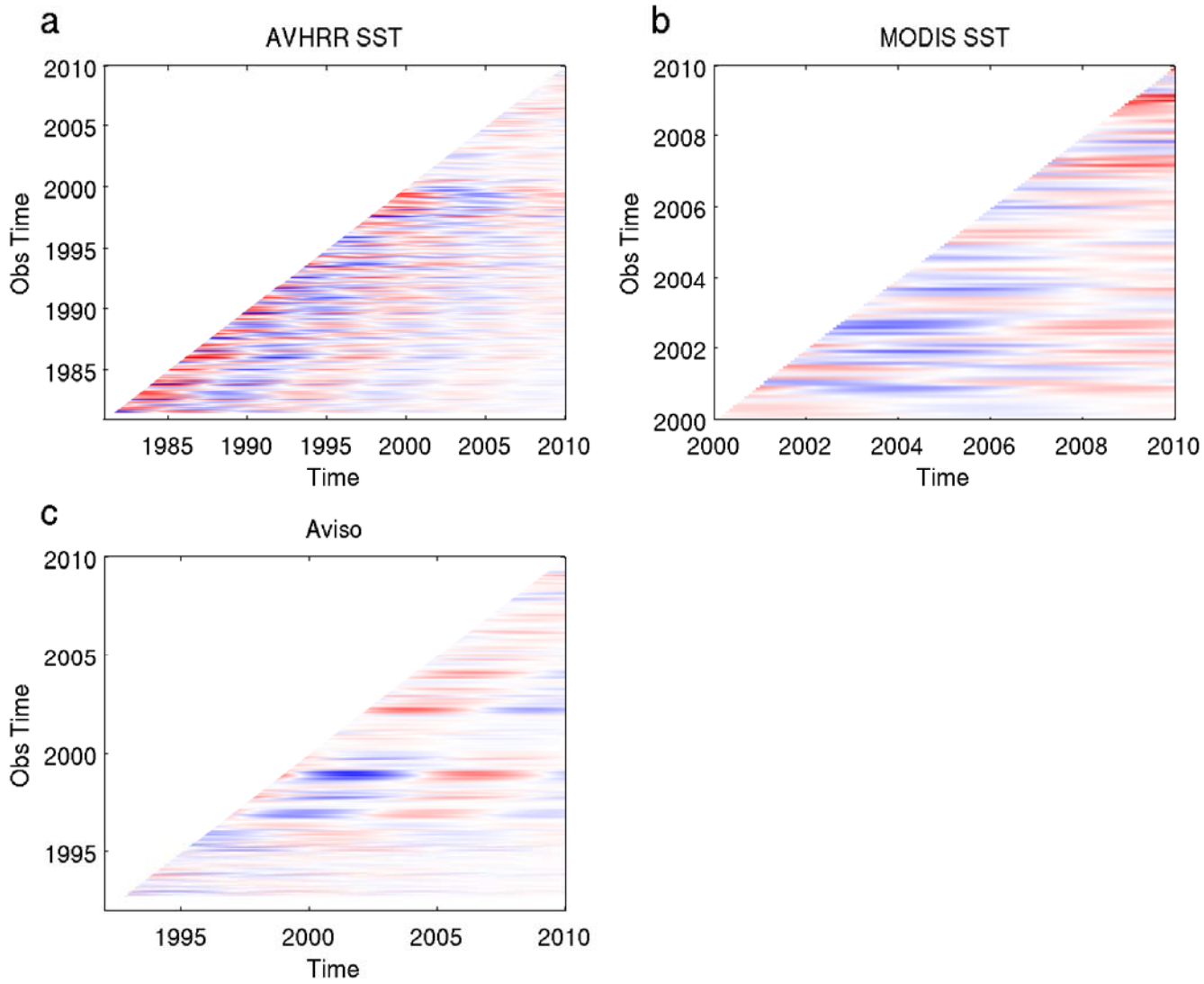

Figure 18: Hovmoller plots of $f_3(t,\tau)$ for central CA upwelling where t is time and τ is the observation time for (a) AVHRR SST, (b) MODIS SST and (b) Aviso SSH.