Observation Impact Notes for OTT

Analysis Cycle Obs Impacts

The methodology for computing the observation impacts for the analysis cycle is discussed here. The observation impacts of the forecast cycle are a bit more complicated and require higher-order approximations, but nevertheless are straightforward. Forecast cycle impacts will be covered in a separate document once we have the forecast cycle impacts driver in place. That said then, let's begin by defining a metric of the circulation, I. The metrics used in the examples provided take one of two possible forms:

$$I = \boldsymbol{h}^T \boldsymbol{x} \tag{1}$$
$$I = \boldsymbol{x}^T \boldsymbol{E} \boldsymbol{x} \tag{2}$$

where, as usual, $\boldsymbol{x}$ denotes the ROMS state-vector. The metric in (1) is linear (e.g. volume transport), while the metric in (2) is quadratic, and hence non-linear (e.g. kinetic energy). The vector $\boldsymbol{h}$ and the matrix $\boldsymbol{E}$ dictate the quantity of interest. One can of course define higher-order functions of $\boldsymbol{x}$ but they are more unusual, and they cannot be expressed in a general form like (1) and (2). Nonetheless, higher-order functions *can* be used in obs impact calculations if desired – we just need to express the math in terms of functions rather matrices and vectors.

In all of the examples provided, the metrics are defined in terms of a time average over the analysis cycle, so it is more convenient to express (1) and (2) as:

$$I = 1/N \sum_{l=1}^{N} \boldsymbol{h}_l^T \boldsymbol{x}_l \tag{3}$$
$$I = 1/N \sum_{l=1}^{N} \boldsymbol{x}_l^T \boldsymbol{E}_l \boldsymbol{x}_l \tag{4}$$

where the subscript l refers to time, and N is the number of times. More specifically, $l=1,N$ could refer to the number of time steps during the analysis cycle (not recommended!) or it could refer to the number of records saved in the netcdf history file (this is the usual case).

The idea behind the observation impact calculations then is to compute the contribution of each observation that is assimilated during 4D-Var to the increments in I. So, if we let I_{n+1} represent the metric computed from the analysis $\boldsymbol{x}_{n+1}$ at the end of outer-loop n, and I_n represent the metric computed from the starting estimate $\boldsymbol{x}_n$, then the metric increment is given by $\delta I_n = I_{n+1} - I_n$. Now the analysis is given by:

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \boldsymbol{K}_n \boldsymbol{d}_n \tag{5}$$

where n denotes the outer-loop number, the innovation vector $\boldsymbol{d}_n = \boldsymbol{y} - H(\boldsymbol{x}_n)$, the vector of observations is $\boldsymbol{y}$, and the observation operator is H (which includes the NL model). When n=1, then $\boldsymbol{x}_1 = \boldsymbol{x}_b$, the usual background. The dual form of the Kalman gain matrix is given by $\boldsymbol{K}_n = \boldsymbol{B}\boldsymbol{G}_n^T(\boldsymbol{G}_n\boldsymbol{B}\boldsymbol{G}_n^T + \boldsymbol{R})^{-1}$ where $\boldsymbol{B}$ and $\boldsymbol{R}$ are the background and observation error covariance matrices respectively, and $\boldsymbol{G}_n$ is the tangent linearization of H (and includes the TL model) for outer-loop n. During each *outer-loop* of ROMS 4D-Var, an approximation for $\boldsymbol{K}_n$ is computed according to:

$$\boldsymbol{K}_n = \boldsymbol{B}\boldsymbol{G}_n^T\boldsymbol{V}_{m,n}\boldsymbol{T}_{m,n}^{-1}\boldsymbol{V}_{m,n}^T\boldsymbol{G}_n\boldsymbol{B}\boldsymbol{G}_n^T \tag{6}$$

where $\boldsymbol{V}_{m,n}$ is the matrix of m Lanczos vectors generated by the m inner-loops during outer-loop n. The Lanczos vectors are orthonormal according to $\boldsymbol{V}_{m,n}^T\boldsymbol{G}_n\boldsymbol{B}\boldsymbol{G}_n^T\boldsymbol{V}_{m,n} = \boldsymbol{I}$, and $\boldsymbol{T}_{m,n}$ is a symmetric tridiagonal matrix. For the ROMS dual 4D-Var (i.e. *define W4DPSAS* and *define RPCG*), the vectors $\boldsymbol{V}_{m,n}$ and $\boldsymbol{U}_{m,n} = \boldsymbol{G}_n\boldsymbol{B}\boldsymbol{G}_n^T\boldsymbol{V}_{m,n}$ are archived in the netcdf *mod* file, as are the elements of the matrix $\boldsymbol{T}_{m,n}$.

Returning now to the increment in the circulation metric $I(\boldsymbol{x})$, we can express δI_n in terms of the state-vector increment $\boldsymbol{\delta x} = \boldsymbol{x}_{n+1} - \boldsymbol{x}_n = \boldsymbol{K}_n\boldsymbol{d}_n$ as follows:

$$\delta I_n = I(\boldsymbol{x}_{n+1}) - I(\boldsymbol{x}_n) = I(\boldsymbol{x}_n + \boldsymbol{K}_n\boldsymbol{d}_n) - I(\boldsymbol{x}_n) \approx (dI/d\boldsymbol{x})|_{\boldsymbol{x}_n}^T\boldsymbol{K}_n\boldsymbol{d}_n \tag{7}$$

where the last approximate equality arises from a 1st-order Taylor expansion of $I(\boldsymbol{x})$. Consider now the linear metric in (1), for which $dI/d\boldsymbol{x} = \boldsymbol{h}$. For the quadratic metric in (2), we have $dI/d\boldsymbol{x} = \boldsymbol{E}\boldsymbol{x} + \boldsymbol{E}^T\boldsymbol{x}$.

Turning now to the time integrated metrics in (3) and (4), we can see that in this case we need to consider the time history of $\boldsymbol{x}_n$ and $\boldsymbol{x}_{n+1}$ over the analysis window. Specifically, we require $\mathcal{M}(\boldsymbol{x}_{n-1})$ and $\mathcal{M}(\boldsymbol{x}_n)$, where $\mathcal{M}$ denotes the NL model. In the case of (3):

$$\delta I_n = I(\boldsymbol{x}_{n+1}) - I(\boldsymbol{x}_n) = 1/N\sum_{l=1}^{N}\boldsymbol{h}_i^T\left(\boldsymbol{x}_{n+1_l} - \boldsymbol{x}_{n_l}\right) = 1/N\sum_{l=1}^{N}\boldsymbol{h}_i^T\left(\mathcal{M}(\boldsymbol{x}_{n+1})_l - \mathcal{M}(\boldsymbol{x}_n)_l\right)$$
$$= 1/N\sum_{l=1}^{N}\boldsymbol{h}_l^T\left(\mathcal{M}(\boldsymbol{x}_n + \boldsymbol{K}_n\boldsymbol{d}_n)_l - \mathcal{M}(\boldsymbol{x}_n)_l\right) \approx 1/N\sum_{l=1}^{N}\boldsymbol{h}_l^T\boldsymbol{M}_{n_l}\boldsymbol{K}_n\boldsymbol{d}_n = \boldsymbol{d}_n^T\boldsymbol{K}_n^T\sum_{l=1}^{N}\boldsymbol{M}_{n_l}^T\boldsymbol{h}_l \tag{8}$$

where $\boldsymbol{M}_n$ is the tangent linear model (distinct from $\boldsymbol{G}_n$ which is the tangent linear model *sampled* at the observation points). The summation in the last equality in (8) represents the adjoint model *forced* by the vectors $\boldsymbol{h}_l$, which is simply $dI/d\boldsymbol{x}_l$. To evaluate the last expression in (8), we need $\boldsymbol{K}_n^T$, but this can be reconstructed from the archived Lanczos vectors according to (6) so that:

$$\delta I_n \approx 1/N\,\boldsymbol{d}_n^T\boldsymbol{U}_{m,n}\boldsymbol{T}_{m,n}^{-1}\boldsymbol{V}_{m,n}^T\boldsymbol{G}\boldsymbol{B}\sum_{l=1}^{N}\boldsymbol{M}_{n_l}^T\boldsymbol{h}_l \tag{9}$$

where $\boldsymbol{U}_{m,n} = \boldsymbol{G}_n\boldsymbol{B}\boldsymbol{G}_n^T\boldsymbol{V}_{m,n}$. Both $\boldsymbol{V}_{m,n}$ and $\boldsymbol{U}_{m,n}$ are archived by ROMS dual 4D-Var. Similarly, for the quadratic metric of the form given by (4), we can show that:

$$\delta I_n \approx 1/N\,\boldsymbol{d}_n^T\boldsymbol{U}_{m,n}\boldsymbol{T}_{m,n}^{-1}\boldsymbol{V}_{m,n}^T\boldsymbol{G}\boldsymbol{B}\sum_{l=1}^{N}\boldsymbol{M}_{n_l}^T(\boldsymbol{E}_l + \boldsymbol{E}_l^T)\boldsymbol{x}_{n_l} \tag{10}$$

where now the adjoint model is forced by $(\boldsymbol{E}_l + \boldsymbol{E}_l^T)\boldsymbol{x}_{n_l}$ which depends now on the state-vector estimate $\boldsymbol{x}_{n_l}$.

In either case, (9) and (10) show that the increment in the metric δI_n can be written as $\delta I_n = \boldsymbol{d}_n^T\boldsymbol{g}_n$, where the form of $\boldsymbol{g}_n$ follows from (9) and (10). In other words, δI_n is simply the dot-

product of the innovation vector with the impact vector $\boldsymbol{g}_n$. Since each element of $\boldsymbol{d}_n$ is uniquely associated with a single observation, so then are the elements of $\boldsymbol{g}_n$.

Some practicalities

(1) In order to compute the observation impacts for a given metric, it is necessary to compute the forcing for the adjoint model (i.e. $\boldsymbol{h}_l$ or $(\boldsymbol{E}_l + \boldsymbol{E}_l^T)\boldsymbol{x}_{n_l}$). This is created as a netcdf file *before* running the observation impact driver and is referred to in *ocean.in* as *ADSname*.

(2) If you perform multiple outer-loops, you need to run the observation impact calculations for each outer-loop *separately*. For multiple outer-loops, there exists a Kalman gain matrix $\boldsymbol{K}_n$ for each outer-loop, and a corresponding sequence of Lanczos vectors $\boldsymbol{V}_{m,n}$. To specify which outer-loop you want to consider, you must set the parameter *Nimpact* in the *s4dvar.in* input file.

(3) The obs impact calculations require information about the dual Lanczos vectors that is archived in the *mod* netcdf file generated during the 4D-Var calculation. You must point to this file using the *LCZname* in the *s4dvar.in* input file.

(4) For the specific form of the time averaged metrics in (3) and (4) is important to *define AD_IMPULSE*. If you want to do something different in time, talk to me first.

(5) Regardless of the metric used, you will also need to provide a history file for the background solution used during the 4D-Var cycle. This is needed for $\boldsymbol{G}_n$, $\boldsymbol{G}_n^T$, and $\boldsymbol{M}_n$ in equations (9) and (10). If you are computing the observation impacts for outer-loop *n*, you will need to provide the file *fwd_00(n-1).nc*. So for outer-loop 1 impacts you will need *fwd_000.nc*, for outer-loop 2 impacts *fwd_001.nc*, etc. Therefore, be sure that you save *all* of your *fwd/his* files from each 4D-Var cycle.

(6) If you are adjusting the surface forcing and/or the open boundary conditions during 4D-Var, there is an option to compute the observation impacts for each component of the control vector separately (*define OBS_IMPACT_SPLIT*). However, the computational cost increases so you may not want to do this initially.

(7) If computing the impacts for multiple outer-loops, you must also *define SKIP_NLM*.

(8) There is also a new feature, described below, that computes, at no extra cost, the impacts for each inner-loop. This can be a useful diagnostic also. To activate this option, you need to *define IMPACT_INNER*.

Examples

Alongshore transport

The time averaged, integrated, transport crossing a section is an example of linear metric of the form (3). In (3) (and (4)), the dot-products represent summations of the ROMS state-vector variables multiplied by appropriate coefficients. With this in mind, let's rewrite (3) in the form:

$$I = 1/N \sum_{l=1}^{N} \sum_{k=1}^{K} \sum_{j=1}^{J} \sum_{i=1}^{I} h_{i,j,k,l} x_{i,j,k,l} \qquad (11)$$

where i and j denote grid-cell locations in the horizontal, k is the vertical, and time is represented by l as before. The transport example that is presented here is from the UCSC west coast (WC12) model, and is the time-averaged transport along a zonal section at 37N, from the coast to 127W, over the upper 500m of the water column. Therefore in this case, all of the elements $h_{i,j,k,l}$ are zero, except those that correspond to v-points along the section. At these points $h_{i,j,k,l}=dx_{i,j}dz_k$, where $dxdz$ represents the area of each grid cell along the section. So in this example, when computing the observation impacts, the adjoint model v variable is forced by $dx_{i,j}dz_k/N$ at those points along the section, while *all* other v-points and other state variables are unforced (in practice they are forced by zeros).

After running the observation impact calculations for a number of cycles, it is important to check how good the tangent linear assumption is based on (9) and (10). These estimates of δI_n should be compared with the estimate $\Delta I_n = I(\boldsymbol{x}_{n+1}) - I(\boldsymbol{x}_n)$ computed from the difference between the non-linear model solutions for $\boldsymbol{x}_{n+1}$ and $\boldsymbol{x}_n$. To illustrate, Fig. 1 shows time series of δI_n and ΔI_n for 37N transport for a sequence of 8-day 4D-Var cycles in WC12 spanning the period Jan-Dec, 2011. The analyses were computed using n=2 outer-loops and m=7 inner-loops, and Fig. 1 shows the metric increments for outer-loop 1. Clearly, the tangent linear assumption in (9) is an excellent approximation in this case. The agreement for outer-loop 2 is equally remarkable (not shown).

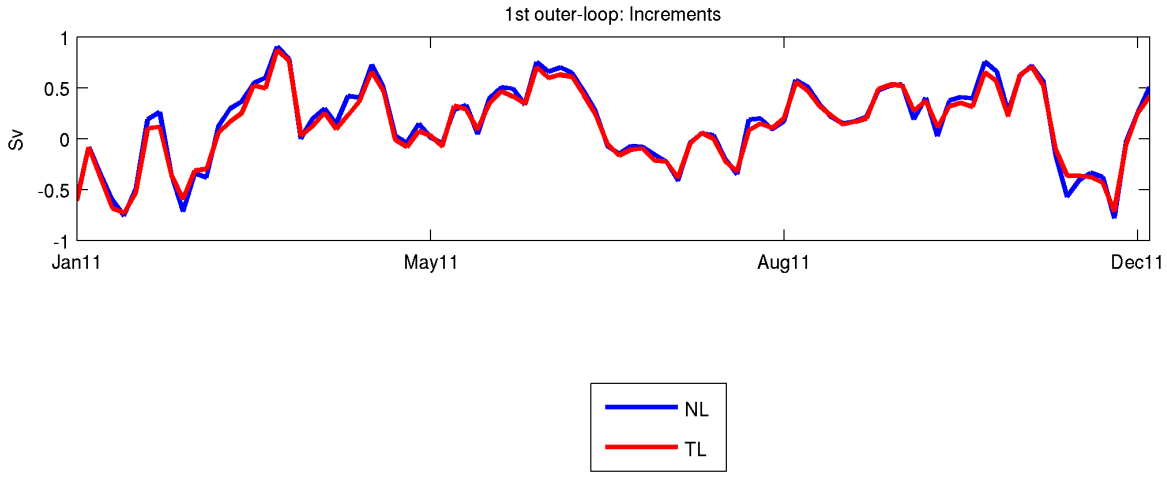


Figure 1: Time series of the 37N transport increment ΔI_1 computed directly from the analysis and background solutions of the non-linear model (blue curve labelled NL), and δI_1 based on the tangent linear assumption in (9) (red curve labelled TL). Results shown are for the 1st outer-loop.

There are many ways that the observation impacts for a particular metric can be presented graphically and analyzed. Some example *Matlab* scripts are provided for the each of the example metrics discussed here, but these scripts are by no means exhaustive! After some thought you will undoubtedly come up with other, more imaginative ways of presenting the data, but for now these examples will get you started.

As an example, Fig. 2 shows a time series of the observation impacts for the 37N transport metric during the 1st outer-loop. In this case the impacts are displayed as the contribution of each observation *type* to the metric increment δI_1. In this example, the impacts are fairly evenly distributed across all of the observation types on average, with *in situ* temperature and salinity observations contributing least. The impacts can be plotted separately for each platform, as illustrated in Fig. 3 in the case of remotely sensed observations, and in Fig. 4 for temperature observations from various *in situ* observing systems. For example, Fig. 4 reveals the relatively large impact that Scripps glider observations (labelled "gliders UCSD") have on the transport increments.

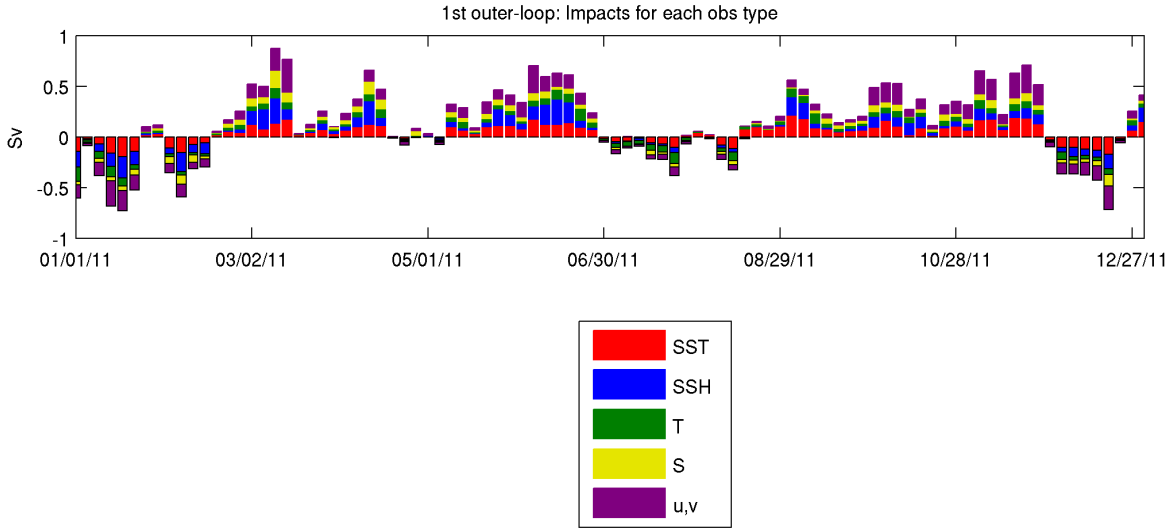


Figure 2: Summary of the impact of each observing platform on 37N transport increment for each 8-day 4D-Var cycle during 2011. In situ hydrographic observations are referred to in the legend as "T" and "S." Results shown are for the 1st outer-loop.

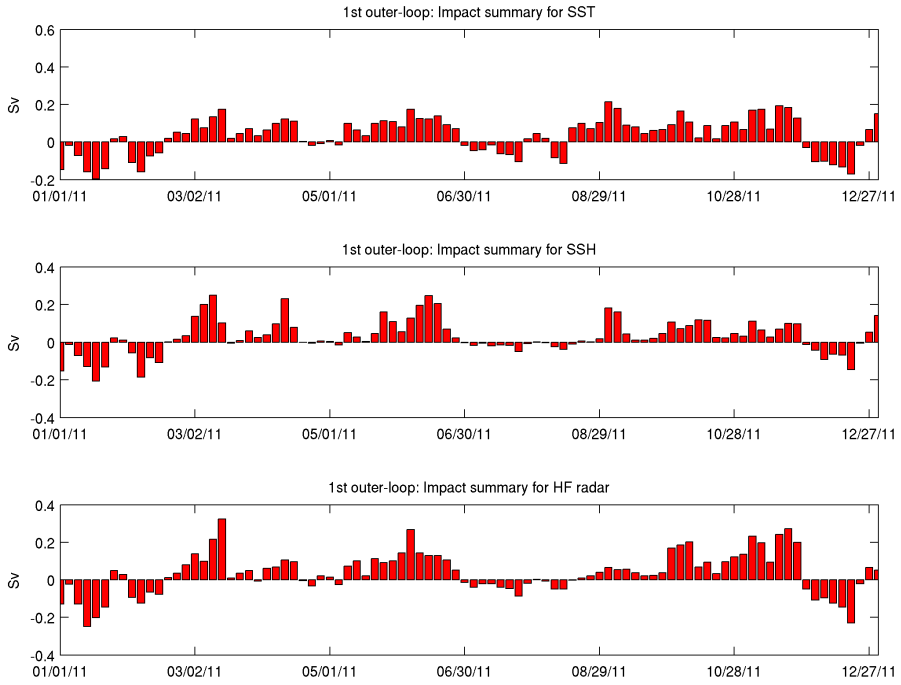


Figure 3: Time series of the impact of each remote sensing platform on increments in the 37N transport. Results shown are for the 1st outer-loop.

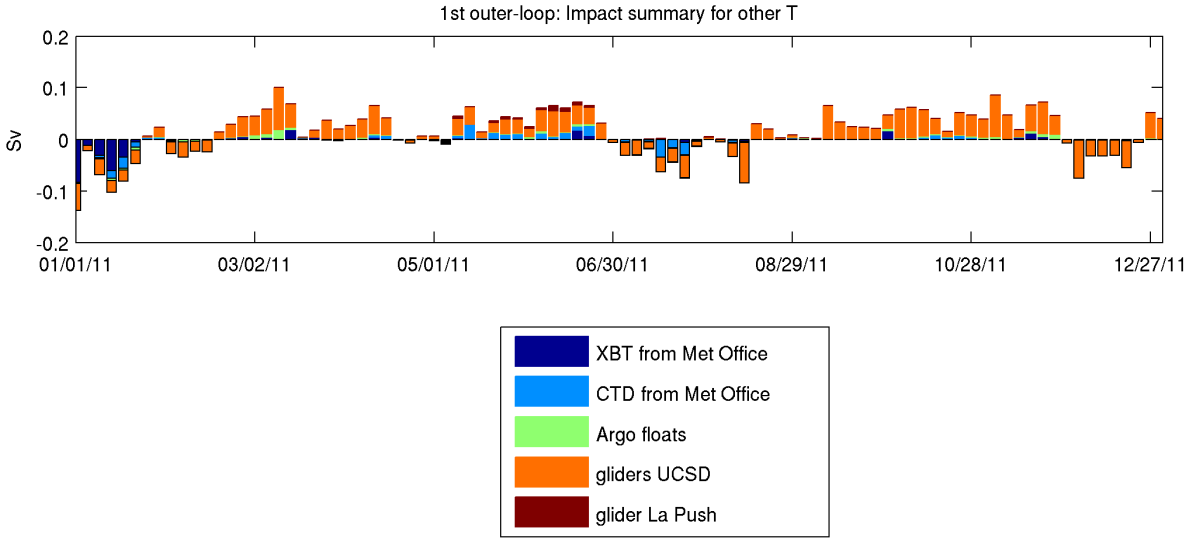


Figure 4: Time series of the contribution of temperature observations from various *in situ* observing platforms on the 37N transport increments. Results shown are for the 1st outer-loop.

Another useful way to display and analyze the observation impacts is as a scatter plot, as illustrated in Fig. 5. In this case the impact (i.e. the contribution to δI_n) of each individual observation is plotted against the value of the innovation (i.e. the difference between the observation and the background). Figure 5 shows scatter plots for *all* observations and 4D-Var cycles during 2011.

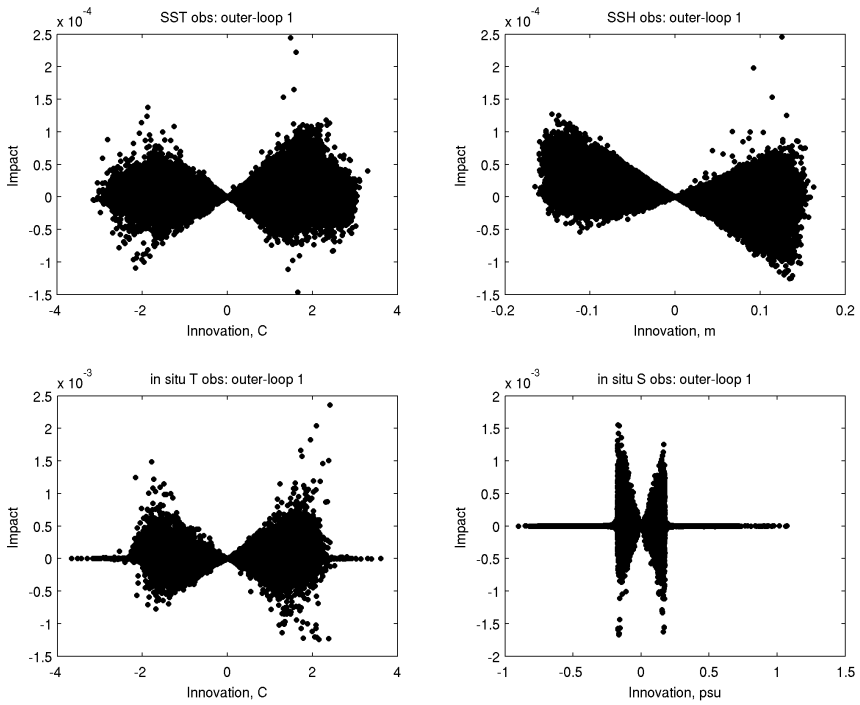


Figure 5: Scatter plots of the impact on 37N transport (Sv) versus the innovations of the individual observations for different observation types for *all* 4D-Var cycles spanning 2011. Results are shown for outer-loop 1.

It is gratifying to see in Fig. 5 that in general small innovations are associated with small impacts, while the range of the impacts possible increases at the innovations increases – this is obvious from the "butterfly" shape of the plots. It is also reassuring to see in Fig. 5 that for very large innovations (i.e. observations that deviate significantly from the background) generally have small impacts. For example, in Fig. 5 there are some *in situ* observations in T and S with large innovations that exceed 2 C and 0.25 psu respectively, which have almost no impact on the transport. If we divide each scatter plot into four quadrants centered at the origin, this yields a contingency diagram, and ideally what we would like is for the four quadrants to be equally populated. This would reveal that there is no mean bias in the impacts or the innovations in any particular direction. This is generally not the case here. In particular, the scatter plot for SSH observations shows that, in general, positive (negative) sea surface height (SSH) innovations tend to favor equatorward (poleward) transport increments. This is an aspect of the model, observation and data assimilation system that can be monitored and investigated further. Such biases are sometimes eliminated or reduced in subsequent outer-loops (not shown).

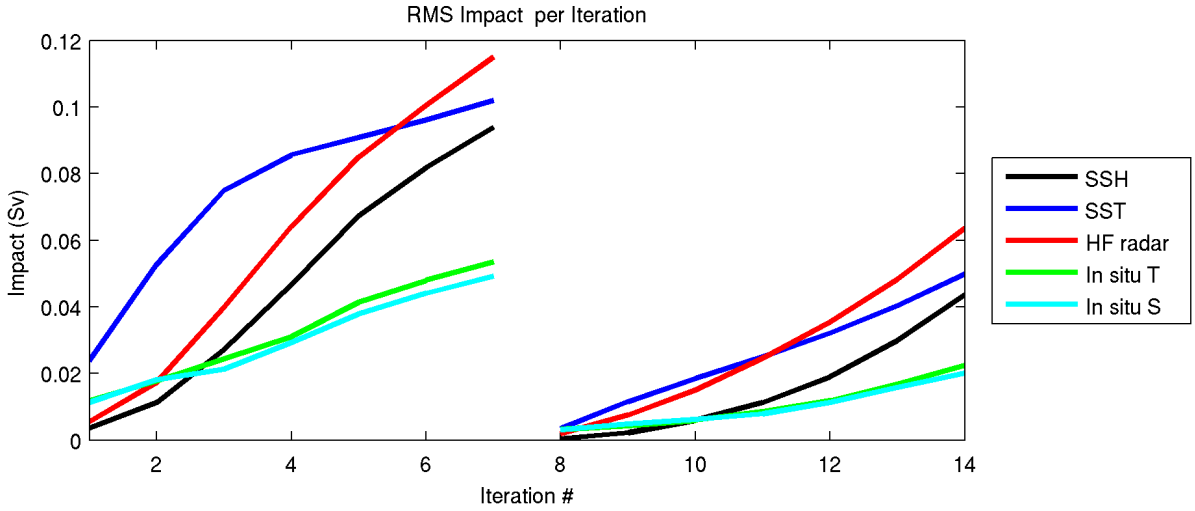


Figure 6: The RMS impact on 37N transport increments of each observation type versus inner-loop number for each outer-loop. Outer-loop 1 ends at iteration 7 and outer-loop 2 begins at iteration 8.

The degree to which each observation type is being utilized in the assimilation system can be assessed by looking at the impacts during each inner-loop *and* outer-loop. Figure 6 shows the RMS 37N transport increments due to different observation types during each inner-loop for all the 4D-Var cycles of 2011 in WC12. Recall in this case that 7 inner-loops and 2 outer-loops were employed, which is a total of 14 iterations of the 4D-Var minimization algorithm. The start and end of each outer-loop is very obvious in Fig. 6. Looking first at outer-loop 1, we can see that the impact of the SST observations increases quickly during the first few inner-loops, but then appears to begin to asymptote to a near constant value toward the end of the 1st outer-loop. This behavior suggests that SST observations are being used very effectively by the 4D-Var system during the 1st outer-loop, and toward the end of the outer-loop most of the information about 37N transport available in the SST observations has been mined. Conversely, Fig. 6 show that during the 1st outer-loop, the impacts for the other observation types (SSH, HF radar, and *in situ* observations) all continue on a rapid upward trajectory with no obvious sign of asymptotic behavior, which suggests that there is still useful information about 37N transport that can be extracted from these data. During the 2nd outer-loop, Fig. 6 shows that the impacts are generally lower than they are in the 1st outer-loop. The lack of any asymptotic behavior for any of the observation types at the end of the 2nd outer-loop suggests that estimates of 37N transport may benefit from additional inner-loops.

It is also useful to monitor the observation impacts during each individual analysis cycle while the near real-time systems are running. For example, Fig. 7 shows the impact of each observing system on the 37N transport during the final 4D-Var cycle of 2011. Obviously, scatter plots like those in Fig. 5 can be plotted for a single cycle, as can the impacts per inner-loop. As one final

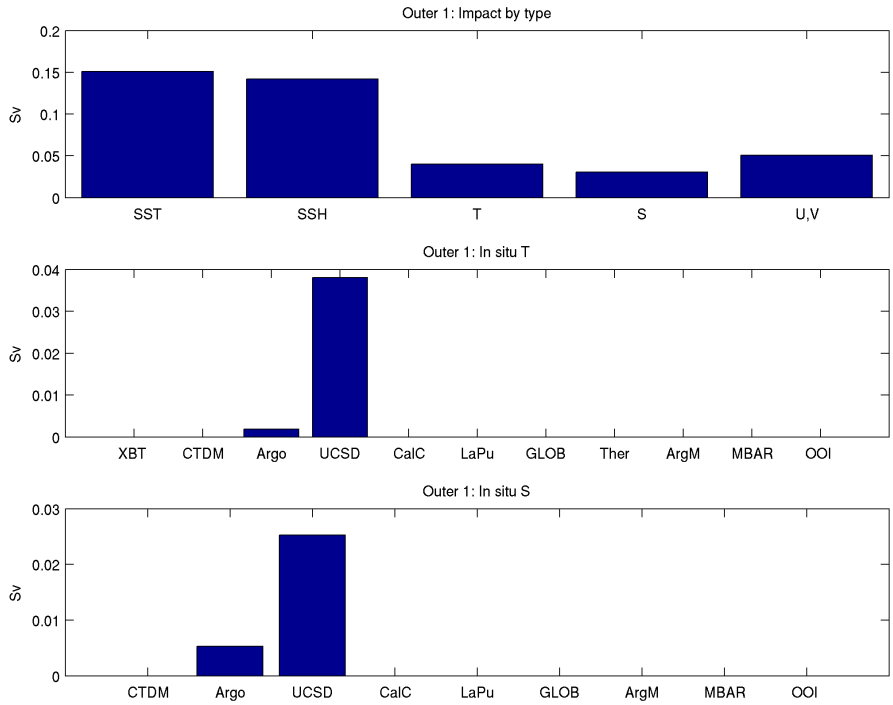


Figure 7: The contribution of each observing platform to the increment in 37N transport during the final 4D-Var cycle of 2011. Results shown are for outer-loop 1.

example for the 37N transport metric, Fig. 8 shows the contribution of each observation type to the 37N transport increment during each inner-loop of both outer-loops during the last 4D-Var cycle of 2011. Interestingly during this particular cycle, all of the transport increments are poleward during the 1st outer-loop, while during the 2nd outer-loop they are close to zero and often equatorward. Unraveling the reason behind this kind of behavior is, of course, the next challenge, but we now at least have a tool for discovering the behavior in the first place!

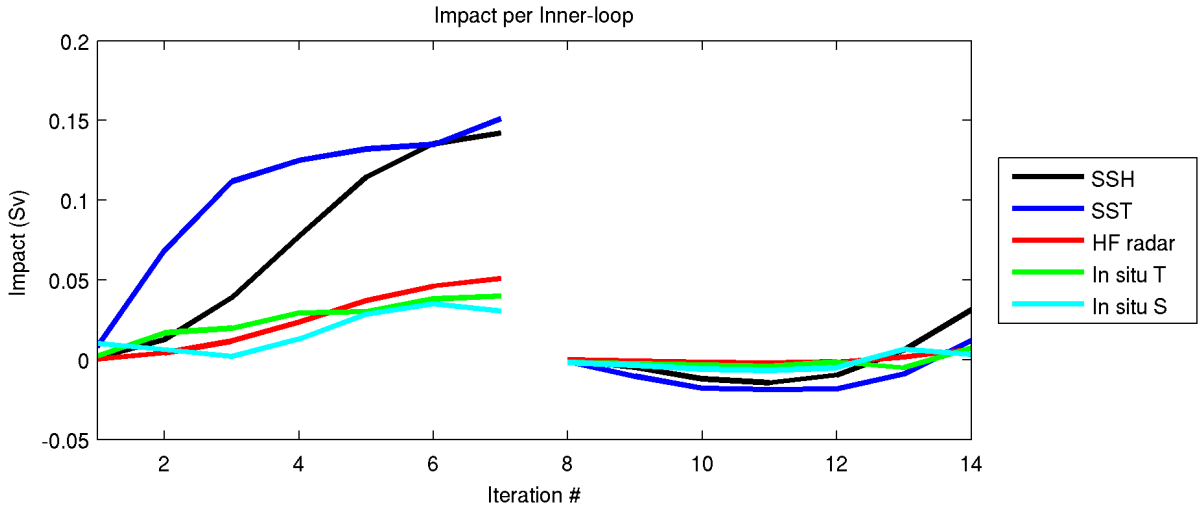


Figure 8: The contribution of each observation type to the increments in 37N transport during each iteration of 4D-Var during the last cycle of 2011. Outer-loop 1 ends at iteration 7, and outer-loop 2 begins at iteration 8.

It is also very illuminating to plot the observation impacts as a function of geographic position, as a function of depth, as a function of offshore distance, and the cumulative impacts over time. I have found all of these to be very illuminating. Clearly, there are many other ways of presenting this information, so think creatively!

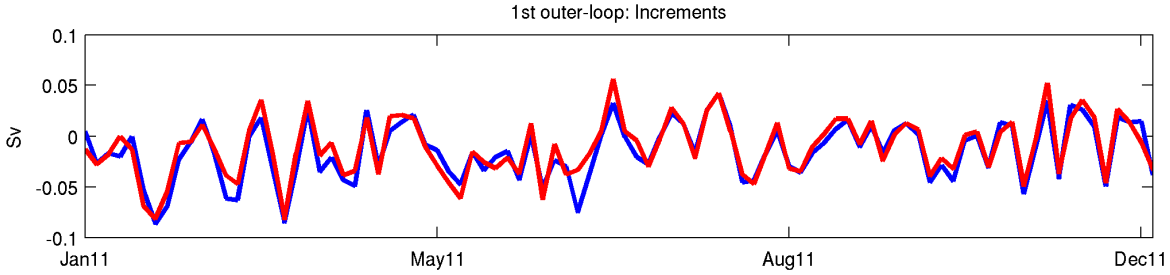
Upwelling transport

The previous example highlighted several different ways that the observation impact information can be displayed and presented. For the remaining examples, only a few figures will be shown, but a complete set of all figures is available in the tar file that accompanies this document.

In this second example, we use a metric which is based on the ROMS *w* field, which is a diagnostic variable. Specifically, we have:

$$I = 1/N \sum_{l=1}^{N} \boldsymbol{h}_l^T \boldsymbol{E}_l \boldsymbol{x}_l = 1/N \sum_{l=1}^{N} \sum_{k=1}^{K} \sum_{j=1}^{J} \sum_{i=1}^{I} h_{i,j,k,l} f_{i,j,k,l} x_{i,j,k,l} \qquad (12)$$

where the matrix $\boldsymbol{E}$ represents the ROMS diagnostic routine *wvelocity.F*. So, in this case the matrix multiplication $\boldsymbol{Ex}$ returns the vector of *w* values at *w-points*. The vector $\boldsymbol{h}$ then selects the values of *w* that will be used to define the metric, and rescales them appropriately. In the example presented here, $\boldsymbol{h}$ defines the vertical transport passing through a horizontal surface at 40 m depth, integrated over an area that extends 200 km offshore between Cape Mendocino and Point Conception on the central California coast. In this case, the elements of $\boldsymbol{h}$ are simply the grid-cell areas *dxdy* at ρ-points. In the second equality in (12), the factors $f_{i,j,k,l}$ represent the coefficients that each element of the ROMS state-vector is effectively multiplied by to compute vertical velocity. This is of course determined by the incompressibility condition $\partial u/\partial x + \partial v/\partial y + \partial w/\partial z = 0$. Recall for the observation impact calculations we require $\partial I/\partial \boldsymbol{x}_l = (1/N)\boldsymbol{E}_l^T \boldsymbol{h}_l$ (cf equation (7)), where $\boldsymbol{E}^T$ represents the adjoint of the routine *wvelocity.F*, namely *ad_wvelocity.F*. This new ROMS routines forces the *u* and *v* fields appropriately based on the correct form of the incompressibility condition used in ROMS. To activate this capability in ROMS, you need to *define AD_WVEL*.



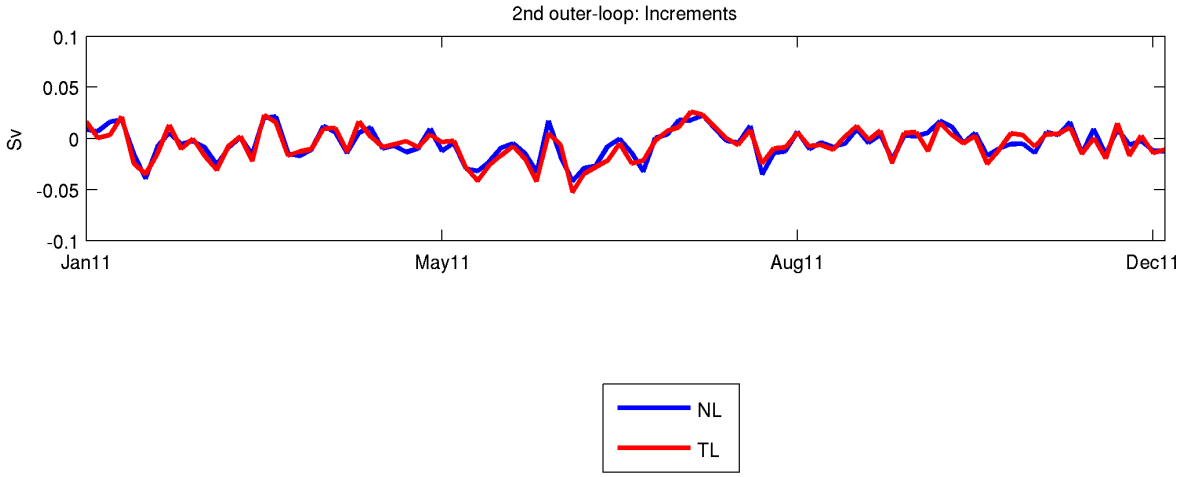


Figure 9: Time series of the coastal upwelling transport increment ΔI_n computed directly from the analysis and background solutions of the non-linear model (blue curve labelled NL), and δI_n based on the tangent linear assumption in (9) (red curve labelled TL). Results shown are for the 1st outer-loop (upper panel) and the 2nd outer-loop (lower panel).

To illustrate, Fig. 9 shows time series of the central California coastal upwelling index increments ΔI_n computed directly from the difference between the non-linear model solutions for $\boldsymbol{x}_{n+1}$ and $\boldsymbol{x}_n$, and δI_n using the tangent linear approximation (9). The agreement is clearly very good for the majority of the 4D-Var cycles, particularly during the 2nd outer-loop, although there are a few significant discrepancies during the 1st outer-loop when, presumably, non-linearity is important.

Eddy Kinetic Energy

The eddy kinetic energy (EKE) metric discussed here is an example of a quadratic (*i.e.* non-linear) circulation metric of the form (4). More specifically, we consider:

$$I = 1/N \sum_{l=1}^{N} (\boldsymbol{x}_l - \overline{\boldsymbol{x}})^T \boldsymbol{E}_l (\boldsymbol{x}_l - \overline{\boldsymbol{x}}) = 1/N \sum_{l=1}^{N} \sum_{k=1}^{K} \sum_{j=1}^{J} \sum_{i=1}^{I} e_{i,j,k,l} (x_{i,j,k,l} - \bar{x}_{i,j,k})^2 \quad (13)$$

where $\overline{\boldsymbol{x}}$ denotes the *seasonal mean* climatology of $\boldsymbol{x}$. The matrix $\boldsymbol{E}$ in this case is diagonal with elements of the form $e_{i,j,k,l} = dz_{i,j,k}/d$, where $dz_{i,j,k}$ is the thickness of each *u* or *v* grid-cell in the target region, and *d* is a depth of the selected portion of the water column. All other elements of $\boldsymbol{E}$ are zero. In the example considered here, the EKE is integrated over the target region 36N-40N, 124W-130W, to a depth of *d*=500m. This region is known from observations to be a region of elevated EKE. Following the usual convention, the factor of $\rho_o/2$ is omitted, so the units of EKE are m^2s^{-2}. In this case, since $\boldsymbol{E}$ is symmetric, $\partial I/\partial \boldsymbol{x}_l = (2/N)\boldsymbol{E}_l(\boldsymbol{x}_l - \overline{\boldsymbol{x}})$. To compute the observation impacts of the 1st outer-loop, we choose $\boldsymbol{x}_l = \boldsymbol{x}_{bl}$, the background, while for the 2nd outer-loop $\boldsymbol{x}_l = \boldsymbol{x}_{1l}$ where $\boldsymbol{x}_{1l}$ is the NL model trajectory at the end of the 1st outer-loop (*i.e.* the *fwd_001* solution used to in the TL and AD models during the 2nd outer-loop). In other words, the forcing fields for the adjoint model are different during each outer-loop observation impact calculation.

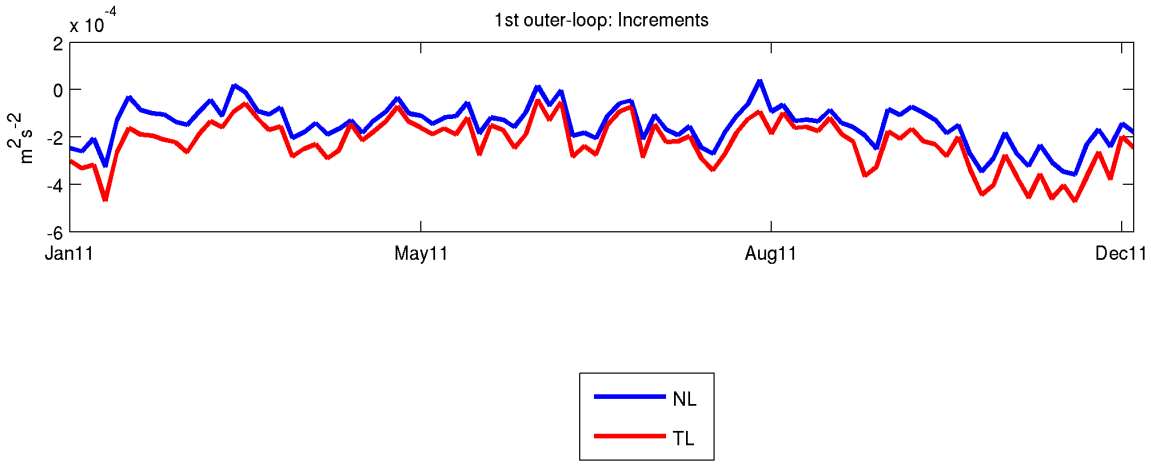


Figure 10: Time series of the EKE increment ΔI_1 computed directly from the analysis and background solutions of the non-linear model (blue curve labelled NL), and δI_1 based on the tangent linear assumption in (10) (red curve labelled TL). Results shown are for the 1st outer-loop.

Time series of the EKE increments ΔI_n computed directly from the NL model solutions and δI_n using the tangent linear approximation (10) are shown in Fig. 10. Notice in this case, that while the two time series are clearly highly correlated, the TL approximation δI_n is always systematically smaller than the EKE increment ΔI_n computed directly from the NL model solutions. This is to be expected in this case because by making the TL approximation, we are effectively ignoring the 2nd order terms $(\delta u^2 + \delta v^2)$ which are always positive. The EKE increment ΔI_n computed directly as the difference EKE of the analysis and the EKE of the background will always include these higher-order terms, and therefore ΔI_n always be larger than the EKE increment δI_n computed from (10). Clearly though, Fig. 10 shows that the time variations of the 1st-order variations in the EKE increment are described very well by (10). The same applies in subsequent outer-loops (not shown).

Isopycnal Depth

This next example serves to illustrate how we can compute observation impacts for metrics that, at first sight, do not appear to be explicit functions of the state-vector. In this case, we consider the depth of a particular isopycnal averaged over some target region as the circulation metric *I*. Therefore, it would appear that *I* is a function only of depth *z*, which of course is *not* a state variable. Specifically, we will consider:

$$I = 1/N \sum_{l=1}^{N} \boldsymbol{h}_l^T P_l(\rho_0 \boldsymbol{I} + \boldsymbol{E}_l \boldsymbol{x}_l) \qquad (14)$$

where in this case $\boldsymbol{E}$ is a matrix such that $(\rho_0 \boldsymbol{I} + \boldsymbol{E}\boldsymbol{x})$ yields a vector of ocean density. For simplicity, we will assume that $\boldsymbol{E}$ describes the linear equation of state. We will see that this works exceedingly well even though a non-linear equation of state was used during 4D-Var. The operator *P* is an interpolation operator that isolates the depth of a particular isopycnal. In the example considered here, we consider the $\sigma = 26$ kg m^{-3} isopycnal since it is a good proxy in the NE Pacific for the depth of the main thermocline. Thus $P(\rho_0 \boldsymbol{I} + \boldsymbol{E}_l \boldsymbol{x}_l)$ yields a vector of the depth of the isopycnal. Finally, the vector $\boldsymbol{h}$ is a mask which defines the geographic region so

that $\boldsymbol{h}^T P(\rho_0 \boldsymbol{I} + \boldsymbol{E}_l \boldsymbol{x}_l)$ yields the average isopycnal depth over the target region. In order to proceed with the observation impacts for this metric, we need to linearize *P*. With this in mind, consider the following. For the outer-loop *n* estimate $\boldsymbol{x}_n$ we can compute the density $\rho_n = \rho_0 - \alpha(T_n - T_0) + \beta(S_n - S_0)$. (Recall that for n=1, $T_l = T_b$ and $S_l = S_b$, the starting background values.) We can then use *P* to compute the depth z_n of our target isopycnal, in this case $\sigma = 26$ kg m^{-3}. We can also compute $(\partial \rho_n / \partial z)|_{z_n}$, the vertical derivative of the density gradient based on the outer-loop *n* estimate at the depth z_n of the target isopycnal depth. Now suppose that at this depth, the difference between the next outer-loop estimate and the current outer-loop density is $\delta\rho$. We can then estimate the change in depth δz of the target isopycnal according to $\delta z_n \approx \delta\rho\ (\partial z / \partial \rho_n)|_{z_n}$ where $(\partial z / \partial \rho_n)|_{z_n} = (\partial \rho_n / \partial z)^{-1}|_{z_n}$. Therefore, the isopycnal depth metric increment has been expressed as a function of a state-variable. From (14), we have:

$$\delta I = I(\boldsymbol{x}_{n+1}) - I(\boldsymbol{x}_n) = 1/N \sum_{l=1}^{N} \boldsymbol{h}_l^T (P_l(\rho_0 + \boldsymbol{E}_l \boldsymbol{x}_{n+1l}) - P_l(\rho_0 + \boldsymbol{E}_l \boldsymbol{x}_{nl}))$$

$$= 1/N \sum_{l=1}^{N} \boldsymbol{h}_l^T (P_l(\rho_0 + \boldsymbol{E}_l(\mathcal{M}(\boldsymbol{x}_{n+1} + \boldsymbol{K}_n \boldsymbol{d}_n))_l) - P_l(\rho_0 + \boldsymbol{E}_l \mathcal{M}(\boldsymbol{x}_n)_l))$$

$$\approx 1/N \sum_{l=1}^{N} \boldsymbol{h}_l^T (\partial P / \partial \boldsymbol{x})_l \, \boldsymbol{E}_l \boldsymbol{M}_{n_l} \boldsymbol{K}_n \boldsymbol{d}_n$$

$$= \boldsymbol{d}_n^T \boldsymbol{K}_n^T \, 1/N \sum_{l=1}^{N} \boldsymbol{M}_{n_l}^T \boldsymbol{E}_l (\partial P / \partial \boldsymbol{x})_l^T \boldsymbol{h}_l \qquad (15)$$

where $(\partial P / \partial \boldsymbol{x})$ represents the tangent linearization of *P* and yields the estimated change in depth $\delta z_n \approx \delta\rho\ (\partial z / \partial \rho_n)|_{z_n}$ as described above.

To illustrate the efficacy of this approach, we consider the $\sigma = 26$ kg m^{-3} isopycnal depth averaged over the central California region between Cape Mendocino and Point Conception out to a distance of 200 km offshore. Figure 11 shows time series of the background isopycnal depth index during 2011, and the seasonal shallowing of the main thermocline during spring and summer associated with coastal upwelling and Ekman pumping due to wind stress curl is clearly evident. Time series of the thermocline depth increments δI_1 and ΔI_1 during the 1st outer-loop, following the now familiar format, are also shown in Fig. 11, and the illustrate that the aforementioned tangent linear approximation for the interpolation operator *P* works very well. The same is true during the 2nd outer-loop (not shown).

The target region can of course be expanded to encompass the entire model grid, providing a measure of the average depth of the isopycnal (or thermocline) over the whole model domain. This is one of the proposed common metrics that we should use for all three IOOS regions, and the tangent linear approximation of *P* works very well in this case for WC12 (not shown) during both outer-loops.

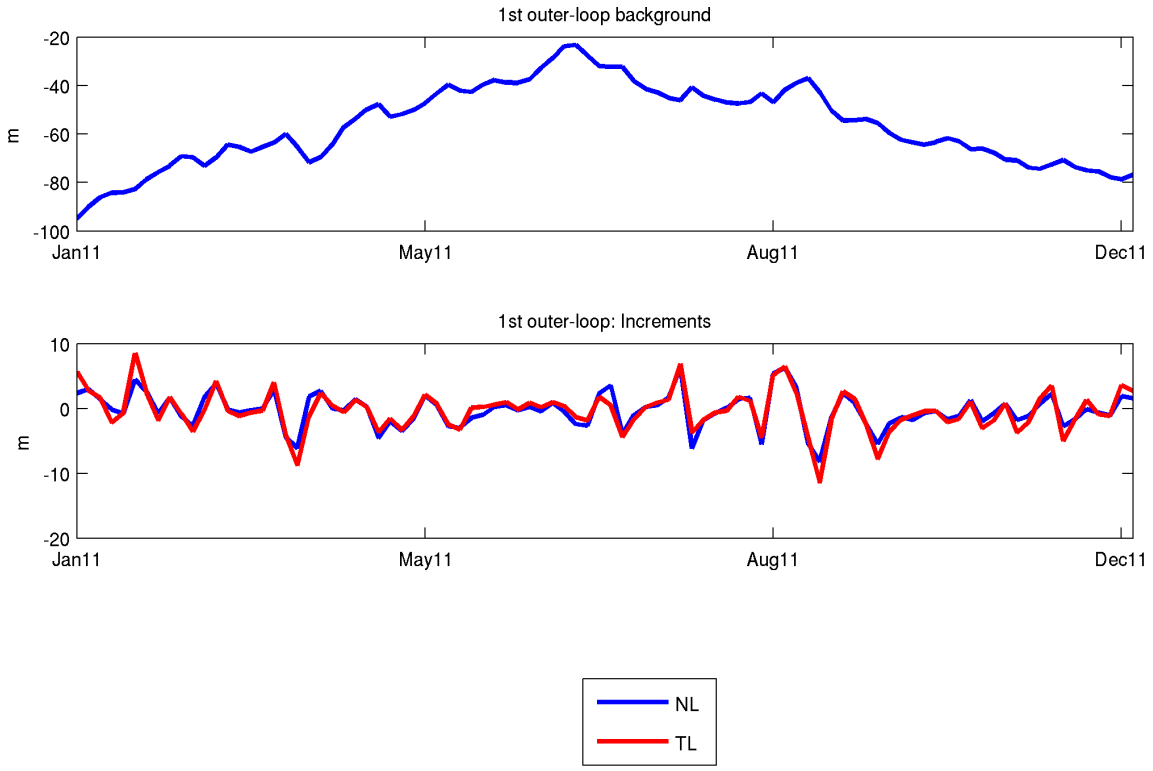


Figure 11: The upper panel shows time series of the depth of the $\sigma = 26$ kg m^{-3} isopycnal averaged over the central California region computed from the 4D-Var background (upper panel). The lower panel shows time series of isopycnal increment δI_1 computed directly from the analysis and background solutions of the non-linear model (blue curve labelled NL), and ΔI_1 based on the tangent linear assumption in (9) (red curve labelled TL). Results shown are for the 1st outer-loop.

An Energy Density Metric

The metric described here represents an example of the common energy metric that we agreed upon for each of the IOOS domains. There are several possibilities based on the various terms in the ocean energy balance. Specifically, we can consider the energy densities:

$$E_I = \rho_0 c_p/(NV) \sum_{l=1}^{N} \sum_{k=1}^{K} \sum_{j=1}^{J} \sum_{i=1}^{I} (T_{i,j,k,l} + 273) dV_{i,j,k} \qquad (16)$$

$$E_{PE} = g/(NV) \sum_{l=1}^{N} \sum_{k=1}^{K} \sum_{j=1}^{J} \sum_{i=1}^{I} \rho_{i,j,k} z_{i,j,k} dV_{i,j,k} \qquad (17)$$

$$E_{KE} = \rho_0/(2NV) \sum_{l=1}^{N} \sum_{k=1}^{K} \sum_{j=1}^{J} \sum_{i=1}^{I} (u^2 + v^2) dV_{i,j,k} \qquad (18)$$

$$E_{APE} = g/(2\mathcal{N}_0^2 \rho_0 NV) \sum_{l=1}^{N} \sum_{k=1}^{K} \sum_{j=1}^{J} \sum_{i=1}^{I} (\rho_{i,j,k} - \rho_0)^2 dV_{i,j,k} \qquad (19)$$

where E_I is the internal energy, E_{PE} is the potential energy, E_{KE} is the kinetic energy, E_{APE} is the available potential energy, dV is the volume of each grid cell, and V is the total volume of the ocean domain. The total energy density is, by definition, $E = E_I + E_{PE} + E_{KE}$. However, this is not a very useful metric though because E is dominated (by many orders of magnitude) by E_I. Since

E_I does not change a great deal between the background and the analysis, we will not include it in our energy metric. Similarly, E_{PE} is typically a few orders of magnitude larger than E_{KE} so a combination these two terms does not lend itself to a useful metric. Instead, it is more useful, and dynamically more relevant, so consider available potential energy E_{APE}. The form of equation (19) for E_{APE} is actually a simplification which makes it more practical for use here. The available potential energy, E_{APE} is usually defined in terms of departures of the in situ potential density from the density ρ_r at some reference depth z_r. In (19), ρ_r is replaced by the constant ρ_0. In addition, the buoyancy frequency $\mathcal{N}^2(z)$ has been replaced by a representative value $\mathcal{N}_0^2$ in an effort to make the expression for E_{APE} less non-linear.

The available potential energy is typically about an order of magnitude larger than the kinetic energy. So, with all this in mind, the common circulation metric that I propose that we use for the observation impact calculations is $I = E_{APE} + E_{KE}$ given by:

$$I = g/(2\mathcal{N}_0^2\rho_0 NV)\sum_{l=1}^{N}\sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{i=1}^{I}(\rho_{i,j,k}-\rho_0)^2 dV_{i,j,k} + \rho_0/(2NV)\sum_{l=1}^{N}\sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{i=1}^{I}(u^2+v^2)dV_{i,j,k}. \qquad (20)$$

The metric given by (20) is clearly non-linear and of the general form given by (4). As in the case of the isopycnal depth metric, density should be computed using a linear equation of state.

To illustrate, Fig. 12 shows time series of the energy density metric (20) computed from the background circulation during the 1st outer-loop of all 4D-Var cycles in WC12 during 2011. In this case, (20) encompasses the *full* model domain and the entire water column. Also shown in Fig. 12 are time series of the energy density metric increment ΔI_1 computed from the difference between I evaluated directly using $\boldsymbol{x}_2$ and $\boldsymbol{x}_1 = \boldsymbol{x}_b$, and δI_1 computed using the tangent linear approximation (10). Clearly, the tangent linear approximation of (20) works very also. Figure 13 shows the impact of each observation type on the energy density metric increments δI_1, and indicates that temperature observations typically have the largest impact, as one might anticipate.

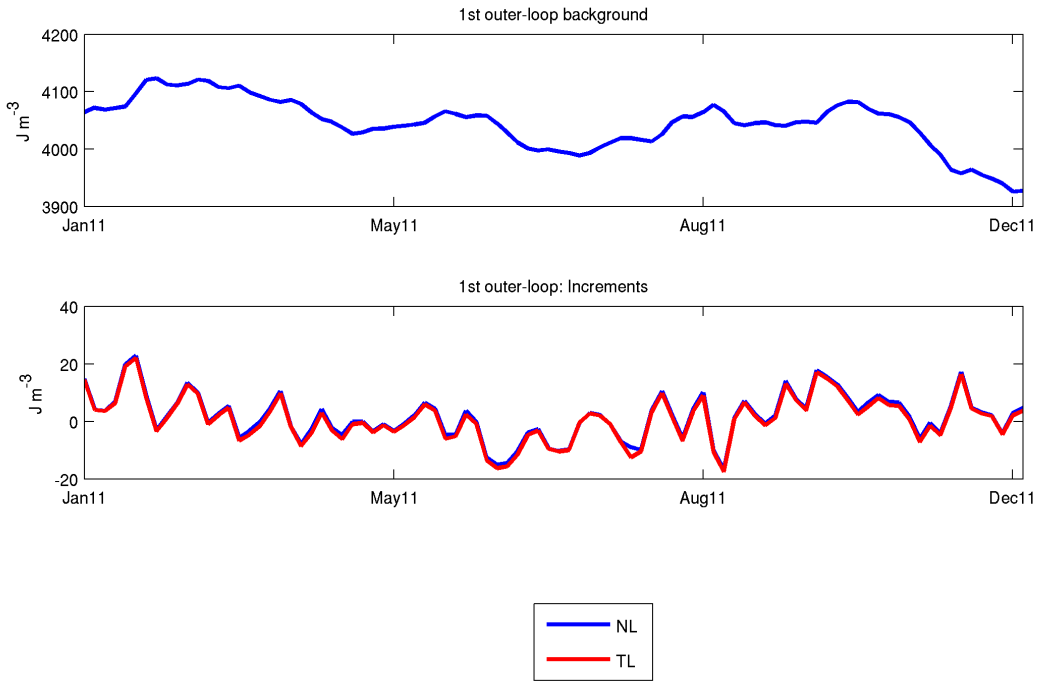


Figure 12: The upper panel shows time series of the energy density metric (20) averaged over the entire model domain, and computed from the 4D-Var background (upper panel). The lower panel shows time series of energy density metric increment ΔI_1 computed directly from the analysis and background solutions of the non-linear model (blue curve labelled NL), and δI_1 based on the tangent linear assumption in (10) (red curve labelled TL). Results shown are for the 1st outer-loop.

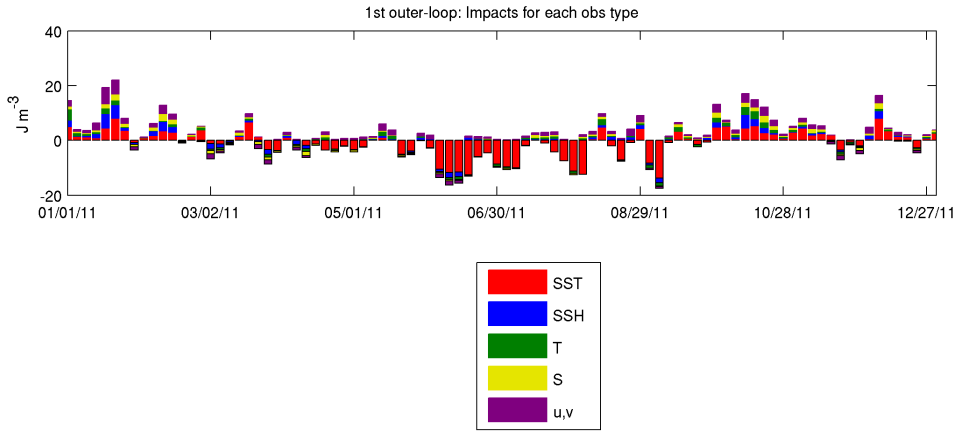


Figure 13: A summary of the impact of each observing platform on the energy density metric based on E_{APE}+E_{KE} for each 8-day 4D-Var cycle during 2011. In situ hydrographic observations are referred to in the legend as "T" and "S." Results shown are for the 1st outer-loop.

Metrics Based on Arbitrary Tracks and Sections using *interpolator.m*

All of the metrics considered so far have been based on simple targets such as boxes, rectangles, or grid-point masks. However, the ROMS *Matlab* routine *interpolator.m* (and its cousin *ad_interpolator.m*) allow us to consider more complex target shapes and circulation metrics. Specifically, *interpolator.m* will return any ROMS state variable interpolated along any arbitrary

track or section. A couple of examples will be presented here, in this case from the Rutgers *DOPPIO* configuration of ROMS for the Mid-Atlantic Bight (MAB). The model domain is shown in Fig. 14, and the target section considered here is highlighted in red, which corresponds to a short section of the 200 m isobath. The motivation for this target is to explore the impact of observations on circulation metrics focused on the MAB shelf-break front in the vicinity of the NSF OOI Pioneer Endurance Array.

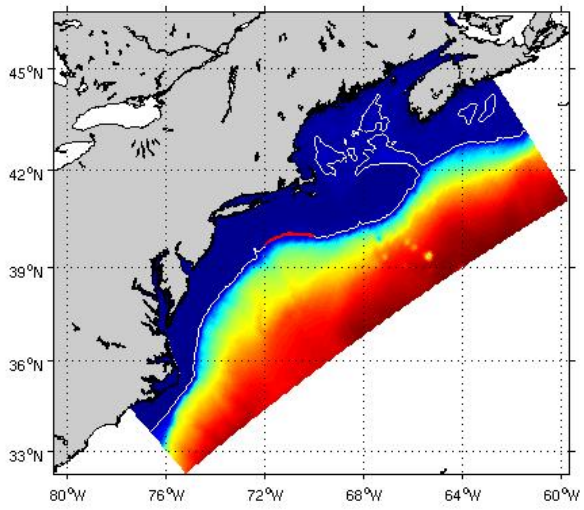


Figure 14: The ROMS Mid-Atlantic Bight and Gulf of Maine domain. The white contour shows the 200 m isobath (a useful proxy for the shelf-break), and the 200 m isobath section highlighted in red approximately coincides with the location of the NSF Ocean Observatories Initiative (OOI) Pioneer Endurance Array, and is used to define the transport metrics described here.

Two specific examples will be presented here, namely:

$$I = \frac{1}{\tau}\int_0^{\tau}\int_{200m}^{0}\oint u_n dsdzdt \tag{21}$$

$$I = \frac{1}{\rho c_p \tau}\int_0^{\tau}\int_{200m}^{0}\oint (u_n - \bar{u}_n)(T - \bar{T})dsdzdt. \tag{22}$$

The metric described by (21) is the time averaged cross-shelf volume transport across the 200m isobath target section in Fig. 14, where u_n now represents the component of velocity *normal* to the section (*not* be confused with u during outer-loop n). The contour integral $\oint ds$ is along the section of the 200m isobath highlighted in Fig. 14, and $\tau = N\Delta t$ is the length of the assimilation window. This metric is clearly linear and of the general form (3), where the elements of $\boldsymbol{h}$ are computed using the *Matlab* routines *romsgenslice.m* and *interpolator.m*. The metric described by (22), represents the time averaged cross-shelf eddy heat transport across the 200m isobath target section in Fig. 14, where an overbar denotes the seasonal mean climatology. This metric is clearly of the general form (4), where the elements of $\boldsymbol{E}$ are computed using *romsgenslice.m* and *interpolator.m*.

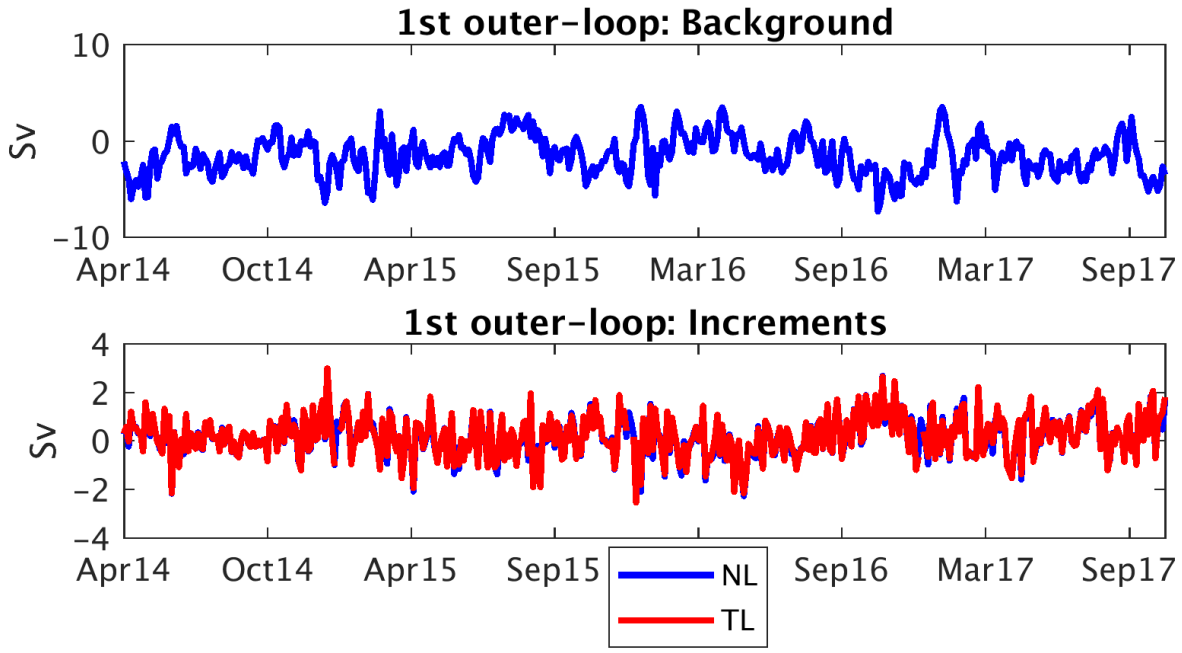


Figure 15: The upper panel shows time series of the cross-shelf volume transport metric (21) computed from the 4D-Var background (upper panel) during the period Apr 2014 to Dec 2017. The lower panel shows time series of cross-shelf volume transport metric increment ΔI_1 computed directly from the analysis and background solutions of the non-linear model (blue curve labelled NL), and δI_1 based on the tangent linear assumption in (9) (red curve labelled TL). Results shown are for the 1st outer-loop.

The examples presented in this case are for the period April 2014-Oct 2017 using 3-day 4D-Var analysis cycles. A time series of the background MAB cross-shelf volume transport metric (21) during the 1st outer-loop is shown in Fig. 15. The corresponding 4D-Var increments δI_1 in the cross-shelf transport computed directly from the analysis and background estimates of outer-loop 1, and using the tangent linear approximation (9) are also shown in Fig. 15. Clearly the tangent linear approximation of (21) performs very well. Similar results were obtained for the 2nd outer-loop (not shown).

To illustrate the level of detail that is possible in observation impact calculations, Fig. 16 shows the observation impacts on the cross-shelf volume transport metric (21) associated with each individual instrument within the OOI Pioneer Array. This level of detail is made possible by assigning a unique provenance to each instrument, in this case crawling CTD profilers and ocean gliders.

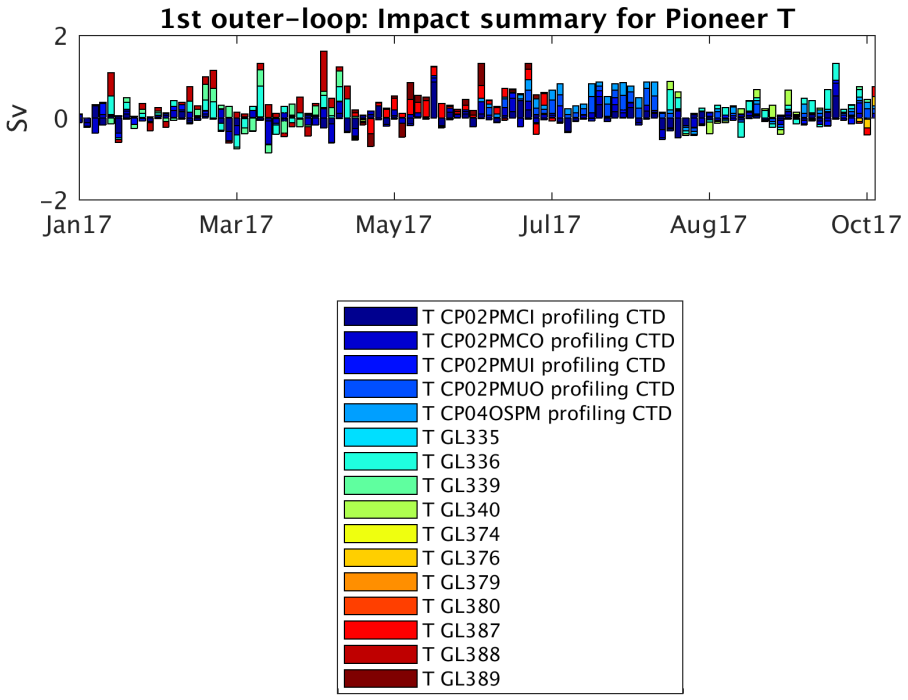


Figure 16: Time series of the observation impacts on the cross-shelf volume transport metric (21) associated with each individual instrument within the OOI Pioneer Endurance Array, during each 3-day 4D-Var analysis cycle in 2017. Two types of observing platforms were available: crawling CTDs profilers at five fixed mooring sites, and profiles from a constellation of gliders ("T GL").

A time series of the background MAB cross-shelf eddy heat transport metric (22) form the same 4 year period is shown in Fig. 17. Also shown are the cross-shelf eddy heat transport increments computed in the usual way. On the whole, the tangent linear approximation (10) of the metric (22) works well, and the two increment time series are highly correlated (0.89). However, there are several cycles where the amplitudes of the two increment time series different significantly, presumably due to non-linearity of both the circulation and the metric (22). Nevertheless, the agreement is good. The observation impacts on the cross-shelf heat transport metric (22) associated with each individual observing platform in the OOI Pioneer array are shown in Fig. 18 for completeness.

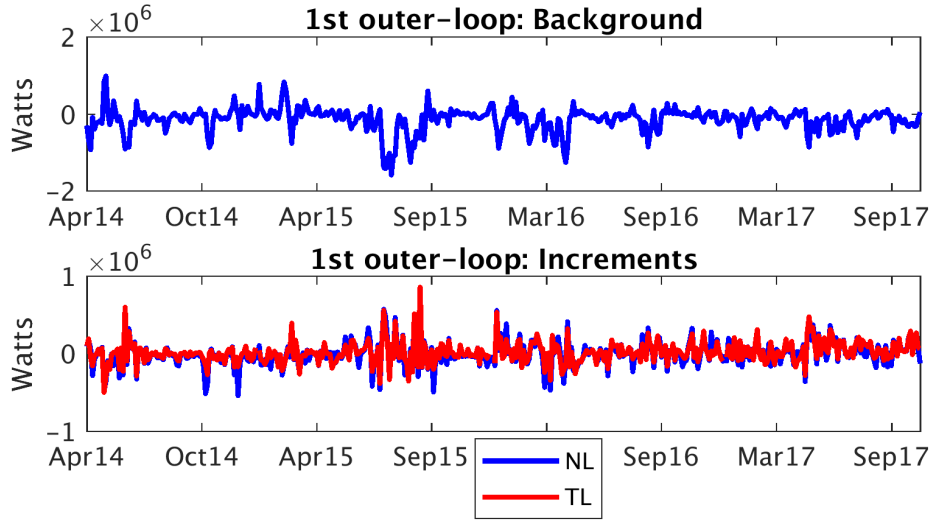


Figure 17: The upper panel shows time series of the cross-shelf eddy heat transport metric (22) computed from the 4D-Var background (upper panel) during the period Apr 2014 to Dec 2017. The lower panel shows time series of cross-shelf eddy heat transport metric increments ΔI_1 computed directly from the analysis and background solutions of the non-linear model (blue curve labelled NL), and δI_1 based on the tangent linear assumption in (10) (red curve labelled TL). Results shown are for the 1st outer-loop.

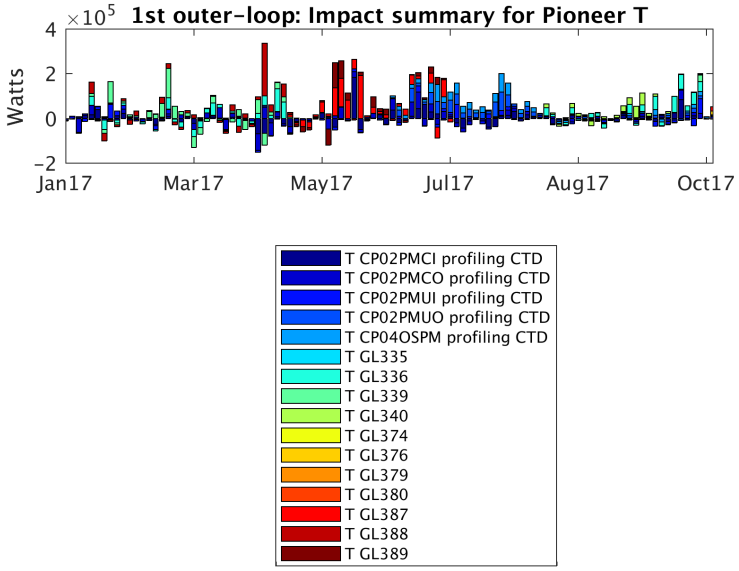


Figure 18: Time series of the observation impacts on the cross-shelf eddy heat transport metric (22) associated with each individual instrument within the OOI Pioneer Endurance Array, during each 3-day 4D-Var analysis cycle in 2017. Two types of observing platforms were available: crawling CTDs profilers at five fixed mooring sites, and profiles from a constellation of gliders ("T GL").

Practicalities

While some of the metrics used here may appear quite involved and complicated, the procedure to follow is *always* the same.

(1) Write a Matlab (or Python script) that computes the desired metric *I*.
(2) Tangent linearize, if necessary, the script in (1).
(3) Write the adjoint of the tangent linear script in (2), following the usual rules, and this will provide the forcing fields required by the adjoint model (*ADSname*) to compute the observation impacts of the metric in step (1).