Supplementary Material

Here, we explain how the steric height was estimated from the iterated integrals, how we computed the cost function gradient, and how we implemented Nesterov's acceleration scheme. We also show an example of changes in temperature-salinity (T-S) diagrams.

1 STERIC HEIGHT IN TERMS OF ITERATED INTEGRALS

Based on the state equation in the MRI.com manual (Tsujino et al., 2010), in-situ density has the following polynomial expression with respect to potential temperature T, salinity S, and depth P. By applying symbolic operations with `SymPy` (Meurer et al., 2017), we expand the logarithm as follows:

$$
\begin{aligned}
-\log(\rho) = \beta_0 &+ \beta_P \Delta P + \beta_S \Delta S + \beta_T \Delta T \\
&+ \beta_{P^2} \Delta P^2 + \beta_{S^2} \Delta S^2 + \beta_{T^2} \Delta T^2 \\
&+ \beta_{ST} \Delta S \Delta T + \beta_{SP} \Delta S \Delta P + \beta_{TP} \Delta T \Delta P \\
&+ O(X^3),
\end{aligned}
\tag{S1}
$$

where $\Delta P := P_u - P_0$, $\Delta S := S_u - S_1 = S_u - 34.7$, $\Delta T := T_u - T_1 = T_u - 2.65$, $X = P, S, T$, and the coefficients (in cm, psu, K units) are

$$
\begin{aligned}
\beta_0 &= -2.72 \cdot 10^{-2}, & \beta_P &= -4.47 \cdot 10^{-8}, \\
\beta_S &= -7.77 \cdot 10^{-4}, & \beta_T &= 8.50 \cdot 10^{-5}, \\
\beta_{P^2} &= 6.03 \cdot 10^{-15}, & \beta_{S^2} &= 2.25 \cdot 10^{-7}, \\
\beta_{T^2} &= 6.08 \cdot 10^{-6}, & \beta_{ST} &= 2.72 \cdot 10^{-6}, \\
\beta_{SP} &= 1.15 \cdot 10^{-10}, & \beta_{TP} &= 2.59 \cdot 10^{-10}.
\end{aligned}
$$

The reference point of the above Taylor expansion was set to a typical value of $(S_1, T_1) = (34.7, 2, 65)$, at 2000m.

The vertical integral from 2000m to the sea level in Eq. (S1) reads

$$
\begin{aligned}
\int_0^1 -\log(\rho_u) dP_u = \beta_0(P_1 - P_0) &+ \beta_P \frac{1}{2}(P_u - P_0)^2 + \beta_S \int_0^1 (S_u - S_1) dP_u \\
&+ \beta_T \int_0^1 (T_u - T_1) dP_u + \beta_{P^2} \frac{1}{3}(P_1 - P_0)^3 + \beta_{S^2} \int_0^1 (S_u - S_1)^2 dP_u \\
&+ \beta_{T^2} \int_0^1 (T_u - T_1)^2 dP_u + \beta_{ST} \int_0^1 (S_u - S_1)(T_u - T_1) dP_u \\
&+ \beta_{SP} \int_0^1 (S_u - S_1)(P_u - P_0) dP_u + \beta_{TP} \int_0^1 (T_u - T_1)(P_u - P_0) dP_u \\
&+ O(X^4),
\end{aligned}
\tag{S2}
$$

where $X = P, S, T$.

Using the shuffle relation $\mathcal{I}^{A}\mathcal{I}^{B} = \mathcal{I}^{A \sqcup\!\sqcup B}$ for any multiindices A and B (e.g., Lyons et al., 2007), each integral term in Eq. (S2) can be translated into a canonical form:

$$\int_0^1 (T_u - T_1) dP_u = -\mathcal{I}^{PT}, \tag{S3}$$

$$\int_0^1 (T_u - T_1)(S_u - S_1) dP_u = \mathcal{I}^{PST} + \mathcal{I}^{PTS}, \tag{S4}$$

$$\int_0^1 (T_u - T_1)^2 dP_u = 2\mathcal{I}^{PTT}, \tag{S5}$$

$$\int_{0 \le u \le 1} (T_u - T_1)(P_u - P_0) dP_u = -\mathcal{I}^{PPT}. \tag{S6}$$

where the canonical form of the iterated integrals is expressed as $\mathcal{I}^{PST} := \int_{0 \le u_1 \le u_2 \le u_3 \le 1} dP_{u_1} dS_{u_2} dT_{u_3}$. Plugging Eqs. (S3)–(S6) into Eq. (S2), The expansion at each horizontal point m is expressed as a linear combination of the iterated integrals:

$$\begin{aligned} h_m := \overline{-\int_0^1 \log(\rho_u) dP_u}^{m} &= -\beta_S \overline{\mathcal{I}^{PS}}^{m} - \beta_T \overline{\mathcal{I}^{PT}}^{m} + 2\beta_{S^2} \overline{\mathcal{I}^{PSS}}^{m} + 2\beta_{T^2} \overline{\mathcal{I}^{PTT}}^{m} \\ &\quad + \beta_{ST} \left(\overline{\mathcal{I}^{PST}}^{m} + \overline{\mathcal{I}^{PTS}}^{m} \right) - \beta_{SP} \overline{\mathcal{I}^{PPS}}^{m} - \beta_{TP} \overline{\mathcal{I}^{PPT}}^{m} + C, \end{aligned} \tag{S7}$$

where $\overline{\mathcal{I}^{\bullet}}^{m}$ denotes the average over the iterated integrals of profiles in mesh m, and C is a constant along time.

Finally, the global mean steric sea level was calculated as an area-weighted average:

$$\bar{h} = \frac{\sum_m h_m A_m}{\sum_m A_m}, \tag{S8}$$

where A_m is the geographical area of mesh m. Steric height anomaly is a deviation from the global mean:

$$\Delta h_m = h_m - \bar{h}. \tag{S9}$$

Figure S1 shows that the estimated steric height anomaly mostly explains the sea surface height anomaly in a region with a depth greater than 2000m, which confirms the validity of the estimations using Eq. (S7).

2 GRADIENT OF THE COST FUNCTION

Recall our cost function:

$$J(\psi) = \frac{1}{2}(\psi - \psi_b)^\top B^{-1} (\psi - \psi_b) + \lambda \left(\sum_{\text{observed } m} J_m(\psi) + \sum_{\text{model } m} J_{m,\text{cyc}}(\psi) \right). \tag{S10}$$

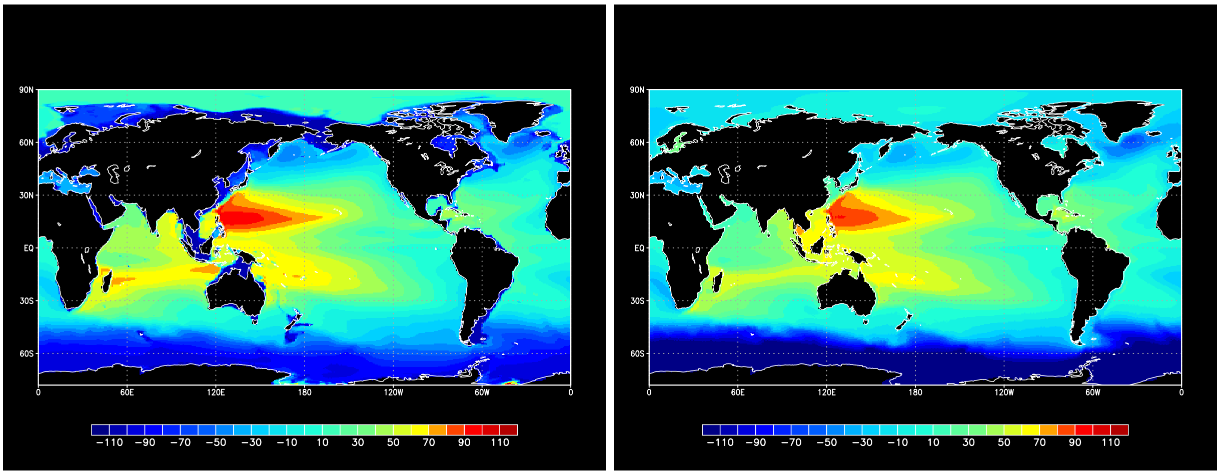

Figure S1. The temporal averages in cm of steric height minus global mean (left) , compared to that of sea surface height minus global mean (right) for Sig-case.

By changing variable $\psi = B^{\frac{1}{2}}\phi + \psi_b$, the original cost function is rewritten with respect to ϕ as

$$\mathcal{J}(\phi) = \frac{1}{2}\phi^\top\phi + \lambda\left(\sum_{\text{observed } m} J_m(B^{\frac{1}{2}}\phi + \psi_b) + \sum_{\text{model } m} J_{m,\text{cyc}}(B^{\frac{1}{2}}\phi + \psi_b)\right). \tag{S11}$$

The gradient of $\mathcal{J}$ with respect to ϕ is

$$\left(\frac{\partial \mathcal{J}}{\partial \phi}(\phi)\right)^\top = \phi + \lambda B^{\frac{\top}{2}}\left[\sum_{\text{observed } m}\left(\frac{\partial J_m}{\partial \psi}(\psi)\right)^\top + \left(\sum_{\text{model } m}\frac{\partial J_{m,\text{cyc}}}{\partial \psi}(\psi)\right)^\top\right], \tag{S12}$$

which can be used as a preconditioned gradient in the gradient method. In Eq. (S12), the gradient of the observational cost (similarly, that of the cyclicity cost) is derived as follows. The observational cost can be rewritten as

$$J_m(\psi) = \frac{1}{2}\sum_{\tau\in T_m}\sum_{k=1}^{n}\left(\sum_{i_1,\cdots,i_k} F_{m,\tau}^{(i_1\cdots i_k)}(\psi)^2\right)^{\frac{1}{k}}, \tag{S13}$$

$$F_{m,\tau}^{(i_1\cdots i_k)}(\psi) := \frac{1}{|M_{m,\tau}|}\sum_{X\in M_{m,\tau}} S_k^{(i_1\cdots i_k)}(X(\psi)) - \frac{1}{|N_{m,\tau}|}\sum_{Y\in N_{m,\tau}} S_k^{(i_1\cdots i_k)}(Y), \tag{S14}$$

where $\mathcal{S}^{(i_1\cdots i_k)}(Y) = \int_{0\le u_1\le\cdots\le u_k\le 1} dY_{u_1}^{(i_1)}\cdots dY_{u_k}^{(i_k)}$. By differentiating with respect to ψ and then taking the transpose, the gradient is obtained as follows:

$$\left(\frac{\partial J_m}{\partial \psi}(\psi)\right)^{\top} = \sum_{\tau\in T_m}\sum_{k=1}^{n}\frac{1}{k}\left(\sum_{i_1,\cdots,i_k} F_{m,\tau}^{(i_1\cdots i_k)}(\psi)^2\right)^{\frac{1}{k}-1}\sum_{i_1,\cdots,i_k} F_{m,\tau}^{(i_1\cdots i_k)}(\psi)\left(\frac{\partial F_{m,\tau}^{(i_1\cdots i_k)}}{\partial\psi}(\psi)\right)^{\top}, \tag{S15}$$

$$\left(\frac{\partial F_{m,\tau}^{(i_1\cdots i_k)}}{\partial\psi}(\psi)\right)^{\top} = \frac{1}{|M_{m,\tau}|}\sum_{X\in M_{m,\tau}}\left(\frac{\partial X}{\partial\psi}(\psi)\right)^{\top}\left(\frac{\partial \mathcal{S}_k^{(i_1\cdots i_k)}}{\partial X}(X)\right)^{\top}, \tag{S16}$$

where $\left(\frac{\partial X}{\partial\psi}(\psi)\right)^{\top}$ is the adjoint of the OGCM, and $\left(\frac{\partial \mathcal{S}_k^{(i_1\cdots i_k)}}{\partial X}(X)\right)^{\top}$ is the adjoint of signature transform.

3 NESTEROV ACCELERATION

Nesterov's acceleration method (Nesterov, 1983) is effective in enhancing the speed of iterations in the gradient method. Our implementation, Algorithm 1, is a robust setting with multiple line search trials and is intended to work even if the gradient is inaccurate. The following notations were used:

ϕ: control variables, $\mathcal{J}$: cost function, $\nabla\mathcal{J}$: gradient of the cost function, k: number of iterations, ℓ: number of line searches, $\gamma_0 > 0$: initial increment value, and $r_{k+1} > 0$: Nesterov acceleration value at iteration $k \ge 0$, which is defined by the following recurrence relation:

$$r_{k+1} = \frac{\rho_k - 1}{\rho_{k+1}}, \tag{S17}$$

$$\rho_{k+1} = \frac{1+\sqrt{1+4\rho_k^2}}{2}. \tag{S18}$$

with $\rho_0 = 1$.

4 EXAMPLE OF T-S DIAGRAM

To gain an intuitive understanding of the changes in the temperature-salinity (T-S) plane due to data assimilation, Fig.,S2 presents a comparison of T-S diagrams as an illustrative example. The TS-case aims to align the point-by-point values more closely, while the Sig-case performs better in terms of similarity statistics, specifically the integrated integrals.

REFERENCES

Lyons, T. J., Caruana, M., and Lévy, T. (2007). *Differential Equations Driven by Rough Paths*, vol. 1908 of *Lecture Notes in Mathematics* (Berlin, Heidelberg: Springer)

Meurer, A., Smith, C. P., Paprocki, M., Čertík, O., Kirpichev, S. B., Rocklin, M., et al. (2017). Sympy: symbolic computing in python. *PeerJ Computer Science* 3, e103. doi:10.7717/peerj-cs.103

Nesterov, Y. E. (1983). A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Doklady Akademii Nauk* 269, 543–547

Algorithm 1 Nesterov Acceleration

function NESTEROV(ϕ)
  Set Firstguess: $\widehat{\phi}_0 = \phi_0 \leftarrow \phi$
  for all $k = 0, 1, \cdots, K$ **do**
    Forward and Adjoint Integrations:

$$\widehat{c}_k \leftarrow \mathcal{J}(\widehat{\phi}_k) \tag{S19}$$

$$g_k \leftarrow \nabla \mathcal{J}(\widehat{\phi}_k) \tag{S20}$$

    for all $\ell = 0, 1, \cdots, L$ **do**
      Line Search:

$$\eta_\ell := (-1)^\ell \theta^{\frac{\ell}{2}} \gamma_0 \tag{S21}$$

$$c_{k+1,\ell} \leftarrow \mathcal{J}(\widehat{\phi}_k - \eta_\ell g_k) \tag{S22}$$

      if $c_{k+1,\ell} < \widehat{c}_k$ **then**
        break
      end if
    end for
    $\phi_{k+1} \leftarrow \widehat{\phi}_k - \eta_\ell g_k$
    if $c_{k+1,\ell} < c_k$ **then**
      New cost: $c_{k+1} \leftarrow c_{k+1,\ell}$
      Acceleration:
      if $\eta_\ell > 0$ **then**

$$\widehat{\phi}_{k+1} \leftarrow \phi_{k+1} + r_{k+1}(\phi_{k+1} - \phi_k) \tag{S23}$$

      else

$$\widehat{\phi}_{k+1} \leftarrow \phi_{k+1} \tag{S24}$$

      end if
    else
      Status quo: $c_{k+1} \leftarrow c_k$

$$\phi_{k+1} \leftarrow \phi_k \tag{S25}$$

$$\widehat{\phi}_{k+1} \leftarrow \phi_k + r_{k+1}(\phi_k - \phi_{k-1}) \tag{S26}$$

    end if
  end for
  return $\phi \leftarrow \phi_{K+1}$
end function

Tsujino, H., Motoi, T., Ishikawa, I., Hirabara, M., Nakano, H., Yamanaka, G., et al. (2010). *Reference manual for the Meteorological Research Institute COMmunity ocean model (MRI.COM) version 3.* Tech. rep., Meteorological Research Institute. doi:10.11483/mritechrepo.59

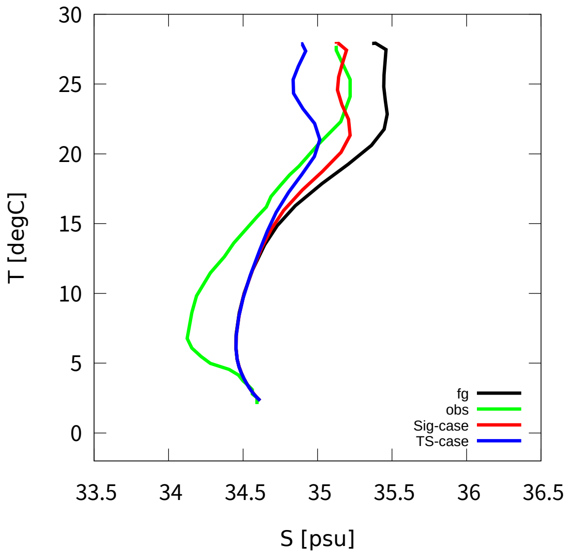

Figure S2. Comparison of T-S diagrams in the subtropical region. The spatio-temporal average values in December 2013 from 157 to 159 degrees east and 19 to 21 degrees north are shown for observation (green), firstguess (black), Sig-case (red), and TS-case (blue). The units are practical salinity unit (psu) for salinity and degree Celsius for temperature.