Ocean Data Assimilation Focusing on Integral Quantities Characterizing Observation Profiles

Nozomi Sugiura

Japan Agency for Marine-earth Science and TECnology

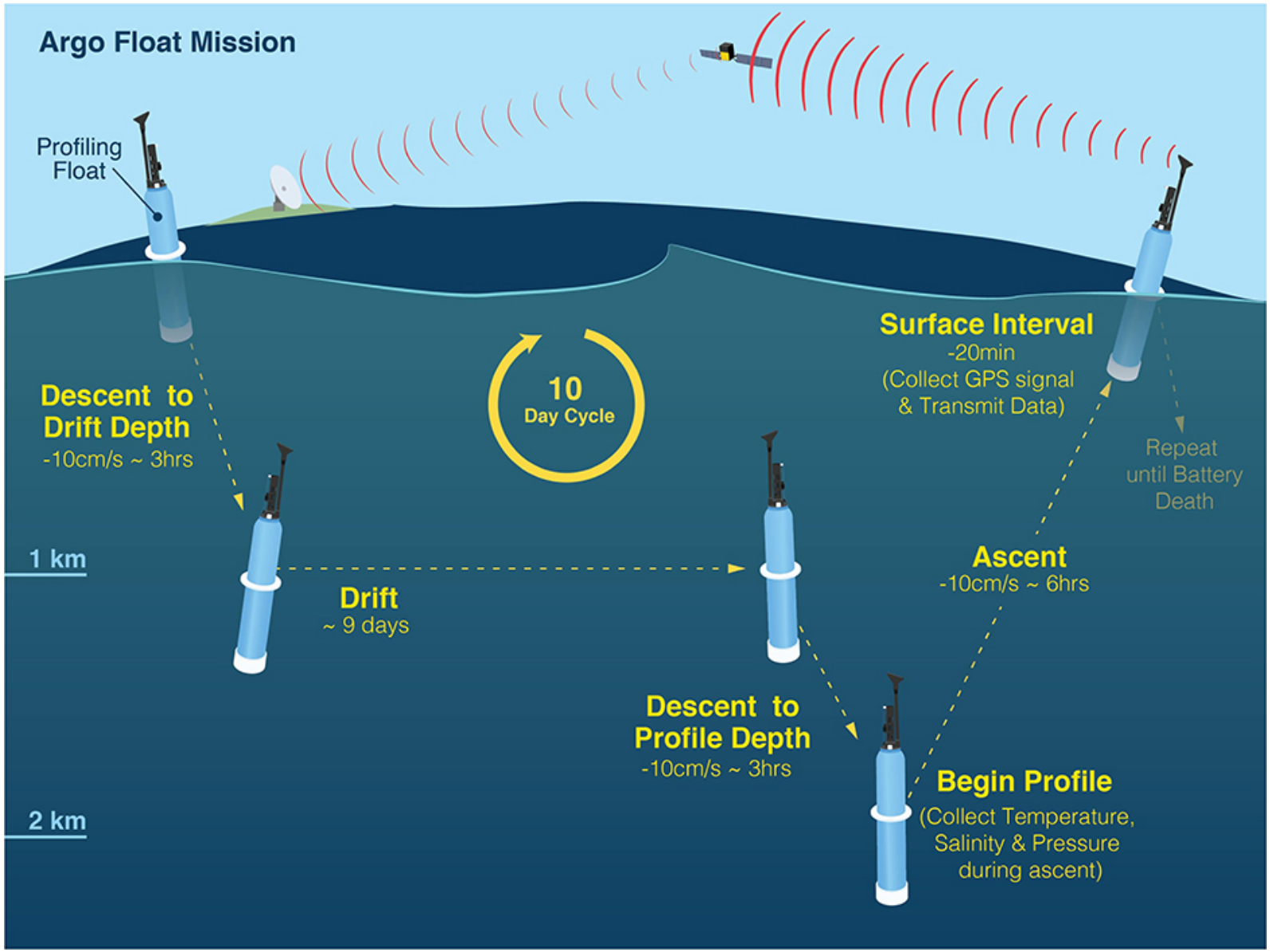
Apr. 2024

Background

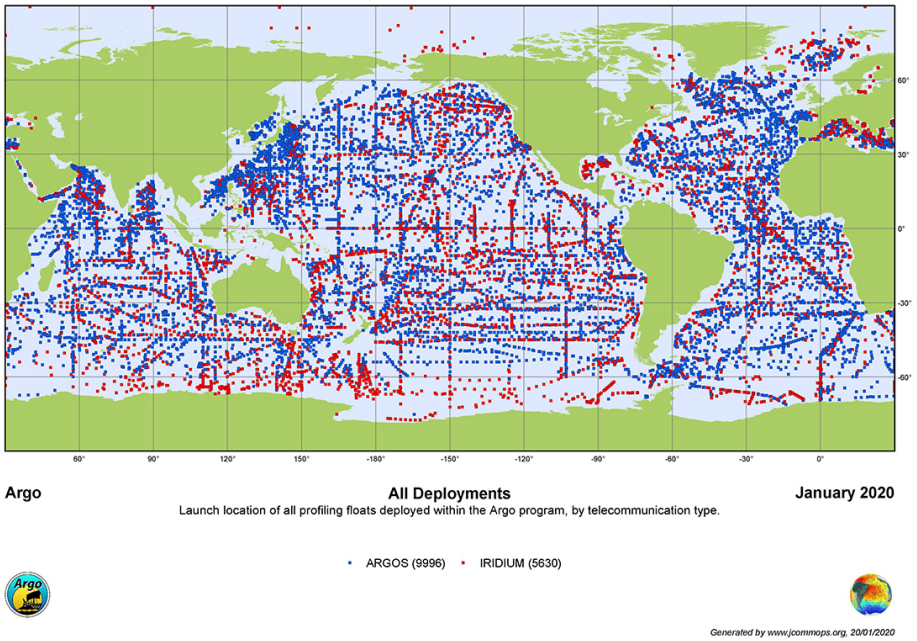
- Ocean circulation is described by the primitive equations, which consist of the three-dimensional Navier-Stokes equations under the Coriolis force and the advection-diffusion equations for salinity and temperature, along with an algebraic equation for density, $\rho = f(P, S, T)$.
- Long-term oceanic variations (ranging from several years to decades) have been studied thorough ocean data assimilation using the 4-dimensional variational method with a long assimilation window. Typically, oceanic initial condition and the time series of atmospheric fluxes are employed as control variables.
- On the other hand, the observed data to be assimilated consist mainly of sea-surface data, such as satellite data, and vertical profile data, such as from Argo array (as shown below) or ship observations. Observed elements are mostly temperature (T), salinity (S), and pressure (P).
- Traditionally, when we assimilate a vertical profile at a certain horizontal position, the salinity and temperature and at each depth (i.e., pressure P_k) have been regarded as a point, (S_k, T_k).
- In this study, we explore the possibility of extracting more information from observation data by treating a profile as a function: $[0, 1] \ni u \mapsto (P_u, S_u, T_u) \in \mathbb{R}^3$,

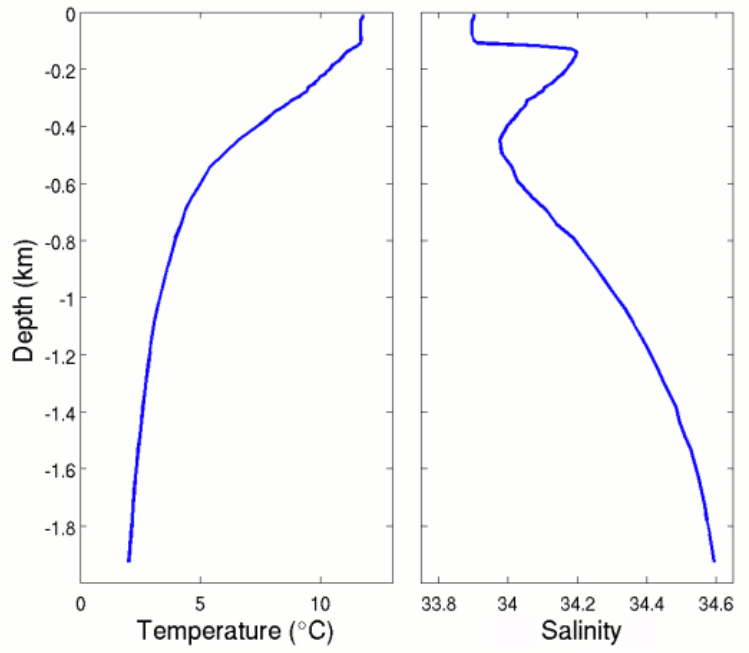
.

Argo array retrieves the profiles of temperature, salinity, and pressure in the global ocean (Wong et al., 2020)



Argo array retrieves the profiles of temperature, salinity, and pressure in the global ocean (Wong et al., 2020)





The signature represents a path as an element in tensor algebra (Lyons et al., 2007)

- The signature of a continuous path $X = (X^{(1)}, \dots, X^{(d)}) : [0, \tau] \to \mathbb{R}^d$ is a collection of real numbers indexed by "words" composed of the alphabet $\{1, \dots, d\}$. The term corresponding to the word $w = w_1 \dots w_k$ is given by

$$\mathcal{S}(X)^{(w)}_{0,\tau} := \langle \mathcal{S}(X)_{0,\tau}, w \rangle = \int_{0 \le t_1 \le \dots \le t_k \le \tau} \mathrm{d}X^{(w_1)}_{t_1} \dots \mathrm{d}X^{(w_k)}_{t_k} \in \mathbb{R}, \tag{1}$$

  which is called a k-th iterated integral. The signature is assumed to be truncated up to the n-th order.

- For example, when $d = 2$, the signature up to the 2nd order is

$$\mathcal{S}(X)_{0,1} = 1 + \begin{bmatrix} \int_{0 \le t_1 \le 1} dX^{(1)}_{t_1} \\ \int_{0 \le t_1 \le 1} dX^{(2)}_{t_1} \end{bmatrix} + \begin{bmatrix} \int_{0 \le t_1 \le t_2 \le 1} dX^{(1)}_{t_1} dX^{(1)}_{t_2} & \int_{0 \le t_1 \le t_2 \le 1} dX^{(1)}_{t_1} dX^{(2)}_{t_2} \\ \int_{0 \le t_1 \le t_2 \le 1} dX^{(2)}_{t_1} dX^{(1)}_{t_2} & \int_{0 \le t_1 \le t_2 \le 1} dX^{(2)}_{t_1} dX^{(2)}_{t_2} \end{bmatrix}$$
$$\in \mathbb{R} \oplus \mathbb{R}^2 \oplus (\mathbb{R}^2 \otimes \mathbb{R}^2) = \bigoplus_{k=0}^{2} (\mathbb{R}^2)^{\otimes k}. \tag{2}$$

- Signatures are closed with respect to the product in the tensor algebra[1]. This enables the construction of a homomorphism between the concatenation operation of paths and the product of signatures (Chen's identity):

$$\mathcal{S}(X)_{0,1} = \mathcal{S}(X)_{0,\sigma} \otimes \mathcal{S}(X)_{\sigma,1}. \tag{3}$$

- In the case of Argo profiles, the 3 letters are $\{P, S, T\}$ (pressure, salinity, temperature), and we assume that the series is truncated up to the 4-th order.

[1] ex. $a \otimes b = (1 + a_1 + a_2) \otimes (1 + b_1 + b_2) = 1 + (a_1 + b_1) + (a_2 + a_1 \otimes b_1 + b_2)$.

Signatures captures a profile from various angles, and measure the areas viewed from each of them

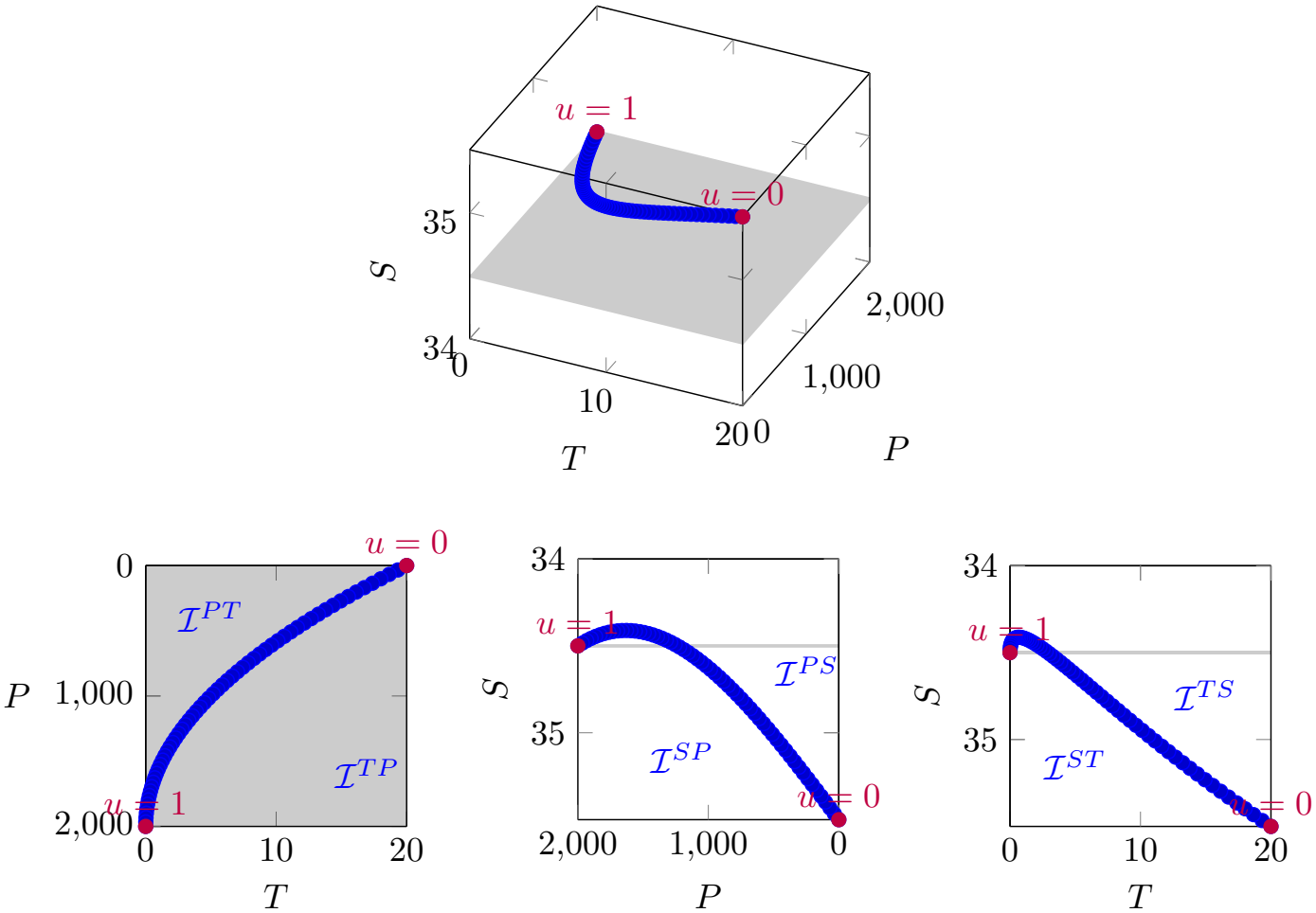


Figure: Grasping the shape of a profile $[0, 1] \ni u \mapsto (P_u, S_u, T_u) \in \mathbb{R}^3$ by the second-order iterated integrals, $(\mathcal{I}^{PT}, \mathcal{I}^{TP}, \mathcal{I}^{PS}, \mathcal{I}^{SP}, \mathcal{I}^{ST}, \mathcal{I}^{TS})$, in signature.

⋆ e.g., $\mathcal{I}^{PS} := \mathcal{S}(X)^{(PS)}_{0,1}$.

For each mesh, we compare model and observation with respect to the arithmetic mean of signatures

- Each mesh, $(m_x, m_y, t) = (m, t)$, covers a range of 1 degree longitude, 0.5 degree latitude, and 1 month.
- Since the expectated value of the signatures serves as a mean embedding of probability measures (Chevyrev and Oberhauser, 2022), the distance for comparing the empirical measures $\tilde{P}_m, \tilde{Q}_m$ of the model and observations can be set as follows.

$$\mathrm{MMD}^2(\tilde{P}_{m,t}, \tilde{Q}_{m,t}) = \left\| \frac{1}{|x_{m,t}|} \sum_{X \in x_{m,t}} \mathcal{S}(X) - \frac{1}{|y_{m,t}|} \sum_{Y \in y_{m,t}} \mathcal{S}(Y) \right\|_{\mathcal{T}}^2 . \tag{4}$$

  Here, the norm is the sum of the Frobenius norms corresponding to the projections π_k onto each order: $\|a\|_{\mathcal{T}}^2 := \sum_{k=1}^{n} (|\pi_k(a)|_F^2)^{1/k}$. Furthermore, $x_{m,t}$ and $y_{m,t}$ represent the sets of profile data from the model and observations, respectively, and $|\quad|$ denotes the number of elements.
- For mesh m, let T_m be the set of calendar months that have at least one observation profile. The cost function can then be set as follows.

$$J_m(\psi) = \frac{1}{2} \sum_{t \in T_m} \mathrm{MMD}^2(\tilde{P}_{m,t}, \tilde{Q}_{m,t}), \tag{5}$$

$$J(\psi) = \frac{1}{2} (\psi - \psi_b)^\top B^{-1} (\psi - \psi_b) + \lambda \left(\sum_{\text{observed } m} J_m(\psi) + \sum_{\text{model } m} J_{m,\mathrm{cyc}}(\psi) \right), \tag{6}$$

  where ψ is the control vector consisting of initial condition and sea surface fluxes, B is the corresponding background error covariance matrix, $\lambda > 0$ is the weight for the observational term, and $J_{m,\mathrm{cyc}}$ is a signature-based cyclicity cost ensuring that the annual mean fields at the beginning and end of the computation are similar.

Data Assimilation Experiment

- The model is the Meteorological Research Institute Ocean General Circulation Model, which is coupled with a sea ice model (Tsujino et al., 2010).
- The spatial resolution is 1 degree in longitude, 0.5 degrees in latitude, and 55 layers in depth. The period of data assimilation spans 10 years, from January 2004 to December 2013.
- The firstguess of atmospheric fluxes are the daily mean fluxes from the Japan Meteorological Agency's JRA-25.
- The control variables are the ocean's initial condition and increments of atmospheric fluxes (5-day average).
- The assimilated data consist of profiles (pressure, salinity, temperature) obtained by Argo, with each profile's pressure ranging from 0 to 2000m.
- The data assimilation method used is the four-dimensional variational method.
- The acceleration method by Nesterov is used as the gradient method.
- The comparative experiment involves a case using the signature-based cost (Eq. (4); Sig case) and a case using the temperature-salinity-based cost at each depth (Eq. (7); TS case). In the latter, the model $X = (T_{\text{model}}, S_{\text{model}})^\top$ and observations $Y = (T_{\text{obs}}, S_{\text{obs}})^\top$ are compared at each vertical level P_k:

$$d^2(x_{m,t}, y_{m,t}) = \sum_k \left| \frac{1}{|x_{m,t}|} \sum_{X \in x_{m,t}} X|_{P_k} - \frac{1}{|y_{m,t}|} \sum_{Y \in y_{m,t}} Y|_{P_k} \right|^2 . \tag{7}$$

Variation of Cost Functions (Top: Signature-Based, bottom: Temperature-Salinity-Based)

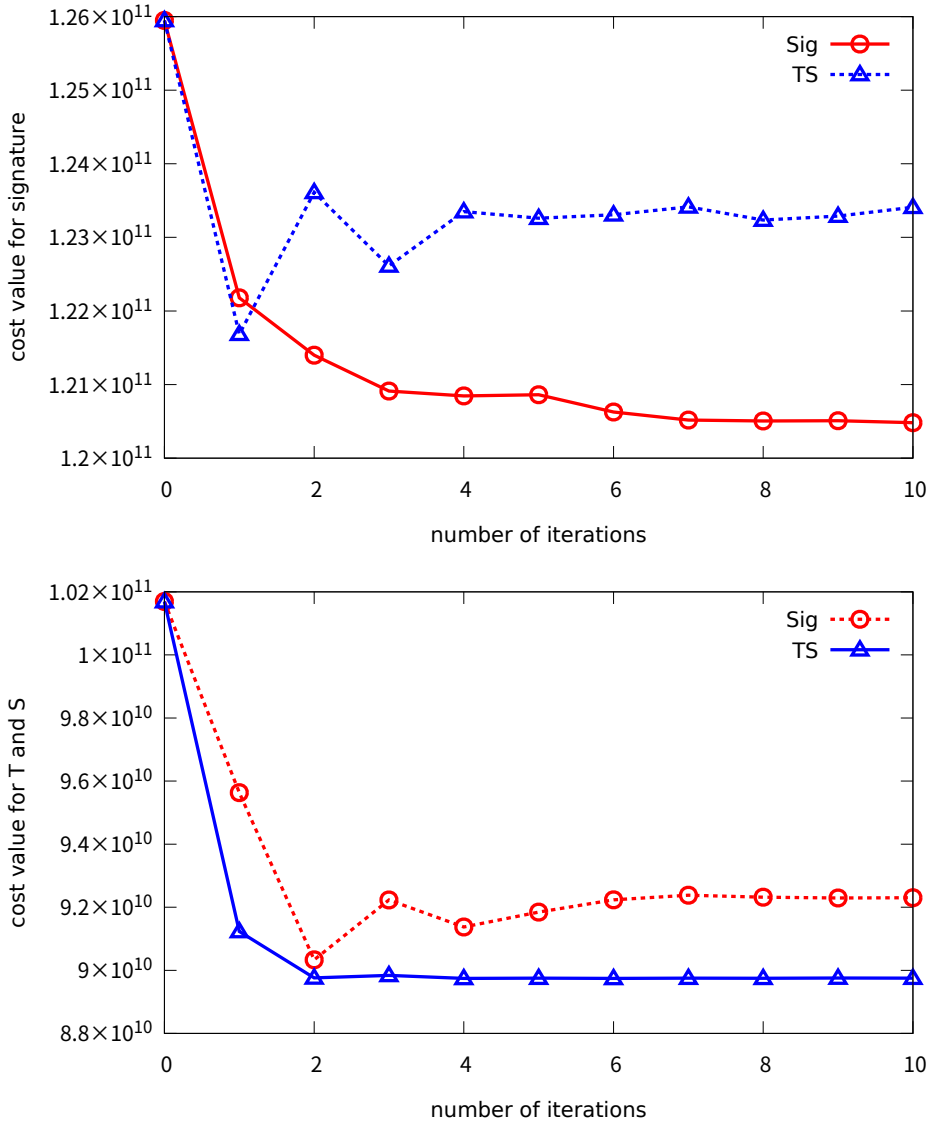


Figure: Variation of the cost functions in terms of signature-based cost (left) and TS-based cost (right). Sig-case is shown by red circles, TS-case by blue triangles.

Breakdown of Cost Function Reduction (Relative Cost for Each Iterated Integral)

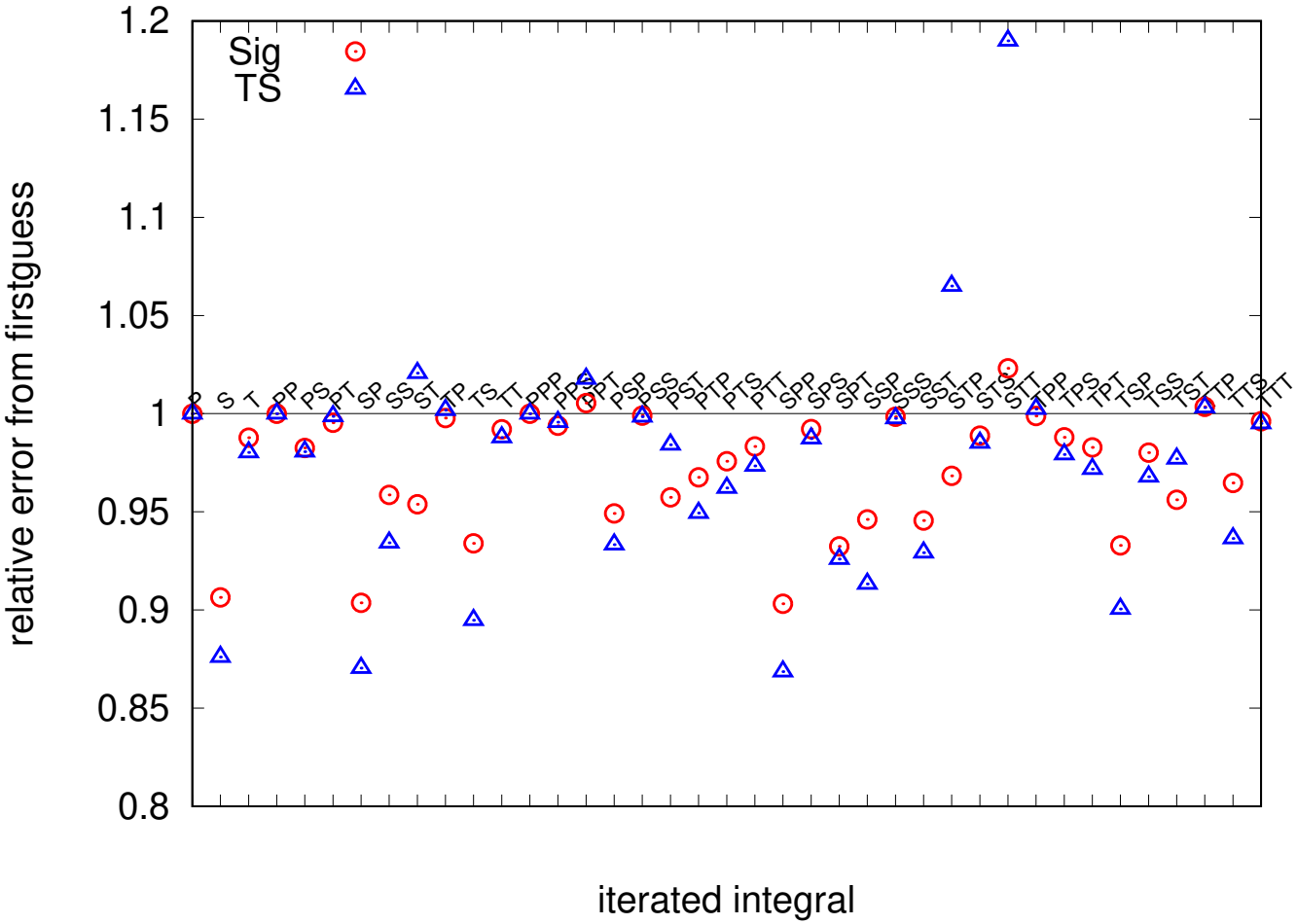
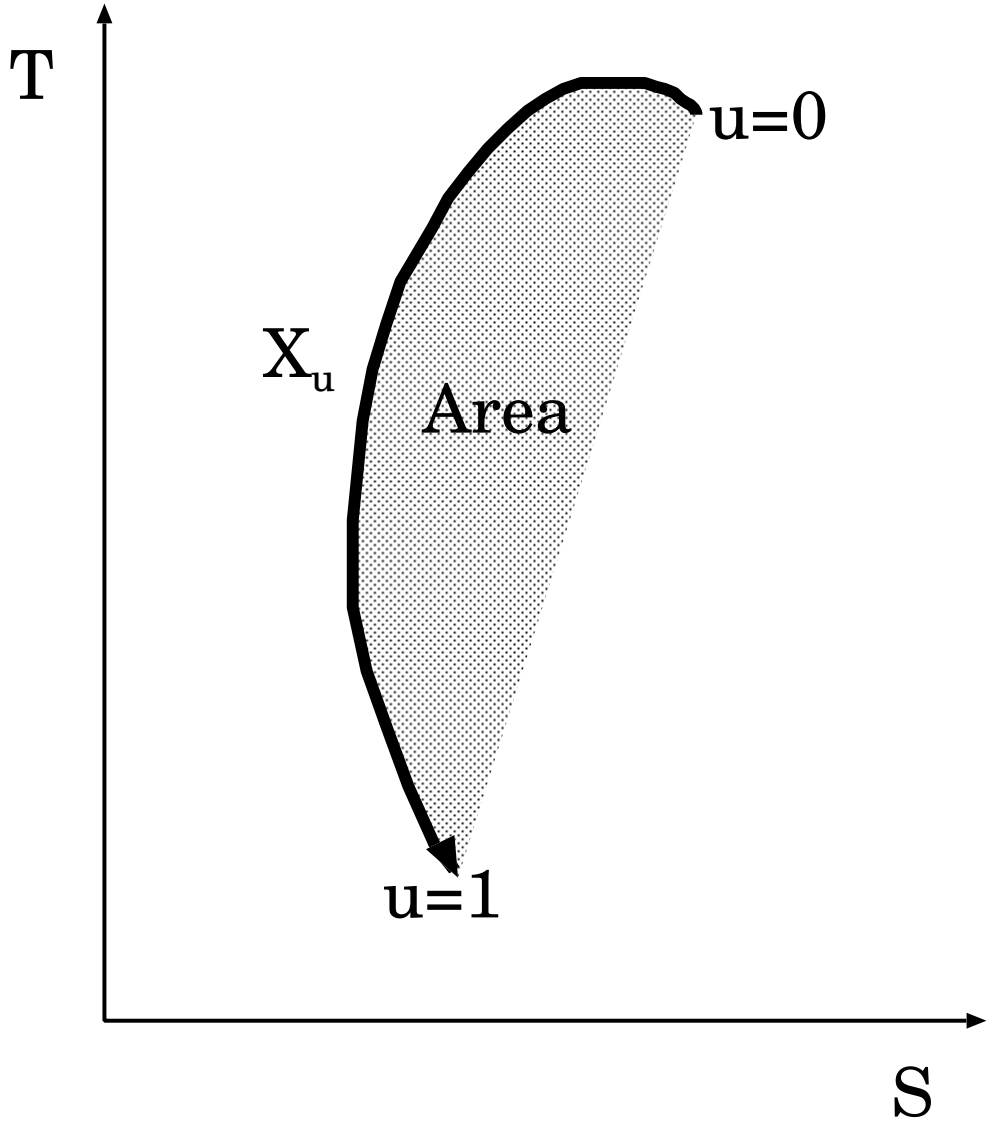


Figure: Relative error to firstguess for each iterated integral. Sig-case is denoted by red circles and TS-case by blue triangles. Horizontal axis is the index of iterated integral, for example, TS corresponds to $\mathcal{S}(X)_{0,1}^{(TS)}$.

Regarding words ST, STP, and STT, the cost in the TS case has deteriorated from the firstguess.

Lévy area on the T-S Plane

In each horizontal mesh, each profile defines:



Spatial Distributions of Relative Errors in the Lévy area on the T-S Plane (Left: Sig Case, Right: TS Case)

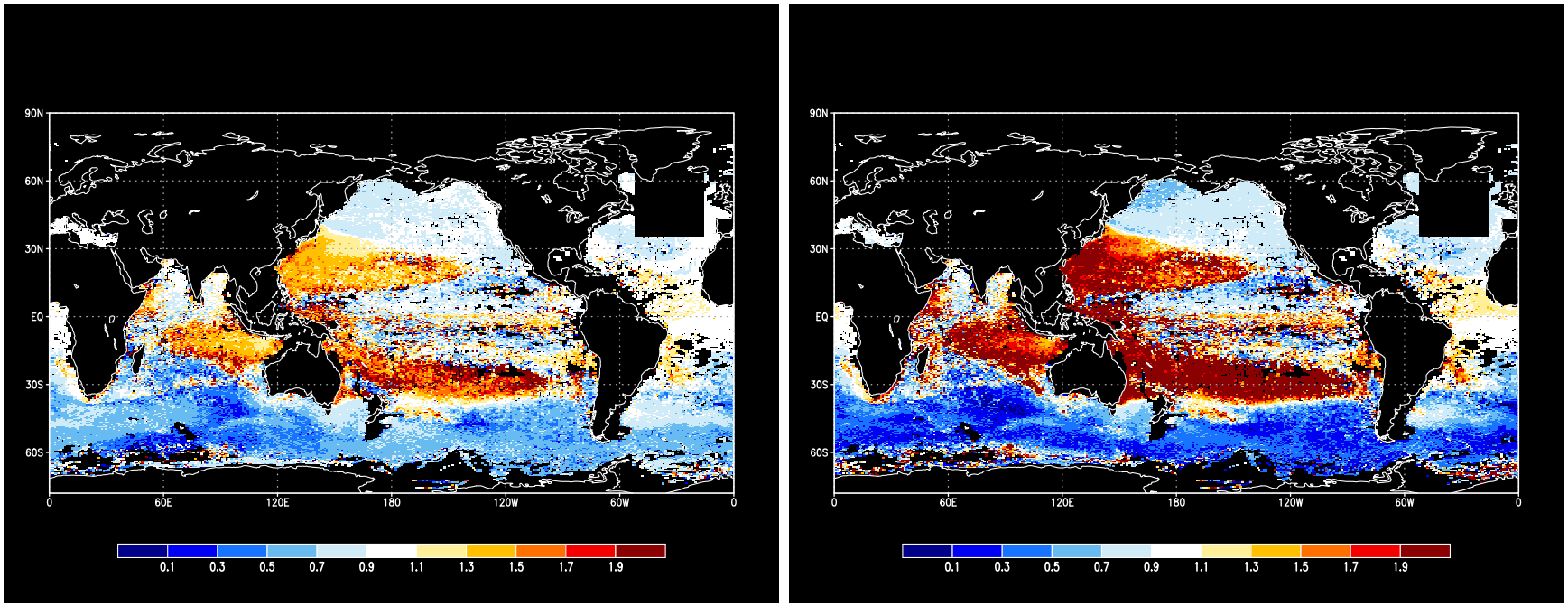


Figure: Relative observational error of TS-area, $(\mathcal{S}(X)^{(ST)} - \mathcal{S}(X)^{(TS)})/2$, to firstguess. Overall relative error is 0.967 for Sig-case (left) and 1.027 for TS-case (right). The colder the color, the better the improvement.

Conclusion

- In this study, we explored the possibility of extracting more information from observation data by considering each profile as a function: $[0, 1] \ni u \mapsto (P_u, S_u, T_u) \in \mathbb{R}^3$.
- In the ocean data assimilation experiment, the comparison was made between the case using a signature-based cost (Sig case) and the case using a temperature-salinity-based cost at each depth (TS case).
- While the changes in the cost functions for both cases are similar, the result regarding each iterated integral sometimes differ depending on whether signature-based or temperature-salinity-based costs are used.
- In particular, in the TS case, there was little improvement in the profile shapes when viewed from the T-S plane, and in some cases, there was a significant deterioration compared to the firstguess.
- On the other hand, although the Sig case showed slightly worse relative errors regarding several iterated integrals compared to the TS case, we found almost no deteriorations from the firstguess, including the terms viewed from the T-S plane.
- This suggests that the Sig case generally yields a more balanced analysis field. Notably, an improvement in the profile shapes on the T-S plane, which is crucial for water mass analysis, is a distinctive feature of the Sig case.
- In other words, in the traditional comparison based on matching temperature and salinity at each depth (TS case), some information in the observational profile seems to have been overlooked.

Acknowledgements: This study was supported by the JST, AIP Japan-Germany-France AI Research, under the grant number JPMJCR20G5.

References

Chevyrev, I. and Oberhauser, H. (2022). Signature moments to characterize laws of stochastic processes. *The Journal of Machine Learning Research*, 23(1):7928–7969.

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

Tsujino, H., Motoi, T., Ishikawa, I., Hirabara, M., Nakano, H., Yamanaka, G., Yasuda, T., and Ishizaki, H. (2010). Reference manual for the meteorological research institute community ocean model (mri.com) version 3. Technical report, Meteorological Research Institute.

Wong, A. P. S., Wijffels, S. E., Riser, S. C., Pouliquen, S., Hosoda, S., Roemmich, D. n., Gilson, J., Johnson, G. C., Martini, K., Murphy, D. J., Scanderbeg, M., Bhaskar, T. V. S. U., Buck, J. J. H., Merceur, F., Carval, T., Maze, G., Cabanes, C., André, X., Poffa, N., Yashayaev, I., Barker, P. M., Guinehut, S., Belbéoch, M., Ignaszewski, M., Baringer, . M. O., Schmid, C., Lyman, J. M., McTaggart, K. E., Purkey, S. G., Zilberman, N., Aĩkire, M. B., Swift, D., Owens, W. B., Jayne, S. R., Hersh, C., Robbins, P., West-Mack, D. b., Bahr, F., Yoshida, S., Sutton, P. J. H., Cancouët, R., Coatanoan, C., Dobbler, D. e., Juan, A. G., Gourrion, J., Kolodziejczyk, N., Bernard, V., Bourlès, B., Claustre, . H., D'Ortenzio, F., Le Reste, S., Le Traon, P.-Y., Rannou, J.-P., Saout-Grit, C., Sp̃eich, S., Thierry, V., Verbrugge, N., Angel-Benavides, I. M., Klein, B., Notarstefano, G. u., Poulain, P.-M., Vélez-Belchí, P., Suga, T., Ando, K., Iwasaska, N., Kobayashi, T., Masuda, S., Oka, E., Sato, K., Nakamura, T., Sato, K., Takatsuki, Y., Yoshida T., Cowley, R., Lovell, J. L., Oke, P. R., van Wijk, E. M., Carse, F., Donnelly, M., Gould, W. J., Gowers, K., King, B. A., Loch, S. G., Mowat, M., Turton, J., Rama Rao, E. P. t., Ravichandran, M., Freeland, H. J., Gaboury, I., Gilbert, D., Greenan, B. J. W., Ouellet, . M., Ross, T., Tran, A., Dong, M., Liu, Z., Xu, J., Kang, K., Jo, HyeongJun a nd Kim, S.-D., and Park, H.-M. (2020). Argo data 1999 - 2019: Two million temperature-salinity profiles and subsurface velocity observations from a global array of profiling floats. *Frontiers in Marine Science*, 7.