



Reconstructing monthly 20th century salinity fields using a data-driven method and Argo data

Erwan Oulhen¹ · Nicolas Kolodziejczyk¹ · Pierre Tandeo^{2,3,4} · Bruno Blanke¹ · Florian Sévellec^{1,3}

Received: 14 October 2024 / Accepted: 29 April 2025
© Springer-Verlag GmbH Germany, part of Springer Nature 2025

Abstract

The distribution of ocean salinity is influenced by factors such as ocean currents, freshwater exchanges with the atmosphere, continental inputs, and sea ice formation and melting. Salinity is a fundamental variable in the seawater's equation of state, influencing density, ocean dynamics, and stratification. To understand its distribution and impact, salinity variability must be accurately estimated from in situ observations. In the 20th century, temperature observations were scarce, but salinity data were even more limited, as many instruments measured only temperature. In 2002, the implementation of the Argo observing system improved sampling and reduced the disparity between both variables. Salinity variability can now be better assessed from the 20-year Argo period. In this study, information from the Argo period and estimates of salinity covariability with temperature are used to reconstruct monthly subsurface salinity fields. The reconstruction uses in situ measurements and employs the data-driven RedAnDA scheme, which combines Data Assimilation, Analog Prediction, and Reduced-space Interpolation, a method previously validated for temperature reconstruction. The coupling of temperature and salinity is achieved using density-weighted three-dimensional empirical orthogonal functions. The method is applied to the tropical Pacific to evaluate its ability to reconstruct variability at monthly, interannual and decadal timescales, with emphasis on ENSO-related variability, which is the primary mode of variability in the region. Compared to Optimal Interpolation results, salinity analysis by RedAnDA shows significant improvements. The results yield a salinity product that provides valuable insights into historical salinity variability, including fresh pool displacements and salinity-driven stratification changes. Changes in subsurface salinity associated with El Niño and La Niña events are estimated from 1930 to 2001.

Keywords Salinity · Tropical pacific · ENSO · Analog · Data assimilation · Multivariate EOFs

1 Introduction

The global distribution of ocean salinity reflects the balance between ocean diffusion, advection, and freshwater fluxes (precipitation, runoff, evaporation, and ice melting), providing insights into changes in the global freshwater cycle (Lagerloef 2002). Salinity serves as a crucial indicator of variations in the hydrological cycle (Zika et al. 2015; Skliris et al. 2016; Cheng et al. 2020; Olmedo et al. 2022). Approximately 80% of Earth's surface freshwater fluxes occur over the ocean (Peixoto and Oort 1992; Durack 2015). At a broad scale, subtropical regions tend to be saltier due to intense evaporation, while high latitudes are fresher due to predominant precipitation. In recent decades, climate change has

amplified this contrast by intensifying the hydrological cycle (Hosoda et al. 2009; Helm et al. 2010; Durack and Wijffels 2010; Durack et al. 2012; Fox-Kemper 2021; Basinski-Ferris and Zanna 2023). However, the link between changes in freshwater fluxes and ocean salinity distribution is complex, as additional processes such as advection and mixing interfere. Some processes result from natural ocean variability (Delcroix and Hénin 1991; Vinogradova and Ponte 2017; Ge et al. 2023; Yu 2023), while others are due to climate change-driven effects, such as ocean warming and ice melt (Zika et al. 2018).

Salinity influences temperature, marine ecosystems, and ocean dynamics by modulating water density, mixed layer depth, and stratification (Murtugudde and Busalacchi 1998; Maes et al. 2002; Fedorov et al. 2004; Wang and Yin 2021). For example, the barrier layer (BL), formed by salinity-driven stratification, significantly impacts ocean mixing and air-sea interactions (Miller 1976; Lukas and Lindstrom 1991; Sprint-

Responsible Editor: Emma Young

Extended author information available on the last page of the article

all and Tomczak 1992; de Boyer Montégut et al. 2007; Qu et al. 2014). This process is particularly pronounced in the Pacific Ocean, where it favors upper-ocean heat uptake that can trigger an El Niño event (Cronin and McPhaden 2002; Maes et al. 2005, 2006; Mignot et al. 2007; Maes 2008; Yim et al. 2008; Bosc et al. 2009; Zhang et al. 2015, 2021; Guan et al. 2023). During El Niño, the increased eastward propagation of fresh anomalies strengthens the BL, intensifying El Niño's surface temperature signature (Zheng et al. 2014; Zheng and Zhang 2015; Zhi et al. 2019; Zhang et al. 2021). Conversely, stratification destabilization due to negative freshwater fluxes can enhance La Niña events through vertical mixing and sea surface cooling (Zheng and Zhang 2012).

Therefore, accurate analysis of the salinity field is crucial. Monitoring salinity using in situ measurements has undergone successive technological developments, improving sampling and observation quality. Before the Argo observing system (Wong et al. 2020), the quality of salinity measurements reached its peak during the World Ocean Circulation Experiment (WOCE, Thompson et al. 2001) in the 1980s and 1990s. Historically, ocean salinity observations have been sparse in both space and time. In the 20th century, most cruises occurred during the summer months due to unfavorable winter conditions (harsh weather and sea ice) and were primarily located in the Northern Hemisphere and coastal areas (Abraham et al. 2013; Durack 2015; Cheng et al. 2020). Furthermore, high-quality salinity measurements were even scarcer than temperature data, as most Mechanical and Expendable Bathythermographs (MBTs & XBTs) did not measure salinity. Additionally, CTD (conductivity, temperature, depth) profilers are prone to drift in their salinity measurements, necessitating careful delayed-mode quality control and corrections (Romero et al. 2021; Wong et al. 2023; Tang et al. 2024).

Since the late 1990s, the Argo global observing system has been deployed, reaching its targeted quasi-global coverage of the upper 2000 m by 2006 (Roemmich et al. 2019). Over two decades, it has become the backbone of the Global Ocean Observing System and the main source of ocean temperature and salinity measurements in the upper 2000 m for many studies (see Johnson et al. 2022, for a recent review). It addresses the aforementioned biases in ocean observing systems, enabling more accurate estimations of changes in freshwater content, stratification, and oceanic heat content (Hosoda et al. 2009; Hansen et al. 2011; von Schuckmann et al. 2014; Roemmich et al. 2015; Riser 2016; Li et al. 2020a).

In a previous study, we validated and applied RedAnDA, a data-driven mapping method (Oulhen et al. 2024) that combines the Analog forecasting method (Lorenz 1969) and ensemble Data Assimilation techniques (Tandeo et al. 2015; Lguensat et al. 2017), adapted for in situ data process-

ing following the Reduced-Space Interpolation framework (Kaplan et al. 1997). In Oulhen et al. (2024), we aimed to reconstruct the three-dimensional temperature field in the tropical Pacific during the second half of the 20th century, leveraging dynamical knowledge from a learning period (i.e., Argo data) and recent advances in applied mathematics. Proper mapping of sparsely distributed data facilitates the study of regional and temporal physical patterns. RedAnDA successfully estimated the temperature field and provided an uncertainty estimate, accounting for observation sampling, unresolved scales, and system non-linearity.

In the present study, RedAnDA is applied to reconstruct salinity maps between the surface and 500 m depth, from 1930 to 2001, using the Argo period as the learning period. Since Argo floats collect data every 10 days, we aim to produce monthly snapshots to capture seasonal-to-interannual variations.

A common approach is to exploit the relationship between salinity and temperature to compensate for the lack of salinity observations and generate oceanic T/S vertical profiles, for example, by using satellite observations (Maes and Behringer 2000; Fujii and Kamachi 2003; Chang et al. 2011; Goes et al. 2018; He et al. 2021; Tang et al. 2022). Fujii and Kamachi (2003) estimated the salinity field in the equatorial Pacific from temperature observations using vertically coupled T/S empirical orthogonal functions constructed from the World Ocean Database 2001 (WOD01). Goes et al. (2018) employed multivariate linear regression to obtain salinity from in situ observations where only temperature was available, calibrating their parameters on Argo profiles in the Atlantic Ocean from the surface to 2000 m depth. In the northern South China Sea, Tang et al. (2022) obtained the daily, high-resolution three-dimensional temperature field from Sea Surface Height and Sea Surface Temperature, using a polynomial relationship calibrated to a monthly climatology derived from Argo and satellite data. They determined the three-dimensional salinity distribution at the same scales using their temperature estimate and its correlation with salinity.

Alternatively, salinity can be inferred directly from observational and simulated data. For example, Good et al. (2013) reconstructed the past variability of temperature and salinity from 1900 onward, using an optimal interpolation method that combined a priori climatology with local observational information limited in its spatio-temporal extent, emphasizing the evaluation of observation quality. Li et al. (2023) produced a high-resolution salinity dataset from 1940 to the present, allowing long-term estimation of salinity changes while capturing smaller-scale signals. To achieve this, they combined in situ observations with several climate or Earth system model simulations (including CMIP6) in an ensemble optimal interpolation procedure.

Here, we investigate inferring salinity by exporting temperature information through the application of RedAnDA TS, the coupled version of the method. Both variables are linked by seawater density through the thermodynamic equation of state (McDougall et al. 2010). As temperature has been more extensively observed than salinity, it is expected that salinity reconstruction will benefit from this procedure.

Our reconstruction focuses on the tropical Pacific Ocean, between 25°N and 25°S. In this region, west of 160°W, lies the West Pacific fresh pool (WFPF), characterized by salinity below 34.8 PSS and a permanent east-west haline gradient, often described as a salinity front. This front shifts eastward to the dateline during El Niño events because of anomalous surface currents and intensified rainfall caused by enhanced atmospheric convection in the central Pacific (Picaud et al. 1996). In this region, salinity-driven stratification in the BL allows heat accumulation and displacement toward the central Pacific, potentially favoring extreme El Niño events (Maes et al. 2002; Zhang et al. 2021). In the central Pacific, the Argo observing system has provided valuable insights into tropical BL variability (Bosc et al. 2009; Qu and Yu 2014; Steffen and Bourassa 2018). Therefore, using the Argo period to learn typical monthly salinity changes enables a more accurate representation of past salinity-driven stratification in the RedAnDA TS reconstructed product.

This paper is organized as follows: Section 2 describes the datasets used for reconstruction and analysis. Section 3 outlines the RedAnDA method, explains the temperature-salinity coupling via the EOFs, and presents the validation and implementation details. Section 4 presents validation results, first with an Observing System Simulation Experiment (OSSE), then with real observations. Section 5 discusses the results and Section 6 concludes with future research directions.

2 Data and products

In this section, we first describe the OCCIPUT model outputs used to validate our approach within an OSSE framework. Then we focus on real datasets: the EN.4.2.2 in situ observations, used to estimate the 20th century salinity field, followed by the ISAS20 temperature and salinity products from which the learning period catalog is constructed. Finally, we introduce the IAP and EN4 objective analysis products, which are used for comparison with different results obtained with RedAnDA.

2.1 Model outputs

The OCCIPUT (Penduff et al. 2014; Bessi eres et al. 2017) simulations correspond to an ensemble of 50 realizations of a global 1/4° ocean/sea-ice simulation, based on the NEMO

3.5 model. The simulations provide synthetic profiles, which are simulated in situ observations, extracted at specific times and locations corresponding to the historical database. They span the period from 1961 to 2015, employing the DFS atmospheric forcing over the same period (Dussin et al. 2016). The forcing updates heat, freshwater content and momentum of sea surface using the NEMO bulk formula. SSS is restored to Levitus et al.'s (1998) climatology with a piston velocity of 167 mm/day (Bessi eres et al. 2017).

2.2 In situ observations

The pre-Argo period is reconstructed by RedAnDA by assimilating profiles from the EN.4.2.2 dataset (Good et al. 2013). This dataset represents the latest version of the Met Office Hadley Centre's 'EN' dataset series, offering quality-controlled potential temperature and salinity profiles from 1900 to the present. Most profiles originate from the World Ocean Database 2018 (WOD18, Boyer et al. 2006), the Global Temperature and Salinity Profile Program (GTSP, Sun et al. 2010), and Argo floats (Roemmich et al. 2012). For this study, in both the synthetic profiles of the OSSE and EN.4.2.2, only depths where temperature and salinity observations have a quality flag of 1 are used.

Observations are selected if sufficiently close to the target depth, without modification. The instrumental error is evaluated following Cowley et al. (2021) and Atkinson et al. (2014). It ranges from 5×10^{-3} PSS for CTD and profiling floats to 2×10^{-2} PSS for moored buoys. For temperature, it ranges from 0.3°C for MBTs and moorings to 0.002°C for the most accurate CTDs. These values are used to construct the observation error matrix employed in our analyses (see Section 3).

2.3 Learning period dataset

For the analysis of real data, the Argo period serves as the learning period from which the catalog and coupled EOFs are calculated. The ISAS20 (Gaillard et al. 2016, In Situ Analysis System) gridded product is used here to represent the Argo period. ISAS20 analysis focuses only on Argo measurements, preserving as much as possible the time and space sampling capabilities of the Argo network (Kolodziejczyk et al. 2023). It provides $0.5^\circ \times 0.5^\circ$ gridded temperature and salinity fields, constructed with an Optimal Interpolation method. It includes 187 standard levels distributed between 0 m and 5500 m. The product covers the period from 2002 to 2020, with monthly maps.

2.4 Temperature and salinity products

RedAnDA's temperature reconstruction is compared to two state-of-the-art reanalysis products that use optimal inter-

polation techniques. EN4: A gridded product from the Hadley Centre of the UK Meteorological Office, assimilating EN.4.2.2 in situ profiles. EN4 employs an optimal interpolation technique, using a smoothed 1971–2000 climatology from the EN2 version. It provides monthly temperature and salinity fields from 1900 to the present day over 42 depth levels.

IAP: An Ensemble Optimal Interpolation of the CODC-GOSD dataset (Cheng and Zhu 2016; Cheng et al. 2017). This method utilizes an ensemble of CMIP5 models to reconstruct global and gridded temperature and salinity at a monthly resolution, from 1940 to the present, at a spatial resolution of $1^\circ \times 1^\circ$ in the upper 2000 m, over 41 levels for salinity and 119 levels down to 6000 meters for temperature (Cheng et al. 2020). Our study uses the sixth and standard versions of the IAP product for temperature and salinity, respectively.

3 Method

3.1 RedAnDA

RedAnDA extends AnDA (Analog Prediction Data Assimilation, Tandeo et al. 2015; Lguensat et al. 2017) by incorporating Reduced-Space Data Assimilation as introduced by Kaplan et al. (1997) to process observations that are sparse in time and three-dimensional space.

The model is based on the Analog forecast method (Lorenz 1969), which was later adapted into the Analog backcast in Oulhen et al. (2024) to enable predictions backward in time. This data-driven approach requires a reference database (a catalog) of previously recorded system trajectories, consisting of state analogs and their temporal predecessors. For a new instance of the same system, its evolution toward an earlier time is predicted using a data-driven linear model statistically built by analyzing 170 analog situations and their temporal predecessors. For further details, we refer the reader to Oulhen et al. (2024). The catalog used in this study was built from the Argo period. The backward-in-time approach implies that the reconstruction starts at the most recent time step, with the model estimating trajectories toward preceding months. The method facilitates the propagation of information from periods with more observations to those with fewer observations.

In this study, RedAnDA is applied to reconstruct past temperature and salinity fields in the tropical Pacific, between 120°E and 70°W , 25°S and 25°N . Depth levels are restricted to 5, 10, 50, 100, 200, 300, 400, and 500 m. The final result consists of $3^\circ \times 3^\circ$ gridded temperature and salinity fields, representing several thousand spatial grid points. This high dimensionality challenges the identification of analogs, which justifies the use of a space reduction technique (Kaplan et al. 1997), here EOF decomposition.

We construct the three-dimensional EOFs using the temperature and salinity covariance estimated from the maps of the learning period, retaining only the most significant EOFs. The number of EOFs selected for RedAnDA (see Section 4.1 for justification) is approximately $L = 15$.

In the context of historical reconstruction, EOF decomposition can be interpreted as a change of variables, where the new variables to be optimally estimated at each desired time step are the principal components at time t , i.e., $\text{PC}(t)$, of the retained EOFs (Kaplan et al. 1997; Emery and Thomson 2001). Once $\text{PC}(t)$ for the 15 EOFs are obtained through a 250-member Ensemble Data Assimilation procedure, three-dimensional gridded temperature and salinity fields are reconstructed using:

$$X(t) \simeq \text{PC}(t) \text{ EOF}, \quad (1)$$

where X represents the target three-dimensional temperature or salinity field, and **EOF** contains the L retained EOFs. Ensemble Data Assimilation provides a robust mathematical framework that combines Analog Prediction and Reduced-Space Interpolation to assimilate sparse observations, using a Kalman smoother (Tandeo et al. 2015; Lguensat et al. 2017; Oulhen et al. 2024). An additional procedure automatically extends the temporal window for assimilation (Oulhen et al. 2024) to mitigate insufficient observational sampling in the early part of the time series.

Analysis uncertainty is estimated from the spread of a 250-member ensemble, reflecting chaos in the system, as well as sampling and instrumental errors in the observations. The variance of the unretained EOFs is included in the uncertainty estimate, providing insight into misrepresented scales.

Climatologies for salinity and temperature are calculated over the Argo period, using ISAS20 (In Situ Analysis System) temperature and salinity fields (Gaillard et al. 2016). The corresponding $\text{PC}(t)$ values from 2002 to 2020 are incorporated into the RedAnDA catalog.

The reconstruction covers the 1930–2001 period, assimilating EN.4.2.2 profiles (Good et al. 2013). In the following, RedAnDA salinity anomalies are compared with those from EN4 and IAP products, both estimated with optimal interpolation schemes. Their respective monthly climatologies were computed for the period 2002–2015. The final comparison examines the impact of salinity on stratification and its inter-annual variability.

3.2 RedAnDA TS

RedAnDA TS is the application of RedAnDA (Oulhen et al. 2024) to temperature and salinity simultaneously. It uses three-dimensional coupled EOFs that express main covariance patterns for salinity and temperature. This approach

leverages in situ temperature observations to compensate for the lower availability of salinity data.

3.2.1 Coupled EOFs

To standardize temperature and salinity units, both variables must be weighted before the EOF calculation. To preserve the physical relationship, both variables are transformed into their density contributions. We consider the linearized seawater equation of state (Bryan and Cox 1972):

$$\rho(t) = \rho_0 + \rho_0 \left[\frac{1}{\rho_0} \frac{\partial \rho}{\partial T} |_{S(t)} \Delta T(t) + \frac{1}{\rho_0} \frac{\partial \rho}{\partial S} |_{T(t)} \Delta S(t) \right]$$

$$= \rho_0 [1 - \alpha(t) \Delta T(t) + \beta(t) \Delta S(t)],$$

where t is the time, ρ is the density, ρ_0 is the reference mean density, T is the potential temperature, S is the practical salinity, ΔT the potential temperature anomaly (from the monthly climatology), ΔS the practical salinity anomaly (from the monthly climatology), α and β are the first derivatives of density with respect to temperature and salinity, respectively, when the other variable is fixed (they are the thermal expansion coefficient and haline contraction coefficient, respectively, McDougall et al. 2010). The Python package gsw 3.4.0 (McDougall and Barker 2011) is used to compute $-\alpha T$, βS , and density.

We express changes in temperature and salinity anomalies as their contributions to changes in density. The coupled

EOFs are estimated from these thermodynamically normalized variables:

$$[-\bar{\alpha} T(t), \bar{\beta} S(t)] \simeq \text{PC}(t) \text{ EOF}, \quad (2)$$

where $\bar{x}$ designates the temporal average of x .

3.2.2 Description of the EOFs

The EOFs are calculated from ISAS20 temperature and salinity fields, weighted as $-\bar{\alpha} \Delta T(t)$ and $\bar{\beta} \Delta S(t)$. Variables $\alpha(t)$ and $\beta(t)$ are first calculated at all time steps of the learning period, then temporally averaged.

The coupled EOF decomposition captures a substantial portion of both variables' variability while significantly reducing the dimensionality of the system. With $L = 15$, $\sim 72\%$ and $\sim 66\%$ of the total variances of the initial true temperature and salinity fields are explained, respectively. Temperature signals at 5, 10, 50, 100, 300, and 500 m depths are reconstructed up to $\sim 75\%$, 75% , 73% , 75% , 65% , and 52% , respectively. For salinity signals, percentages are 67% , 67% , 68% , 57% , 56% , and 46% .

The EOF patterns show structures associated with ENSO variability in oceanic temperature and salinity (Fig. 1a–l). The first coupled EOF shows a temperature tongue extending from America to 140°E and a strong salinity anomaly around the dateline, in the ITCZ and South Pacific Convergence Zone (SPCZ), due to changes in precipitation and ocean dynamics (Fig. 1a, g). The ENSO signature in temperature penetrates

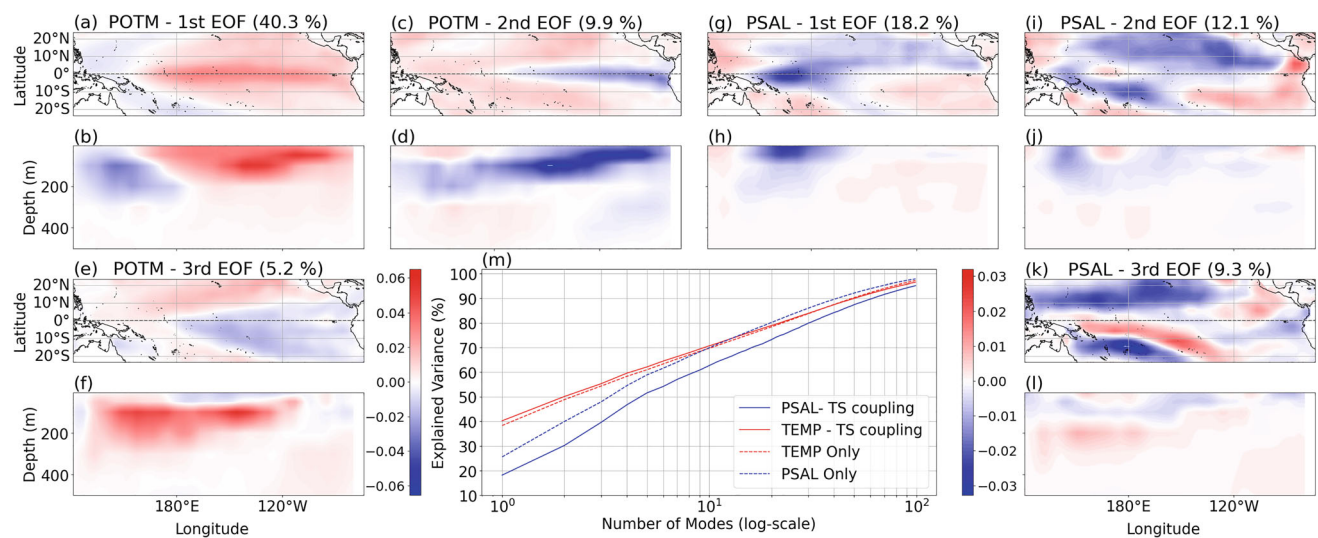


Fig. 1 The potential temperature component of the first three coupled EOFs at 5 m depth and as a zonal section along the equator is shown in panels (a, b), (c, d), and (e, f), with each pair associated with the first, second, and third mode, respectively. Panels (g, h), (i, j), and (k, l) display practical salinity anomalies associated with the coupled EOFs. Temperature and salinity panels use different colormaps. Associated

colorbars are found on each side of panel m. Panel m shows the percentage of explained variance in the initial temperature and salinity time series as a function of the number of EOFs used for their approximation. Solid lines represent the explained variance in the coupled EOF decomposition, while dotted lines are associated with univariate EOFs. In this panel, the x-axis increases logarithmically

deeper than it does for salinity. Salinity variability is confined to the surface (Fig. 1b and h).

The second EOF (Fig. 1c, d and i, j) is associated with changes (positive in the figure) concentrated in the center of the basin around the dateline and extending poleward in a horseshoe shape (Fig. 1c), while changes of opposite signs occur in the eastern part, along the equator, associated with upwelling (Fig. 1c and d). Salinity changes are more widespread over the basin, and the contrast between regions of opposite changes is stronger (Fig. 1i). The third EOF (Fig. 1e, f and k, l) is associated with a significant increase in the subsurface below 50 m (Fig. 1f), SST decrease east of the dateline, and well-marked salinity changes, with a western salinity decrease and its augmentation along the equator and east of the SPCZ region.

It is important to note that the structures of these EOFs only determine the spatial distribution of changes in salinity and temperature. Temporally, the changes are modulated by the coefficients $PC(t)$, notably with changes in sign. For the first EOF, $PC(t)$ would be positive during El Niño and negative during La Niña.

However, relating the EOFs to specific dynamics in the tropical Pacific is likely not straightforward; instead, several modes may be needed to represent them. Therefore, ENSO variability may be expressed by a combination of the first, second, and other EOFs. Eastern and Central Pacific El Niño events (Kao and Yu 2009; Chi et al. 2022) represent two types of spatial distributions for SST anomalies. In the literature, they were found to be well represented by their respective combinations of the first and second surface temperature EOFs (Takahashi et al. 2011). How the variability is shared among three-dimensional EOFs is more uncertain. The influence of the thermocline and halocline variability may spread ENSO-related changes over more modes, as subsurface heat content and SST do not evolve simultaneously (Zhao et al. 2021).

The third, fourth, and higher EOFs are thus important to consider, as they add degrees of freedom when expressing the system's state. Also, the EOF space reduction applied to a nonlinear system may preserve long-living structures but neglect the stochastic components that modulate them if the truncation is too severe (Farrell and Ioannou 2001; Wang and Scott 2005).

Each additional mode contributes progressively less to the explained variance (Fig. 1m), which is a characteristic feature of EOF decomposition. The first EOFs for temperature explain a greater amount of the initial variance than they do for salinity. Coupled EOFs explain less salinity variance than the univariate version. Conversely, the differences between the two EOF versions for temperature are minimal. For salinity, the coupled EOFs express about 5–8% less of the initial signal variance than univariate EOFs.

3.3 OSSE

To validate salinity reconstruction by RedAnDA TS, its efficiency is first assessed through an OSSE (Masutani et al. 2010), using OCCIPUT data (Penduff et al. 2014; Bessi eres et al. 2017), as in Oulhen et al. (2024). Analogs and predecessors are evaluated during the 1995–2015 learning period (red 1, Fig. 2). The period spanning 1961–1994 is reconstructed at a monthly resolution, based on in situ synthetic observations, using RedAnDA (red 2, Fig. 2). The results are expressed as $PC(t)$, which are projected back into real space to reconstruct both temperature and salinity anomalies (red 3, Fig. 2).

The target truth is represented by the full-grid maps. To simplify the calculations and focus on large spatial scales, the OSSE reconstruction grids are coarsened to $3^\circ \times 3^\circ$ through horizontal averaging, with depth levels limited to 5, 10, 50, 100, 300, and 500 m.

In the OSSE experiment, the optimal value of L is determined through an initial application of RedAnDA TS using a relatively high number of EOFs. The effective contributions of each reconstructed $PC(t)$ are then evaluated based on their temporal and spatial correlation with the reference truth, helping to identify the required number of useful EOFs (see Section 4.1). This optimal number, L , is selected for a second RedAnDA TS analysis. If the results remain robust, i.e., if the temporal and spatial correlation scores are still maximized for L EOFs, the configuration is adopted and the same number of EOFs is used for real data reconstruction.

Within the OSSE framework, the RedAnDA TS method is compared with the previous version of RedAnDA (Oulhen et al. 2024), referred to as UnivarS when applied to salinity reconstruction and UnivarT for temperature (Fig. 2).

UnivarS uses only salinity data from the learning period to derive its EOFs and catalog (blue 1, Fig. 2). RedAnDA is then applied (blue 2, Fig. 2), and the analyzed $PC(t)$ are projected onto the univariate EOFs to obtain the gridded salinity analysis (blue 3, Fig. 2).

To explore the potential of leveraging temperature information for salinity estimation, an S-from-T approach is also implemented. This method relies on UnivarT analyses (green 1, Fig. 2). The full-grid temperature results are projected onto the temperature component of the coupled EOFs (green 2, Fig. 2), then transformed back into full-grid space using the salinity component (green 3, Fig. 2).

As in Oulhen et al. (2024), an Optimal Interpolation (OI) method is implemented, following Good et al.'s (2013) methodology. This approach performs a sequential analysis of temperature or salinity anomalies relative to a monthly climatology, derived from in situ profiles. Two horizontal correlation radii (300 km and 400 km) and two vertical correlation radii (100 m and 200 m) are used. Temporal persistence propagates forward $\sim 82\%$ of the analysis from time $t - dt$ and incorporates it into the background field at time t .

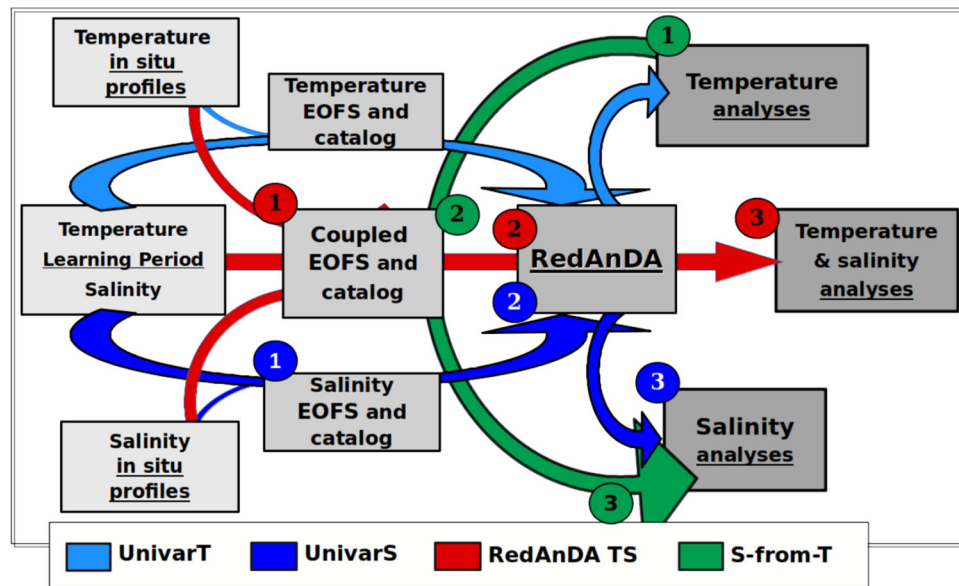


Fig. 2 Summary of data sources (lighter gray), types of EOFs and catalogs (intermediate gray), and final outputs (darker gray) for different RedAnDA approaches: UnivarT for temperature (light blue), UnivarS

for salinity (dark blue), RedAnDA TS (red), and S-from-T (green). The numbers enumerate the steps in each approach

Correlation is used to quantify the performance of the reconstructions. It is defined as follows:

$$\text{Correlation}(\mathbf{T}, \mathbf{T}^{\text{true}}) = \frac{\overline{\mathbf{T}(t) \times \mathbf{T}^{\text{true}}(t)}}{\sqrt{\overline{\mathbf{T}(t)^2} \times \overline{\mathbf{T}^{\text{true}}(t)^2}}}, \quad (3)$$

where $\mathbf{T}(t)$ and $\mathbf{T}^{\text{true}}(t)$ are the anomalies at time t , respectively, from the reconstructed signal and from the truth; and $\overline{\mathbf{T}(t)}$ designates the temporal average of $\mathbf{T}(t)$. The correlation score ranges from -1 to 1 , where -1 indicates opposite phase variations, 1 indicates similar variations, and 0 indicates there is no consistency between the signals.

4 Results

4.1 OSSE – Selection of the number of EOFs

Within the OSSE framework, temperature and salinity fields are reconstructed from 1961 to 1994 using synthetic in situ measurements and compared with the truth, i.e., the model outputs. We construct the catalog for the different RedAnDA approaches using model outputs from 1995 to 2015 to mimic the length of the Argo period (20 years) (Oulhen et al. 2024).

The OSSE allows estimation of the best tuning for RedAnDA, notably the optimal number of EOFs (L). This optimum is set by considering two criteria: the ability to accurately represent temporal fluctuations in salinity and the spatial distribution of the variability. The first criterion is the

spatially averaged temporal correlation between the reconstructed anomalies and the true values. The second is the spatial correlation of the interannual standard deviation with that of the truth. RedAnDA TS, UnivarS, and S-from-T are applied with $L = 40$ EOFs, and we examine the evolution of these correlation scores (Fig. 3) depending on the number of EOFs used when reconstructing the gridded field (we have the analyses for the 40 temporal coefficients, but the gridded fields are evaluated by combining the temporal coefficients with their respective EOFs). This enables us to determine which L optimizes the reconstruction for each RedAnDA configuration.

Increasing the number of EOFs generally improves the spatial distribution of variability, following a logarithmic evolution. However, the temporal correlation increases sharply with the first three EOFs, then grows more gradually for L values in the ~ 3 – 15 interval. Beyond this range, adding more modes degrades the quality of the reconstruction for salinity anomalies.

Based on these results, we set two conditions for choosing L : the spatial correlation must exceed 0.9 , and the temporal correlation must be maximized. The 0.9 threshold (horizontal bar in Fig. 3 b) is arbitrary but allows reaching the higher correlations when L is between 10 and 20 (Fig. 3a). Beyond $L = 20$, the correlation diminishes, likely due to the inadequacy of the EOFs to represent the changes over the reconstruction period.

For RedAnDA TS and S-from-T, we set $L = 15$ and for UnivarS $L = 14$. The values of the correlations for

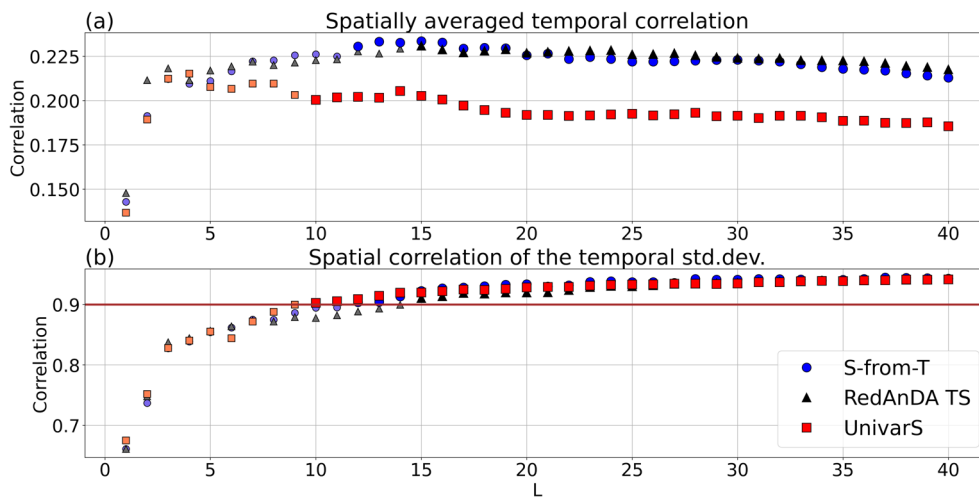


Fig. 3 (a) The spatially averaged temporal correlation between the anomalies of the true salinity field (1961–1994) and these from S-from-T (blue dots), RedAnDA TS (black triangles), and UnivarS (red squares) analyses, as a function of the number of EOFs used to reconstruct the gridded field. (b) The spatial correlation of the temporal standard deviation of the analyses with the true standard deviation. Monthly anomalies are used when estimating the standard deviations. The horizontal black line indicates the 0.9 threshold. Markers appear duller and smaller when the spatial correlation of the standard deviation falls below 0.9, highlighting when this threshold is reached

RedAnDA TS are 0.23 and 0.91 for temporal and spatial correlations, respectively. For UnivarS, they are 0.2 and 0.92, and for S-from-T, they are 0.23 and 0.93.

4.2 OSSE – Validation

We then compare the true salinity field with reconstructions by the Optimal Interpolation (OI) method and the three RedAnDA configurations (i.e., TS, UnivarS, S-from-T) applied with the selected L values.

Table 1 presents spatially averaged temporal correlation scores of the different reconstructions for three timescales: decadal, interannual, and seasonal.

Correlation scores at interannual scales indicate the superiority of RedAnDA TS (0.27) and S-from-T (0.27) over UnivarS (0.2) and OI (0.12). Regarding the decadal anomalies, the correlation scores are lower for all products (around 0.05). The ranking differs, with UnivarS having the highest score (0.11) and OI the lowest (0.01). The monthly climatol-

Table 1 Spatially averaged temporal correlations with the detrended truth for, from left to right, RedAnDA TS, UnivarS, S-from-T, and OI

	RedAnDA TS	UnivarS	S-from-T	OI
Interannual	0.3	0.27	0.31	0.12
Decadal	0.09	0.11	0.09	0.03
Seasonal	0.85	0.88	0.87	0.87

Three lines of scores indicate the scores obtained when comparing the interannual anomalies (salinity anomalies with removal of the decadal signal), the decadal signal (10-year running-mean filtered anomalies) between 1966 and 1990, and the monthly climatology of the reconstructions

ogy estimated over the reconstruction period remains similar to that of the learning period, with high correlation scores (about 0.85 for RedAnDA TS and 0.87 for other analyses).

To better assess the reasons for the rankings in Table 1, we study our various reconstructions from different perspectives, including structures and dynamics at depth, ENSO representation, and interannual variability in stratification.

The different salinity reconstructions at 300-m depth during the La Niña event of 1973–74 are shown in Fig. 4, for October 1973, January 1974, and April 1974.

Between 10°N and 20°N and 120°E and 170°W, near Box A (northern boxed region), an intense high-salinity anomaly of about 0.15 PSS persisted throughout the three displayed months. The OI method fails to adequately reconstruct this structure (Fig. 4d, e, and f). Overall, at 300 m depth, the OI interannual anomalies barely compare with the truth due to the lack of observations in most of the basin. Even with observations along the dateline in October 1973, around 15°S and 170°W in January 1974, and along 20°N between 150°E and 160°W in April 1974, the reconstructed field displays anomalies that neither capture the zonally extended structures of the truth nor maintain consistent signs, as evidenced by the band of fresh anomaly (−0.15 PSS) located along 15–20°S (Fig. 4a–f). In situ observations may capture small-scale variability, which can alias the large-scale salinity reconstruction without sufficient additional observations to correct these spurious deviations.

RedAnDA TS, UnivarS, and S-from-T reconstruct the high-salinity anomaly near Box A, though with a reduced intensity of ~0.1 PSS (Fig. 4g–o), likely due to variance loss from spatial compression.

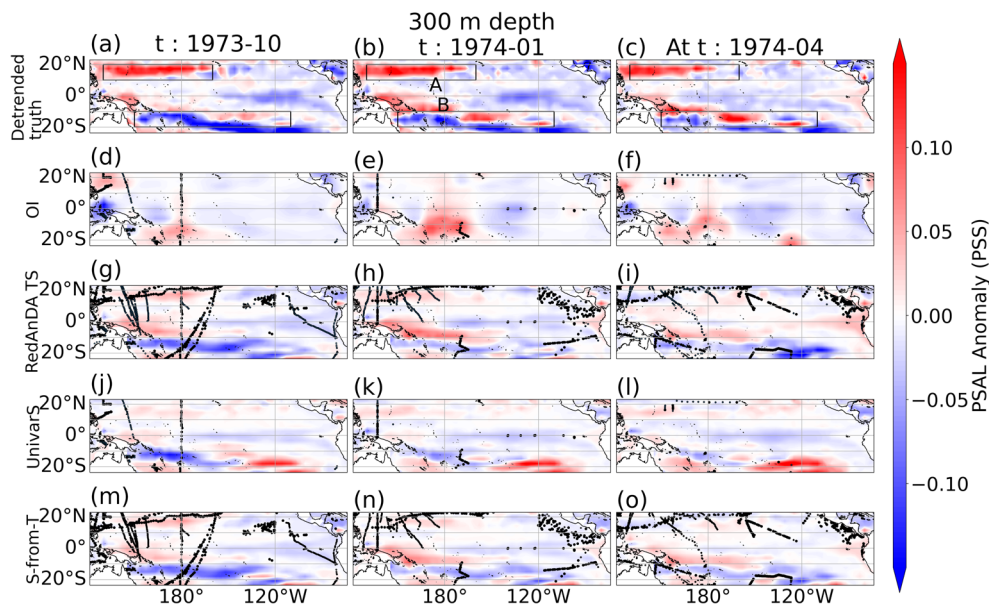


Fig. 4 Salinity anomaly snapshots at 300-m depth in October 1973, January 1974, and April 1974. From the first row to the fifth, displayed salinity fields are the OCCIPUT output (considered the truth, detrended in this figure to better highlight interannual anomalies), OI (implemented for comparison), RedAnDA TS, UnivarS, and S-from-T,

respectively. Black dots represent available observations at the current depth (larger dots) and at other depth levels (smaller dots). The black boxes in a, b, and c represent two areas of high variance, referred to as “Box A” and “Box B”. Salinity variability in these regions is associated with the Northern and Southern Equatorial Currents (Li et al. 2020b)

In the southern part of the basin, in Box B, the true salinity field shows zonal strips of strong anomalies with alternating signs (with absolute values exceeding 0.15 PSS), featuring a northern salty strip and a fresh southern strip. This structure is associated with the Southern Equatorial Current (SEC), and its interannual variability relates to ENSO variability (Kessler and Cravatte 20130. In October 1973, while this strip is not well reconstructed by RedAnDA TS and S-from-T, it is more accurately estimated by UnivarS (Fig. 4g, j, and m). In January and April 1974, the warm side of the anomaly strip is satisfactorily reconstructed by all three methods (Fig. 4b, c, h, i, k, l, n, and o). However, around 120°W and along 20°S, UnivarS reconstructs a spurious southern salty anomaly instead of the fresh extension (Fig. 4g-o).

Between October 1973 and April 1974, at the equator between 170°W and 100°W, a fresh anomaly appears and propagates westward (Fig. 4a, b, and c). This event is not captured by the OI method (Fig. 4d, e, and f). RedAnDA approaches show more skill in reconstructing this change, particularly UnivarS, which more accurately reproduces the contrast between the central fresh anomaly and the salty patch eastward (about 0.1 PSS difference) (Fig. 4a-c, g-o). Again, the intensity of the anomalies is slightly underestimated by RedAnDA approaches by about ~0.05 PSS.

The temporal evolution of salinity in “Box A” (Fig. 4a, b, and c) is presented in Fig. 5a. It is strongly related to the interannual variability of the NEC, which carries saltier or fresher waters from the subtropics. The most intense events, occur-

ring in 1970-1972, 1973-1976, 1984-1986, and 1988-1989, corresponded to La Niña events. They are accurately captured by the different reconstructions, although underestimated by about 0.05 PSS, except for the 1973-1976 and 1988-1989 events. Overall, S-from-T and RedAnDA TS yield very similar results, with S-from-T being smoother than RedAnDA TS, and their correlations with the truth are 0.59 and 0.61, respectively. Despite fewer salinity observations compared to the temperature, the results from UnivarS are accurate, with a correlation reaching 0.63. The OI method fails to reconstruct the variability between 1976 and 1983. Most of the time the OI fluctuations are noisy, resulting in a relatively low correlation score of 0.1.

The temporal variability of “Box B” (150°E-110°W, 10°S-20°S, Fig. 4a, b, and c) is associated with the SEC (Fig. 5b). Salinity fluctuations are more intense in this region than in the northern part of the basin (up to twice the greatest amplitude, as in 1983). Additionally, decadal variability appears significant, with a roughly positive phase between 1961 and 1974, a negative phase between 1976 and 1987, and a positive phase again from 1988 onward. RedAnDA TS and S-from-T fail to capture the decadal variability, despite seeming to capture the interannual fluctuations. Their correlation scores are 0.42 for RedAnDA TS and 0.44 for S-from-T, higher than OI’s score (0.3). UnivarS achieves the highest score of 0.46 though it does not efficiently capture certain changes, such as those in 1964 and 1983. Its higher score may be due to its consistent positive sign during the positive phases of decadal variabil-

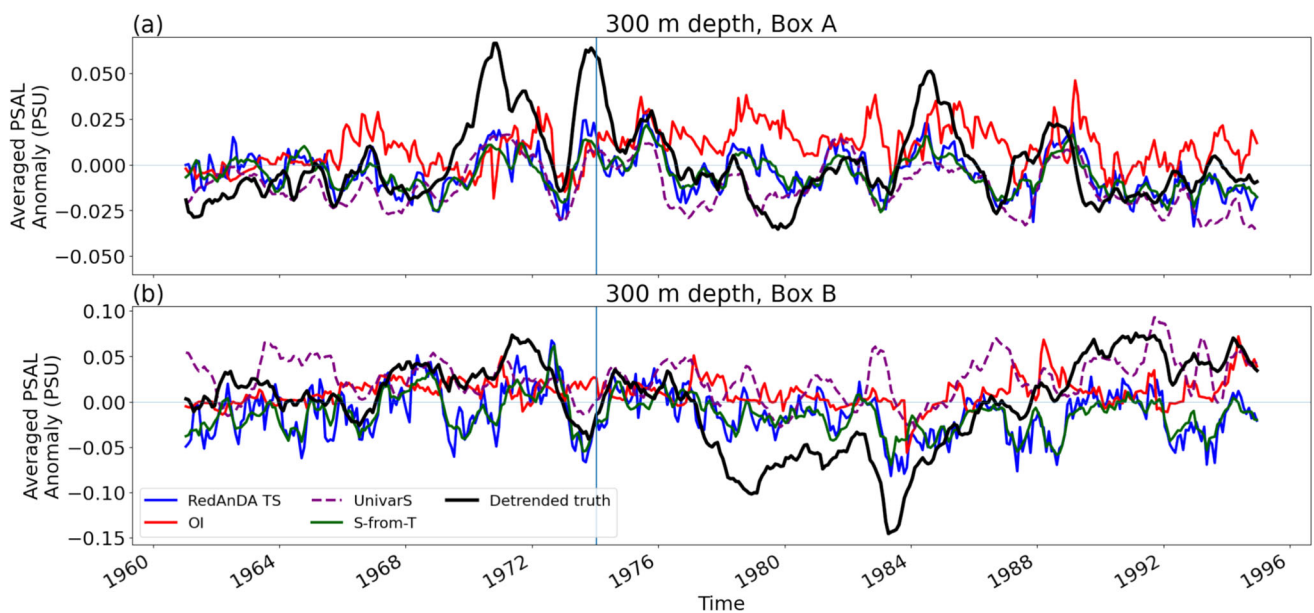


Fig. 5 (a) and (b): averaged salinity anomalies at 300 m depth within “Box A” and “Box B” from Fig. 4a–c. The RedAnDA TS average is shown as a solid blue line, UnivarS as a dashed purple line, OI as a red

line, and S-from-T as a green line. The detrended signals of the truth are shown in black. The vertical blue line indicates the selected month in Fig. 4b, e, h, k, and n

ity. However, since UnivarS fails to reconstruct the negative phase, it likely does not accurately estimate this variability.

We now assess the ENSO variability at the surface, with Fig. 6 showing SST Niño 3.4 indices and SSS anomalies in a longitude-time system. First, SST estimated by all the reconstructions satisfactorily represents the ENSO events (Fig. 6a).

The relationship between the western Pacific fresh pool position and ENSO is clear, as the Niño 3.4 index aligns well with the zonal displacements of the fresh anomalies throughout the period (Fig. 6a, b). The development of El Niño is accompanied by an eastward displacement of the fresh pool, as trade winds weaken, allowing warm waters to extend further eastward, inducing changes in oceanic currents and enhanced precipitation in the central Pacific around 180°E. During La Niña events, saltier anomalies appear near the dateline, associated with the opposite behavior. This is consistent with Maes et al. (2005).

In the central tropical Pacific (around 140°W), the OI reconstruction fails to reconstruct salinity sufficiently, as most of the in situ data along the equator are located near the coasts of the tropical Pacific Ocean, west of 150°E and east of 110°W (Fig. 6c). Consequently, anomalies are mostly found west of the dateline and the easternmost part of the basin, and between 160°W and 110°W, no signal is satisfactorily reconstructed (Fig. 6b and c). The western anomalies roughly correspond between the truth and the OI, but the zonal distribution of the changes is underestimated in the latter. Eastern anomalies are poorly resolved. The OI fails to reconstruct most of the other annual salty events, such as those in 1988–

1993 (Fig. 6b and c). However, events in 1983 and 1994 are captured.

Both RedAnDA TS and UnivarS approaches provide more robust reconstructions of surface salinity (spatially averaged temporal correlation scores reaching 0.43 and 0.41, respectively, and 0.41 for S-from-T against 0.26 for OI). Most of the fresh pool displacements associated with ENSO, west of 160°W, are reconstructed, e.g., in 1963, 1966, 1969, 1972, 1976, 1983, and 1988 (Fig. 6b, d, and e). In 1983, the displacement was strong enough to create an intensified fresh signature in the eastern Pacific. This event was reconstructed successfully, though the saltier anomalies west of 150°E appear overestimated. A similar overestimation is also found during the 1972 El Niño event (Fig. 6b, d, and e). UnivarS shows a tendency toward persistence, which may mask some La Niña events, such as before 1970 (Fig. 6b and e). The persistence issue is also noticeable between 1976 and 1980, hiding more subtle salinity fluctuations west of the dateline.

East of 160°W, the true salinity exhibits annual salty events, some of which are particularly intense, such as in 1972 (Fig. 6b). Some short-lived and less frequent fresh events also occur, notably in 1973, 1974, and 1975. Neither OI nor UnivarS successfully reconstructs this variability (Fig. 6c and e). RedAnDA TS captures some of it, but with underestimated intensity as in 1972 (Fig. 6d). In the western Pacific, correlations between the true equatorial SSSA values and the reconstructions are approximately 0.54, 0.5, 0.53, and 0.32, for RedAnDA TS, UnivarS, S-from-T, and OI, respectively. In contrast, east of 160°W, the correlations are around 0.29, 0.30, 0.27, and 0.21.

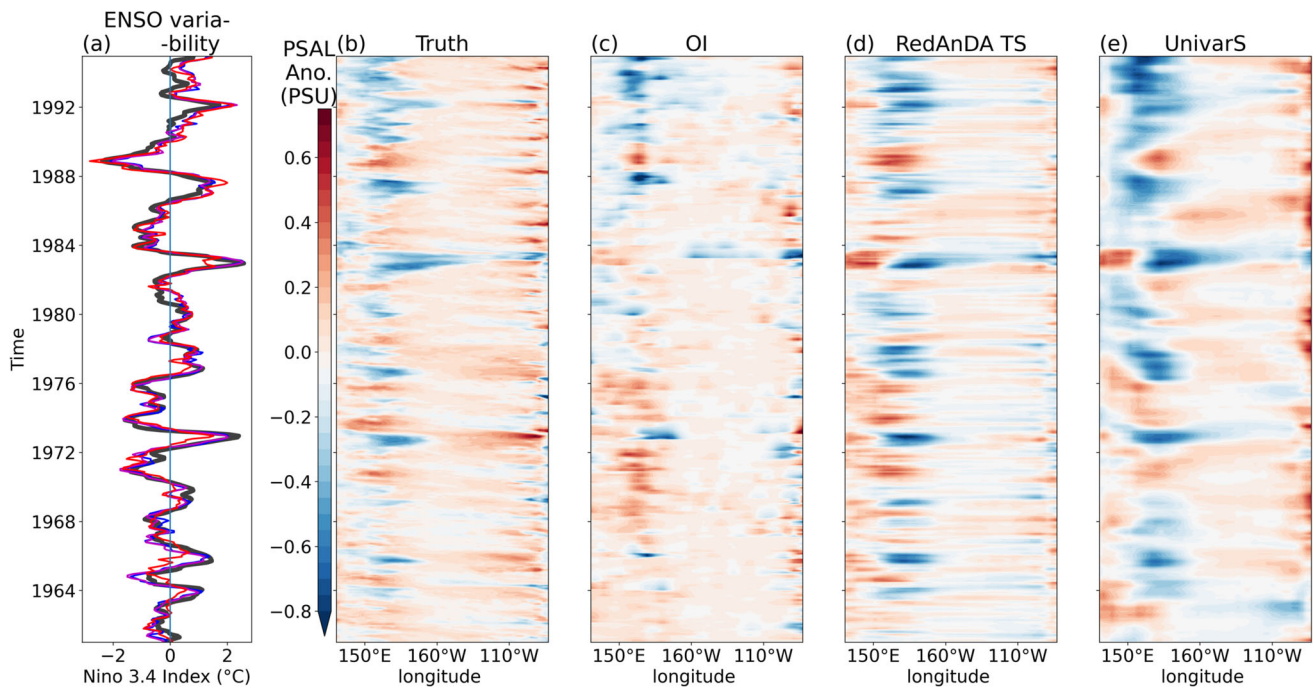


Fig. 6 (a) The filtered temporal evolution of Niño 3.4 indices obtained with temperature anomalies from OCCIPUT (thick black lines), OI (red), RedAnDA TS (blue), and UnivarS (magenta) with time on the y-axis. (b), (c), (d), and (e) Salinity anomalies averaged between 5°S and 5°N at 5 m depth over time from OCCIPUT, OI, RedAnDA TS, and

UnivarS, respectively. A third-order low-pass Butterworth filter is used to remove the annual signals from Niño 3.4 indices to facilitate their interpretation and readability. All anomalies are calculated relative to the 1995–2015 period

RedAnDA TS and UnivarS tend to propagate strong western anomalies eastward (Fig. 6b, d, and e), which can introduce spurious changes, such as a non-existent fresh anomaly in 1992–1993, around 110°W (Fig. 6d and e).

Reconstructions of temperature and salinity allow for the assessment of density and ocean stratification. Here, stratification is quantified by the Brunt–Väisälä frequency $N^2(T, S)$. It is calculated as follows:

$$N^2(T, S) = -\frac{g}{\rho} \frac{\partial \rho}{\partial z} \approx g \left(\bar{\alpha} \frac{\partial T}{\partial z} - \bar{\beta} \frac{\partial S}{\partial z} \right), \quad (4)$$

where g is the gravitational constant, ρ is the density, z is the depth oriented upward, $\bar{\alpha}$ and $\bar{\beta}$ are the temporally averaged thermal expansion coefficient and haline contraction coefficient, respectively. We compute $N^2(T, S)$ using absolute temperature (T) and salinity (S) values from all reconstructions. Anomalies are derived by subtracting the corresponding monthly climatology for 1961–1994. UnivarT temperature product was used for S-from-T and UnivarS. The skill of the reconstruction with respect to stratification is estimated through their correlations with the truth (Fig. 7).

Surface (5 m and 10 m) and 500-m depth stratification are the most poorly reconstructed levels by the four methods, with OI showing the lowest correlations: 0.08, 0.21, and 0.20 on average, respectively (Fig. 7a–e). Between 50 and 300-m depth, the correlations are higher (0.3–0.5), with the

300-m level being the best reconstruction for all products (~ 0.46) (Fig. 7a–e). S-from-T, UnivarS, and RedAnDA TS share similar correlation scores.

Two regions are better evaluated (correlations between 0.3–0.6 against 0.1–0.3 around 135°W) (Fig. 7b–e). These correspond to regions where the regional observational sampling is higher.

To further assess the accuracy of reconstructed stratification at the interannual timescale and to visualize the salinity contribution, the stabilizing and destabilizing effects of Ocean Salinity Stratification (OSS) in the upper 500 m layers are determined following Maes and O’Kane (2014) and O’Kane et al. (2016). We consider $N_T^2(T, S)$ and $N_S^2(T, S)$, the respective contributions of temperature and salinity to the overall stratification. We approximate them as:

$$N_T^2(T, S) \approx g \bar{\alpha} \frac{\partial T}{\partial z}, \quad (5)$$

$$N_S^2(T, S) \approx N^2(T, S) - N_T^2(T, S). \quad (6)$$

The stabilizing impact of salinity on water masses, with amplification of N^2 , denoted $OSS+$, is determined as the average over the upper 500 m of the positive values of $N_S^2(T, S)$. The destabilizing part is $OSS-$ and is the average over the upper 500 m of the negative values of $N_S^2(T, S)$. Temporal means and interannual standard deviations of

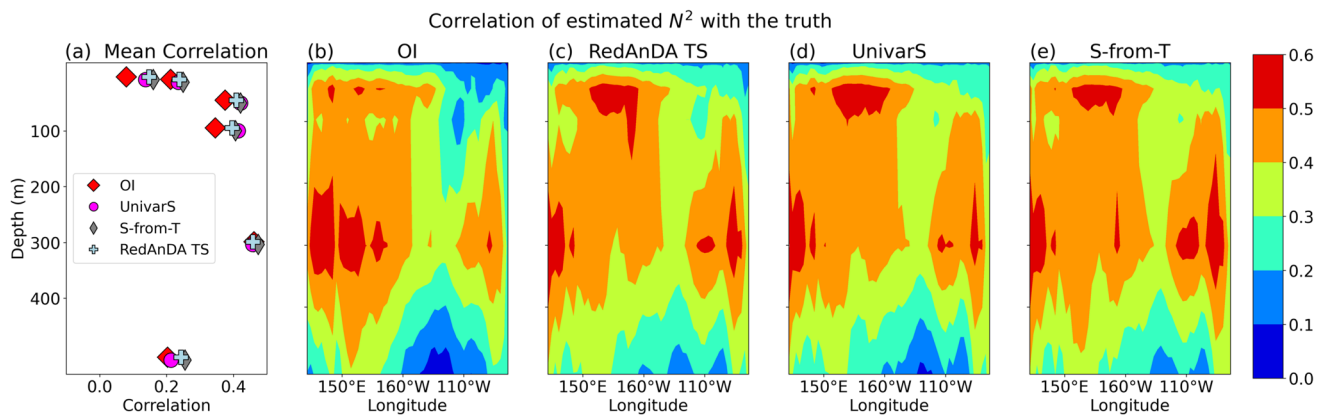


Fig. 7 (a) Meridionally (25°S – 25°N) averaged correlation of the true N^2 anomaly with N^2 anomalies estimated from the UnivarS, OI, S-from-T, and RedAnDA TS products, with depth. (b) to (e) Merid-

ionally averaged correlations of the true N^2 to N^2 estimated from OI, RedAnDA TS, UnivarS, and S-from-T, respectively, with depth

$OSS+$ and $OSS-$ for the different products are shown in Fig. 8.

In the model truth, in the Southern Hemisphere, the average $OSS-$ distribution has a greater amplitude: lower than $-8 \times 10^{-6} \text{ s}^{-2}$ (Fig. 8a). Along 10°N , it is lower: around $-4 \times 10^{-6} \text{ s}^{-2}$. This global distribution is reconstructed by OI, RedAnDA TS, and UnivarS, partly due to a priori seasonal climatology (Fig. 8c, e, and g). However, along the equator, east of 180°E , the reconstructions overestimate the $OSS-$.

Regarding the mean $OSS+$ true distribution (Fig. 8b), two regions with values exceeding $1.5 \times 10^5 \text{ s}^{-2}$ are found,

corresponding to the Western Pacific Fresh Pool (west of 180°E), the SPCZ (along 10°S), and the Eastern Pacific Fresh Pool (east of 120°W , north of 5°S). A local minimum is found on the equator, between 180°E and 120°W . The mean $OSS+$ distribution in OI, RedAnDA TS, and UnivarS reconstructions correlates well with the true $OSS+$, exhibiting no significant differences (Fig. 8).

Areas of high interannual variability (shading in Fig. 8) for $OSS-$ and $OSS+$ roughly align with regions where salinity significantly affects the upper 500 m stratification (Fig. 8a and b). For $OSS-$, south of 10°S and north of 8°N between 180°E and 120°W , the RedAnDA TS and UnivarS

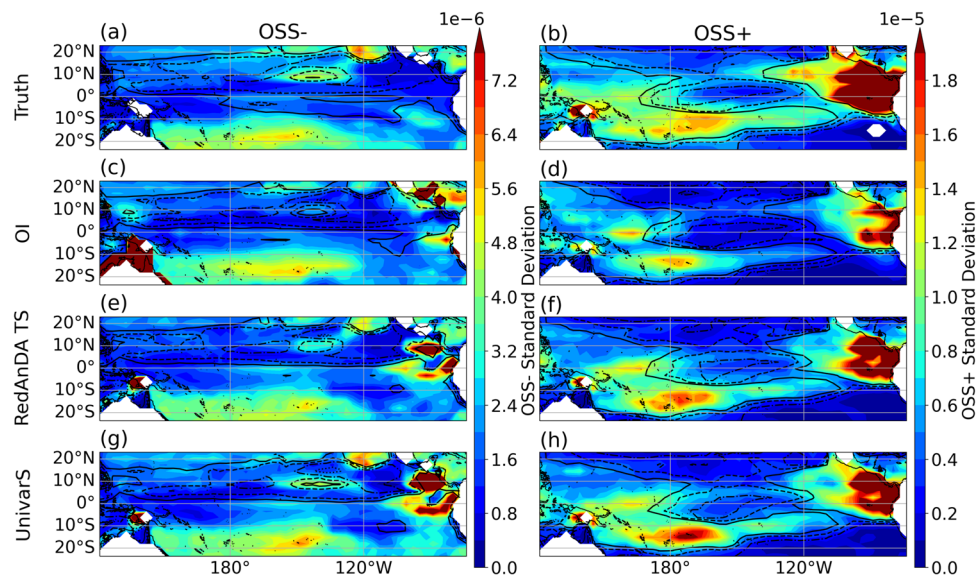


Fig. 8 (a, c, e, g) $OSS-$ averaged values (contours) and interannual standard deviations (shading) estimated from true salinity and temperature, OI T/S products, RedAnDA TS, and UnivarT/UnivarS, respectively. Dash-dot lines, dashed lines, and solid lines correspond to $[-4 \times 10^{-6}, -6 \times 10^{-6}, -8 \times 10^{-6}] \text{ s}^{-2}$. (b, d, f, h) $OSS+$ averaged

values (contours) and interannual standard deviations (shading) estimated from true salinity and temperature, RedAnDA TS, and UnivarT/UnivarS, respectively. Dash-dot lines, dashed lines, and solid lines correspond to $[5 \times 10^{-6}, 1 \times 10^{-5}, 1.5 \times 10^{-5}] \text{ s}^{-2}$

OSS— standard deviation distributions match the truth, with similar values: up to $5 \times 10^{-6} \text{ s}^{-2}$ (Fig. 8a, e, and g). The OI reconstructs only partially the *OSS*— variability (Fig. 8a and c). East of 120°W , both RedAnDA reconstructions and, to a lesser extent, the OI reconstruction, express significant spurious *OSS*— variability (Fig. 8a, c, e, and g). The spatial correlation of *OSS*— standard deviation is higher for RedAnDA TS (0.43) than for OI (-0.13) and UnivarS (0.33). The negative correlation observed in OI is primarily due to spurious results north of Australia at 10°S .

For *OSS*+, in the Western Pacific Fresh Pool, Eastern Pacific Fresh Pool and SPCZ regions, interannual standard deviations in the OI reconstruction appear underestimated compared to RedAnDA TS and UnivarS (Fig. 8b, d, f, and h). The correlation scores for OI, RedAnDA TS and UnivarS are 0.73, 0.92, and 0.92, respectively.

4.3 Application to real data

We apply RedAnDA TS and UnivarT/UnivarS to real oceanographic data, maintaining the same number of EOFs as in their OSSE versions. We reconstruct salinity and temperature for 1930–2001. The IAP and EN4 products are used for intercomparison. First, we compare the three different SSS results in terms of absolute salinity. Their respective climatologies are calculated for 2002–2020. This period aligns with the ISAS20 product, which serves as the catalog (learning period) in our experiment. We calculate salinity-based ENSO indices and compare them to the Niño 3.4 index and the Southern Oscillation Index (SOI) from the National Oceanic and Atmospheric Administration (NOAA, Ropelewski and Jones 1987; Rayner et al. 2003). The impact of salinity on stratification is then assessed for each product and the results are compared with results from O’Kane et al. (2016).

Surface salinity variations in the Pacific are shown for the different products in Fig. 9 in a longitude-time coordinate system, alongside the histogram of salinity observations over time. Between 180°E and 110°W , IAP shows higher absolute salinity on average (by approximately 0.16 PSS) than RedAnDA TS, which in turn is moderately saltier than EN4 (around 0.02 PSS). Similar values are found for UnivarS biases (Fig. 9b–e). This difference in salinity magnitude results in smaller fresh pools in the western and eastern Pacific for the IAP product (Fig. 9c). After 2002, the bias between IAP and ISAS is approximately 0.18 PSS, while it is 0.05 PSS between ISAS and EN4 and 0.23 PSS between IAP and EN4.

EN4’s relaxation to climatology in the absence of observations is apparent, notably before 1960 (Fig. 9a and d). During this period, EN4’s salinity field exhibits mostly the seasonal cycle related to the first guess climatology.

Following Qu and Yu (2014), we define the separation of the WFPF from the central Pacific salinity as the 34.8 isohaline front (Fig. 9b, c, d, and e). In their study, they defined the Niño S34.8 index, i.e., the longitude of the front, to express ENSO variability in terms of sea surface salinity (from SST and SSS analyses of satellite and in situ observations). For the IAP product, which has higher average SSS than other products, the 34.8 isohaline is located around 159°E , while it lies between 167°E and 165°E for EN4, UnivarS, and RedAnDA TS. UnivarS shows greater variability east of 120°W than RedAnDA TS. For example, in 1962, it captures a fresh event reaching 34 PSS that is absent in RedAnDA TS. Similarly, UnivarS reconstructs more persistent high-salinity events (reaching 34.8 PSS) in 1973, 1988, 1989, 1992, 1993, 1994, and 1998 (Fig. 9b and e).

We observe larger extensions of the fresh pool during the major El Niño events of 1965, 1972, 1987, and 1997. The associated retraction of the pool during major La Niña events is observed in 1954, 1975, 1988, and 1998. Disagreements between the salinity reconstructions, highlighted by the 34.8 isohalines, are unsurprisingly intensified during poorly observed periods, such as before 1965 (Fig. 9a, b, c, d, and e). During these periods, RedAnDA TS and UnivarS still reconstruct interannual variability.

To assess the ENSO-related variability of the 34.8 isohaline, we compare correlations of the different Niño S34.8 indices with the Niño 3.4 index from Rayner et al. (2003) (Table 2). A second salinity-based ENSO index complements the comparison. It is calculated as $C - (A + B)$, where A is the surface salinity anomaly averaged over 2°S – 2°N and 150°E – 170°E , B over 2°S – 2°N and 170°E – 170°W , and C over 160°E – 160°W and 25°S – 10°S (Singh et al. 2011). In their study, they used Delcroix et al.’s (2011) salinity analysis from in situ observations (Voluntary Observing Ships, TAO/TRITON moorings, CTD and Argo profiles), which was representative of 5 m depth variability. The correlation of their index with the Southern Oscillation Index (SOI) monthly anomalies from Ropelewski and Jones (1987) is analyzed.

To validate this approach, we calculate correlation scores (Table 2) in a way similar to Singh et al. (2011); Qu and Yu (2014), with equivalent indices, temporal windows and lags, enabling direct comparison with their former results (Ref. Value in Table 2). Correlation scores of the RedAnDA products concatenated with ISAS, IAP and EN4 are calculated in running temporal windows (31 and 8 years long for the indices from Singh et al. (2011) and Qu and Yu (2014), respectively). The results show increasing correlation scores from the 1930s to 2020 (not shown), suggesting an expected improvement in surface salinity reconstruction over time. The standard deviations reveal how correlation scores vary

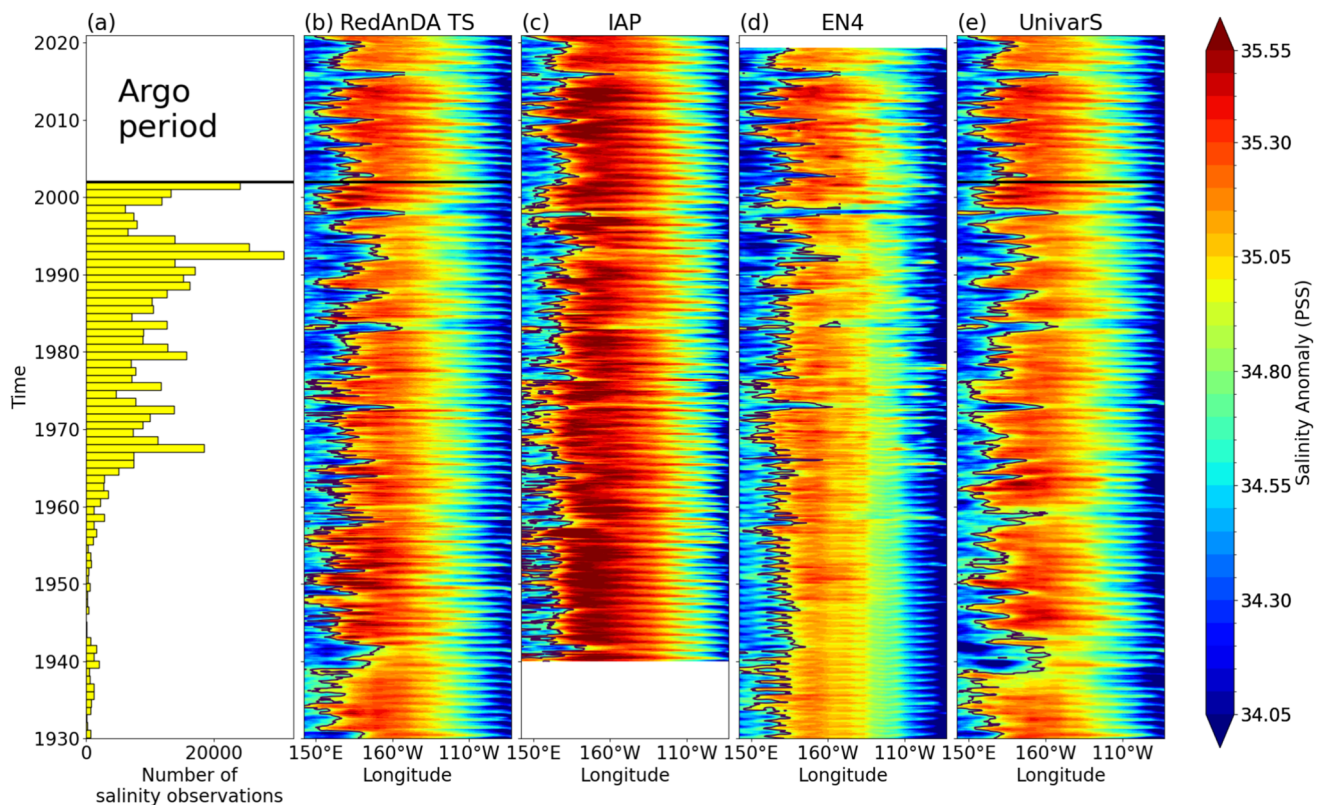


Fig. 9 (a) Annual histogram of in situ salinity observations in the tropical Pacific between 5 and 500-m depth. (b), (c), (d), and (e) Absolute salinity averaged between 5°S and 5°N at 5-m depth over time from RedAnDA TS assembled with ISAS, IAP, EN4, and UnivarS assembled

with ISAS, respectively. The thick black line highlights the 34.8 PSS isohaline, an ENSO index used by Qu and Yu (2014). The horizontal black line separates ISAS reconstructed salinity from that of RedAnDA TS (b) and UnivarS (e)

over time, with high standard deviations indicating a deterioration of correlation in the past.

RedAnDA TS indices show the highest correlations of about 0.74 and -0.56 , with relatively low standard deviations. These values are close to the reference values from Singh et al. (2011) and Qu and Yu (2014), respectively, suggesting good reconstruction skill of RedAnDA TS compared to other methods. Standard deviations for EN4 and UnivarS are high, indicating discrepancies between the correlation scores during the mid-20th century and the 21st century. On average, IAP scores are lower than those of RedAnDA TS and UnivarS. While UnivarS correlation scores are satisfactory, poor observational sampling significantly impacts the reconstruction. The difference between RedAnDA TS scores and

those from Singh et al. (2011) may result from differences in their respective analysis methods.

The estimates of salinity control on stratification ($OSS-$ and $OSS+$) by RedAnDA TS, EN4, IAP, and UnivarT/UnivarS are compared in Fig. 10. The results are compared with those of O’Kane et al. (2016), whose Fig. 9 shows the upper 500 m averaged $OSS+$ and $OSS-$ fields, estimated from the CSIRO reanalysis dataset Oke et al. (2013), an EnOI scheme, using the ACCESS-O model, and in situ and remote sensing observations.

The averaged distributions (contours) for $OSS-$ and $OSS+$ from RedAnDA TS and UnivarS resemble those obtained in the OSSE (Fig. 8) and the results of O’Kane et al. (2016, Fig. 9). RedAnDA products and EN4 fields show sim-

Table 2 First row: Temporal mean and standard deviation of 8-year running correlations between Qu and Yu’s (2014) salinity-based ENSO index, with Niño 3.4 leading by 1 month, for RedAnDA TS-ISAS20 (1930–2020), IAP (1940–2021), EN4 (1930–2019), and UnivarS-ISAS20 (1930–2020)

	Ref. Value	RedAnDA TS	IAP	EN4	UnivarS
Qu & Yu, 2014	0.75	0.74 ± 0.17	0.48 ± 0.22	0.46 ± 0.28	0.64 ± 0.28
Singh et al., 2011	-0.88	-0.56 ± 0.09	-0.53 ± 0.06	-0.51 ± 0.11	-0.55 ± 0.15

Second row: Mean and standard deviation of 31-year running correlations between Singh et al.’s (2011) salinity-based ENSO index with the SOI index leading by 2 months. The first column (Ref. Value) shows the scores obtained in both reference studies during their respective periods: 2005–2013 for Qu and Yu (2014) and 1977–2008 for Singh et al. (2011)

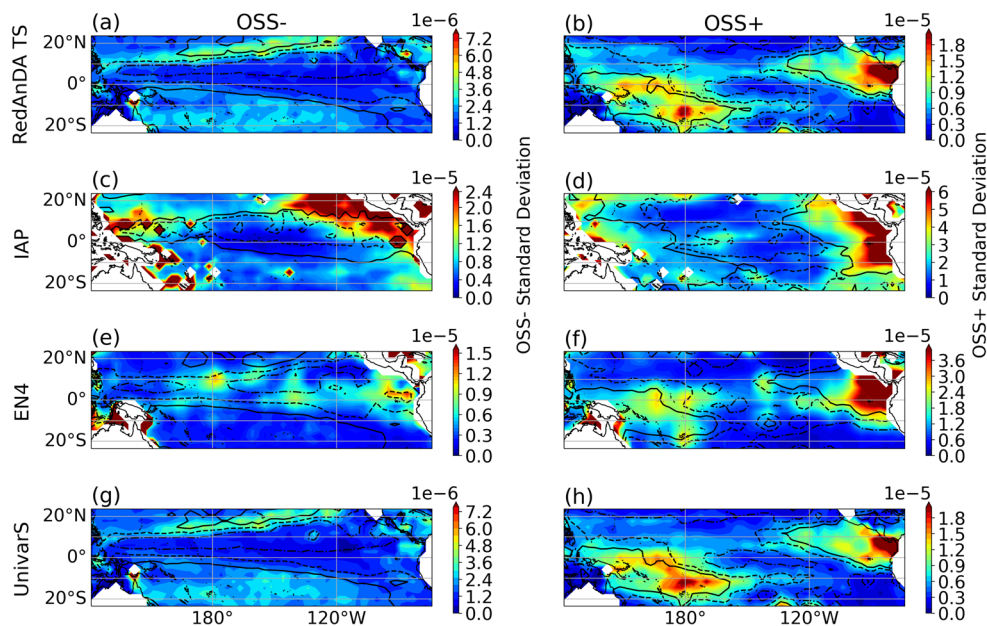


Fig. 10 (a, c, e, g) *OSS*− averaged values (contours) and interannual standard deviations (colors) estimated from RedAnDA TS, EN4, IAP, and UnivarS, respectively. Dashed and dotted lines, dashed lines, and solid lines correspond to $[-0.5 \times 10^{-5}, -1 \times 10^{-5}, -1.5 \times 10^{-5}]$

s^{-2} . (b, d, f, h) *OSS*+ averaged values (contours) and interannual standard deviations (colors) estimated from RedAnDA TS, EN4, IAP, and UnivarS, respectively. Dashed and dotted lines, dashed lines, and solid lines correspond to $[1 \times 10^{-5}, 2 \times 10^{-5}, 3 \times 10^{-5}] s^{-2}$

ilar patterns, particularly for *OSS*+, while IAP's absolute salinity appears notably intense.

The *OSS*− and *OSS*+ interannual variability (color shades) estimated by RedAnDA TS and UnivarS follows the areas of pronounced salinity effects on stratification, a previously observed tendency. Both products reproduce the variations observed in O'Kane et al. (2016) (Fig. 10a, b, g and h). In the EN4 and IAP products, the reconstructed variability differs from the analysis of O'Kane et al. (2016) (Fig. 10c, d, e and f).

5 Discussion

Using the OSSE, we calculate RedAnDA's performance compared to the classical OI approach. RedAnDA variations (RedAnDA TS, S-from-T, UnivarS) can estimate the interannual changes in salinity anomalies more accurately than a classical OI approach. Applied to real data, RedAnDA TS and UnivarS are compared with IAP and EN4 products. The interannual scale is better represented by the RedAnDA approaches (Tables 1, 2, Figs. 6, 8, 9, 10). Notably, RedAnDA better estimates the complexity of the spatial distribution of these anomalies through the use of EOFs. The vertical structure of salinity in RedAnDA products is more consistent with its description in O'Kane et al.'s (2016) dedicated study.

ENSO-related surface salinity variability is particularly well represented in coupled EOFs approaches (RedAnDA TS, S-from-T). This type of space-reduction captures tem-

perature and salinity interannual covariability and filters out some salinity small-scale features, likely intensifying the ENSO-related signature in salinity reconstruction.

The interannual variability of SSS derived from RedAnDA TS shows a stronger correlation with ENSO-related temperature indices than UnivarS, IAP, and EN4. This correlation value is consistent with reference values from other studies (Singh et al. 2011; Qu and Yu 2014).

S-from-T's satisfactory results demonstrate the accuracy of coupled EOFs approach in representing interannual changes. Its results are close to those of RedAnDA TS. This likely occurs because salinity observations are significantly fewer than temperature observations, thus attenuating their impact when assimilated in RedAnDA TS.

The choice of space-reduction and variable coupling techniques is essential in determining the final result. In this study, the retained component of TS covariability favors the reconstruction of ENSO-related changes at the expense of reconstructing variability in the Eastern Pacific and at depth. Future work should explore different approaches to estimate rotated EOFs (Lian and Chen 2012) and other space-reduction techniques like Dynamical Modes (Gavrilov et al. 2019).

The seasonal scale is well estimated in all different products of the OSSE (OI, RedAnDA TS, UnivarS and S-from-T). The climatology is well estimated over the learning period. However, all methods struggle to resolve decadal variability. This low-frequency variability is likely associated with changes in the North and South Pacific gyres, and the tropical

Pacific EOF decomposition for RedAnDA may not include them. Additionally, the length of the learning period may limit the extent to which decadal variance can be incorporated into EOFs. Considering EOFs for the whole Pacific could help capture this aspect of variability.

A 20-year learning period limits the EOFs' ability to capture long-term variability. Compared to IAP and EN4, RedAnDA TS trends in salinity differ, with increasing SSS west of 170°E, north of 5°N. A limited catalog might not adequately represent the diversity of system dynamics and its long-term variability. Investigating the sensitivity of temperature and salinity reconstruction quality to the size of the catalog is necessary to fully understand these implications.

The RedAnDA approaches effectively capture the (de-)stabilizing effects of salinity on stratification on average. At the interannual scale, these effects were well reconstructed, though with an overestimation of the destabilizing effect OSS^- and an underestimation of the stabilizing effect OSS^+ in the Eastern Pacific fresh pool in the OSSE.

The better reconstructed interannual variations in stratification by UnivarS and S-from-T stem from the inclusion of UnivarT, which should be more thoroughly compared with the temperature component of RedAnDA TS in future work. UnivarT was studied in a previous study (Oulhen et al. 2024) and showed promising results.

From the salinity reconstructions by RedAnDA TS, the interannual changes appear more similar to EN4 results than to IAP. As IAP uses CMIP5 outputs to calculate the background and covariance used in its analysis, this may indicate the limitations of using model-based information when reconstructing interannual patterns of salinity variability with real data.

6 Conclusion

RedAnDA was used to reconstruct monthly salinity (and temperature) in the upper 500 m of the tropical Pacific during the 20th century, assimilating data from Ocean Stations, CTDs, moored, and drifting buoys. It has also been extended to the coupled reconstruction of temperature and salinity (RedAnDA TS), leveraging the greater abundance of temperature observations (MBTs, XBTs) during this period and the common dynamics governing their distribution. To exploit this mutual relationship, their covariance was calculated over the Argo period, from 2002 to 2020, and employed to construct Empirical Orthogonal Functions (EOFs) that optimally reduce the dimensionality of the system while preserving the main features of temperature and salinity covariability.

In an Observing System Simulation Experiment (OSSE), the performance of the reconstructions was assessed and compared, notably with a state-of-the-art Optimal Interpolation product following Good et al.'s (2013) methodology.

The results showed that for reconstructing ENSO-related salinity variability, RedAnDA TS (salinity reconstruction coupled with temperature) is more effective than the OI and UnivarS, particularly during periods with fewer in situ observations. Including temperature information when reconstructing salinity gives satisfactory results, as evidenced by a better representation of major ENSO events in the western tropical Pacific. The S-from-T approach was also implemented, corresponding to the projection of UnivarT reconstruction onto coupled EOFs to determine the salinity field through correlation-based equivalence. The vertical structure of salinity is also enhanced, showing better correlation scores in stratification reconstruction. However, decadal variability may still be poorly estimated in our products, and caution should be exercised when using them to study long-term changes.

RedAnDA TS and UnivarS were applied to real data, enabling the reconstruction of regional temperature and salinity patterns and time series since 1930. Salinity results were compared with EN4 and IAP data, showing notable differences in salinity mean state, stratification mean state, and fluctuations.

The new TS product spanning 1930–2001, combined with its learning period (ISAS20), constitutes an unprecedented 90-year super-product for assessing the modulation of ENSO by longer components of variability and vice versa (Power et al. 2021; Hu et al. 2022). As ENSO patterns evolve in a warming climate, precipitation may intensify and shift eastward (Cai et al. 2021), potentially freshening salinity in the western equatorial Pacific and SPCZ (Singh et al. 2011). The combination of RedAnDA TS and ISAS20 could provide a valuable database for studying such shifts in ENSO regimes, provided low-frequency variability is improved.

Acknowledgements ISAS temperature and salinity monthly gridded field products are made freely available by SNO Argo France at LOPS Laboratory (supported by UBO/CNRS/Ifremer/IRD) and IUEM Observatoire des Sciences de l'Univers (UBO/CNRS/IRD) at <https://doi.org/10.17882/52367>. The authors express gratitude to Thierry Penduff and Jean-Marc Molines for providing OCCIPUT outputs and their invaluable assistance. Appreciation is extended to Gaël Alory, Aida Alvera Azcárate, Laurent Bertino, Sophie Cravatte, Christophe Maes, Anne-Marie Tréguier, Jérôme Vialard, and Pascal Yiou for their constructive discussions and their suggestions.

Author Contributions E.O designed the study, conducted the analyse and wrote the main manuscript text. All other authors (N.K, P.T, B.B, and F.S) have participated to design the study, have edited the manuscript, and have looked for funding.

Funding Erwan Oulhen was funded by the Région Bretagne and the Université de Bretagne Occidentale. This study was conducted within the framework of the LEFE-GMMC OASIS project (CNRS-INSU/GMMC).

Data Availability ISAS20 data are available through <https://doi.org/10.17882/52367>; EN.4.2.2 profiles and objective analyses are available through <https://www.metoffice.gov.uk/hadobs/en4/download-en4-2-2>.

[html](#); OCCIPUT data are available upon demand; IAP data are available through <http://www.ocean.iap.ac.cn/>; NOAA Niño 3.4 and SOI indices are available through <https://psl.noaa.gov/>; RedAnDA codes are available through <https://github.com/eoulhen/RedAnDA>. RedAnDA TS, UnivarS and UnivarT products, as well as other datasets from our experiments are available through 10.17882/102598.

Declarations

Ethical Approval and Consent to participate Not applicable.

Human Ethics Not applicable.

Consent for publication All authors consent for publication.

Competing Interests The authors declare no competing interests.

Clinical Trial Number in the manuscript Not applicable.

References

- Abraham JP, Baringer M, Bindoff NL et al (2013) A review of global ocean temperature observations: implications for ocean heat content estimates and climate change. *Rev Geophys* 51(3):450–483. <https://doi.org/10.1002/rog.20022>
- Atkinson CP, Rayner NA, Kennedy JJ et al (2014) An integrated database of ocean temperature and salinity observations. *J Geophys Res Oceans* 119(10):7139–7163. <https://doi.org/10.1002/2014JC010053>
- Basinski-Ferris A, Zanna L (2023) Estimating freshwater flux amplification with ocean tracers via linear response theory. *Earth Syst Dyn Discuss* 2023:1–2. <https://doi.org/10.5194/esd-2023-14>
- Bessi eres L, Leroux S, Brankart JM et al (2017) Development of a probabilistic ocean Modelling system based on NEMO 3.5: Application at eddy resolution. *Geosci Model Dev* 10(3):1091–1106. <https://doi.org/10.5194/gmd-10-1091-2017>
- Bosc C, Delcroix T, Maes C (2009) Barrier layer variability in the western Pacific warm pool from 2000 to 2007. *J Geophys Res Oceans* 114(C6). <https://doi.org/10.1029/2008JC005187>
- Boyer TP, Antonov JJ, Garcia H et al (2006) NOAA Atlas NESDIS 60. World Ocean Database 2005
- Bryan K, Cox MD (1972) An approximate equation of state for numerical models of ocean circulation. *J Phys Oceanogr* 2(4):510–514. [https://doi.org/10.1175/1520-0485\(1972\)002<0510:AAEOSF>2.0.CO;2](https://doi.org/10.1175/1520-0485(1972)002<0510:AAEOSF>2.0.CO;2)
- Cai W, Santos A, Collins M et al (2021) Changing El Ni o–Southern Oscillation in a warming climate. *Nat Rev Earth Environ* 2. <https://doi.org/10.1038/s43017-021-00199-z>
- Chang YS, Rosati A, Zhang S (2011) A construction of pseudo salinity profiles for the global ocean: Method and evaluation. *J Geophys Res Oceans* 116(C2). <https://doi.org/10.1029/2010JC006386>
- Cheng L, Zhu J (2016) Benefits of CMIP5 multimodel ensemble in reconstructing historical ocean subsurface temperature variations. *J Clim* 29(15):5393–5416. <https://doi.org/10.1175/JCLI-D-15-0730.1>
- Cheng L, Trenberth KE, Fasullo J et al (2017) Improved estimates of ocean heat content from 1960 to 2015. *Sci Adv* 3(3). <https://doi.org/10.1126/sciadv.1601545>
- Cheng L, Trenberth KE, Gruber N et al (2020) Improved estimates of changes in upper ocean salinity and the hydrological cycle. *J Clim* 33(23):10357–10381. <https://doi.org/10.1175/JCLI-D-20-0366.1>
- Chi J, Qu T, Du Y et al (2022) Ocean salinity indices of interannual modes in the tropical Pacific. *Clim Dyn* 5. <https://doi.org/10.1007/s00382-021-05911-9>
- Cowley R, Killick RE, Boyer T et al (2021) International quality-controlled ocean database (IQuOD) v0.1: The temperature uncertainty specification. *Front Mar Sci* 8. <https://doi.org/10.3389/fmars.2021.689695>
- Cronin MF, McPhaden MJ (2002) Barrier layer formation during westerly wind bursts. *J Geophys Res Oceans* 107(C12):SRF 21–1–SRF 21–12. <https://doi.org/10.1029/2001JC001171>
- de Boyer Mont g t C, Mignot J, Lazar A et al (2007) Control of salinity on the mixed layer depth in the world ocean: 1. General description. *J Geophys Res Oceans* 112(C6). <https://doi.org/10.1029/2006JC003953>
- Delcroix T, H nin C (1991) Seasonal and interannual variations of sea surface salinity in the tropical Pacific ocean. *J Geophys Res Oceans* 96(C12):22135–22150. <https://doi.org/10.1029/91JC02124>
- Delcroix T, Alory G, Cravatte S et al (2011) A gridded sea surface salinity data set for the tropical Pacific with sample applications (1950–2008). *Deep Sea Research Part I: Oceanographic Research Papers* 58(1):38–48. <https://doi.org/10.1016/j.dsr.2010.11.002>
- Durack PJ (2015) Ocean salinity and the global water cycle. *Oceanography* 28:20–3. <https://doi.org/10.5670/oceanog.2015.03>
- Durack PJ, Wijffels SE (2010) Fifty-year trends in global ocean salinities and their relationship to broad-scale warming. *J Clim* 23(16):4342–4362. <https://doi.org/10.1175/2010JCLI3377.1>
- Durack PJ, Wijffels SE, Matear RJ (2012) Ocean salinities reveal strong global water cycle intensification during 1950 to 2000. *Science* 336(6080):455–458. <https://doi.org/10.1126/science.1212222>
- Dussin R, Barnier B, Brodeau L et al (2016) Drakkar forcing set DFS5. MyOcean Report. https://www.drakkar-ocean.eu/publications/reports/report_DFS5v3_April2016.pdf
- Emery WJ, Thomson RE (2001) Chapter 4 - The spatial analyses of data fields. Elsevier Science, Amsterdam, pp 305–370. <https://doi.org/10.1016/B978-044450756-3/50005-8>
- Farrell BF, Ioannou PJ (2001) Accurate low-dimensional approximation of the linear dynamics of fluid flow. *J Atmos Sci* 58(18):2771–2789. [https://doi.org/10.1175/1520-0469\(2001\)058<2771:ALDAOT>2.0.CO;2](https://doi.org/10.1175/1520-0469(2001)058<2771:ALDAOT>2.0.CO;2)
- Fedorov AV, Pacanowski RC, Philander SG et al (2004) The effect of salinity on the wind-driven circulation and the thermal structure of the upper ocean. *J Phys Oceanogr* 34(9):1949–1966. [https://doi.org/10.1175/1520-0485\(2004\)034<1949:TEOSOT>2.0.CO;2](https://doi.org/10.1175/1520-0485(2004)034<1949:TEOSOT>2.0.CO;2)
- Fox-Kemper B (2021) Ocean, cryosphere and sea level change. In: AGU Fall Meeting Abstracts, pp U13B–09. <https://doi.org/10.1017/9781009157896.011>
- Fujii Y, Kamachi M (2003) Three-dimensional analysis of temperature and salinity in the equatorial Pacific using a variational method with vertical coupled temperature-salinity EOF modes. *J Geophys Res* 10. <https://doi.org/10.1029/2002JC001745>
- Gaillard F, Reynaud T, Thierry V et al (2016) In situ-based reanalysis of the global ocean temperature and salinity with ISAS: Variability of the heat content and steric height. *J Clim* 29(4):1305–1323. <https://doi.org/10.1175/JCLI-D-15-0028.1>
- Gavrilov A, Seleznev A, Mukhin D et al (2019) Linear dynamical modes as new variables for data-driven ENSO forecast. *Clim Dyn* 52:2199–2216. <https://doi.org/10.1007/s00382-018-4255-7>
- Ge K, Li Y, Lyu Y et al (2023) Surface salinity changes of the tropical and subtropical oceans since 1970 and their relationship with surface freshwater fluxes. *J Geophys Res Oceans* 128(12). <https://doi.org/10.1029/2023JC020207>
- Goes M, Christophersen J, Dong S et al (2018) An updated estimate of salinity for the Atlantic Ocean sector using temperature-salinity relationships. *J Atmos Oceanic Tech* 35. <https://doi.org/10.1175/JTECH-D-18-0029.1>

- Good SA, Martin MJ, Rayner NA (2013) EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J Geophys Res Oceans* 118(12):6704–6716. <https://doi.org/10.1002/2013JC009067>
- Guan C, Tian F, McPhaden MJ, et al (2023) Zonal structure of tropical Pacific surface salinity anomalies affects the eastern and central Pacific El Niños differently. *Geophys Res Lett* 50(21). <https://doi.org/10.1029/2023GL105554>
- Hansen J, Sato M, Kharecha P et al (2011) Earth's energy imbalance and implications. *Atmos Chem Phys* 11:13421–13449. <https://doi.org/10.5194/acp-11-13421-2011>
- He Z, Wang X, Wu X et al (2021) Projecting three-dimensional ocean thermohaline structure in the North Indian ocean from the satellite sea surface data based on a variational method. *J Geophys Res Oceans* 126. <https://doi.org/10.1029/2020JC016759>
- Helm KP, Bindoff NL, Church JA (2010) Changes in the global hydrological cycle inferred from ocean salinity. *Geophys Res Lett* 37:L18701. <https://doi.org/10.1029/2010GL044222>
- Hosoda S, Suga T, Shikama N et al (2009) Global surface layer salinity change detected by Argo and its implication for hydrological cycle intensification. *J Oceanogr* 65:579–586. <https://doi.org/10.1007/s10872-009-0049-1>
- Hu J, Wang H, Gao C et al (2022) Interdecadal wind stress variability over the tropical Pacific causes ENSO diversity in an intermediate coupled model. *Clim Dyn* 6. <https://doi.org/10.1007/s00382-022-06414-x>
- Johnson GC, Hosoda S, Jayne SR et al (2022) Argo-two decades: Global oceanography, revolutionized. *Ann Rev Mar Sci* 14(1):379–403. <https://doi.org/10.1146/annurev-marine-022521-102008>
- Kao HY, Yu JY (2009) Contrasting eastern-Pacific and central-Pacific types of ENSO. *J Clim* 22(3):615–632. <https://doi.org/10.1175/2008JCLI2309.1>
- Kaplan A, Kushnir Y, Cane M et al (1997) Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. *J Geophys Res* 102:27835–27860. <https://doi.org/10.1029/97JC01734>
- Kessler WS, Cravatte S (2013) ENSO and short-term variability of the South equatorial current entering the Coral Sea. *J Phys Oceanogr* 43(5):956–969. <https://doi.org/10.1175/JPO-D-12-0113.1>
- Kolodziejczyk N, Prigent-Mazella A, Gaillard F (2023) ISAS temperature, salinity, dissolved oxygen gridded field. <https://doi.org/10.17882/52367>
- Lagerloef GSE (2002) Introduction to the special section: The role of surface salinity on upper ocean dynamics, air-sea interaction and climate. *J Geophys Res Oceans* 107(C12):SRF 1–1–SRF 1–2. <https://doi.org/10.1029/2002JC001669>
- Lguensat R, Tandeo P, Ailliot P et al (2017) The analog data assimilation. *Mon Weather Rev* 145(10):4093–4107. <https://doi.org/10.1175/MWR-D-16-0441.1>
- Li G, Cheng L, Zhu J et al (2020a) Increasing ocean stratification over the past half-century. *Nat Clim Chang* 10. <https://doi.org/10.1038/s41558-020-00918-2>
- Li X, Yang Y, Li R et al (2020b) Structure and dynamics of the Pacific North equatorial subsurface current. *Sci Rep* 10(1):1175. <https://doi.org/10.1038/s41598-020-68605-y>
- Li G, Cheng L, Pan Y et al (2023) A global gridded ocean salinity dataset with 0.5° horizontal resolution since 1960 for the upper 2000 m. *Front Mar Sci* 1. <https://doi.org/10.3389/fmars.2023.1108919>
- Lian T, Chen D (2012) An evaluation of rotated EOF analysis and its application to tropical Pacific SST variability. *J Clim* 25(15):5361–5373. <https://doi.org/10.1175/JCLI-D-11-00663.1>
- Lorenz EN (1969) The predictability of a flow which possesses many scales of motion. *Tellus* 21(3):289–30. <https://doi.org/10.1111/j.2153-3490.1969.tb00444.x>
- Lukas R, Lindstrom E (1991) The mixed layer of the western equatorial Pacific ocean. *J Geophys Res Oceans* 96(S01):3343–3357. <https://doi.org/10.1029/90JC01951>
- Maes C (2008) On the ocean salinity stratification observed at the eastern edge of the equatorial Pacific warm pool. *J Geophys Res Oceans* 113(C3):1. <https://doi.org/10.1029/2007JC004297>
- Maes C, Behringer D (2000) Using satellite-derived sea level and temperature profiles for determining the salinity variability: A new approach. *J Geophys Res Oceans* 105(C4):8537–8547. <https://doi.org/10.1029/1999JC900279>
- Maes C, O'Kane TJ (2014) Seasonal variations of the upper ocean salinity stratification in the tropics. *J Geophys Res Oceans* 119(3):1706–1722. <https://doi.org/10.1002/2013JC009366>
- Maes C, McPhaden MJ, Behringer D (2002) Signatures of salinity variability in tropical Pacific Ocean dynamic height anomalies. *J Geophys Res Oceans* 107(C12):SRF 13–1–SRF 13–13. <https://doi.org/10.1029/2000JC000737>
- Maes C, Picaut J, Belamari S (2005) Importance of the salinity barrier layer for the buildup of el nino. *J Clim* 18(1):104–118. <https://doi.org/10.1175/JCLI-3214.1>
- Maes C, Ando K, Delcroix T et al (2006) Observed correlation of surface salinity, temperature and barrier layer at the eastern edge of the western Pacific warm pool. *Geophys Res Lett* 33. <https://doi.org/10.1029/2005GL024772>
- Masutani M, Schlatter TW, Errico RM et al (2010) Observing System Simulation Experiments, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 647–679. https://doi.org/10.1007/978-3-540-74703-1_24
- McDougall T, Barker P (2011) Getting started with TEOS-10 and the gibbs seawater (GSW) oceanographic toolbox. *SCOR/IAPSO WG 127*:1–28
- McDougall T, Feistel R, Millero F et al (2010) The international thermodynamic equation of seawater 2010 (TEOS-10): calculation and use of thermodynamic properties
- Mignot J, de Boyer Montégut C, Lazar A et al (2007) Control of salinity on the mixed layer depth in the world ocean: 2. Tropical areas. *J Geophys Res Oceans* 112(C10). <https://doi.org/10.1029/2006JC003954>
- Miller JR (1976) The salinity effect in a mixed layer ocean model. *J Phys Oceanogr* 6(1):29–35. [https://doi.org/10.1175/1520-0485\(1976\)006<0029:TSEIAM>2.0.CO;2](https://doi.org/10.1175/1520-0485(1976)006<0029:TSEIAM>2.0.CO;2)
- Murtugudde R, Busalacchi AJ (1998) Salinity effects in a tropical ocean model. *J Geophys Res Oceans* 103(C2):3283–3300. <https://doi.org/10.1029/97JC02438>
- O'Kane TJ, Monselesan DP, Maes C (2016) On the stability and spatiotemporal variance distribution of salinity in the upper ocean. *J Geophys Res Oceans* 121(6):4128–4148. <https://doi.org/10.1002/2015JC011523>
- Oke PR, Sakov P, Cahill ML et al (2013) Towards a dynamically balanced eddy-resolving ocean reanalysis: BRAN3. *Ocean Model* 67:52–70. <https://doi.org/10.1016/j.ocemod.2013.03.008>
- Olmedo E, Turiel A, González-Gambau V et al (2022) Increasing stratification as observed by satellite sea surface salinity measurements. *Sci Rep* 12(1):627. <https://doi.org/10.1038/s41598-022-10265-1>
- Oulhen E, Kolodziejczyk N, Tandeo P et al (2024) Reconstructing the ocean state using Argo data and a data-driven method. *J Atmos Oceanic Tech* 41(12):1165–1179. <https://doi.org/10.1175/JTECH-D-23-0157.1>
- Peixoto JP, Oort AH (1992) Physics of climate. American Institute of Physics, New York, NY (United States)
- Penduff T, Barnier B, Terray L et al (2014) Ensembles of eddy-resolving ocean simulations for climate. *CLIVAR Exchanges, Special Issue on High Resolution Ocean Climate Modelling* 19:26–29. https://www.researchgate.net/publication/283497595_Ensembles_of_eddy-resolving_ocean_simulations_for_climate

- Picaut J, Ioualalen M, Menkes C et al (1996) Mechanism of the zonal displacements of the Pacific warm pool: Implications for ENSO. *Science* 274(5292):1486–1489. <https://doi.org/10.1126/science.274.5292.1486>
- Power S, Lengaigne M, Capotondi A et al (2021) Decadal climate variability in the tropical Pacific: Characteristics, causes, predictability, and prospects. *Science* 374(6563):eaay9165. <https://doi.org/10.1126/science.aay916>
- Qu T, Yu JY (2014) ENSO indices from sea surface salinity observed by Aquarius and Argo. *J Oceanogr* 70:367–375. <https://doi.org/10.1007/s10872-014-0238-4>
- Qu T, Song YT, Maes C (2014) Sea surface salinity and barrier layer variability in the equatorial Pacific as seen from Aquarius and Argo. *J Geophys Res Oceans* 119(1):15–29. <https://doi.org/10.1002/2013JC009375>
- Rayner NA, Parker DE, Horton EB et al (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J Geophys Res Atmos* 108(D14). <https://doi.org/10.1029/2002JD002670>
- Riser SC (2016) Fifteen years of ocean observations with the global Argo array. *Nat Climate Change* 6:145–153. <https://doi.org/10.1038/nclimate2872>
- Roemmich D, John Gould W, Gilson J (2012) 135 years of global ocean warming between the Challenger expedition and the Argo programme. *Nat Clim Chang* 2(6):425–428. <https://doi.org/10.1038/nclimate1461>
- Roemmich D, Church J, Gilson J et al (2015) Unabated planetary warming and its ocean structure since 2006. *Nat Clim Chang* 5:240–245. <https://doi.org/10.1038/nclimate2513>
- Roemmich D, Alford MH, Claustre H et al (2019) On the future of Argo: A global, full-depth, multi-disciplinary array. *Front Mar Sci*. <https://doi.org/10.3389/fmars.2019.00439>
- Romero E, Tenorio-Fernandez L, Castro I et al (2021) Filtering method based on cluster analysis to avoid salinity drifts and recover Argo data in less time. *Ocean Sci* 17(5):1273–128. <https://doi.org/10.5194/os-17-1273-2021>
- Ropelewski CF, Jones PD (1987) An extension of the Tahiti-Darwin Southern Oscillation Index. *Mon Weather Rev* 115(9):2161–2165. [https://doi.org/10.1175/1520-0493\(1987\)115<2161:AEOTTS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<2161:AEOTTS>2.0.CO;2)
- von Schuckmann K, Sallée JB, Chambers D et al (2014) Monitoring ocean heat content from the current generation of global ocean observing systems. *Ocean Sci* 10:547–557. <https://doi.org/10.5194/osd-10-923-2013>
- Singh A, Delcroix T, Cravatte S (2011) Contrasting the flavors of El Niño-Southern Oscillation using sea surface salinity observations. *J Geophys Res Oceans* 116(C6). <https://doi.org/10.1029/2010JC006862>
- Skiris N, Zika JD, Nurser G et al (2016) Global water cycle amplifying at less than the Clausius-Clapeyron rate. *Sci Rep* 6:38752. <https://doi.org/10.1038/srep38752>
- Sprintall J, Tomczak M (1992) Evidence of the barrier layer in the surface layer of tropics. *J Geophys Res* 97:7305–7316. <https://doi.org/10.1029/92JC00407>
- Steffen J, Bourassa M (2018) Barrier layer development local to tropical cyclones based on Argo float observations. *J Phys Oceanogr* 48(9):1951–1968. <https://doi.org/10.1175/JPO-D-17-0262.1>
- Sun C, Thresher A, Keeley R et al (2010) The data management system for the global temperature and salinity profile programme. *Proceedings OceanObs* 9:86
- Takahashi K, Montecinos A, Goubanova K et al (2011) ENSO regimes: Reinterpreting the canonical and Modoki El Niño. *Geophys Res Lett* 38(10):1. <https://doi.org/10.1029/2011GL047364>
- Tandeo P, Ailliot P, Ruiz J et al (2015) Combining analog method and ensemble data assimilation: Application to the Lorenz-63 chaotic system. In: Machine learning and data mining approaches to climate science: Proceedings of the 4th international workshop on climate informatics. Springer, pp 3–12
- Tang B, Zhao D, Cui C et al (2022) Reconstruction of ocean temperature and salinity profiles in the Northern South China Sea using satellite observations. *Front Mar Sci* 9:945835. <https://doi.org/10.3389/fmars.2022.945835>
- Tang L, Zhou H, Li J et al (2024) Effect of Argo salinity drift since 2016 on the estimation of regional steric sea level change rates. *Remote Sens* 16(11). <https://doi.org/10.3390/rs16111855>. <https://www.mdpi.com/2072-4292/16/11/1855>
- Thompson B, Crease J, Gould J (2001) Chapter 1.3 The origins, development and conduct of WOCE. In: Siedler G, Church J, Gould J (eds) *Ocean Circulation and Climate*, International Geophysics, vol 77. Academic Press, p 31–VIII. [https://doi.org/10.1016/S0074-6142\(01\)80110-8](https://doi.org/10.1016/S0074-6142(01)80110-8)
- Vinogradova NT, Ponte RM (2017) In search of fingerprints of the recent intensification of the ocean water cycle. *J Clim* 30(14):5513–5528. <https://doi.org/10.1175/JCLI-D-16-0626.1>
- Wang F, Scott R (2005) On the prediction of linear stochastic systems with a low-order model. *Dynamic Meteorology and Oceanography*, Tellus A. <https://doi.org/10.3402/tellusa.v57i1.14604>
- Wang K, Yin K (2021) Barrier effect of the Pearl River estuarine plume on wind-induced coastal upwelling of nutrients. *J Geophys Res Biogeosci* 126(11). <https://doi.org/10.1029/2020JG006067>
- Wong APS, Wijffels SE, Riser SC et al (2020) Argo data 1999–2019: Two million temperature-salinity profiles and subsurface velocity observations from a global array of profiling floats. *Front Mar Sci* 7. <https://doi.org/10.3389/fmars.2020.00700>
- Wong APS, Gilson J, Cabanes C (2023) Argo salinity: Bias and uncertainty evaluation. *Earth Syst Sci Data* 15(1):383–393. <https://doi.org/10.5194/essd-15-383-2023>
- Yim B, Yeh SW, Noh Y et al (2008) Sea surface salinity variability and its relation to El Niño in a cgm. *Asia-Pac J Atmos Sci* 44:173–189
- Yu L (2023) Connecting subtropical salinity maxima to tropical salinity minima: synchronization between ocean dynamics and the water cycle. *Prog Oceanogr* 219:103172. <https://doi.org/10.1016/j.pocean.2023.103172>
- Zhang R, Gao C, Kang X et al (2015) ENSO modulations due to interannual variability of freshwater forcing and ocean biology-induced heating in the tropical Pacific. *Sci Rep*. <https://doi.org/10.1038/srep18506>
- Zhang X, Sprintall J, Zeng L (2021) What role does the barrier layer play during extreme El Niño events? *J Geophys Res Oceans* 126(5). <https://doi.org/10.1029/2020JC017001>
- Zhao S, Jin FF, Stuecker MF (2021) Understanding lead times of warm water volumes to ENSO sea surface temperature anomalies. *Geophys Res Lett* 48(19). <https://doi.org/10.1029/2021GL094366>
- Zheng F, Zhang R (2015) Interannually varying salinity effects on ENSO in the tropical Pacific: a diagnostic analysis from Argo. *Ocean Dyn* 65. <https://doi.org/10.1007/s10236-015-0829-7>
- Zheng F, Zhang RH (2012) Effects of interannual salinity variability and freshwater flux forcing on the development of the 2007/08 La Niña event diagnosed from Argo and satellite data. *Dyn Atmos Oceans* 57:45–57. <https://doi.org/10.1016/j.dynatmoce.2012.06.002>
- Zheng F, Zhang R, Zhu J (2014) Effects of interannual salinity variability on the barrier layer in the western-central equatorial Pacific: A diagnostic analysis from Argo. *Adv Atmos Sci* 31:532–542. <https://doi.org/10.1007/s00376-013-3061-8>
- Zhi H, Zhang RH, Lin P et al (2019) Effects of salinity variability on recent El Niño events. *Atmosphere* 10(8). <https://doi.org/10.3390/atmos10080475>

- Zika JD, Skliris N, Nurser AJG et al (2015) Maintenance and broadening of the ocean's salinity distribution by the water cycle. *J Clim* 28(24):9550–9560. <https://doi.org/10.1175/JCLI-D-15-0273.1>
- Zika JD, Skliris N, Blaker AT et al (2018) Improved estimates of water cycle change from ocean salinity: The key role of ocean warming. *Environmental Research Letters* 13(7):074036. <https://doi.org/10.1088/1748-9326/aace42>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Erwan Oulhen¹ · Nicolas Kolodziejczyk¹ · Pierre Tandeo^{2,3,4} · Bruno Blanke¹ · Florian Sévellec^{1,3}

✉ Erwan Oulhen
erwan.oulhen@univ-brest.fr

Nicolas Kolodziejczyk
nicolas.kolodziejczyk@univ-brest.fr

Pierre Tandeo
pierre.tandeo@imt-atlantique.fr

Bruno Blanke
bruno.blanke@univ-brest.fr

Florian Sévellec
florian.sevellec@univ-brest.fr

¹ Laboratoire d'Océanographie Physique et Spatiale (LOPS, UMR 6523), University Brest, CNRS, IRD, Ifremer, Brest 29200, France

² Lab-STICC, UMR CNRS 6285, IMT Atlantique, Brest 29238, France

³ Odyssey, Inria/IMT/CNRS, Rennes 35042, France

⁴ Cluster for Pioneering Research, RIKEN, Kobe 650-0047, Japan