# SynObs Flagship OSEs

# Guideline Version 1.02

**Ver 1.00 October 22, 2023**

**Ver 1.01 December 13, 2023**

**Ver 1.02 August 30, 2024**


**Lead Authors:**

**Yosuke Fujii (JMA/MRI, SynObs Co-chair),**

**Magdalena Balmaseda (ECMWF)**

**Elisabeth Rémy (Mercator Ocean International, SynObs chair)**

**(in consultation with Prediction Centers, Diagnostics Groups, and other SynObs members)**

## 1. Purpose and expected outcome

Ocean observation data are essential for ocean predictions. However, developing and sustaining the ocean observation network requires a huge amount of human and financial resources. Therefore, the ocean observation network should be designed to efficiently acquire effective observation data, and their adequacy should be continually evaluated. Especially nowadays, many ocean observing systems require scientific support for their usefulness in order to sustain them.

Observing System Experiments (OSEs) are widely conducted in order to evaluate the importance of the ocean observation data and effectiveness of ocean observing systems for ocean predictions. Observing System Simulation Experiments (OSSEs) are also used for evaluating the future observing system or a new design of an observation network. However, the results of OSEs and OSSEs severely depend on the property of the system, including systematic errors of the system (i.e., model bias), physical parameterization and data assimilation scheme applied in the system, as demonstrated in previous studies (e.g., Fujii et al. 2014; 2019). Therefore, it is preferable to conduct the OSEs/OSSEs using multiple prediction systems in order to mitigate the differences in evaluations caused by the system dependency and to reach robust and reliable conclusions.

The UN Decade of Ocean Science for Sustainable Development provides us with a good opportunity to make a close collaboration between ocean observing and ocean prediction communities, as well as among various groups in the ocean prediction community, for evaluating the importance of ocean observation systems and co-designing the future evolution of the ocean observation network. Taking advantage of the collaborative opportunity, SynObs (Synergistic Observing Network for Ocean Prediction) plans “the SynObs flagship OSEs/OSSEs”, a collaborative OSEs/OSSEs using multiple ocean and subseasonal-to-seasonal (S2S) prediction systems. In order to fully exploit the collaborative opportunity, SynObs adopts the strategy that allows as many ocean prediction and observation groups to participate as possible. The purposes and outcomes of this activity are as follows:

- To demonstrate the importance of major ocean observing systems (Argo array, tropical mooring arrays, satellite altimetry, SWOT, etc.) for ocean and coupled predictions. Robust and reliable evaluations can be made through OSEs/OSSEs with multiple ocean prediction systems.
- To evaluate the synergy among different observing systems, typical between satellite and in situ observing systems, in ocean predictions.
- To contribute to future design of the ocean observing network for improving its effectiveness and efficiency.
- To build a framework to co-design the future ocean observing network among ocean observing and prediction communities.
- To establish the best practice to evaluate the current ocean observing network and to design its future evolution.

The SynObs flagship OSEs/OSSEs includes Ocean Prediction (OP) OSEs, OP OSSEs, and Subseasonal-to-Seasonal (S2S) OSEs. In the OP OSEs, relatively high-resolution ocean data assimilation and prediction systems are expected to participate in. The period of data assimilation (analysis) runs for OP OSEs will be at least 1 year during 2020, and preferably extended to 3 years until 2022 (possible further extension in the future), and 10-day ocean forecasts will be performed. The OP OSSEs are also expected to be joined by relatively high-resolution ocean data assimilation and prediction systems. For the OP OSSEs, the GEOS/NASA high resolution coupled atmosphere-ocean simulation of approximately one year will be used as the Nature Run from which virtual observation data will be synthesized, and 10-day ocean forecasts will be performed. The S2S OSEs are mainly expected to be joined by S2S forecasting systems which consists of a coupled atmosphere-ocean model and an ocean data assimilation system, but participation of uncoupled ocean data assimilation system only in the analysis part of the OSEs is also expected. It will adopt a long analysis period (2003-2022, with possible extension to 2023) and S2S forecasts (1-4 month forecasts) with coupled ocean-atmosphere models will be performed. The S2S OSEs is designed such that there is complementarity with the OP OSEs. It should be noted that the setting of S2S OSEs is determined such that their analysis parts can be also used for the OP OSEs. Output data from the analysis and forecast runs performed for the flagship OSEs/OSSEs will be collected and shared as the SynObs common database with volunteer groups responsible for the diagnostics (Diagnostic Groups, see section 4) as well as made publicly available. Table 1 provides a summary of the time span of the analysis and forecasts length/frequency for OSEs (both OP and S2S).

**Table 1.** Guidelines for the time span of OSE analysis experiments, and frequency and lead time of the OSE forecasts, both for OP and S2S.

| OSE Type | Analysis phase | Forecast phase | Priority |
|---|---|---|---|
| **OP OSE**<br>Ocean Prediction | **Name**: OP-AN<br>**Period** :<br>Required: 1st Jan 2020 -31st Dec 2020<br>Optional: 1st Jan 2020- 31st Dec 2022 | **Name**: OP-FC<br>**Forecast length**: 10 days<br>**Initialization**:<br>● From OP-AN<br>● Every 5 days starting on 6th of Jan (6,11,16…) | Left column in Table 2 |
| **S2S OSE**<br>Subseasonal to Seasonal Prediction | **Name**: S2S-AN<br>**Period** :<br>Required: 1st Jan 2003 -31st Dec 2022<br>Optional: 1st Jan 1993- 31st Dec 2023 | **Name**: S2S-FC<br>**Forecast length**:<br>● Extended range: 35 days<br>● Seasonal range: 4 months<br>**Initialization**:<br>● From S2S-AN<br>● Extended range: 1st of each month<br>● Seasonal range: 1st of May and 1st of Nov<br>**Ensemble size**:<br>at least 10 members per initial date | Right column in Table 2 |

This document describes the outline of OP OSEs and S2S OSEs. The detailed plan for OP OSSEs is currently under consideration, and will appear in a separate document. Section 2 describes the setting of observation data in OSEs (data denial experiments). Section 3 provides summaries of the specific features of the OP OSEs and the S2S OSEs, respectively. Section 4 describes how to analyze the OSE results. Section 5 describes the schedule and expected publication. Section 6 explains about the common database.

## 2. Proposed observation denial settings in OSEs

Table 2 lists the different data denial experiments or OSEs, following the priority order of the experiments for the OP group. The rightmost column indicates the recommendations for the S2S OSEs. However, it is left up to each prediction center to decide the order of implementing the OSEs on their own according to their own priority (or choose the experiments they are interested in). Each OSE in Table 2 contains a data assimilation (or analysis) experiment, followed by the corresponding forecast set. In each OSE, a specific data type is withheld from the assimilation. The impact of that data type will be assessed by comparing the specific OSE experiment with the CNTL experiment, which contains all data types. It is recommended that the World Ocean Database (WOD) be used to collect the in-situ temperature and salinity profiles which will be assimilated because Argo, mooring and other profiles are already sorted in different files and, thus, it is easy to identify Argo and mooring profiles. However, it is approved to use other datasets if necessary. In any case, the source of observation data and the method of sorting Argo, mooring, and other in-situ profiles should be clarified. The remainder of this section describes the experiments in more detail.

**Table 2.** Summary of observation data setting for the OP and S2S OSEs. The number of the left column indicates the order of priority for the OP-OSE protocol, while the right column lists the guidance for the S2S-OSE protocol.

| Configuration of OSEs | | | | | | | | S2S Priority |
|---|---|---|---|---|---|---|---|---|
| **1** | CNTL | | SST | Argo 80% | Mooring | Other TS | Alt. (optional) | Required |
| **2** | NoAlt | | SST | Argo 80% | Mooring | Other TS | | Optional |
| **3** | NoArgo | | SST | | Mooring | Other TS | Alt. (optional) | Recommended |
| **4** | NoMoor | | SST | Argo 80% | | Other TS | Alt. (optional) | Recommended |
| **5** | NoSST | | | Argo 80% | Mooring | Other TS | Alt. (optional) | Recommended |
| **6** | NoInsitu | | SST | | | | Alt. (optional) | Required |
| **7** | SSTonly | | SST | | | | | Required |
| **8** | Free | | | | | | | Optional |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **9** | HalfArgo | | SST | Argo 40% | Mooring | Other TS | Alt. (optional) | Optional |
| **10** | Oper | Oper. Setting | SST | Argo 100% | Mooring | Other TS | Nadir Altimeter | Optional |

(1) CNTL

CNTL is the reference run in the set of OSEs. In CNTL, SST data, 80% of Argo data, Mooring data, and other in-situ temperature and salinity profile data are assimilated. Data of nadir altimeters can be also assimilated by the choice of each operational center. Other settings of the model and assimilation system can be different from the operational settings. 20% of the Argo data, which are expected to distribute homogeneously in the global ocean as shown in Fig. 1, are withheld from assimilation in order to use as independent data for validation. Which data are withheld is determined by the WMO number of the floats, that is, the data for floats whose last digit of the WMO number is 8 or 9 are withheld. Identification of the reference Argo data is based on the Argo GDAC snapshot file on Mar. 2023 ([http://doi.org/10.17882/42182#100487](http://doi.org/10.17882/42182#100487)).

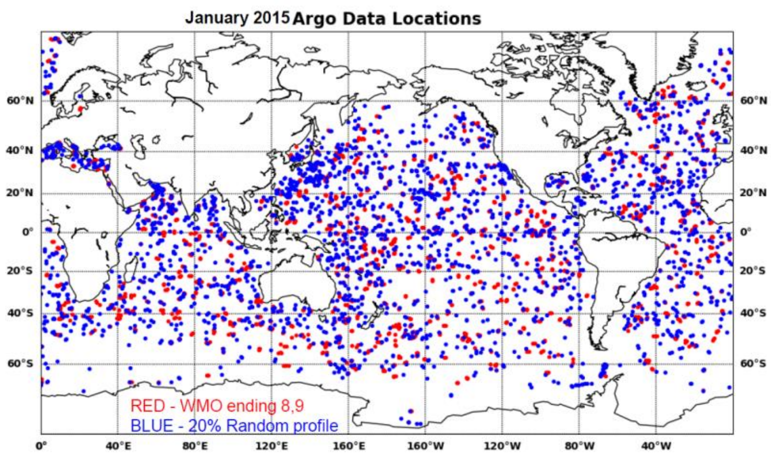


**Fig. 1.** Distributions of the Argo floats whose last digits of WMO number is 8 or 9 (red) and 20% random profiles (blue). Example for January 2015. Temperature and salinity profiles observed by the Argo floats whose last digits of WMO number is 8 or 9 will be withheld in all OSEs and used for the validation as the reference data (Produced by Li Ren, NASA/GMAO.)

(2) NoAlt (or AddAlt)

NoAlt is implemented for evaluating the impact of Nadir Altimeter data. NoAlt is performed with the same setting as CNTL but the nadir altimeter data are withheld. If the nadir altimeter data are not assimilated in CTL, NoAlt is not necessary. In that case, the OP center can perform “AddAlt” in which the nadir altimeter data are added to the observation data being assimilated in CNTL instead.

(3) NoArgo

NoArgo is implemented for evaluating the impact of Argo data. NoArgo is performed with the same setting as CNTL but all Argo data are withheld.

(4) NoMoor

NoMoor is implemented for evaluating the impact of Mooring data (mainly those in the tropical ocean). NoMoor is performed with the same setting as CNTL but all data observed by moorings are withheld.

(5) NoSST

NoSST is implemented for evaluating the impact of SST data observed by satellites. NoSST is performed with the same setting as CNTL but all SST data are withheld.

(6) NoInsitu

NoInsitu is implemented for evaluating the impact of all in-situ temperature and salinity profile data. NoInsitu is performed with the same setting as CNTL but all in-situ temperature and salinity profile data, including Argo and Mooring data, are withheld.

(7) SSTonly

SSTonly is implemented for checking the performance of the system when only SST data are performed. SSTonly is performed with the same setting as CNTL but only SST data observed by satellites are assimilated. If only SST and in-situ TS profiles are assimilated in CNTL, this experiment is the same as NoInsitu, and can be omitted.

(8) Free

Free is implemented for checking the model performance without assimilating data. Free is performed with the same setting of the ocean model as CNTL and without any procedures related to data assimilation (including climatological nudging, bias correction, etc.)

(9) HalfArgo

HalfArgo is implemented for evaluating the impact of changing the density of Argo floats. HalfArgo is performed with the same setting as CNTL but the number of Argo data being assimilated is reduced from 80% to 40 %. The data for floats whose last digit of the WMO number is 4, 5, 6, 7, 8 or 9 are withheld.

(10) Oper

Oper is implemented for checking the performance of the operational system with the actual operational setting. In Oper, the data assimilation run is performed with the completely same setting as in the operation. Or prediction centers can just provide their operational results.

## 3. Outline of OSEs

### 3.1 Ocean Prediction (OP) OSEs

For Ocean Prediction (OP) OSEs, prediction centers (including operational centers and research Institute) are expected to perform the OSEs using their own ocean data assimilation and prediction systems in accordance with the following instruction, and provide the OSE output data to the SynObs common database according to the requested format (See Section 6), as well as to the groups in charge of diagnostics (diagnostic groups, see section 4) according to the requests of the groups.

#### 3.1.1 Setting for OP-AN runs

Prediction centers are requested to provide output of the assimilation run for each OSE for the 3-year period in 01 Jan. 2020 - 31 Dec. 2022, with particular priority given to the period in 01 Jan. 2020 - 31 Dec. 2020. All assimilation runs with the OSE settings must start at least three months before the initial of the period in which the data will be provided (longer spin-up period is acceptable) from the regular data assimilation or free simulation result at that day. What initial conditions are used needs to be clarified. And, it is also preferable to start the OSEs before 2000 in order to provide data for 2003-2022 (or 2003-2023) requested for S2S OSEs.

Data producers are requested to provide the data described in section 6.1 (OP-AN) from the assimilation runs for OP OSEs. But it is recommended to output data as close as possible to the system's native grid data, or as high resolution as possible, and store them at each center. In particular, it is preferable to store the data in the system's native vertical levels in order to calculate vertically integrated values accurately. In addition, prediction centers are also expected to provide data specifically requested by diagnostic groups. And if the data assimilation runs are started before Jan. 2000, it is also requested to provide the data described in section 6.3 (S2S-AN).

#### 3.1.2 Setting for OP-FC runs

The forecast runs of the OP OSEs (OP-FC) are requested to be performed from the end of every pentad (00 UTC of Jan. 6th, 11th, 16th, 21th, …) in the assimilation period (here 29 Feb. in leap years are ignored, i.e., the initial time next to 00 UTC on 25 Feb. is always 00 UTC on 2 Mar), with particular priority given to the period in 5 Jun. 2021 - 31 May 2022. The forecast length should be 10 days (even if Feb. 29th is included in the period). Ideally, ocean models should be forced by results of atmospheric forecasts in ocean forecasts. However, atmospheric forcing calculated from atmospheric reanalysis data can be used in OP OSEs if the atmospheric forecast data are not available. If an OP center uses a coupled atmosphere-ocean prediction system, the center can use the oceanic part of the forecast results for OP OSEs.

Prediction centers are requested to provide the data described in section 5.2 (OP-FC). But it is recommended to output data as close as possible to the system's native grid data, or as high resolution as possible, and store them at each center. In particular, it is preferable to store the data in the system's native vertical levels in order to calculate vertically integrated values. In addition, prediction centers are also expected to provide data specifically requested by diagnostic groups.

### 3.2 Subseasonal-to-seasonal (S2S) OSEs

The ocean is now an active component of weather and subseasonal-to-seasonal (S2S) prediction systems. Therefore, the ocean observations have the potential to influence the ocean and atmospheric predictions at different time scales, from days to seasons ahead. Therefore, S2S OSEs attempt to quantify how the ocean observation network affects the predictions of ocean and atmospheric variables from weeks to season ahead.

As for all the OSEs in this project, the S2S OSEs will contain 2 components: the analysis OSEs for the preparation of initial conditions (S2S-AN), sometimes also referred as ocean reanalysis OSES or ORA OSES); and the forecast part (S2S-FC), sometimes also referred as reforecast OSEs. The S2S-FC will be used to assess the impact of the observation data on the forecasts of ocean and atmosphere variables.

#### 3.2.1 General Setting

It is recommended that the analysis and forecasts OSEs are conducted with systems as close as possible to the operational systems, but it does not need to be the same. The minimum subset of S2S OSEs comprises the CNTL, the NoInsitu and SSTonly experiments in Table 2. Note that in case the CNTL does not contain altimeter assimilation, NoInsitu and SSTonly are identical. In addition to this minimal set, experiments NoSST, NoArgo and NoMoorings are recommended, while NoAlti and Free are optional.

#### 3.2.2 Setting for S2S-AN

The period for S2S-AN is 2003-2022, with possible extension to 2023, which is equivalent to the Argo period. To avoid spin up problems that can contaminate the assessment, it is advised to start this reanalysis from 2000 or before. Reanalysis going back to 1993 are optional. The initial conditions and spin up procedure for these S2S-AN reanalyses should be documented.

Prediction centers are requested to provide the data described in section 6.3 in the netCDF format for S2S-AN. But it is recommended to output data as close as possible to the system's native grid data, or as high resolution as possible, and store it at each center. In particular, it is preferable to store the data in the system's native vertical levels in order to calculate vertically integrated values accurately. Prediction centers are also expected to provide data in the period of 2020 (possibly extended to 2022) described in Section 6.1 (OP-RA), and data specifically requested by diagnostic groups.

#### 3.2.3 Setting for S2S-FC

The S2S-AN will provide initial conditions for the S2S-FC experiments, which will consist of reforecasts sets, with initial conditions spanning the S2S-AN period. The forecasts for each initial condition will comprise at least 10 ensemble members. The experiments will target two forecast ranges:

- Extended range: 35 days, initialized on the 1st day of every month for the S2S-AN period (2003-2022, with possible extension to 2023)
- Seasonal range: 4 months (126 days), initialized on the 1st of May and November for the S2S-AN period (2003-2022, with possible extension to 2023)

The output variables for S2S-FC should contain part of the variables defined in the S2S database and some additional 2D variables. The data in the S2S database is currently 1.5x1.5 deg. grid for forecasts, but in this project, it is requested to output in 1x1 deg. grids, i.e., the same used by other OSEs. Data for all ensemble members should be provided. The data requested to be provided for the common database are described in Section 6.4 (all data, including atmospheric data) should be provided in the netCDF format.

## 4. Diagnostics

SynObs will ask several volunteer diagnostic groups to examine the impacts of observation data on several targeted variables or diagnostics (typically, the difference in the variables/diagnostics among OSEs) in the global ocean using the flagship OSE output data (including both OP and S2S OSEs). SynObs will also have some regional diagnostic groups who focus on the data impacts in specific targeted regions. The diagnostic groups will use the data on the common database or request the data required for their diagnostics to the prediction centers and conduct the diagnostics by their own idea. The results will be shared among SynObs members and possibly be submitted to the 2nd SynObs special issue as well as be presented in relevant academic meetings. We will collaborate for the diagnostics with several international groups, such as OceanPredict task teams, UN Ocean Decade "Ocean Observing Co-Design" exemplar groups, Tropical Pacific Observing System (TPOS) Science Advisory Committee, Argo Science team, WMO S2S group, etc.

Investigation of the following diagnostics at global scale is currently expected to be conducted by diagnostics groups.

- Impact of Argo on Heat budget and surface flux imbalances
- Representation of mesoscale Eddies in the SSH fields
- Impact of Argo on Sound Speed Profiles (Shallow water ducts/local minimums, depth of deep sound channel/absolute sound speed minimum).
- TC-related quantities (0-50mT, Z20, Z26, TCHP, MLD)
- Innovation statistics, class 4 assessment
- Lagrangian drift comparisons to drifters
- Ocean heat budget analysis and marine heatwaves
- Verification and intercomparison of near-surface ocean currents

In addition, following regional analyses are expected.

- Upper ocean impact on tropical climate variability (MJO, ENSO) forecasts. Tropical Pacific upper ocean analysis and forecasts.
- Vertical mode decomposition of equatorial Pacific variability (estimate of Kelvin/Rossby wave contribution to SSH)
- Predictive skill (S2S) for SSH along the coast of Peru/Chile
- Predictive skill (seasonal) for ENSO indices and other eastern Pacific indices.
- Calculation of the Brazil Current volume transport in few latitudes
- Evaluation of SST, SSH and T/S structure in the Brazil Current Regin
- Analysis around the Greater Agulhas Current region
- Current fields (e.g., the Kuroshio Path) in the western North Pacific.
- Marine Heatwaves monitoring and forecast skills in the western North Pacific.
- Diagnostics in the western North Atlantic
- Diagnostics in the Indian Ocean
- Diagnostics in polar/subpolar regions

## 5. Schedule and publication

Ideally, analysis results of OP and S2S OSEs from each prediction center to the SynObs common database are expected to be provided until the end of Dec. 2023, and forecast results until the end of Apr. 2024. Diagnostic results by each volunteer diagnostic group will be conducted after 2024. It is planned to submit a paper which introduces the flagship OSEs (both OP and S2S OSEs) to an academic journal. Early results of the flagship OSEs are planned to be introduced in these papers, as well as several academic meetings planned to be held in 2024, including the WMO observation impact workshop (27-30 May 2024 in Norrköping, Sweden) and OceanPredict Symposium 2024 (18-22 Nov 2024 in Paris, France). The final diagnostic results will be published in the 2nd SynObs special issue of a relevant academic journal (which journal will be used has not been decided yet). The special issue is planned to be published after the SynObs project ends (around 2027).

## 6. Common database for the OP and S2S OSEs

SynObs will build a database from the output of the flagship OSEs/OSSEs (including OP OSEs, S2S OSEs, and OP OSSEs) on the server prepared by JAMSTEC Application Laboratory (APL). Part of the database might be placed separately on servers prepared by other SynObs supporters. In the database, the data will be stored in netCDF files. The detailed configuration of the data archived in the common database is described in this section.

The data files will be created by prediction centers with their output data. The python code which creates sample data files are provided from https://github.com/shokido/SynOBS/tree/main/SCRIPTS. The database can be accessed at least by diagnostic groups, and will probably be full-open. Diagnostic groups can download the data from the common database if the property of the data (resolution, frequency, etc.) is suitable for their purpose. In order to manage the size of the JAMSTEC common archive, the number of variables, frequency of output and resolution has been limited. Scientists interested in specific diagnostics requiring variables that are not in the common archive are encouraged to obtain it from the individual prediction centers. The prediction centers do not necessarily provide the requested data, but are expected to provide as much as possible to the diagnostic groups.

There are three types of output: 3D-fields, 2D-fields and 1D fields at Point Locations (PL). PL are defined as observation locations by specific Argo floats and moorings. The horizontal grids, vertical discretization, and point locations are given in Table 3 below.

**Table 3.** Specification of horizontal and vertical grids and point locations for output

| Grid specification | | Comments |
|---|---|---|
| Vertical discretization ocean | **1m, 5m, 10m, 15m, 20m, 25m, 30m, 35m, 40m, 45m, 50m, 60m, 70m, 80m, 90m, 100m, 120m, 140m, 160m, 180m, 200m, 220m, 240m, 270m, 300m, 330m, 360m, 400m, 450m, 500m, 550m, 600m, 700m, 800m, 900m, 1000m,**1200m, 1350m, 1500m, 1750m, 2000m, *2500m, 3000m, 3500m, 4000m, 4500m, 5000m, 5500m* | Mooring PL (first 36 levels): upper 1000m (**in bold**)<br>Argo floats: upper 2000m (first 42 levels)<br>Model Output: all depths (49 levels) |
| Vertical discretization atmosphere | Pressure levels (hPa) at 1000,925,850,700,500,300,200,100,50,10 | As specified in the S2S database.<br>Only for coupled S2S-FC |
| Model Horizonal grids (regular latitude/longitude starting at 0E, 90S) | Low: 1° x 1° | S2S-OSEs |
| | Medium: 0.25° x 0. 25° | OP-OSEs |
| | High: 0.1° x 0. 1° | Optional for OP-OSEs |
| Point locations | **Moorings:**<br>0°N-110°W, 0°N-140°W, 0°N-170°W, 0°N-165°E, 0°N-156°E, 1°S-140°W, 1°N-140°W, 30.32°S-71.76°W[(*1)], 13°N-137°E [(*2)], 2°N-140°W<br><br>**Argo:**<br>● Those observed by the reference Argo[(*3)] floats (whose last digit of WMO number is "8" or "9") in the analysis or forecast period.<br>● (Optional) Those observed by the Argo[(*3)] floats other than the reference Argo floats in the analysis or forecast period. | [*1] COSMOS Site near the Chile Coast. Observing TSUV above 800m depth<br>[*2] TRITON buoy site at the northern boundary of the western tropical Pacific warm water pool. Observing SWHF, LWHF, and TS above 300m and UV above 120m.<br>[*3] Identification of the positions and the times of Argo observations is based on the Argo GDAC snapshot on Mar, 2023 (http://doi.org/10.17882/42182#100487). |

The output requirements for 3D, 2D and PL, as well as its temporal frequency and horizontal grid depends on the type of OSEs. Table 4 lists the groups of the ocean and atmospheric variables that will be used to specify the output requirements of the different OSEs, summarized in Table 5.

**Table 4. Summary of output variable. Those in bold are common to the S2S database**

| Group | Included Variables |
|---|---|
| OP-1 | 3D-TSUV, SSH, SIC, SIT, SWHF, LWHF, LAHF, SNHF, NetHF, NetWF, Taux, Tauy<br>Optional: TotalHF, TotalWF, Analysis Increment (3D-TSUV, SSH) |
| OP-2 | SST, SSS, SSU, SSV, SSH, 0-50mT, Z20, Z26, TCHP, MLD005, 15mU, 15mV |
| S2S-1 | Ocean:<br>3D-TSUV**, SSH, 0-300mT, Z20, MLD001, 0-300mS, SIC, SIT,** 0-50mT, Z17, Z26, Z28, TCHP,MLD005, ILD05, SWHF, NetHF<br>Atmosphere:<br>**3D-TZUVQ, T2m, U10m, V10m, Precip, LWHF, SWHF, LAHF, SNHF,** Taux, Tauy, **MSLP, Total Cloud Cover, OLR** |
| S2S-2 | Ocean:<br>**SST ,SSS, SSU, SSV, SSH, 0-300mT, Z20, MLD001, 0-300mS, SIC, SIT,** 0-50mT, Z17, Z26, Z28, TCHP, MLD005, ILD05, SWHF, NetHF<br>Atmosphere:<br>**3D-TZUVQ, T2m, U10m, V10m, Precip, LWHF, SWHF, LAHF, SNHF,** Taux, Tauy, **MSLP, Total Cloud Cover, OLR** |
| S2S-3 | Ocean:<br>**SST, SSH, SIC, MLD001**, ILD05<br>Atmosphere:<br>**OLR, U200, U850** |
| Point Location | **Argo:** TS<br>**Mooring:** TSUV, SWHF, NetHF |

**Table 5: Summary of dataset on the common database**

| | Daily | Pentad/Weekly | Monthly | Point Location |
|---|---|---|---|---|
| OP-OSE Analysis (OP-AN) | **OPA-D (OPA-DH)**<br>Variables: OP-2<br>Resolution: 0.25°, 0.1°<br>Frequency: Daily | **OPA-P**<br>Variables: OP-1<br>Resolution: 0.25°<br>Frequency: Pentad | | **OPA-PL**<br>Argo (Daily)<br>Mooring (Hourly) |
| OP-OSE Forecast (OP-FC) | **OPF-D (OPF-DH)**<br>Variables: OP-2<br>Resolution: 0.25°, 0.1°<br>Lead Times: D1, D3, D7 | **OPF-P**<br>Variables: OP-1<br>Resolutions: 0.25°<br>Lead Times: P1, P2 | | **OPF-PL**<br>Argo (Daily, D1-D10)<br>Mooring<br>(Hourly, H1-H240) |
| S2S-OSE Analysis (S2S-AN) | **S2SA-D**<br>Variables: S2S-2 Oc<br>Resolution: 1°<br>Frequency: Daily | | **S2SA-M**<br>Variables: S2S-1 Oc+Flx<br>Resolution: 1°<br>Frequency: Monthly | **S2SA-PL**<br>Argo (Daily)<br>Mooring (Daily) |
| S2S-OSE Forecast (S2S-FC) | **S2SF-D**<br>Variables: S2S-3 Oc+Atm<br>Resolution: 1°<br>Lead Times: D1-D35 | **S2SF-W**<br>Variables: S2S-2 Oc+Atm<br>Resolution: 1°<br>Lead Times: W1-W18 | **S2SF-M**<br>Variables: S2S-1 Oc+Atm<br>Res: 1°<br>Lead Times: M1-M4 | **S2SF-PL**<br>Argo (Daily, D1-D126)<br>Mooring<br>(Daily, D1-D126) |

The output variables in Table 4 are classified into different groups according to their frequency and type of OSE experiments. *Groups OP-1 and S2S-1* contain 3D and selected 2D oceanic variables. *Group S2S-1* also includes atmospheric variables. They are suitable for low frequency output (pentad for OP-OSEs, monthly for S2S-OSEs). *Groups OP-2 and S2S-2* contain only 2D oceanic variables of the OP-1 and S2S-1 groups plus T, S, U, V at specific levels, selected isotherm positions, average temperature at different ranges, and variables related with mixed layer processes. *Group S2S-2* also includes the same atmospheric variables as *Group S2S-1*. It should be noted that the majority of variables included in *Groups S2S-1 and S2S-2* are those in the S2S database. *Group S2S-3* contains a very limited set of variables that will be output at a daily frequency in S2S-FC.

The variables for the different groups are listed below:

*Group OP-1*:
- Oceanic Prognostic Variables: **T, S, U, V, SSH**
- Sea Ice Prognostic Variables: **SIC, SIT**
- Surface Fluxes: **SWHF**, LWHF, LAHF, SNHF, **NetHF**, NetWF, Taux, Tauy
- Surface Fluxes (Optional): TotalHF, TotalWF
- Analysis Increments (Optional, only AN types): Tinc, Sinc, Uinc, Vinc, SSHinc

*Group S2S-1:*
- Ocean variables: those written in bold characters in OP-1 plus
  - Vertically integrated variables: 0-300mT, 0-300mS, 0-50mT, TCHP
  - Mixed layer related variables: MLD001, MLD005, ILD05.
  - Isotherm variables: Z20, Z17, Z26, Z28
- Atmosphere variables (only for coupled experiments):
  - 3D-TZUVQ
  - T2m, U10m, V10m, Precip, LWHF, SWHF, LAHF, SNHF, Taux, Tauy, MSLP, Total Cloud Cover, OLR

*Group OP-2:*
- Surface ocean Variables: SST, SSS, SSU, SSV, SSH
- Variables related to temperature fields: 0-50mT, Z20, Z26, TCHP
- Mixed Layer Depth: MLD005
- Near-surface velocity: 15mU, 15mV

*Group S2S-2 :*
- Ocean variables (2D variables in *S2S-1* + surface ocean variables):
  - Surface ocean/ice variables: SST, SSS, SSU, SSV, SSH, SIC, SIT
  - Vertically integrated variables: 0-300mT, 0-300mS, 0-50mT, TCHP
  - Mixed layer related variables: MLD001, MLD005, ILD05.
  - Isotherm variables: Z20, Z17, Z26, Z28
  - Surface Fluxes: SWHF, NetHF
- Atmosphere variables (only for coupled experiments): The same as in *S2S-1*

*Group S2S-3*:

- Ocean: SST, SSH, SIC, MLD001, ILD05, NetHF
- Atmosphere:  OLR, U200, U850

The configuration of data from data assimilation (analysis) and forecast runs in OP OSEs are described in Section 6.1 and 6.2, and data assimilation and forecast runs (Reanalysis OSEs and reforecast OSEs) in S2S OSEs are described in Section 6.3 and 6.4. Acronyms and detailed explanation of output variables are described in Section 6.5.

### 6.1 Data for data assimilation (analysis) runs in OP OSEs (OP-AN)

#### 6.1.1 Pentad data for OP-AN (OPA-P)

Variables: *Group OP-1*

Horizontal Grids: 0.25˚ grids started from 0˚E, 90˚S. (Medium in Table 3)

Vertical Levels: All 49 levels in Table 3

Time interval: Average (or integrated) values for each pentad (5-day periods of 1-5 Jan., 6-10 Jan., 11-15 Jan., ... 27-31 Dec. The pentad period becomes 6 days when Feb. 29th is included, that is the pentad period is always from Feb. 25th to Mar. 1st.). Daily output of the surface flux variables (SWHF, LWHF, LAHF, SNHF, NetHF, NetWF, Taux, Tauy, TotalHF, TotalWF) is optional.

Data Period: 1-5 Jan 2020 - 27-31 Dec 2022 (to 27-31 Dec 2020 at least)

netCDF File Configuration: Create a netCDF file per variable and per pentad

#### 6.1.2 Daily Data for OP-AN (OPA-D)

Variables: *Group OP-2*

Horizontal Grids: 0.25˚ grids started from 0˚E, 90˚S. (Medium in Table 3)

Time interval: Averaged for each day

Data Period: 1 Jan 2020 - 31 Dec 2022 (to 31 Dec 2020 at least)

netCDF File Configuration: Create a netCDF file per variable and per day

#### 6.1.3 High-Resolution Daily Data (OPA-DH) (Optional)

Same as OPA-D but horizontal grids are 0.1˚ grids starting from 0˚E, 90˚S (High in Table 3).

#### 6.1.4 Point Location Data (OPA-PL)

##### 6.1.4.1 Argo floats

Variables: T, S (Vertical profiles 1-2000m at specific locations and times)

Horizontal Locations and times:

- Those observed by the reference Argo floats (whose last digit WMO number is "8" or "9")
- (Optional) Those observed by the Argo floats other than the reference Argo floats

Vertical Levels: Upper 2000m (first 42 levels in Table 3)

Data Period: 1 Jan 2020 - 31 Dec 2022 (to 31 Dec 2020 at least)

netCDF File Configuration: Create one netCDF file per month for each reference and assimilated Argo observations (All observations by reference or Assimilated Argo floats in the targeted month are collected in one file.)

Caution: It is recommended that the data are generated offline from the daily averaged output.

#### 6.1.4.2 Mooring Sites

Variables: T, S, U, V (Time series of vertical profiles of 1-1000m at specific locations)

SWHF, NetHF (Time Series at specific locations)

Horizontal Locations: As in Table 3.

Vertical Levels: Upper 1000m, first 36 levels in Table 3

Time Interval: Hourly (1st data: 00:00-01:00 UTC averages, 2nd data: 01:00-02:00 UTC averages, 3rd data: 02:00-03:00 UTC averages, ...). If it is impossible, it can be daily.

Data Period: 00UTC 1 Jan 2020 - 23UTC 31 Dec 2022 (to 23UTC 31 Dec 2020 at least)

netCDF File Configuration: Create a netCDF file per month. Data for the 10 locations are collected in one file.

Caution: It is recommended that the data are written out at the run time.

### 6.2 Data for forecast runs in OP OSEs (OP-FC)

#### 6.2.1 Pentad Data (OPF-P)

Variables: *Group OP-1* (except for analysis increments)

Horizontal Grids: 0.25˚ grids started from 0˚E, 90˚S. (Medium in Table 3)

Vertical Levels: All 49 levels in Table 3.

Lead Time: Averages of Pentad 1 （the pentad started from the initial time of the forecast) and Pentad 2 (The pentad period is always 5 days even if Feb. 29th is included.) Daily output of the surface flux variables (SWHF, LWHF, LAHF, SNHF, NetHF, NetWF, Taux, Tauy, TotalHF, TotalWF) from Day 1 to Day 10 is optional.

netCDF File Configuration: Create a netCDF file per variable, per valid pentad and per initial date.

#### 6.2.2 Daily Data (OPF-D)

Variables: *Group OP-2*

Horizontal Grids: 0.25˚ grids started from 0˚E, 90˚S. (Medium in Table 3)

Lead time: Averaged for Day 1, Day 3, Day 7.

netCDF File Configuration: Create a netCDF file per variable, per valid day and per initial date.

#### 6.2.3 High-resolution daily data (OPF-DH) (Optional)

Same as OPF-D but horizontal grids are 0.1˚ grids starting from 0˚E, 90˚S (High in Table 3).

#### 6.2.4 Point Location Data (OPF-PL)

##### 6.2.4.1 Argo floats

Variables: T, S (Vertical profiles 1-2000m at specific locations and times)

Horizontal Locations and times:

- Those observed by the reference Argo floats (whose last digit WMO number is "8" or "9")
- (Optional) Those observed by the Argo floats other than the reference Argo floats

Vertical Levels: Upper 2000m, first 42 levels in Table 3.

Data Period: The whole forecast period (Day 1 to Day 10) for each forecast (each initial time)

netCDF File Configuration: Create one netCDF file per an initial time for each reference and assimilated Argo observations (All observations by reference or Assimilated Argo floats in the forecast period are collected in one file.)

Caution: It is recommended that the data are generated offline from the daily averaged output.

##### 6.2.4.2 Mooring Sites

Variables: T, S, U, V (Time series of vertical profiles of 1-1000m at specific locations)

SWHF, NetHF (Time series at specific locations)

Horizontal Locations: As in Table 3.

Vertical Levels: Upper 1000m, first 36 levels in Table 3

Time Interval: Hourly (1st data: 00:00-01:00 UTC averages, 2nd data: 01:00-02:00 UTC averages, 3rd data: 02:00-03:00 UTC averages, ...). If it is impossible, it can be daily.

Data Period: The whole forecast period (Day 1 to Day 10) for each forecast.

netCDF File Configuration: Create a netCDF file per initial date. Data for the 10 locations are collected in one file.

### 6.3 Data for data assimilation (analysis) runs in S2S OSEs (S2S-AN)

#### 6.3.1 Monthly data (S2SA-M)

Variables: *Group S2S-1* (except for Atmospheric variables)

Horizontal Grids: 1˚ grids started from 0˚E, 90˚S. (Low in Table 3)

Vertical Levels (for 3D variables): All 40 levels in Table 3.

Time interval: Averaged for month

Data Period: 1 Jan 2003 - 31 Dec 2022 (possibly extended to 31 Dec 2023)

netCDF File Configuration: Create a netCDF file per variable and per month .

#### 6.3.2 Daily data (S2SA-D)

Variables: *Group S2S-2* (except for Atmospheric variables)

Horizontal Grids: 1˚ grids started from 0˚E, 90˚S. (Low in Table 3)

Time interval: Averages daily files

Data Period: 1 Jan 2003 - 31 Dec 2022 (possibly extended to 31 Dec 2023)

netCDF File Configuration: Create a netCDF file per variable and per day.

#### 6.3.3 Point Location data (S2SA-PL)

##### 6.3.3.1 Argo Floats

Variables: T, S (Vertical Profiles 1-2000m at specific locations and times)

Horizontal Locations and times:

- Those observed by the reference Argo floats (whose last digit WMO number is "8" or "9")
- (Optional) Those observed by the Argo floats other than the reference Argo floats

Vertical Levels: Upper 2000m, first 42 levels in Table 3.

Data Period: 1 Jan 2003 - 31 Dec 2022 (possibly extended to 31 Dec 2023)

netCDF File Configuration: Create one netCDF file per month for each reference and assimilated Argo observations (All observations by reference or Assimilated Argo floats in the targeted month are collected in one file.)

Caution: It is recommended that the data are generated offline from the daily averaged output.

#### 6.3.3.2 Mooring Sites

Variables: T, S, U, V (Time series of vertical profiles of 1-1000m at specific locations)

SWHF, NetHF (Time series at specific locations):

Horizontal Locations: As in Table 3.

Vertical Levels: Upper 1000m, first 36 levels in Table 3.

Time Interval: Daily

Data Period: 1 Jan 2003 - 31 Dec 2022 (possibly extended to 31 Dec 2023)

netCDF File Configuration: Create a netCDF file per month. Data for the 10 locations are collected in one file.

## 6.4 Data for forecast runs in S2S OSEs (S2S-FC)

### 6.4.1 Monthly data (S2SF-M)

Variables: *Group S2S-1*

Horizontal Grids: 1˚ grids started from 0˚E, 90˚S. (Low in Table 3)

Vertical Levels (for 3D variables): All 40 levels in Table 3.

Time interval: Averages for each month

Data Period: From Month 1 to Month 4 for each forecast

Ensemble Members: Data for all members

netCDF File Configuration: Create a netCDF file per variable, per month, per ensemble member, and per initial date

#### 6.4.2 Weekly Data (S2SF-W)

Variables: *Group S2S-2*

Horizontal Grids: 1˚ grids started from 0˚E, 90˚S. (Low in Table 3)

Time interval: Averaged for each week

Data Period: From Week 1 to Week 18 for each forecast

Ensemble Members: Data for all members

netCDF File Configuration: Create a netCDF file per variable, per week, per ensemble member, and per initial date

#### 6.4.3 Daily Data (S2SF-D)

Variables: *Group S2S-3*

Horizontal Grids: 1˚ grids started from 0˚E, 90˚S. (Low in Table 3)

Time interval: Averaged for each day

Data Period: From Day 1 to Day 35 for each forecast

Ensemble Members: Data for all members

netCDF File Configuration: Create a netCDF file per variable, per day, per ensemble member, and per initial date

#### 6.4.4 Point Location Data (S2SF-PL)

##### 6.4.4.1 Argo floats

Variables: T, S (Vertical Profiles 1-2000m at specific locations and times)

Horizontal Locations and times:

- Those observed by the reference Argo floats (whose last digit WMO number is "8" or "9")
- (Optional) Those observed by the Argo floats other than the reference Argo floats

Vertical Levels: Upper 2000m, first 42 levels in Table 3.

Data Period: The whole forecast period (From Day 1 to Day 126) for each forecast (each initial time and each member)

Ensemble Members: Data for all members

netCDF File Configuration: Create one netCDF file per ensemble member and per initial date for each reference and assimilated Argo observations (All observations by reference or Assimilated Argo floats in the forecast period are collected in one file.)

Caution: It is recommended that the data are generated offline from the daily averaged output.

#### 6.4.4.2 Mooring Sites

Variables: T, S, U, V (Time series of vertical profiles of 1-1000m at specific locations)

SWHF, NetHF (Time Series at specific locations)

Horizontal Locations: As in Table 3.

Vertical Levels: Upper 1000m, first 36 levels in Table 3

Time Interval: Daily

Data Period: The whole forecast period (From Day 1 to Day 126) for each forecast

Ensemble Members: Data for all members

netCDF File Configuration: Create a netCDF file per ensemble member and per initial date. Data for the 10 locations are collected in one file.

## 6.5 Definition of the output variables

### 6.5.1 Oceanic prognostic variables

T: Potential temperature with respect to 0m (3D, Units in $^{\circ}$C)

S: Practical Salinity (3D, Units in PSU)

U: Zonal Velocity (3D, Units in m/s)

V: Meridional Velocity (3D, Units in m/s)

SSH: Raw output of the sea surface height from the model (2D, Units in m). It is not required to apply the correction related to global mean steric height, global fresh water mass change, etc.

Cautions:

- Use the linear interpolation in the vertical direction to get the values at the requested levels.
- If the uppermost level of the ocean model is deeper than 1m, the values at the uppermost level are regarded as the surface variables.

**6.5.2 Sea ice prognostic variables**

SIC: Sea Ice Concentration Ratio (2D, no Units)

SIT: Sea Ice Thickness (2D, Units in m)

**6.5.3 Oceanic surface variables**

SST: Sea surface temperature, which is defined as T (potential temperature with respect to 0m) at 1m (2D, Units in °C). If the uppermost level of the ocean model is deeper than 1m, T at the uppermost level is regarded as SST

SSS: S (practical salinity) at 1m. If the uppermost level of the ocean model is deeper than 1m, S at the uppermost level is regarded as SSS (2D, Units in PSU)

SSU: U (zonal velocity) at 1m (2D, Units in m/s). If the uppermost level of the ocean model is deeper than 1m, U at the uppermost level is regarded as SSU.

SSV: V (meridional velocity) at 1m (2D, Units in m/s). If the uppermost level of the ocean model is deeper than 1m, V at the uppermost level is regarded as SSV.

Cautions:

- Use the linear interpolation in the vertical direction to get the values at 1m, when there is no native vertical level at 1m.
- If the uppermost level of the ocean model is deeper than 1m, the values at the uppermost level are regarded as the surface variables.

**6.5.4 Diagnostic variables related temperature fields**

0-50mT: Vertically averaged T (It should be noted that T is potential temperature with respect to 0m) between 0m and 50m (2D Units in °C)

0-300mT: Vertically averaged T between 0m and 50m (2D. Units in °C)

Z17: Depth of the 17°C isotherm estimated from T.  (2D. Units in m)

Z20: Depth of the 20°C isotherm estimated from T.  (2D. Units in m)

Z26: Depth of the 26°C isotherm estimated from T.  (2D. Units in m)

Z28: Depth of the 28˚C isotherm estimated from T.  (2D. Units in m)

TCHP: Tropical Cyclone Heat Potential (2D, Units in $kJ/cm^2$). Calculated as the oceanic heat content relative to 26˚C above Z26, that is, integrate the following HC from the surface to Z26,

HC = Cp x rho x (T - 26)

where Cp is the constant specific heat of seawater, rho is the density of seawater, and T is the potential temperature with respect to 0 m. Use $3.985\times10^4$J/kg K as the constant specific heat of seawater. Use the average of the potential density with respect to 0 m averaged between 0 m and Z26 (calculated for each profile) as the density of seawater. Potential density should be calculated using the sea water state equation of EOS-80 (see UNESCO 1981; https://unesdoc.unesco.org/ark:/48223/pf0000046148).

Cautions:

- It should be noted that not in situ temperature but potential temperature with respect to 0m should be used for the calculation of the diagnostics related to the temperature fields.
- Assume potential temperature linearly changes in vertical direction between model native levels in the calculation of variables above.

### 6.5.5 Diagnostic variables related to mixed later and isothermal layer

MLD001: Mixed Layer Depth with the 0.01 density criteria (2D. Units in m). Depth at which the potential density with respect to 0m is $0.01kg/m^3$ larger than the potential density calculated from SST and SSS).

MLD005: Mixed Layer Depth with the 0.05 density criteria (2D. Units in m). Depth at which the potential density with respect to 0m is $0.05kg/m^3$ larger than the potential density calculated from SST and SSS).

ILD05: Isothermal Layer Depth with 0.5˚C temperature criteria (2D. Units in m). Depth at which the potential temperature with respect to 0m is 0.5˚C smaller than SST.

Cautions:

- Potential density should be calculated using the sea water state equation of EOS-80 (see UNESCO 1981; https://unesdoc.unesco.org/ark:/48223/pf0000046148).
- Assume potential density linearly changes in vertical direction between model native levels in the calculation of MLD001 and MLD005.
- Assume potential temperature linearly changes in vertical direction between model native levels in the calculation of ILD05.

### 6.5.6 Diagnostic variables related to salinity and horizontal velocities

0-300mS: Vertically averaged S between 0m and 300m (2D. Units in PSU)

15mU: Horizontal Velocity at 15m depth

15mV:  Meridional velocity at 15m depth

Cautions:

- Assume S, U, and V are linearly changed in the vertical direction between model native levels in the calculation of variables above.

### 6.5.7 Surface Heat Fluxes

SWHF: Shortwave (solar) heat flux at the sea surface (at the top of sea ice when sea ice exists) (Units in $W/m^2$). Sum of upward and downward fluxes (solar heat flux absorbed into ocean (seawater or sea ice).

LWHF: Longwave heat flux at the sea surface (at the top of sea ice when sea ice exists) (Units in $W/m^2$). Sum of upward and downward fluxes.

LAHF: Latent heat flux at the sea surface (at the top of sea ice when sea ice exists) (Units in $W/m^2$).

SNHF: Sensible heat flux at the sea surface (at the top of sea ice when sea ice exists) (Units in $W/m^2$).

NetHF: Net heat flux at the sea surface (at the top of sea ice when sea ice exists) (Units in $W/m^2$). Sum of LWHF, SWHF, LAHF, and SNHF.

TotalHF: Total Heat Flux at the top of the seawater (Units in $W/m^2$). It includes flux from additional heat sources other than analysis increments (e.g., heat flux from sea ice to the ocean, heat flux with river inflow, etc.). The value should be averaged in each grid. This variable is for the analysis of the seawater heat budget.

Cautions:

- A positive value means downward flux.
- When sea ice and open ocean regions are mixed in a grid, take the average of the heat flux at the seawater surface in the open ocean areas and flux at the sea ice surface at sea ice areas for SWHF, LWHF, LAHF, and SNHF.

### 6.5.8 Surface FreshWater Flux

NetWF: Net freshwater flux at the sea surface (Units in m/s). Precipitation minus evaporation. When sea ice and open ocean regions are mixed in a grid, take the average of the freshwater flux at the seawater surface in the open ocean areas and flux at the sea ice surface at sea ice areas.

TotalWF: Total freshwater flux at the sea surface (Units in m/s). It includes flux from additional freshwater sources other than analysis increments (e.g., freshwater flux from sea ice to the ocean, freshwater flux with river inflow, etc.). The value should be averaged in each grid.

### 6.5.9 Surface wind stress

Taux: Zonal wind stress at the surface (at the sea ice surface in sea ice areas) (Units in $N/m^2$).

Tauy: Meridional wind stress at the surface (at the sea ice surface in sea ice areas) (Units in $N/m^2$).

### 6.5.10 Analysis Increment

Tinc: Analysis Increment of T integrated in the data valid period (3D. Units in degree C)

Sinc: Analysis Increment of S integrated in the data valid period (3D. Units in PSU)

Uinc: Analysis Increment of U integrated in the data valid period (3D. Units in m/s)

Vinc: Analysis Increment of V integrated in the data valid period (3D. Units in m/s)

SSHinc: Analysis Increment of SSH integrated in the data valid period (3D. Units in m)

## 7. Contact

If you have any questions about this document or the flagship OSEs/OSSEs, please contact us at the official mail address of SynObs (synobs@mri-jma.go.jp).

**Acknowledgements**

We appreciate many fruitful comments and support from Prediction Centers, Diagnostic Groups and other SynObs members during the editing process of this document. UN Decade programs ForeSea and Ocean Observing Co-Design officially support this activity. It is also officially supported by OceanPredict, WMO S2S Program, and Tropical Pacific Observing System (TPOS).