Observation System Simulation Experiments in the Atlantic Ocean for enhanced surface ocean pCO_2 reconstructions.

Anna Denvil-Sommer[1,2], Marion Gehlen[2], Mathieu Vrac[2]

[1]School of Environmental Sciences, University of East Anglia, Norwich, UK
[2]Laboratoire des Sciences du Climat et de l'Environnement (LSCE), Institut Pierre Simon Laplace (IPSL), CNRS/CEA/UVSQ/Univ. Paris-Saclay, Orme des Merisiers, Gif Sur Yvette, 91191, France

Correspondence to: Anna Denvil-Sommer (anna.sommer.lab@gmail.com)

Abstract. To derive an optimal observation system for surface ocean pCO_2 in the Atlantic Ocean and the Atlantic sector of the Southern Ocean eleven Observation System Simulation Experiments (OSSEs) were completed. Each OSSE is a Feed-Forward Neural Network (FFNN) that is based on a different data distribution and provides ocean surface pCO_2 for the period 2008-2010 with a 5 day time interval. Based on the geographical and time positions from three observational platforms, volunteering observing ships (VOS), Argo floats and OceanSITES moorings, pseudo-observations were constructed using the outputs from an online-coupled physical-biogeochemical global ocean model with 0.25º nominal resolution. The aim of this work was to find an optimal spatial distribution of observations to supplement the widely used Surface Ocean CO_2 Atlas (SOCAT) and to improve the accuracy of ocean surface pCO_2 reconstructions. OSSEs showed that the additional data from mooring stations and an improved coverage of the Southern Hemisphere with biogeochemical ARGO floats corresponding to least 25% of the density of active floats (2008-2010) (OSSE 10) would significantly improve the pCO_2 reconstruction and reduce the bias of derived estimates of sea-air CO_2 fluxes by 74% compared to ocean model outputs.

1 Introduction

The ocean is a major sink of anthropogenic CO_2 (Ciais et al., 2013; Friedlingstein et al., 2020). For the period 2010-2019 the ocean uptake was 2.5 ± 0.6 GtC/yr with a strong intensification (from 1.9 to 3.1 GtC/yr) along with increasing of CO_2 emissions (Friedlingstein et al., 2020). The ocean carbon sink estimate is derived from Global Ocean Biogeochemical Models (Hauck et al., 2020) and data-based reconstructions of surface ocean partial pressures of carbon dioxide (pCO_2). The data-based reconstructions rely on the interpolation of surface ocean pCO_2 - derived from measurements of surface ocean CO_2 fugacity - by a variety of methods (e.g. Watson et al., 2020; Gregor et al., 2019; Denvil-Sommer et al., 2019; Bittig et al., 2018; Landschützer et al., 2013, 2016; Rödenbeck et al., 2014, 2015; Fay et al., 2014; Zeng et al., 2014; Nakaoka et al., 2013; Schuster et al., 2013; Takahashi et al., 2002, 2009). These methods provide converging estimates of the global ocean carbon sink and its variability at seasonal and interannual time scales (Rödenbeck et al., 2015; Denvil-Sommer et al., 2019). They are, however, sensitive to the observation coverage in space and time which contributes to inconsistent results over regions with sparse data (Denvil-Sommer et al., 2019; Rödenbeck et al., 2015) and to persistent uncertainties at global scale (Gregor et al., 2019; Hauck et al., 2020).

The majority of observations contributing to the Surface Ocean CO_2 Atlas (SOCAT) (Bakker et al., 2016) are still obtained by underway sampling systems on board of volunteering observing ships. The data density is not homogenous, with Southern latitudes being less well sampled in space and also in time (Monteiro et al., 2010). Sparse data coverage and the lack of observations covering the full seasonal cycle challenge mapping methods and result in noisy reconstructions of surface ocean pCO_2 and disagreements between different models (Denvil-Sommer et al., 2019, Rödenbeck et al., 2015). The ship-based sampling effort is progressively complemented by autonomous observing platforms, such as biogeochemical ARGO floats equipped with pH sensors. The expansion of the observing system to autonomous platforms is of particular relevance in regions that are undersampled either because of the presence of fewer regular shipping lines (e.g., South Atlantic) or because adverse weather conditions prevent a year around sampling (e.g., Southern Ocean). The benefits of combining ship-based measurements of pCO_2 and data from biogeochemical ARGO floats was recently demonstrated for the assessment of Southern Ocean CO_2 fluxes (Bushinsky et al., 2019). Majkut et al. (2014) and Kamenkovich et al. (2017) reported on observing system simulations with autonomous biogeochemical profiling floats in the Southern Ocean that improve estimates of carbon dioxide uptake and biogeochemical variables. While Majkut et al. (2014) used a coarse-resolution model and fixed floats, Kamenkovich et al. (2017) extended this work to a more realistic case with moving floats and high-resolution numerical simulations. Based on a coupled climate carbon model and observations, Lenton et al. (2009) proposed sampling strategies to obtain large-scale integrated CO_2 fluxes in the North Pacific and North Atlantic. They show that regular sampling of ocean surface pCO_2 with a 3-month time step and every 6° in latitude and 10° in longitude is sufficient to capture more than 80% of total CO_2 flux variability.

This study extended the scope to the Atlantic basin, including the Atlantic sector of the Southern Ocean. It explored design options for a future augmented Atlantic scale observing system which would optimally combine data streams from various platforms and contribute to reduce the bias in reconstructed surface ocean pCO_2 fields and sea-air CO_2 fluxes. A series of Observation System Simulation Experiments (OSSEs) were carried out in a perfect model framework using output from an online-coupled physical-biogeochemical global ocean model at 1/4° nominal resolution. Since all fields used by the FFNN are produced by the same model run and thus internally consistent, the comparison between reconstructed and modelled pCO_2 distributions allows to assess the theoretical skill for each experiment. Starting from measurements extracted from the SOCAT database, the goal was to identify how and where the new data from biogeochemical ARGO floats can improve surface ocean pCO_2 reconstructions and how to optimally integrate them with other existing platforms. *Pseudo-observations* were obtained by sub-sampling model output at sites of real-word observations. Surface ocean pCO_2 was reconstructed from these pseudo-observations at basin scale by applying a non-linear feed forward neural network (FFNN) (Bishop, 1995; Rumelhart et al., 1986). The choice of the FFNN for our experiments was motivated by its overall performance reported in Denvil-Sommer et al. (2019). The architecture of the FFNN method was adapted to the current problem and differs from the one presented in Denvil-Sommer et al. (2019).

The remainder of the article is structured into Section 2 presenting the model output, the observing systems and observations as well as the design experiments, and the description of the statistical model. Results are presented and discussed in Section 3. Section 4 is dedicated to the conclusion and the presentation of perspectives.

2 Data and methods

Here we present the ensemble of observing systems that either already perform measurements to estimate pCO_2 or have the possibility to be equipped with new sensors to provide biogeochemical measurements (Williams et al., 2017). These datasets provide information on geographical, as well as time positions and hence on the distribution of pCO_2 measurements. In this section we also describe the ocean model output and how we use it in the OSSEs. As mentioned in the introduction the data from the model co-localized with real positions of observing-systems are called *pseudo-observations*.

2.1 Data

a) Observing systems
Three observing systems were selected for the study: (1) volunteering observing ships providing *in situ* measurements of surface ocean CO_2 fugacity (fCO_2), (2) moorings (OceanSITES), and (3) profilers (Argo). These observations form the dataset of geographical and time positions for our experiments. Surface ocean measurements of fCO_2 from multiple platforms are converted to pCO_2 and compiled in the SOCAT database (Bakker et al., 2016). Moorings are not routinely equipped with sensors of CO_2 fugacity. We used their geographical positions to identify possible locations for additional measurements. Biogeochemical ARGO floats are increasingly equipped with pH sensors allowing computing pCO_2 from pH and SST-based alkalinity. For the design experiments, we considered distributions of physical ARGO floats (2008-2011) from Gasparin et al. (2019) and supposed that they were equipped with pCO_2 sensors.

(1) **SOCAT** database v5 (Bakker et al., 2016; (https://www.socat.info/index.php/data-access/)): the database provides a good coverage of the Northern Hemisphere. Data for the period 2001-2010 were used, representing ~60% of data in SOCAT database (Fig.1a). The use of data for the period 2001-2010 allows us to capture

interannual variability from a long historical record of SOCAT data and to explore how SOCAT data can be enhanced by other observational platforms. It also provides more data for the training of the Neural Network. While the data from 2001 to 2010 are used in training, the reconstruction focuses only on the years 2008 to 2010. We used the synthesis files SOCATv5, these are the raw data from which the gridded SOCAT product is derived. There are 24 moorings in SOCATv5 that provided CO_2 fugacity measurements between 2001 and 2010. These moorings were excluded from OceanSITES data (see below).

(2) **Argo profilers:** We used the network of Argo (Gould et al., 2004; Argo 2000) distributions provided by Mercator Ocean (details can be found in Gasparin et al., 2019) for the period 2008-2010. This network provides a synthetic homogeneous distribution of 1 profiler per 3ºx3º grid box per 10 days, amounting to 310-360 measurements per day (Fig.1b) based on real trajectories of Argo floats. This synthetic Argo distribution was built based on the time, date and location of Argo profiles during the 2009–2011 period (Gasparin et al., 2019). To provide a homogeneous coverage Gasparin et al. (2019) removed some float trajectories in well-sampled regions, for example the Gulf Stream, or added floats in the low-sampled Tropical and South Atlantic regions. The target for BioGeoChemical Argo (1/4 of ARGO coverage) (Bittig et al., 2018) was derived from this distribution. It is worth noting that Argo floats provide measurements every 10 days. Floats dive to a depth of 2000 m and then rise to the surface by measuring vertical profiles of ocean variables. In this study we use a 5-day time step (see below section b)) which can be a limitation to apply our results to real observations as it does not represent an average value over 5 days. We paid more attention to the spatial distribution, and we believe that with Argo measurements recorded over a longer period our results can be applied to one-month time steps. In this case, 3 monthly measurements can be representative of a monthly mean.

(3) **OceanSITES:** This dataset combines observations from open ocean Eulerian time series stations providing data since 1999 (Fig.1c). We used all available locations of moorings (except moorings included in SOCATv5) and added this information to the period of reconstruction 2008-2010 (http://www.oceansites.org/). It provided 318 additional positions to our data set.

For this study, the same set of predictors was used as in Denvil-Sommer et al. (2019) for training the Machine Learning (ML) algorithm: sea surface salinity (SSS), sea surface temperature (SST), sea surface height (SSH), mixed layer depth (MLD), chlorophyll *a* concentration (Chl *a*) and atmospheric CO_2 ($pCO_{2,\mathrm{atm}}$). These variables are known to represent the main physical, chemical and biological drivers of surface ocean pCO_2 (Takahashi et al., 2009; Landschützer et al., 2013).

b) Model output and pseudo-observations

Here we used the numerical output from an online-coupled physical-biogeochemical global ocean model, the NEMO/PISCES model, at 5-day resolution. This configuration of the Nucleus for European Modelling of the Ocean (NEMO) framework was implemented on a global tripolar grid. It coupled the ocean general circulation model OPA9 (Madec et al., 1998), the sea ice code LIM2 (Fichefet & Maqueda, 1997), and the biogeochemical model PISCESv1 (Aumont and Bopp, 2006). Information on the simulation is given in Gehlen et al. (2020) and Terhaar et al. (2019), including the evaluation of the modelled mean state and the seasonal cycle of sea surface temperature and air-sea fluxes of CO_2 (Gehlen et al., 2020). The geographical and time positions identified from the data mentioned before were used to create pseudo-observations by sub-sampling NEMO/PISCES model output at sites of real-word observations. Thus, the positions of SOCAT, Argo floats and mooring stations were chosen over 5 days centred on the NEMO/PISCES date and sub-sampled on the model grid. The model grid coordinate closest to the real geographical position was chosen, if several measurements were co-localized at the same grid coordinate and same time step it is counted as one measurement. No Argo floats were added to grid cells if there was already a measurement identified in the SOCAT database. All predictors and target pCO_2 were taken from model output at corresponding coordinates. These outputs served as the reference for validation and evaluation of our experiments and for assessing the ML method's accuracy. The simulation covers the period 1958 to 2010, the last 3 years were retained for the design study.

2.2 Observational System Simulation Experiences

Table 1 summarizes experiments designed for different combinations of observing platforms.
The first test is based on individual sampling data extracted from the SOCAT database. As mentioned before these data provide a good coverage of the Northern Hemisphere. The lesser coverage in the Southern Hemisphere results in a larger dispersion of methods based on these observations only (Denvil-Sommer et al., 2019; Rödenbeck et al., 2015). This has motivated experiments with additional data from Argo profilers limited to the Southern Hemisphere. An experiment based on the full physical ARGO network was included to evaluate the method for a high spatial and temporal coverage (an optimal, yet unrealistic case).

We have tested combinations of SOCAT data and (1) total Argo data, (2) Argo only in the Southern Hemisphere, and (3) 25% or (4) 10% of the initial (total) Argo distribution. Finally, these experiments were repeated with additional mooring data. It is worth noting (Table 1) that OSSE 4 is closest to the target of the BGC-Argo program with a BGC-Argo density corresponding to 25% of the existing Argo distribution. However, we decided to choose OSSE 3 as a benchmark against which to evaluate individual experiments. This experiment has a high data density and provides additional information on a potential future BGC-Argo network.

2.3 Method

We used a Feed-Forward Neural Network (FFNN) based on Denvil-Sommer et al. (2019) to reconstruct surface ocean pCO_2 over the Atlantic Ocean. Compared to the previous study we skipped the first step consisting of the reconstruction of the pCO_2 climatology. The reconstruction covered January 2008 to December 2010 with a 5-day frequency and at the spatial resolution of the tripolar ORCA025 model grid (nominal 1/4º resolution). The approach consisted in a method that reconstructs the non-linear relationships between the target pCO_2 and predictors responsible for pCO_2 variability:

$$pCO_{2,n} = f(SSS_n, SST_n, SSH_n, Chl_n, MLD_n, pCO_{2,atm,n}, SSS_{anom,n}, SST_{anom,n}, SSH_{anom,n}, Chl_{anom,n}, MLD_{anom,n}, pCO_{2,atm,anom,n} lat_n, long_{1,n}, long_{2,n}) \quad (1)$$

As previously (Denvil-Sommer et al., 2019), we use Keras, a high-level neural network Python library ("Keras: the Python Deep Learning library", Chollet, 2015; https://keras.io) to construct and train the FFNN models. We first identified an optimal configuration (number and size of hidden layers, the activation functions etc.) of the FFNN model. Based on our earlier work (Denvil-Sommer et al., 2019), a hyperbolic tangent was chosen as an activation function for neurons in hidden layers, and a linear function was chosen for the output layer. As an optimization algorithm, the mini-batch gradient descent or "RMSprop" was used (adaptive learning rates for each weight, Chollet, 2015; Hinton et al., 2012).

The numbers of hidden layers and parameters/weights depend on the number of data used for training. In this work, the FFNN was applied separately for each month (one model for January, one model for February, etc.). A sub-set of 50% of data was used for training. 25% participated in the evaluation of the model during the training algorithm and 25% were used to validate the model after training. These data were chosen regularly in time and space: each third grid point was kept for evaluation, each forth for validation. Tables S1 presents the numbers of training data for each month and each OSSE. To adjust the number of FFNN parameters/weights we followed the empirical rule that suggests limiting the number of parameters to the number of training data points divided by 10 to avoid overfitting (Amari et al., 1997). The FFNNs for all OSSEs except OSSE 2 have four layers (two hidden layers) with 1116 parameters in total. The input layer has 15 input nodes and 20 output nodes that represent the input for the first hidden layer. The first hidden layer has 25 output nodes and the second hidden layer – 10 output nodes. The OSSE 2 which is based on Argo data for the period 2008-2010, has significantly less data for training and thus, the FFNN for the OSSE 2 is different: 3 layers (one hidden layer with 20 input and 10 output nodes) with 541 total parameters.

All data have to be normalized before their use in the FFNN as exemplified for SSS:

$$SSS_n = \frac{SSS - \overline{SSS}}{STD(SSS)} \quad (2)$$

$\overline{SSS}$ is the total mean of variable SSS, *STD(SSS)* is standard deviation of SSS.

Normalization is required to rank all predictors in the same scale and to avoid the possible influence of one predictor with strong variability (Kallache et al., 2011).

Following Denvil-Sommer et al. (2019) we normalized the geographical positions (lat, long) in the following way:

$$lat_n = sin(lat * \pi/180)$$
$$long_{n,1} = sin(long * \pi/180)$$

$$long_{n,2} = cos(long * \pi/180).$$

A K-fold cross-validation was used to evaluate and validate the FFNN architecture. The cross-validation is based on K=4 different subsamples where 25 % of independent data are chosen for validation. In each of the 4 cases the

25% of data are different and there is no overlap. Thereby, each run has 4 outputs. Different architectures of the FFNN were tested and the final one was chosen based on skill assessed by the root-mean-square difference (RMSD), the r^2 and the bias of 4 outputs for each architecture. To ensure a good accuracy of the method and check that there is no overfitting, we compared the RMSD, r^2 and bias estimated from the validation dataset with those estimated from the training dataset. Denvil-Sommer et al. (2019) provide a detailed description of the model including the accuracy of the ML method and its ability to correctly reproduce the pCO_2 variability.

2.4 Diagnostics

The comparison between OSSEs is done per biome, following Rödenbeck et al. (2015) (Fig. 2, Table 2). Biome 8, North Atlantic Ice, has been omitted due to poor data coverage in all OSSEs. It is expected that reconstructions over this region will yield large biases susceptible to interfere with the interpretation of results from individual OSSEs.

In order to simplify the comparison, we used Taylor and Target Diagrams with standard deviation, biases, correlation and normalized RMSD (uRMSD) of the mean of 4 FFNN outputs for each OSSE. Here uRMSD is estimated as:

$$uRMSD = \sqrt{mean(\{ [pCO_{2\,OSSE} - \overline{pCO_{2\,OSSE}}] - [pCO_{2\,NEMO} - \overline{pCO_{2\,NEMO}}] \}^2)} \qquad (3)$$

For each OSSE and each output of the k-fold cross-validation, we estimated a time mean difference between its pCO_2 and NEMO pCO_2 at each grid point:

$$Diff_{j,i} = mean_T(pCO_{2\,OSSE\,j,i} - pCO_{2\,NEMO}) = \frac{1}{T}\sum_{t=1}^{T}(pCO_{2\,OSSE\,j,i,t} - pCO_{2\,NEMO\,t}),$$

where $mean_T$ is a time mean over the period, T is a number of time steps, j is an index of the OSSE and i is an index of output, from 1 to 4.
Further, the maximum absolute value from 4 outputs, $maxValue_j$, was estimated for each OSSE:

$$maxValue_j = max_i(abs(Diff_{j,i})),$$

where max_i is a maximum value on i, the index of output, for each fixed j, the OSSE index. The index i of the maximum absolute value of FFNN outputs is called i_{max}.
The final mean difference $meanD_j$ was estimated as:

$$meanD_j = sign(Diff_{j,i\,max}) * maxValue_j, \qquad (4)$$

where sign(x) is a function that returns the sign of a value x, -1 or 1.
The STD of the mean difference $Diff_{j,i}$ is estimated for each OSSE as:

$$STD_j = std(Diff_{j,i}), \qquad (5)$$

where j is fixed, and all outputs of FFNN i are included in the estimation of STD.

The time series of the mean value from 4 FFNN outputs for pCO_2 were provided per biome, with the maximum and minimum values from these 4 outputs indicated by shading. The time series of CO_2 sea-air flux are shown in the same way as the ones for pCO_2. The sea–air CO_2 flux, $fgCO_2$, was calculated after Rödenbeck et al. (2015):

$$fgCO_2 = k\rho L(pCO_2 - pCO_{2,atm}), \qquad (6)$$

ρ is seawater density and L is the temperature-dependent solubility (Weiss, 1974). k is the piston velocity estimated as (Wanninkhof, 1992):

$$k = \Gamma u^2 (Sc^{CO_2}/Sc^{ref})^{-0.5}.$$

The global scaling factor Γ was estimated following Rödenbeck et al. (2014) with the global mean CO_2 piston velocity equaling 16.5 cm h−1. *Sc* corresponds to the Schmidt number estimated according to Wanninkhof (1992). The wind speed was computed from 6-hourly NCEP wind speed data (Kalnay et al., 1996). To simplify the interpretation of results, the NEMO/PISCES CO_2 air-sea flux was also calculated by using formula (6) and NCEP wind speed.

3 Results

Figure 3 shows the Taylor Diagram (correlation coefficient between reconstructed pCO_2 and model output, and Standard Deviation of reconstructed fields) of 11 OSSEs in the region of 8 biomes (pink) and in each of these biomes separately (color code corresponds to Fig. 2). The target diagrams per biomes for each OSSE are presented on Figure 4. Over regions well-covered with observations (biomes 9, 10, 11) results of different OSSEs lie close to each other. The OSSE 1 (marker symbol "+"; Fig. 3a) that is based only on SOCAT data has a lower correlation coefficient over the whole region (0.67, pink) and per biomes (Fig. 3a). Over regions with poor observational coverage the results from OSSE 1 lie at a distance from results of all other OSSEs. OSSE 1 also shows the largest normalized RMS differences (uRMSD) (Fig. 4), as exemplified for biome 17 with uRMSD of 17.33 μatm, STD of 21.11 μatm (compared to 24.03 μatm estimated from NEMO/PISCES data) and bias of -11.63 μatm (all values in the Fig. 3 and 4 are presented in Tables 3 and 4). The OSSE 2 (based on all Argo data, “O”) and OSSE 3 (combination of Argo and SOCAT data, “X”) provide comparable results (Fig. 3b and c). OSSE 3 tends to have smaller uRMSD and bias and to lie closer to the STD values from the NEMO/PISCES model (Fig. 4). OSSE 3 is based on the maximum of pseudo-observations for training and represents most likely an unrealistic endmember. However, as mentioned before, OSSE 3 is used as the benchmark to find other OSSEs with similar results and more feasible data coverage.

OSSE 4 (square) and OSSE 5 (rhombus) are based on OSSE 3, the only difference being the number of Argo data: OSSE 3, 100%; OSSE 4, 25% and OSSE 5, 10%. The results of OSSEs 4 and 5 are similar to those obtained for OSSE 3. The largest difference is observed over biome 17 (Fig. 3, Fig. 4i): correlation coefficients are 0.85 (OSSE 3), 0.77 (OSSE 4), 0.75 (OSSE 5); biases are -0.66 μatm, -2.25 μatm, -4.02 μatm; uRMSDs are 10.18 μatm, 11.75 μatm, 11.8 μatm (Tables 3, 4).

OSSEs 6 (triangle), 7 (inverted triangle), 8 (pentahedron) were trained on SOCAT data complemented with Argo data in the Southern Hemisphere. In general, the skill scores are lower compared to OSSE 3, especially for OSSE 8 (10% of Argo data in the Southern Hemisphere) where results approach those of OSSE 1 (Fig. 3). Large differences are obtained for biomes 12 and 17 (Fig. 3, Fig. 4e and i): in biome 12/17, correlation coefficients for OSSE 6, 7, 8 are 0.64/0.86, 0.54/0.8, 0.52/0.66, respectively, compared to 0.79/0.85 for OSSE 3; uRMSDs are 11.46/10.01 μatm, 13.3/11.03 μatm, 13.87/15.16 μatm compared to 8/10.18 μatm for OSSE 3; biases are 3.82/-0.18 μatm, 3.77/-1.8 μatm, 2. 7/-4.12 μatm compared to -0.14/-0.66 μatm for OSSE 3 (Tables 3, 4). Over biome 12 all OSSEs show STD values lower than the one computed for NEMO/PISCES model output (Table 3). This could result from the STD of the mean output being slightly lower than the individual STDs for 4 OSSE FFNN outputs (not shown). However, individual STDs also underestimate the NEMO/PISCES STD which might suggest that the ensemble of predictors does not properly represent the variability over the Equatorial Atlantic.

Reconstruction skill scores are improved by the addition of data from mooring stations to OSSEs 6, 7, and 8 in OSSEs 9 (hexagon), 10 (star) and 11 (triangle centroid) (Fig. 3 and 4, Tables 3 and 4). Over the ensemble of 8 biomes the decrease in the number of Argo data goes along with a general decrease of correlation coefficients, 0.88 (OSSE 9), 0.85 (OSSE 10), 0.83 (OSSE 11), and an increase of uRMSDs, 8.37 μatm (OSSE 9), 8.71 μatm (OSSE 10), 9.16 μatm (OSSE 11) (Fig. 3, 4a, Tables 3 and 4). Statistics are slightly worse for OSSE 11 compared to OSSEs 9 and 10, which have comparable results. While OSSE 10 shows a smaller correlation coefficient over the whole region compared to OSSE 9, its STD (24.89 μatm) lies closer to the NEMO/PISCES STD (25.34 μatm) and it has a smaller bias (-0.39 μatm). Similar results are found over other biomes: in biome 12, OSSEs 9 and 10 have correlation coefficients close to each other (0.68 and 0.63, respectively) and larger than for OSSEs 6, 7 and 8, while for OSSE 11 it is 0.58. The STDs are almost equal (OSSE 9, 12.98 μatm and OSSE 10, 12.9 μatm) and uRMSDs have a small difference compared to the one computed for OSSE 3 (8 μatm) (Tables 3, 4). Thus, the remainder of the discussion will focus on OSSE 10 in comparison to OSSEs 1 and 3. OSSE 10 provides comparable results to OSSE 9 and is in good agreement with OSSE 3 while using less data for training. Figures 3 and 4 are summarized in Figure S1 of the Supplementary Material.

Figures 5a, b and c present the differences between reconstructed $p\text{CO}_2$ distributions (Fig.5 a – OSSE 1; b – OSSE 3; c – OSSE 10) and NEMO/PISCES output. The maximum in absolute value from 4 outputs for each OSSE FFNN is shown (Eq. 4). There is a large improvement in the Southern Hemisphere for OSSEs 3 (Fig. 5b) and 10 (Fig. 5c) compared to OSSE 1 (Fig. 5a): the difference varies mostly between -3 and 3 μatm for OSSEs 3 and 10, and between -15 and 15 μatm for OSSE 1 (Fig. 5). However, the average values of the mean over biomes are not always better for OSSE 3 (Table 5): in biome 13, OSSE 1 shows a small positive difference of 0.11 μatm, while for OSSE 3 negative difference of -0.32 μatm is computed, exceeding 0.11 μatm in its absolute value. This is due to error compensation by averaging, the reduction of the positive difference in the middle of biome 13 in OSSE 3 increases the impact of negative small differences in this region. Error compensation also contributes to positive biases computed for OSSEs 6-11 for biome 12 (Table 4). Additional data from Argo floats correct the negative bias in the southern part of the biome close to the African coast (Fig. 5c). Thus, the strong positive bias in the northern part becomes dominant and results in a total positive bias. A large improvement is obtained in biomes 16 and 17: from -8.04 μatm for OSSE 1 to -1.89 μatm and -1.91 μatm for OSSEs 3 and 10 in biome 16, and from -14.9 μatm for OSSE 1 to -2.05 μatm and -1.55 μatm for OSSEs 3 and 10 in biome 17 (Table 5). Over the whole region, 70ºW-30ºE 80ºS-80ºN, OSSE 1 has a mean difference of -6.57 μatm, it is -1.7 μatm and -2.34 μatm for OSSEs 3 and 10. The difference between OSSEs 3 and 10 results from the Labrador Sea and Baffin Bay: OSSE

10 has fewer data in this region compared to the OSSE 3. However, there is an improvement in OSSE 10 compared to OSSE 1 and 3 in the Greenland Sea (Fig. 5). It results from the addition of mooring data in the Greenland Sea region (Fig. 1c).

Figures 5d, e and f present the standard deviations (STD) of differences for all 4 outputs for each OSSE FFNN (Fig.5 d – OSSE 1; e – OSSE 3; f – OSSE 10) (Eq. 5). Over most of the Atlantic Ocean STD varies between 0 and 10 μatm for OSSEs 3 and 10. In each case there is a strong STD along the coasts and in the Labrador Sea, as well as the Baffin Bay. In general, the mean value of STD tends to decrease (Table 5) from OSSE 1 to OSSEs 3 and 10. In the Southern Hemisphere STD reaches up to 30 μatm (Figures 5d, e and f)) when only SOCAT data are used in the FFNN algorithm (OSSE 1). It is significantly reduced in response to the addition of float data in OSSEs 3 and 10 with also less spatial variability. The results for other OSSEs are added to the Supplementary material (Table S2, Fig. S2, S3).

Figure 6 shows the correlation between the mean value of 4 OSSEs outputs and NEMO/PISCES pCO_2 (a - OSSE 1, b - OSSE 3, c - OSSE 10). The additional data from Argo floats and mooring stations increase the correlation coefficient from 0.68 in the case of OSSE 1 (SOCAT data only) to 0.86 and 0.85 in the case of OSSEs 3 and 10 (Table 6). A higher correlation was also obtained for these two OSSEs compared to OSSE 1 over the region covering the Greenland Sea, the Norwegian Sea and Barents Sea (mostly biome 9). In the Southern Hemisphere the correlation with NEMO/PISCES pCO_2 is also larger when Argo data are included, especially in biomes 16 and 17: 0.7 and 0.57 for OSSE 1, 0.83 and 0.85 for OSSE 3, as well as 0.78 and 0.89 for OSSE 10 (Table 6). However, there is a low correlation along the African coasts which is in agreement with our previous results for mean difference and STD (Fig. 5). It reflects the predominantly open ocean data used for this exercise. A well-pronounced decrease in correlation is observed for biome 15 (Subtropical seasonally stratified Southern Ocean). Such a decrease can result from the spatial distribution of data or from the predictor data set. We will discuss it further in the next section. The results for other OSSEs are presented in the Supplementary material (Table S3, Fig. S4).

In Figure 7, time series of pCO_2 for OSSEs 1, 3 and 10 are compared to corresponding NEMO/PISCES model output. For each OSSE, the mean pCO_2 from 4 FFNN outputs is shown, as well as the mean bias (OSSE - NEMO/PISCES). Figure 7a and b presents the pCO_2 time series over the period of reconstruction 2008-2010 for OSSE 1, 3, 10 compared to NEMO/PISCES pCO_2 used as reference (black) over all biomes. For OSSE 1 (SOCAT data only) a large difference and an underestimation of reconstructed pCO_2 (blue) compared to NEMO/PISCES pCO_2 (black) are found: the maximum error is up to -10 μatm (Fig. 7b). To the contrary, OSSEs 3 and 10 show a good agreement with NEMO/PISCES model output. Averages of pCO_2 over the 8 biomes are 372.18 μatm for OSSE 3, 372.26 μatm for OSSE 10 and 368.39 μatm for OSSE 1, compared to 372.65 μatm for NEMO/PISCES (Table 7). The experiment corresponding to the BGC-Argo distribution target over the entire Atlantic basin, OSSE 4 (Fig. S8, S9), has a basin-wide average pCO_2 equal to 371.8 μatm (Table 7). This corresponds to a larger difference with NEMO/PISCES (-0.84 μatm) compared to OSSEs 3 and 10.

Panels (c) to (h) of Figure 7 illustrate time series of reconstructed pCO_2 for biomes with varying data coverage. Biome 11, the Subtropical permanently stratified North Atlantic, (Figure 7c and d) is well covered by data. All three OSSEs yield pCO_2 reconstructions that are in good accordance with the NEMO/PISCES reference. The amplitude and the phasing of the seasonal cycle are well reproduced. The bias varies within a range of +/-5 μatm for OSSEs 3 and 10. A predominantly negative bias is found for OSSE 1 with values as high as -10 μatm. The pCO_2 averaged over biome 11 for OSSE 10 is close to NEMO/PISCES with, respectively 389.39 μatm and 390.11 μatm (Table 7). OSSE 1 yields a biome-averaged pCO_2 equal to 387.11 μatm, while it is 389.39 μatm for the OSSE 3.

Biome 13, the Subtropical permanently stratified South Atlantic, (Figure 7e and f) corresponds to a region with a low data coverage. This region has a dynamic similar to biome 11 in the Northern Hemisphere, however the data coverage in biome 13 represents only 15% of data coverage in biome 11 (Fig. S5). We observe a large difference between pCO_2 reconstructed by OSSE 1 (blue) and NEMO/PISCES (black). While the phasing of the reconstructed seasonal cycle is satisfying, it is noisy with a systematic overestimation in spring by up to 18 μatm (Table 7). However, the total averaged pCO_2 over biome 13 for OSSE 1 is close to the one of NEMO/PISCES: 391.66 μatm, respectively 389.54 μatm. The preceding suggests that while the variability of the predictors (mainly SST) is sufficient to constrain at first order the biome-average pCO_2 and the phasing of the seasonal cycle, an improved coverage by *in situ* observations is needed for a smooth reconstruction of the seasonal cycle and its amplitude. Reconstructions are largely improved by the addition of data from Argo floats (OSSE 3) and moorings (OSSE 10). Biases mostly range between -3 and 3 μatm for these OSSEs.

The Southern Ocean Ice biome (biome 17) is characterized by a sparse data coverage and a bias towards the ice-free season. The results for biome 17 are presented in Figure 7g and h. OSSE 1 underestimates the pCO_2 in this region over the full seasonal cycle. The maximum difference is obtained in September-October, which also corresponds to the months with the lowest number of available observations (Fig. S5). The biome-wide average is 351.44 μatm, -11.63 μatm below the NEMO/PISCES reference. The reconstruction is much improved for OSSEs 3 and 10, both for the phasing and amplitude of the seasonal cycle, as well as for the biome-wide averages. The latter are 362.42 μatm and 362.87 μatm, respectively for OSSE 3 and OSSE 10, compared to 363.08 μatm computed for NEMO/PISCES (Table 7).
Results for all OSSEs and for all biomes are included in the Supplementary material (Table S4, Fig. S6 – S11).

Figure 8 shows the sea-air CO_2 flux time series (negative, uptake of CO_2 by the ocean). Over all biomes and in the region 70ºW-30ºE 80ºS-80ºN, OSSEs 3 (red) and 10 (green) show a good agreement with NEMO/PISCES $fgCO_2$: the differences vary around zero and mostly do not exceed ± 0.3 Pg/yr (Fig. 8b, d, f and h). The total averaged $fgCO_2$ for OSSE 3 and 10 are -0.74 Pg/yr compared to -0.7 Pg/yr in NEMO/PISCES, while for OSSE 1 it equals -0.99 Pg/yr (Table 8). The mean value over biome 11 is slightly better for OSSE 10 than for OSSE 3 compared to NEMO/PISCES: -0.06 Pg/yr (OSSE 10), -0.07 (OSSEs 3) and -0.03 Pg/yr for NEMO/PISCES. The OSSE 1 (blue) shows again a large difference, it overestimates the ocean sink computed by the NEMO/PISCES model mostly during the whole period (Fig. 8b). In the well data-covered biome 11, OSSE 1 also has a tendency to overestimate the sea-air CO_2 flux (Fig. 8d): the total averaged $fgCO_2$ is -0.18 Pg/yr for OSSE 1 while it is -0.03 Pg/yr in the model. While the phasing and amplitude of the seasonal cycle of sea-air fluxes of CO_2 are well reproduced over biome 13 by OSSEs 3 and 10, the $fgCO_2$ reconstructed by OSSE 1 is noisy with differences with respect to the model reference of up 1 Pg/yr (Fig. 8e). The maximum differences between OSSE 1 and NEMO/PISCES are systematically found in January and June, the months with the lowest number of available observations for training (Fig. S5). The biome-wide mean sea-air flux of CO_2 is close to zero in NEMO/PISCES: -0.004 Pg/yr. This slight uptake of CO_2 by the ocean in the model reference is not reproduced by the OSSEs which yield a source over biome 13, albeit of variable strength: 0.19 Pg/yr for OSSE 1, 0.05 Pg/yr for OSSE 3 and 0.08 Pg/yr for OSSE 10. Over the Southern Ocean biome 17 (Fig.8g and h) OSSE 1 (blue) overestimates $fgCO_2$ by -0.65 g/yr (Table 8). OSSE 10 (green) reproduces the local maxima and minima of the $fgCO_2$ time series slightly better than OSSE 3, with average differences equaling -0.03 Pg/yr and -0.06 Pg/yr, respectively. Results for all OSSEs and for all biomes can be found in the Supplementary material (Table S5, Fig. S12 - S17).

The relationship between the average number of Argo floats (5-day period) and the error in $fgCO_2$ estimates (Table 8, Table S5) is shown in Figure 9 for all biomes (a), biome 11 (b), biome 13 (c) and biome 17 (d). Figure 9a illustrates how the increase of the number of floats usually yields a reduction in the error of $fgCO_2$ estimates. Considering the whole region, OSSE 10 provides the best results with less Argo floats (-0.04 PgC/yr and 48 Argo floats). At the biome-scale, the addition of floats does, however, not systematically reduce the error. This holds for biome 11 (Fig. 9b), which is well-covered by observations, but also for biome 13 with a much sparser data-coverage (Fig. 9d). For biome 11, OSSE 10 has the best trade-off between error reduction and number of floats. The largest error (0.22 PgC/yr) is obtained for OSSE 2 (only Argo data). It suggests that the period chosen for this study is too short to adequately capture the seasonal variability. This hypothesis is supported by the fact that while OSSE 3 and OSSE 2 share the same number of Argo data, OSSE 3 is further constrained by SOCAT data that cover the period 2001-2010. These additional data from SOCAT introduce the information needed for the reconstruction of the seasonal cycle. For biome 13 (Fig. 9c), the combination of SOCAT data and Argo float data improves estimates of $fgCO_2$. The errors in OSSE 10 are comparable to OSSE 3 (benchmark), 0.08 PgC/yr (OSSE 10) and 0.06 PgC/yr (OSSE 3). The error is even lower for OSSE 11 (0.04 PgC/yr), the experiment with the smallest number of Argo floats (19), than for OSSE 3. Unfortunately, results provided by OSSE 11 are less good over the remainder of the biomes. The tendency for a decrease of $fgCO_2$ error with an increase of the number of Argo floats is confirmed for biome 17 (Fig. 9d). The additional data from mooring stations (OSSE 9, 10 and 11) improve in particular OSSEs with smaller numbers of floats. An error of -0.03 PgC/yr is computed for OSSE 10 (49 floats) over biome 17. The results for other biomes can be found in the Supplementary material (Fig. S18).

4 Summary and Conclusion

The aim of this work was to identify an optimal observational network of pCO_2 over the Atlantic Ocean. The analysis was based on results obtained with a Feed-Forward Neural Network model trained on the SOCAT database. The SOCAT database has sparse coverage in the Southern Hemisphere. The approach consisted in adding the position of mooring data and Argo trajectories in the Atlantic Ocean to find an optimal distribution and combination of data to reconstruct pCO_2 with a good accuracy. The advantage of the SOCAT database is the long time period covered by its records, which allows to reconstruct the interannual variability with a good accuray. However, its data coverage is biased towards the North Atlantic, which leads to larger reconstruction errors over the South Atlantic by the Neural Network. As a long-term perspective, the inclusion of data from Argo floats will contribute to a more homogenous data distribution and provide better spatial coverage. The Argo floats and moorings used here do not currently provide pCO_2 measurements, hence only their positions were used to build OSSEs. A series of experiments were performed using outputs from the NEMO/PISCES model. The model simulations were sub-sampled at co-localized sites of observing systems for all predictors (SSS, SST, SSH, CHL, MLD, $pCO_{2,\,atm}$) used in the FFNN and the target (pCO_2) to create pseudo-observations with a 5-day time step. These experiments should be useful for the planning of future deployments of BGC-Argo floats (Biogeochemical-Argo Planning Group, 2016) and moorings equipped with the sensors to measure pCO_2 or CO_2 fugacity.

The results suggest that the addition of data from Argo floats could significantly improve the accuracy of FFNN-based ocean pCO_2 reconstructions over the Atlantic Ocean and the Atlantic sector of the Southern Ocean compared to the case when only SOCAT data are used (OSSE 1). However, even with an improved coverage over the open ocean, additional observations are required in coastal regions and shelf seas which are not accessible to floats, as well as in regions with a strong seasonal variability of pCO_2 and all predictors. This is exemplified by OSSE 2, the experiment based on all Argo data, which yields high RMSDs in biome 9, the Subpolar seasonally stratified North Atlantic (Fig. 3, Fig.4b, Table 4). The RMSD of 17.1 μatm reflects the poor coverage of this region by Argo floats (Fig. 1b), in particular the Greenland Sea and the North Sea, with a large part of the latter not suitable for the deployment of floats. The combination of SOCAT data and Argo floats (OSSE 3) improves the reconstruction with a RMSD reduced to 9.59 μatm (Fig. 4b, Table 4).

The reduction of the number of Argo data used in our experiments slightly decreases the accuracy (Fig. 3 and 4, Tables 3 and 4). A lower number of Argo data corresponds, however, to a more realistic distribution of instruments and to the target of the global BGC-Argo network. The results are still comparable to OSSE 3. The best compromise between the statistics yielded by the comparison between reconstructed pCO_2 and NEMO/PISCES outputs, as well as the feasibility of a future observation network is found for OSSE 10. In this experiment SOCAT data are combined with simulated mooring data and 25% of the initial distribution of Argo floats placed only in the Southern Hemisphere (around 49 floats with a 5-day sampling period). The use of only SOCAT data results in a correlation coefficient of 0.67 compared to NEMO/PISCES output and a standard deviation of 26.08 μatm (25.34 μatm for NEMO/PISCES) over the region of study. While the successful OSSE 10 has a correlation coefficient of 0.85 and a standard deviation of 24.89 μatm. These results are close to the unrealistic benchmark case with total Argo float distribution over 2008-2010: 0.87 and 23.79μatm. The total pCO_2 over the whole region is also close to NEMO/PISCES, ~370 μatm and ~371 μatm, respectively. The air-sea flux $fgCO_2$ is -0.83 Pg/yr (OSSE 10) and -0.76 Pg/yr (NEMO). The bias in sea-air CO_2 fluxes compared to NEMO/PISCES is reduced by 74% in OSSE 10 compared to OSSE 1 ($fgCO_2$ is -1.03 Pg/yr).

The OSSE 10 network could be further improved by instrumenting the Baffin Bay, the Labrador Sea, the Norwegian Sea, as well as regions along the coast of Africa (10ºN to 20ºS), all regions with pronounced biases in all OSSEs, with moorings or gliders as well as sail-drones and sail buoys along the shelf break and on the continental shelf.

The inclusion of errors from *in situ* measurements is one of the next steps of this work. The real measurements contain instrumental and representation errors. The inclusion of errors in pseudo-observations will help to estimate the impact of observations on the reliability of OSSEs presented in this work. It will include the errors for predictor values (SSS, SST, SSH, CHL, MLD, $pCO_{2,\,atm}$) that are measured directly or derived from remote sensing (e.g., SST, chlorophyll, SSH), as well as the errors related to the computation of pCO_2 from pH and alkalinity. The new FFNN runs could provide important information on the effect of biases from observational datasets and identify predictors or targets that have large errors and that must be corrected. The consistent introduction of error estimates for each predictor will provide this information.

Author contribution:

ADS, MG, MV contributed to the development of the methodology and designed the experiments, and ADS carried them out. ADS developed the model code and performed the simulations. ADS prepared the paper with contributions from all coauthors.

Acknowledgments:

The authors would like to thank an anonymous referee and Luke Gregor for their helpful comments and questions. This study was funded by the AtlantOS project (EU Horizon 2020 research and innovation program, grant agreement no. 2014-633211) and GreenGrog PPR (GMMC). At present ADS is under funding from the Royal Society (grant no. RP\R1\191063) at the UEA. MV acknowledges support from the CoCliServ project, which is part of ERA4CS, an ERA-NET initiated by JPI Climate and cofunded by the European Union. The authors thank Florent Gasparin for providing the reference data of Argo distributions.

References

Argo. Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE. https://doi.org/10.17882/42182, 2000.

Aumont, O., and L. Bopp: Globalizing results from ocean in situ iron fertilization studies, Global Biogeochem. Cycles, 20, GB2017, doi:10.1029/2005GB002591, 2006.

Bakker, D. C. E., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., Smith, K., Cosca, C., Harasawa, S., Jones, S. D., Nakaoka, S.-I., Nojiri, Y., Schuster, U., Steinhoff, T., Sweeney, C., Takahashi, T., Tilbrook, B., Wada, C., Wanninkhof, R., Alin, S. R., Balestrini, C. F., Barbero, L., Bates, N. R., Bianchi, A. A., Bonou, F., Boutin, J., Bozec, Y., Burger, E. F., Cai, W.-J., Castle, R. D., Chen, L., Chierici, M., Currie, K., Evans, W., Featherstone, C., Feely, R. A., Fransson, A., Goyet, C., Greenwood, N., Gregor, L., Hankin, S., Hardman-Mountford, N. J., Harlay, J., Hauck, J., Hoppema, M., Humphreys, M. P., Hunt, C. W., Huss, B., Ibánhez, J. S. P., Johannessen, T., Keeling, R., Kitidis, V., Körtzinger, A., Kozyr, A., Krasakopoulou, E., Kuwata, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lo Monaco, C., Manke, A., Mathis, J. T., Merlivat, L., Millero, F. J., Monteiro, P. M. S., Munro, D. R., Murata, A., Newberger, T., Omar, A. M., Ono, T., Paterson, K., Pearce, D., Pierrot, D., Robbins, L. L., Saito, S., Salisbury, J., Schlitzer, R., Schneider, B., Schweitzer, R., Sieger, R., Skjelvan, I., Sullivan, K. F., Sutherland, S. C., Sutton, A. J., Tadokoro, K., Telszewski, M., Tuma, M., van Heuven, S. M. A. C., Vandemark, D., Ward, B., Watson, A. J., and Xu, S.: A multi-decade record of high-quality fCO_2 data in version 3 of the Surface Ocean CO_2 Atlas (SOCAT), Earth Syst. Sci. Data, 8, 383–413, https://doi.org/10.5194/essd-8-383-2016, 2016.

Biogeochemical-Argo Planning Group: The scientific rationale, design and Implementation Plan for a Biogeochemical-Argo float array. Edited by Ken Johnson and Hervé Claustre. doi:10.13155/46601, 2016.

Bishop, C. M.: Neural Networks for Pattern Recognition, Oxford University Press, Cambridge, UK, 1995.

Bittig, H.C., Steinhoff, T., Claustre, H., Fiedler, B., Williams, N.L., Sauzède, R., Körtzinger, A., and Gattuso, J.-P.: An Alternative to Static Climatologies: Robust Estimation of Open Ocean CO_2 Variables and Nutrient Concentrations From T, S, and O_2 Data Using Bayesian Neural Networks, Front. Mar. Sci., 5, 328, https://doi.org/10.3389/fmars.2018.00328, 2018.

Bushinsky, S. M., Landschützer, P., Rödenbeck, C., Gray, A. R., Baker, D., Mazloff, M. R., et al.: Reassessing Southern Ocean air-sea CO_2 flux estimates with the addition of biogeochemical float observations, Global Biogeochemical Cycles, 33, 1370–1388, https://doi.org/10.1029/2019GB006176, 2019.

Chollet, F.: Keras, available at: https://keras.io (last access: 12 May 2019), 2015.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Le Quéré, C., Myneni, R. B., Piao, S., and Thornton, P.: Carbon and other biogeochemical cycles, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Denvil-Sommer, A., Gehlen, M., Vrac, M., and Mejia, C.: LSCE-FFNN-v1: a two-step neural network model for the reconstruction of surface ocean pCO2 over the global ocean, Geosci. Model Dev., 12, 2091–2105, https://doi.org/10.5194/gmd-12-2091-2019, 2019.

Fay, A. R. and McKinley, G. A.: Global open-ocean biomes: mean and temporal variability, Earth Syst. Sci. Data, 6, 273–284, https://doi.org/10.5194/essd-6-273-2014, 2014.

Fay, A. R., McKinley, G. A., and Lovenduski, N. S.: Southern Ocean carbon trends: Sensitivity to methods, Geophys. Res. Lett., 41, 6833–6840, https://doi.org/10.1002/2014GL061324, 2014.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S., Aragão, L. E. O. C., Arneth, A., Arora, V., Bates, N. R., Becker, M., Benoit-Cattin, A., Bittig, H. C., Bopp, L., Bultan, S., Chandra, N., Chevallier, F., Chini, L. P., Evans, W., Florentie, L., Forster, P. M., Gasser, T., Gehlen, M., Gilfillan, D., Gkritzalis, T., Gregor, L., Gruber, N., Harris, I., Hartung, K., Haverd, V., Houghton, R. A., Ilyina, T., Jain, A. K., Joetzjer, E., Kadono, K., Kato, E., Kitidis, V., Korsbakken, J. I., Landschützer, P., Lefèvre, N., Lenton, A., Lienert, S., Liu, Z., Lombardozzi, D., Marland, G., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Niwa, Y., O'Brien, K., Ono, T., Palmer, P. I., Pierrot, D., Poulter, B., Resplandy, L., Robertson, E., Rödenbeck, C., Schwinger, J., Séférian, R., Skjelvan, I., Smith, A. J. P., Sutton, A. J., Tanhua, T., Tans, P. P., Tian, H., Tilbrook, B., van der Werf, G., Vuichard, N., Walker, A. P., Wanninkhof, R., Watson, A. J., Willis, D., Wiltshire, A. J., Yuan, W., Yue, X., and Zaehle, S.: Global Carbon Budget 2020, Earth Syst. Sci. Data, 12, 3269–3340, https://doi.org/10.5194/essd-12-3269-2020, 2020.

Gasparin, F., Guinehut, S., Mao, C., Mirouze, I., Rémy, E., King, R.R., Hamon, M., Reid, R., Storto, A., Le Traon, P.Y. and Martin, M.J.: Requirements for an integrated in situ Atlantic Ocean observing system from coordinated observing system simulation experiments, Frontiers in Marine Science, 6, p.83, 2019.

Gehlen, M., Berthet, S., Séférian, R., Ethé, C., and Penduff, T.: Quantification of chaotic intrinsic variability of sea-air CO2 fluxes at interannual timescales, Geophysical Research Letters, 47, e2020GL088304, https://doi.org/10.1029/2020GL088304, 2020.

Gould, J., Roemmich, D., Wijffels, S., Freeland, H., Ignaszewsky, N., Jianping, X., et al.: Argo profiling floats bring new era of in situ ocean observations, Eos Transactions American Geophysical Union, 85(19): 185–19, https://doi.org/10.1029/2004EO190002, 2004.

Gregor, L., Lebehot, A. D., Kok, S., and Scheel Monteiro, P. M.: A comparative assessment of the uncertainties of global surface ocean CO2 estimates using a machine-learning ensemble (CSIR-ML6 version 2019a) – have we hit the wall? Geosci. Model Dev., 12, 5113–5136, https://doi.org/10.5194/gmd-12-5113-2019, 2019.

Hauck, J., Zeising, M., Le Quéré, C., Gruber, N., Bakker, D.C.E., Bopp, L., Chau, T.T.T., Gürses, Ö., Ilyina, T., Landschützer, P., Lenton, A., Resplandy, L., Rödenbeck C., Schwinger, J., and Séférian, R.: Consistency and Challenges in the Ocean Carbon Sink Estimate for the Global Carbon Budget, Front. Mar. Sci. 7:571720. doi: 10.3389/fmars.2020.571720, 2020.

Kamenkovich, I., A. Haza, A. R. Gray, C. O. Dufour, and Garraffo, Z.: Observing System Simulation Experiments for an array of autonomous biogeochemical profiling floats in the Southern Ocean, J. Geophys. Res. Oceans, 122, 7595–7611, doi:10.1002/2017JC012819, 2017.

Landschützer, P., Gruber, N., Bakker, D. C. E., Schuster, U., Nakaoka, S., Payne, M. R., Sasse, T. P., and Zeng, J.: A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink, Biogeosciences, 10, 7793–7815, https://doi.org/10.5194/bg-10-7793-2013, 2013.

Landschützer, P., Gruber, N., and Bakker, D. C. E.: Decadal variations and trends of the global ocean carbon sink, Global Biogeochem. Cy., 30, 1396–1417, https://doi.org/10.1002/2015GB005359, 2016.

Lenton, A., Bopp, L., and Matear, R. J.: Strategies for high-latitude northern hemisphere CO_2 sampling now and in the future. Deep-Sea Research Part II: Topical Studies in Oceanography, 56(8–10), 523–532. https://doi.org/10.1016/j.dsr2.2008.12.008, 2009.

Majkut, J. D., Carter, B. R., Frölicher, T. L., Dufour, C. O., Rodgers, K. B., and Sarmiento, J. L.: An observing system simulation for Southern Ocean carbon dioxide uptake, Philos. Trans. R. Soc. A, 372, https://doi.org/10.1098/rsta.2013.0046, 2014.

Monteiro, P., Schuster, U., Hood, M., Lenton, A., Metzl, N., Olsen,A., Rogers, K., Sabine, C., Takahashi, T., Tilbrook, B., Yoder,J., Wanninkhof, R., Watson, A.: A global sea surface carbon ob-serving system: Assessment of changing sea surface CO_2 and air-sea CO_2 fluxes, Proceedings of OceanObs'09: Sustained

Ocean Observations and Information for Society, 1, 702–714, https://doi.org/10.5270/OceanObs09.cwp.64, 2010.

Nakaoka, S., Telszewski, M., Nojiri, Y., Yasunaka, S., Miyazaki, C., Mukai, H., and Usui, N.: Estimating temporal and spatial variation of ocean surface pCO_2 in the North Pacific using a self-organizing map neural network technique, Biogeosciences, 10, 6093–6106, https://doi.org/10.5194/bg-10-6093-2013, 2013.

Rödenbeck, C., Bakker, D. C. E., Metzl, N., Olsen, A., Sabine, C., Cassar, N., Reum, F., Keeling, R. F., and Heimann, M.: Interannual sea–air CO_2 flux variability from an observation-driven ocean mixed-layer scheme, Biogeosciences, 11, 4599–4613, https://doi.org/10.5194/bg-11-4599-2014, 2014.

Rödenbeck, C., Bakker, D. C. E., Gruber, N., Iida, Y., Jacobson, A. R., Jones, S., Landschützer, P., Metzl, N., Nakaoka, S., Olsen, A., Park, G.-H., Peylin, P., Rodgers, K. B., Sasse, T. P., Schuster, U., Shutler, J. D., Valsala, V., Wanninkhof, R., and Zeng, J.: Data-based estimates of the ocean carbon sink variability – first results of the Surface Ocean *p*CO2 Mapping intercomparison (SOCOM), Biogeosciences, 12, 7251–7278, https://doi.org/10.5194/bg-12-7251-2015, 2015.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning internal representations by backpropagating errors, Nature, 323, 533–536, 1986.

Schuster, U., McKinley, G. A., Bates, N., Chevallier, F., Doney, S. C., Fay, A. R., González-Dávila, M., Gruber, N., Jones, S., Krijnen, J., Landschützer, P., Lefèvre, N., Manizza, M., Mathis, J., Metzl, N., Olsen, A., Rios, A. F., Rödenbeck, C., Santana-Casiano, J. M., Takahashi, T., Wanninkhof, R., and Watson, A. J.: An assessment of the Atlantic and Arctic sea–air CO_2 fluxes, 1990–2009, Biogeosciences, 10, 607–627, https://doi.org/10.5194/bg-10-607-2013, 2013.

Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wanninkhof, R., Feely, R. A., Sabine, C., Olafsson, J., and Nojiri, Y.: Global sea-air CO_2 flux based on climatological surface ocean pCO_2, and seasonal biological and temperature effects, Deep.-Sea Res. Pt. II, 49, 1601–1622, https://doi.org/10.1016/S0967-0645(02)00003-6, 2002.

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C., Watson, A., Bakker, D. C. E., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Körtzinger, A., Steinhoff, T., Hoppema, M., Olafsson, J., Arnarson, T. S., Tilbrook, B., Johannessen, T., Olsen, A., Bellerby, R., Wong, C. S., Delille, B., Bates, N. R., and de Baar, H. J. W.: Climatological mean and decadal change in surface ocean pCO_2, and net sea-air CO_2 flux over the global oceans, Deep.-Sea Res. Pt. II, 56, 554–577, https://doi.org/10.1016/j.dsr2.2008.12.009, 2009.

Terhaar, J., Orr, J. C., Gehlen, M., Ethé, C., and Bopp, L.: Model constraints on the anthropogenic carbon budget of the Arctic Ocean, Biogeosciences, 16, 2343–2367, https://doi.org/10.5194/bg-16-2343-2019, 2019.

Williams, N. L., et al.: Calculating surface ocean pCO_2 from biogeochemical Argo floats equipped with pH: An uncertainty analysis, Global Biogeochem.Cycles, 31, 591–604, doi:10.1002/2016GB005541, 2017.

Zeng, J., Nojiri, Y., Landschützer, P., Telszewski, M., and Nakaoka, S.: A global surface ocean fCO_2 climatology based on a feed-forward neural network, J. Atmos. Ocean Technol., 31, 1838–1849, https://doi.org/10.1175/JTECH-D-13-00137.1, 2014.

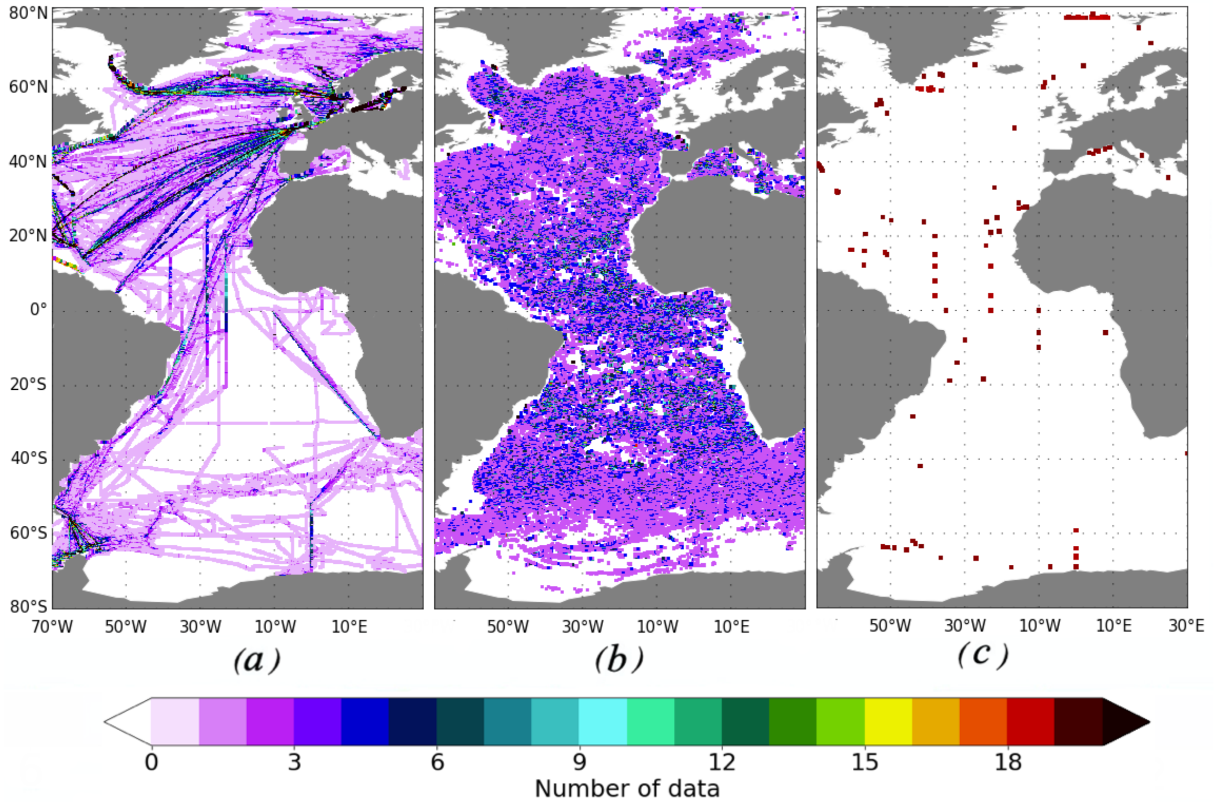

Figure 1: Spatial distribution of data sets used for training (number of measurements per grid points and 5-day time step): (a) SOCAT data for the period 2001-2010; (b) synthetic Argo data for the period 2008-2010; (c) mooring positions modelled for the period 2008-2010.

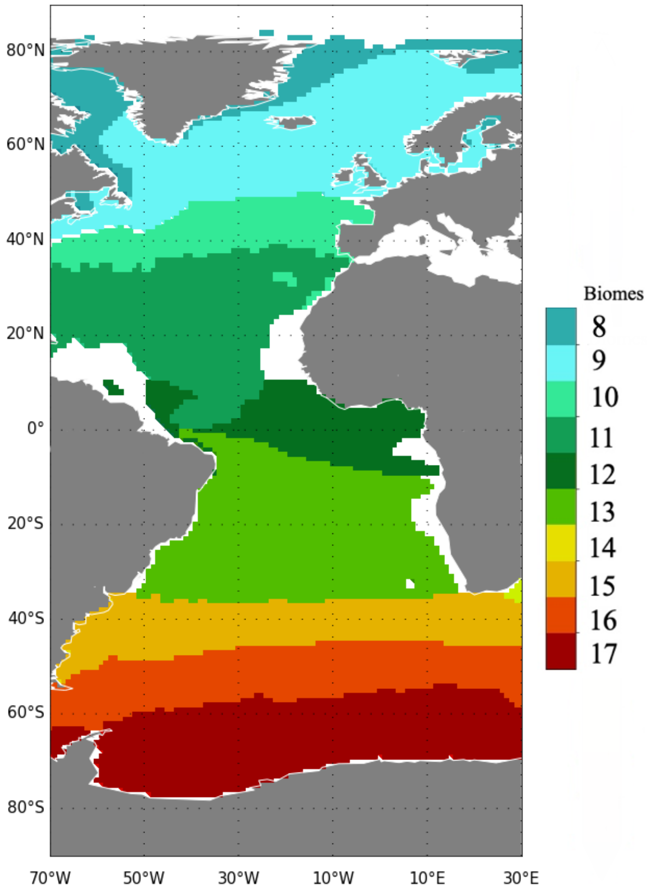

Figure 2: Map of biomes (after Rödenbeck et al., 2015; Fay and McKinley, 2014) focused on the region [70°W-30°E] and used for comparison between OSSEs.

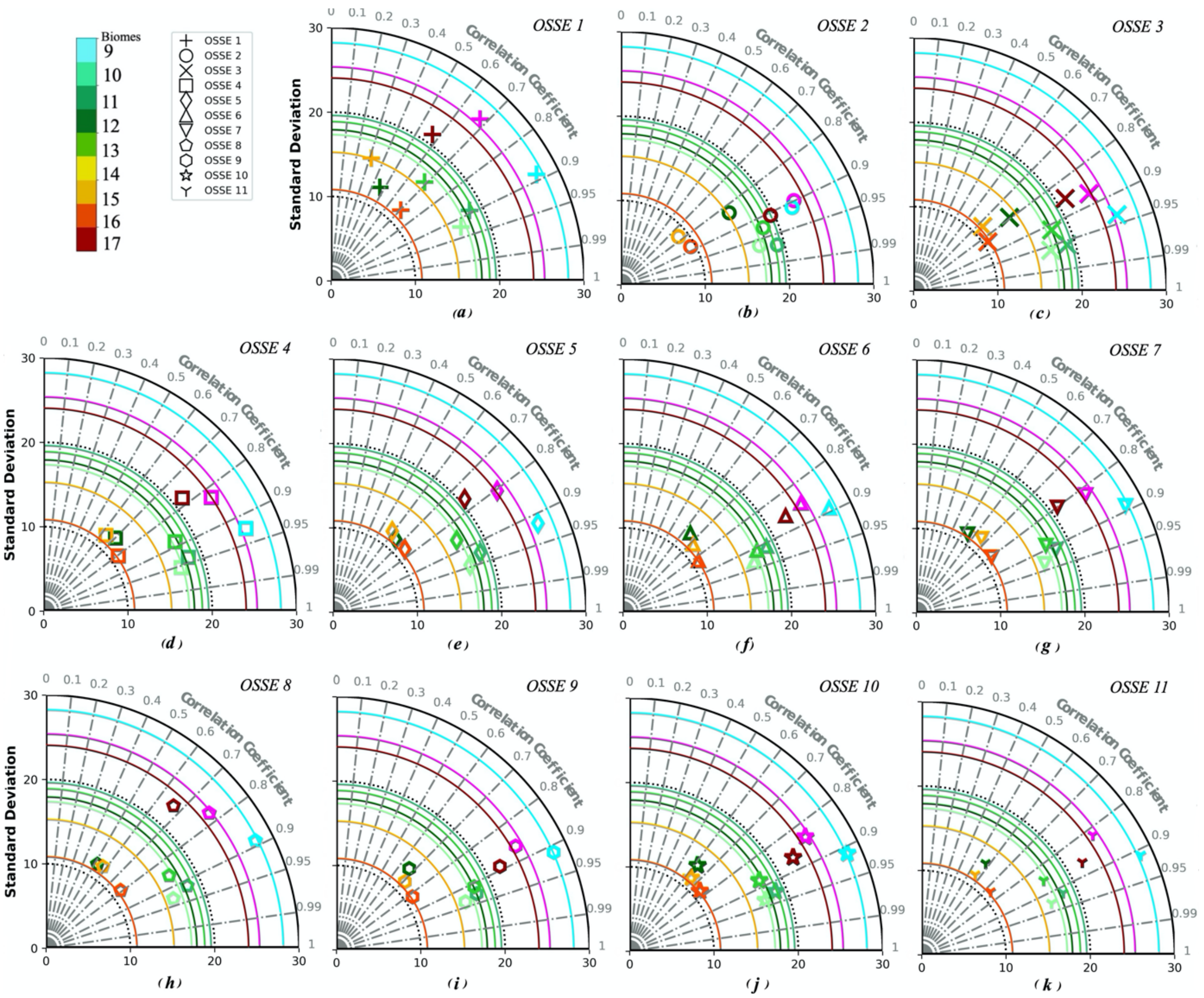

Figure 3: Taylor Diagram of 11 OSSEs summarized in Table 2; the colour code corresponds to Fig. 2, the purple colour represents the whole of the 8 biomes: (a) - OSSE 1: SOCAT data only; (b) - OSSE 2: synthetic Argo data only; (c) - OSSE 3: SOCAT and synthetic Argo data; (d) - OSSE 4: SOCAT data and 25% of original synthetic Argo data; (e) - OSSE 5: SOCAT data and 10% of original synthetic Argo data; (f) - OSSE 6: SOCAT data and synthetic Argo data in the Southern Hemisphere; (g) - OSSE 7: SOCAT data and 25% of original synthetic Argo data in the Southern Hemisphere; (h) - OSSE 8: SOCAT data and 10% of original synthetic Argo data in the Southern Hemisphere; (i) - OSSE 9: SOCAT data, synthetic Argo data in the Southern Hemisphere and data from mooring stations; (j) - OSSE 10: SOCAT data, 25% of original synthetic Argo data in the Southern Hemisphere and data from mooring stations; (k) - OSSE 11: SOCAT data, 10% of original synthetic Argo data in the Southern Hemisphere and data from mooring stations.

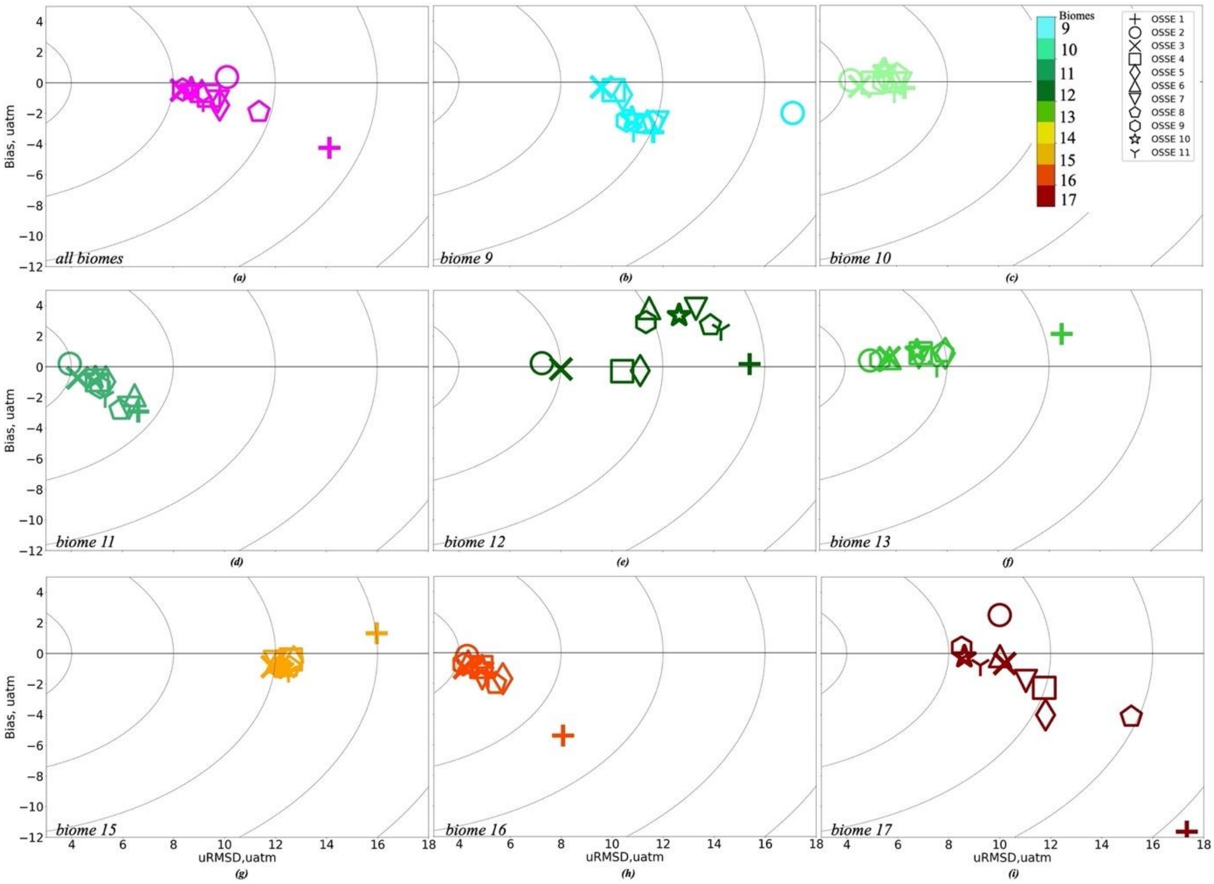

Figure 4: Target Diagram per biome for 11 OSSEs, the colour code corresponds to Fig. 2, the purple colour represents the whole of the 8 biomes: (a) - all 8 biomes, (b) - biome 9 (Subpolar seasonally stratified North Atlantic), (c) - biome 10 (Subtropical seasonally stratified North Atlantic), (d) - biome 11 (Subtropical permanently stratified North Atlantic), (e) - biome 12 (Equatorial Atlantic), (f) - biome 13 (Subtropical permanently stratified South Atlantic), (g) - biome 15 (Subtropical seasonally stratified Southern Ocean), (h) - biome 16 (Subpolar seasonally stratified Southern Ocean), (i) - biome 17 (Southern Ocean ice). OSSE 1: SOCAT data only; OSSE 2: synthetic Argo data only; OSSE 3: SOCAT and synthetic Argo data; OSSE 4: SOCAT data and 25% of original synthetic Argo data; OSSE 5: SOCAT data and 10% of original synthetic Argo data; OSSE 6: SOCAT data and synthetic Argo data in the Southern Hemisphere; OSSE 7: SOCAT data and 25% of original synthetic Argo data in the Southern Hemisphere; OSSE 8: SOCAT data and 10% of original synthetic Argo data in the Southern Hemisphere; OSSE 9: SOCAT data, synthetic Argo data in the Southern Hemisphere and data from mooring stations; OSSE 10: SOCAT data, 25% of original synthetic Argo data in the Southern Hemisphere and data from mooring stations; OSSE 11: SOCAT data, 10% of original synthetic Argo data in the Southern Hemisphere and data from mooring stations. OSSEs 1, 3 and 10 are in bold as we focus our detailed comparison on these three OSSEs.

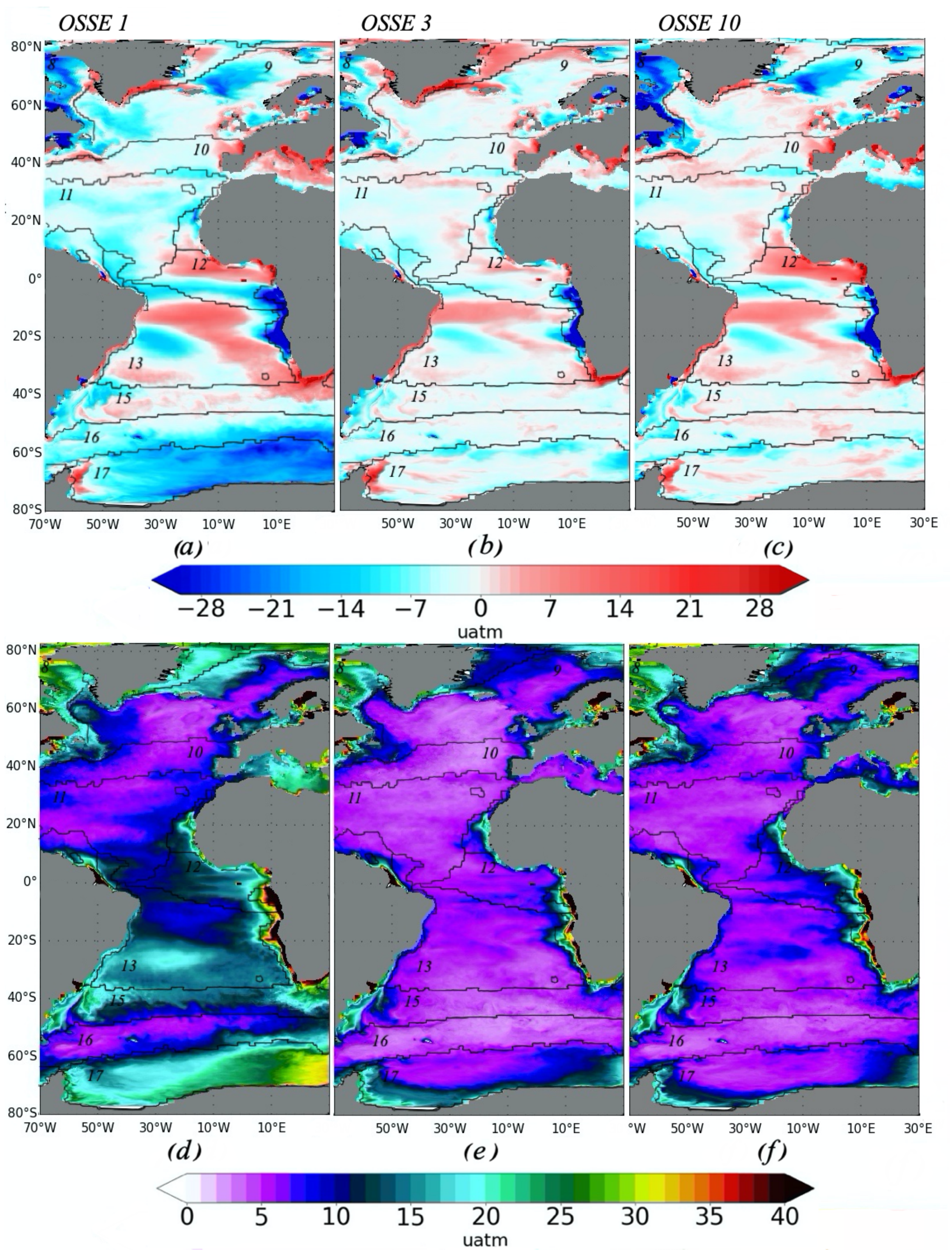

Figure 5: Differences between OSSE FFNN outputs and NEMO/PISCES $p CO_2$ and its standard deviation (STD) in μatm: (a), (b), (c) - its maximum and minimum values from 4 outputs for each OSSE FFNN, Eq. (4); (g), (h) - standard deviation of differences for all 4 outputs for each OSSE FFNN, Eq. (5). (a), (d) – OSSE 1: SOCAT data only; (b), (e) – OSSE 3: SOCAT and synthetic Argo data; (c), (f) – OSSE 10: SOCAT data, 25% of original synthetic Argo data in the Southern Hemisphere and data from mooring stations. Contours and numbers on maps correspond to biomes.

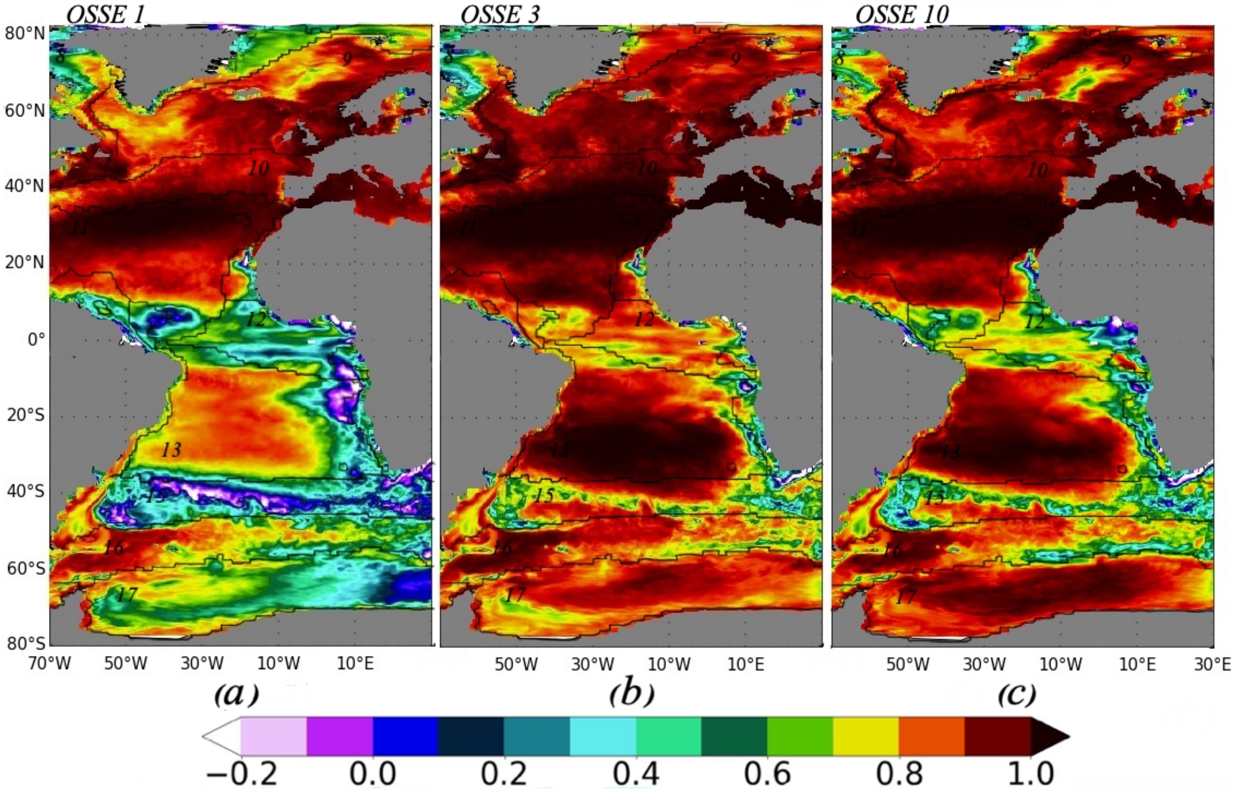

Figure 6: Correlation coefficient between OSSE FFNN outputs and NEMO/PISCES *p*CO_2: (a) - OSSE 1: SOCAT data only, (b) - OSSE 3: SOCAT and synthetic Argo data, (c) - OSSE 10: SOCAT data, 25% of original synthetic Argo data in the Southern Hemisphere and data from mooring stations. Contours and numbers on maps correspond to biomes.

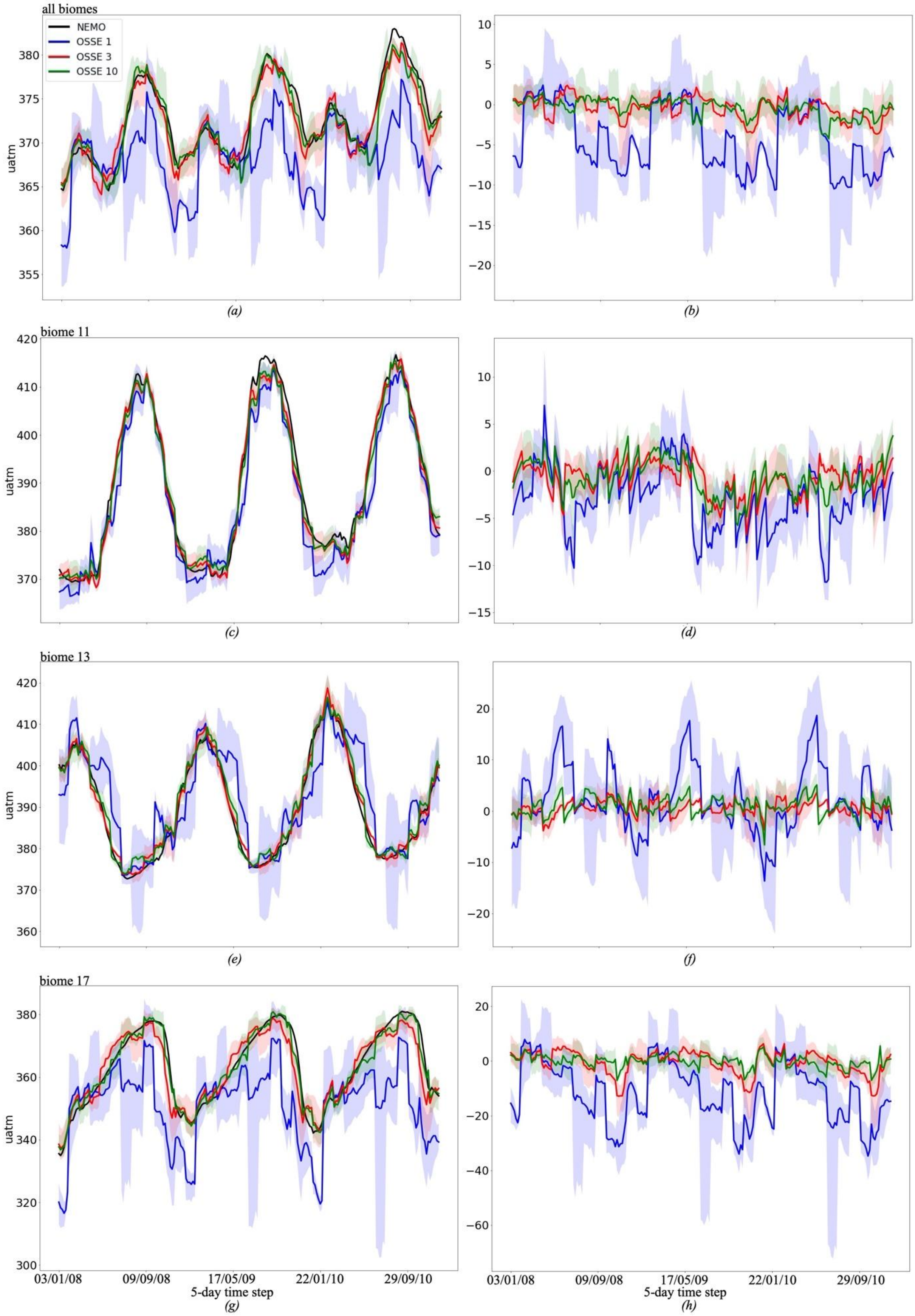

Figure 7: (a), (c), (e) - mean of 4 FFNN outputs for OSSE 1 (blue) (SOCAT data only), 3 (red) (SOCAT and synthetic Argo data), 10 (green) (SOCAT data, 25% of original synthetic Argo data in the Southern Hemisphere and data from mooring stations); shading corresponds to the maximum and minimum values from 4 FFNN outputs for each OSSE. Black curve - NEMO/PISCES pCO_2. (b), (d), (f) - mean of differences of 4 FFNN outputs between OSSE 1 (blue), 3

(red), 10 (green) and NEMO/PISCES pCO$_2$; shading corresponds to the maximum and minimum values of differences from 4 FFNN outputs for each OSSE. (a), (b) - estimates are available over all biomes presented in Figure 2 except biome 8; (c), (d) - biome 11 (Subtropical permanently stratified North Atlantic); (e), (f) - biome 13 (Subtropical permanently stratified South Atlantic); (g), (h) - biome 17 (Southern Ocean ice).

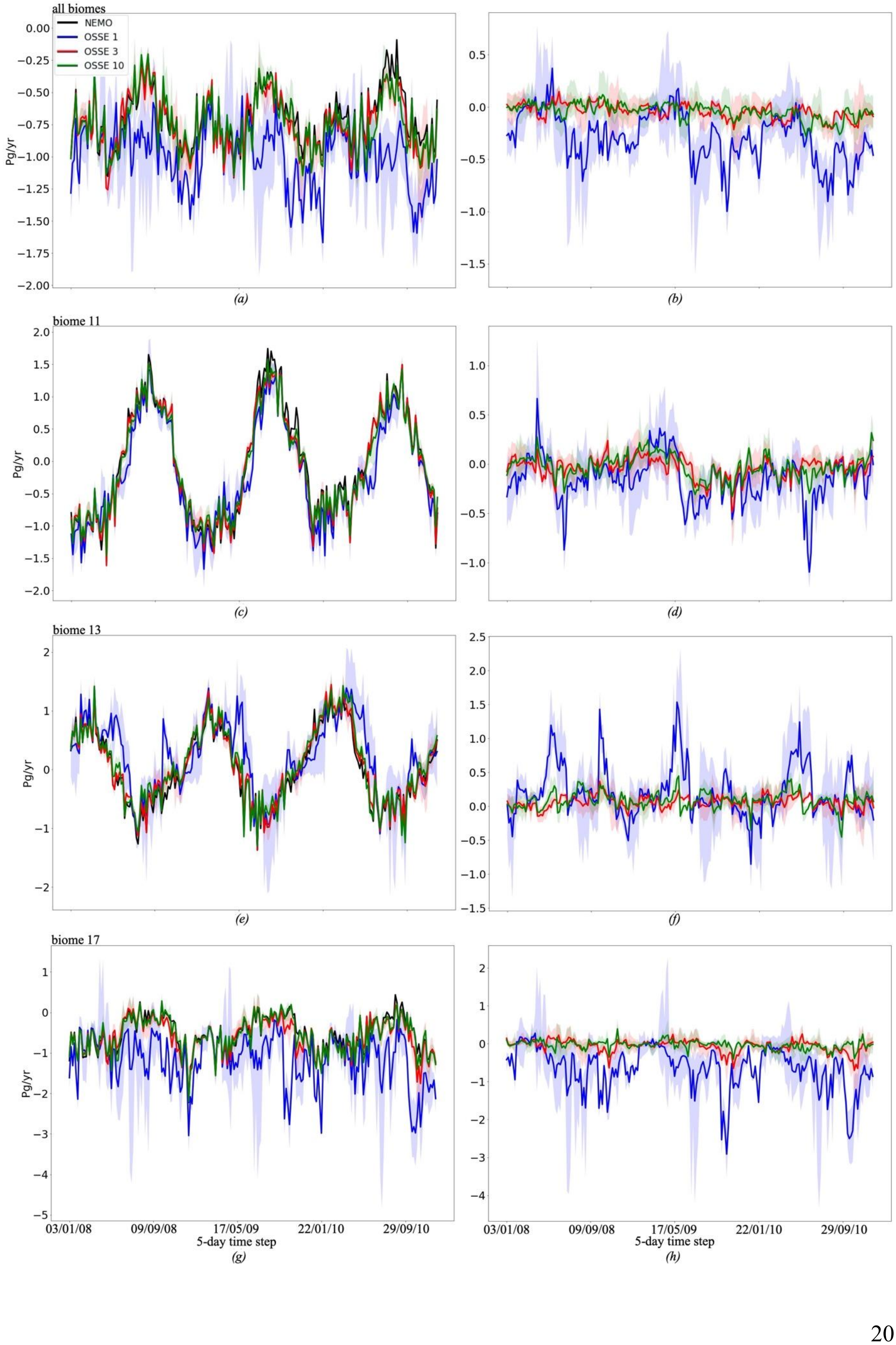

Figure 8: (a), (c), (e) - mean of *fg*CO_2 from 4 FFNN outputs for OSSE 1 (blue) (SOCAT data only), 3 (red) (SOCAT and synthetic Argo data), 10 (green) (SOCAT data, 25% of original synthetic Argo data in the Southern Hemisphere and data from mooring stations); shading corresponds to the maximum and minimum values from 4 FFNN *fg*CO_2 estimates for each OSSE. Black curve - NEMO/PISCES *fg*CO_2. (b), (d), (f) - mean of differences of 4 FFNN outputs between OSSE 1 (blue), 3 (red), 10 (green) *fg*CO_2 and NEMO/PISCES *fg*CO_2; shading corresponds to the maximum and minimum values of differences from 4 FFNN *fg*CO_2 for each OSSE. (a), (b) - estimates are available for all biomes presented in Figure 2 except biome 8; (c), (d) - biome 11 (Subtropical permanently stratified North Atlantic); (e), (f) - biome 13 (Subtropical permanently stratified South Atlantic); (g), (h) - biome 17 (Southern Ocean ice).

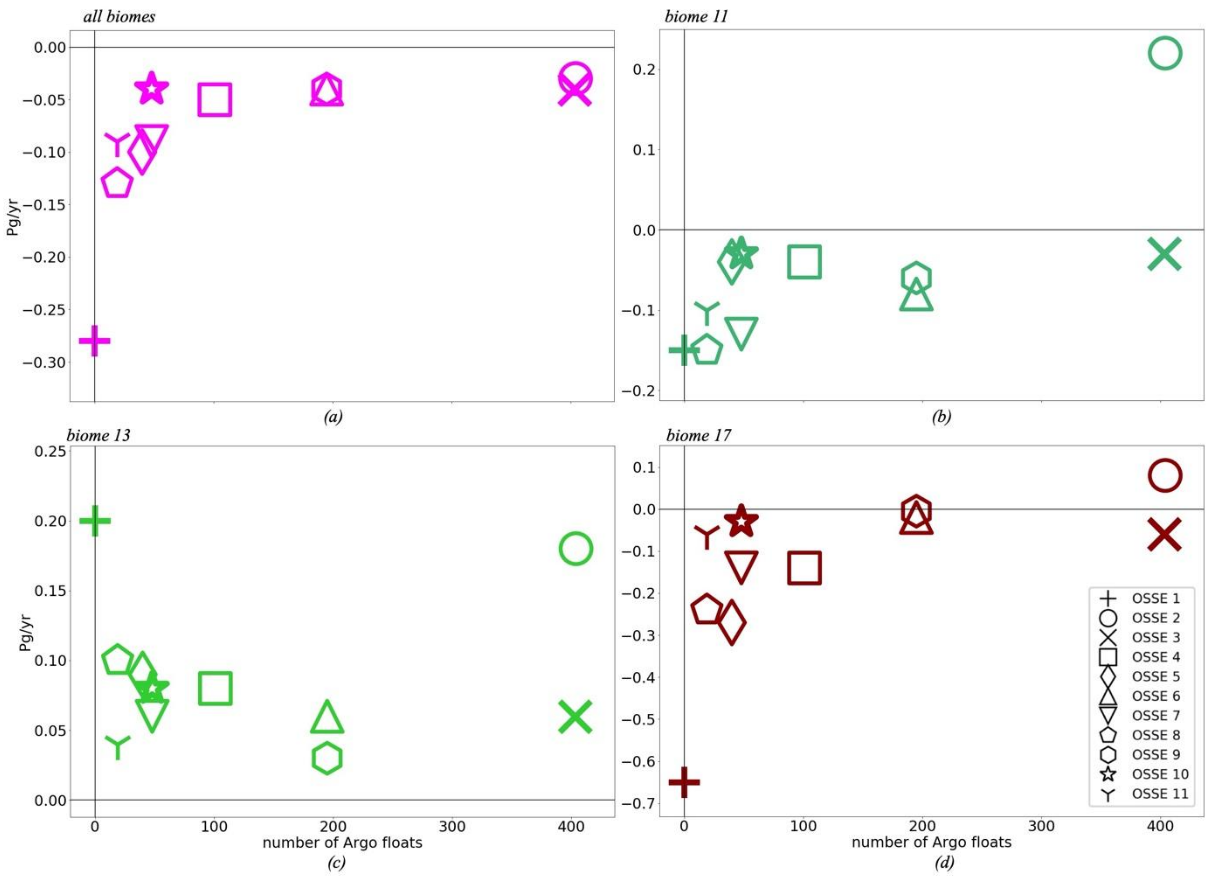

Figure 9: Averaged number of Argo profiles per 5-day time step over 2008-2010 versus averaged differences between each OSSE *fg*CO_2 and NEMO *fg*CO_2 (in Pg/yr), the colour code corresponds to Fig. 2, the purple colour represents the whole of the 8 biomes. (a) - all biomes; (b) - biome 11 (Subtropical permanently stratified North Atlantic); (c) - biome 13 (Subtropical permanently stratified South Atlantic); (d) - biome 17 (Southern Ocean ice). OSSE 1: SOCAT data only; OSSE 2: synthetic Argo data only; OSSE 3: SOCAT and synthetic Argo data; OSSE 4: SOCAT data and 25% of original synthetic Argo data; OSSE 5: SOCAT data and 10% of original synthetic Argo data; OSSE 6: SOCAT data and synthetic Argo data in the Southern Hemisphere; OSSE 7: SOCAT data and 25% of original synthetic Argo data in the Southern Hemisphere; OSSE 8: SOCAT data and 10% of original synthetic Argo data in the Southern Hemisphere; OSSE 9: SOCAT data, synthetic Argo data in the Southern Hemisphere and data from mooring stations; OSSE 10: SOCAT data, 25% of original synthetic Argo data in the Southern Hemisphere and data from mooring stations; OSSE 11: SOCAT data, 10% of original synthetic Argo data in the Southern Hemisphere and data from mooring stations. OSSEs 1, 3 and 10 are in bold as they represent the main OSSEs of our comparisons.

Table 1: Information on Observation System Simulation Experiments.

Data	OSSE number	Period for training	averaged number of Argo floats per 5 days

SOCAT	OSSE 1	2001-2010	0
Argo (3ºx3º)	OSSE 2	2008-2010	404
SOCAT + Argo (3ºx3º)	OSSE 3	2001-2010 (SOCAT) + 2008-2010 (Argo)	403
SOCAT + Argo 25% (3ºx3º)	OSSE 4	2001-2010 (SOCAT) + 2008-2010 (Argo)	101
SOCAT + Argo 10% (3ºx3º)	OSSE 5	2001-2010 (SOCAT) + 2008-2010 (Argo)	40
SOCAT + Argo South (3ºx3º)	OSSE 6	2001-2010 (SOCAT) + 2008-2010 (Argo South)	195
SOCAT + Argo 25% South (3ºx3º)	OSSE 7	2001-2010 (SOCAT) + 2008-2010 (Argo South)	48
SOCAT + Argo 10% South (3ºx3º)	OSSE 8	2001-2010 (SOCAT) + 2008-2010 (Argo South)	19
SOCAT + Argo S + Moorings	OSSE 9	2001-2010 (SOCAT) + 2008-2010 (Argo South, Moorings)	195
SOCAT + Argo S 25% + Moorings	OSSE 10	2001-2010 (SOCAT) + 2008-2010 (Argo South, Moorings)	48
SOCAT + Argo S 10% + Moorings	OSSE 11	2001-2010 (SOCAT) + 2008-2010 (Argo South, Moorings)	19

Table 2: Biomes from Fay and McKinley (2014) used for time series comparison (Fig. 2).

Number	Name
8	(Omitted) North Atlantic ice
9	Subpolar seasonally stratified North Atlantic
10	Subtropical seasonally stratified North Atlantic
11	Subtropical permanently stratified North Atlantic
12	Equatorial Atlantic
13	Subtropical permanently stratified South Atlantic
15	Subtropical seasonally stratified Southern Ocean
16	Subpolar seasonally stratified Southern Ocean
17	Southern Ocean ice

Table 3: Correlation coefficient and Standard Deviation (μatm) of 11 OSSEs from Table 2 estimated over 8 Atlantic Ocean biomes and at basin scale; the results are presented in Fig. 3. OSSEs 1, 3 and 10 are in bold as we focus our detailed comparison on these three OSSEs.

Biome OSSE	All biomes	9	10	11	12	13	15	16	17
NEMO STD	25.34	28.17	17.29	19.59	17.89	18.84	15.20	10.79	24.03
OSSE 1	**0.67/ 26.08**	**0.88/ 27.44**	**0.92/ 16.67**	**0.89/ 18.42**	**0.46/ 12.48**	**0.68/ 16.11**	**0.31/ 15.28**	**0.70/ 11.76**	**0.57/ 21.11**
OSSE 2	0.89/ 22.82	0.91/ 22.28	0.96/ 17.09	0.97/ 19.14	0.83/ 15.42	0.92/ 18.19	0.76/ 8.89	0.87/ 9.43	0.90/ 19.56
OSSE 3	**0.87/ 23.79**	**0.93/ 25.78**	**0.96/ 17.00**	**0.95/ 19.03**	**0.79/ 14.33**	**0.91/ 17.91**	**0.73/ 11.21**	**0.83/ 10.55**	**0.85/ 21.06**
OSSE4	0.82/ 23.99	0.92/ 25.91	0.95/ 17.11	0.93/ 18.31	0.70/ 12.13	0.88/ 17.62	0.63/ 11.62	0.80/ 10.99	0.77/ 21.2
OSSE 5	0.80/ 24.18	0.92/ 26.48	0.94/ 17.16	0.92/ 18.83	0.65/ 11.39	0.86/ 16.95	0.59/ 11.86	0.75/ 11.3	0.75/ 20.58
OSSE 6	0.85/ 24.72	0.89/ 27.40	0.93/ 16.66	0.91/ 18.73	0.64/ 12.34	0.91/ 17.51	0.72/ 11.56	0.82/ 10.84	0.86/ 22.41
OSSE 7	0.82/ 24.48	0.89/ 27.87	0.93/ 16.32	0.91/ 18.19	0.54/ 11.17	0.88/ 17.33	0.66/ 11.71	0.80/ 11.12	0.80/ 20.90
OSSE 8	0.77/ 25.10	0.89/ 27.90	0.93/ 16.19	0.91/ 18.3	0.52/ 11.66	0.86/ 16.92	0.57/ 11.74	0.79/ 11.17	0.66/ 22.63
OSSE 9	0.88/ 24.51	0.92/ 28.17	0.95/ 16.11	0.94/ 17.67	0.68/ 12.98	0.92/ 17.84	0.72/ 11.31	0.84/ 10.89	0.91/ 21.63
OSSE 10	**0.85/ 24.89**	**0.91/ 28.28**	**0.94/ 17.10**	**0.94/ 18.41**	**0.63/ 12.90**	**0.88/ 17.36**	**0.65/ 11.35**	**0.78/ 11.01**	**0.89/ 22.25**
OSSE 11	0.83/ 24.67	0.91/ 28.39	0.93/ 16.4	0.93/ 18.10	0.58/ 13.20	0.86/ 16.79	0.56/ 11.29	0.74/ 10.96	0.88/ 21.92

Table 4: Normalised RMS differences and Biases (μatm) of 11 OSSEs from Table 2 estimated over 8 Atlantic Ocean biomes and at basin scale; the results are presented in Fig. 4. OSSEs 1, 3 and 10 are in bold as we focus our detailed comparison on these three OSSEs.

Biome OSSE	All biomes	9	10	11	12	13	15	16	17
OSSE 1	**14.13/ -4.25**	**11.63/ -3.26**	**6.32/ -0.39**	**6.63/ -2.93**	**15.41/ 0.17**	**12.5/ 2.12**	**15.97/ 1.32**	**8.08/ -5.41**	**17.33/ -11.63**
OSSE 2	10.11/ 0.36	17.10/ -2.02	4.21/ 0.09	3.94/ 0.19	7.26/ 0.22	4.98/ 0.38	12.63/ -0.43	4.31/ -0.21	10.00/ 2.50
OSSE 3	**8.32/ -0.46**	**9.59/ -0.32**	**4.56/ -0.30**	**4.24/ -0.71**	**8.00/ -0.14**	**5.73/ 0.57**	**11.87/ -0.85**	**4.20/ -0.97**	**10.18/ -0.66**
OSSE 4	9.40/	10.08/	5.08/	5.01/	10.41/	6.96/	12.59/	4.87/	11.75/

	-0.84	-0.53	-0.05	-0.88	-0.29	0.85	-0.40	-0.93	-2.25
OSSE 5	9.82/ -1.46	10.43/ -0.83	5.50/ 0.50	5.35/ -0.98	11.11/ -0.25	7.93/ 0.85	12.72/ -0.54	5.71/ -1.69	11.80/ -4.02
OSSE 6	9.12/ -0.54	11.40/ -2.57	5.93/ 0.02	6.48/ -1.86	11.46/ 3.82	5.75/ 0.53	12.06/ -0.51	4.35/ -0.56	10.01/ -0.18
OSSE 7	9.75/ -1.22	11.79/ -2.64	6.16/ -0.10	6.26/ -2.68	13.30/ 3.77	6.90/ 0.58	11.97/ -0.56	4.90/ -1.68	11.03/ -1.80
OSSE 8	11.36/ -1.89	11.62/ -2.59	6.02/ 0.49	5.91/ -2.80	13.87/ 2.70	7.84/ 0.90	12.55/ -0.89	5.42/ -2.03	15.16/ -4.12
OSSE 9	8.37/ -0.44	10.58/ -2.52	5.47/ -0.001	5.13/ -1.33	11.34/ 2.91	5.37/ 0.41	12.18/ -0.88	4.16/ -0.75	8.51/ 0.37
OSSE 10	**8.71/ -0.39**	**10.79/ -2.35**	**5.54/ 0.79**	**4.94/ -0.71**	**12.64/ 3.35**	**6.82/ 1.01**	**12.25/ -0.92**	**4.89/ -0.90**	**8.61/ -0.21**
OSSE 11	9.16/ -1.18	10.85/ -3.21	5.91/ -0.68	5.32/ -1.97	14.28/ 2.41	7.59/ 0.002	12.49/ -1.18	5.13/ -1.56	9.23/ -0.77

Table 5: Differences (Eq. 4) between OSSE FFNN outputs and NEMO/PISCES *p*CO_2 and its standard deviation (STD) (Eq. 5) in μatm.

Biome OSSE	Region 70ºW-30ºE 80ºS-80ºN	All 8 biomes	9	10	11	12	13	15	16	17
OSSE 1	-6.57/ 14.49	-6.57/ 13.54	-4.84/ 10.17	-1.46/ 6.98	-4.21/ 7.62	-2.03/ 13.88	0.11/ 13.88	-1.35/ 14.96	-8.04/ 8.99	-14.90/ 20.83
OSSE 3	-1.70/ 8.12	-1.50/ 7.15	-1.36/ 7.52	-0.90/ 4.62	-1.48/ 4.64	-1.49/ 7.09	-0.32/ 5.58	-1.93/ 7.16	-1.89/ 4.42	-2.05/ 10.59
OSSE 10	-2.34/ 8.64	-1.54/ 7.50	-3.54/ 8.59	-0.10/ 6.18	-1.52/ 5.42	1.93/ 9.38	-0.04/ 6.51	-2.15/ 8.18	-1.91/ 5.21	-1.55/ 8.99

Table 6: Correlation coefficient between OSSEs and NEMO/PISCES *p*CO_2.

Biome OSSE	Region 70ºW-30ºE 80ºS-80ºN	All 8 biomes	9	10	11	12	13	15	16	17
OSSE 1	0.68	0.67	0.88	0.92	0.89	0.46	0.68	0.31	0.70	0.57
OSSE 3	0.86	0.87	0.93	0.96	0.95	0.79	0.91	0.73	0.83	0.85
OSSE 10	0.85	0.85	0.92	0.94	0.94	0.63	0.88	0.65	0.78	0.89

Table 7: pCO_2 averaged over the region 70ºW-30ºE 80ºS-80ºN and biomes from Fig. 2 for the NEMO/PISCES model and OSSEs 1, 3 and 10, as well as the corresponding averaged differences between OSSEs and NEMO/PISCES (in μatm).

Biome / OSSE	Region 70ºW-30ºE 80ºS-80ºN	All 8 biomes	9	10	11	12	13	15	16	17
NEMO	371.13	372.65	350.36	373.18	390.11	397.18	389.54	376.14	376.99	363.08
OSSE 1	367.09/ -4.04	368.39/ -4.25	347.10/ -3.26	372.78/ -0.39	387.17/ -2.93	397.36/ 0.17	391.66/ 2.12	377.46/ 1.32	371.58/ -5.41	351.44/ -11.63
OSSE 3	370.62/ -0.51	372.18/ -0.46	350.04/ -0.32	372.88/ -0.30	389.39/ -0.71	397.04/ -0.14	390.10/ 0.57	375.29/ -0.85	376.02/ -0.97	362.42/ -0.66
OSSE 10	370.14/ -0.99	372.26/ -0.39	348.01/ -2.35	373.98/ 0.79	389.39/ -0.71	400.53/ 3.35	390.55/ 1.01	375.22/ -0.92	376.09/ -0.90	362.87/ -0.21

Table 8: $fgCO_2$ averaged over the region 70ºW-30ºE 80ºS-80ºN and biomes from Fig. 2 for the NEMO/PISCES model and OSSEs 1, 3, 4 and 10, as well as the corresponding averaged differences between each OSSEs and NEMO/PISCES (in Pg/yr).

Biome / OSSE	Region 70ºW-30ºE 80ºS-80ºN	All 8 biomes	9	10	11	12	13	15	16	17
NEMO	-0.76	-0.70	-2.34	-1.14	-0.03	0.53	-0.004	-0.74	-0.50	-0.52
OSSE 1	-1.03/ -0.26	-0.99/ -0.28	-2.57/ -0.23	-1.17/ -0.03	-0.18/ -0.15	0.42/ -0.10	0.19/ 0.20	-0.68/ 0.06	-1.15/ -0.64	-1.17/ -0.65
OSSE 3	-0.80/ -0.04	-0.74/ -0.04	-2.36/ -0.02	-1.16/ -0.02	-0.07/ -0.03	0.49/ -0.04	0.05/ 0.06	-0.82/ -0.07	-0.61/ -0.10	-0.59/ -0.06
OSSE 10	-0.83/ -0.06	-0.74/ -0.04	-2.50/ -0.15	-1.09/ 0.04	-0.06/ -0.03	0.56/ 0.03	0.08/ 0.08	-0.82/ -0.07	-0.60/ -0.09	-0.56/ -0.03