An in situ-satellite blended analysis of global sea surface salinity

P. Xie¹, T. Boyer², E. Bayler³, Y. Xue¹, D. Byrne², J. Reagan²,4, R. Locarnini², F. Sun¹, R. Joyce¹, and A. Kumar¹

¹NOAA/NWS Climate Prediction Center, College Park, Maryland, USA, ²NOAA/NESDIS National Oceanographic Data Center, Silver Spring, Maryland, USA, ³NOAA/NESDIS Center for Satellite Applications and Research, Camp Springs, Maryland, USA, ⁴Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland, USA

Abstract The blended monthly sea surface salinity (SSS) analysis, called the NOAA “Blended Analysis of Surface Salinity” (BASS), is constructed for the 4 year period from 2010 to 2013. Three data sets are employed as inputs to the blended analysis: in situ SSS measurements aggregated and quality controlled by NOAA/NODC, and passive microwave (PMW) retrievals from both the National Aeronautics and Space Administration’s (NASA) Aquarius/SAC-D and the European Space Agency’s (ESA) Soil Moisture-Ocean Salinity (SMOS) satellites. The blended analysis comprises two steps. First, the biases in the satellite retrievals are removed through probability distribution function (PDF) matching against temporally spatially colocated in situ measurements. The blended analysis is then achieved through optimal interpolation (OI), where the analysis for the previous time step is used as the first guess while the in situ measurements and bias-corrected satellite retrievals are employed as the observations to update the first guess. Cross validations illustrate improved quality of the blended analysis, with reduction in bias and random errors over most of the global oceans as compared to the individual inputs. Large uncertainty, however, remains in high-latitude oceans and coastal regions where the in situ networks are sparse and current-generation satellite retrievals have limitations. Our blended SSS analysis shows good agreements with the NODC in situ-based analysis over most of the tropical and subtropical oceans, but large differences are observed for high-latitude oceans and along coasts. In the tropical oceans, the BASS is shown to have coherent variability with precipitation and evaporation associated with the evolution of the El Niño-Southern Oscillation (ENSO).

1. Introduction

Observing oceanic salinity has been a challenging task. Traditionally, measurements of oceanic salinity have been performed using instruments mounted on in situ platforms, including moored buoys, profiling floats, and research vessels [Lagerloef et al., 2009]. These in situ salinity observations have been very sparse, temporally and spatially, until the launch of the Argo array in 2000 [Boutin and Martin, 2006; Gould and Turton, 2006]. The goal of Argo is to deploy and maintain 3000 floats to provide 10 day coverage at a 300 km spatial resolution for temperature and salinity measurements in the upper 2000 m for the ice-free open ocean. Since 2006, Argo has maintained more than 3000 floats, meeting spatial and temporal goals over much of the World’s ocean. Argo, coupled with near-real-time Conductivity-Temperature-Depth (CTD) profiles from research ships, the tropical moored buoy arrays, gliders, drifters, and other instrumentation has provided an unprecedented ability to monitor the salinity structure of the global oceans.

Despite an increase in salinity coverage provided by Argo and other in situ based sources, the spatial delineation of salinity fronts and some other small-scale features is not always achieved due to the nominal 300 km scale (optimal) of Argo measurements and the tendency of floats to concentrate in certain ocean areas and disperse from others. Temporally, the evolution of fronts and short time scale salinity variability events cannot be effectively tracked with the 10 day temporal sampling of Argo measurements.

The dependency of ocean surface microwave emissivity at certain frequencies on SSS provides an opportunity to observe SSS from a spaceborne platform. In recent years, two such satellite missions have been launched with the objective of exploiting this dependency to measure SSS. The European Space Agency’s (ESA) Soil Moisture-Ocean Salinity (SMOS) satellite was launched in November 2009 on a sun-synchronous...
orbit, scanning the earth at approximately 06/18 Local Standard Time (LST). SSS retrievals are generated by ESA and available on a near-real-time basis since January 2010 [Berger et al., 2002]. Aquarius/SAC-D, the joint U.S. National Aeronautics and Space Agency (NASA) and Comisión Nacional de Actividades Espaciales (CONAE, Space Agency of Argentina), was launched in June 2010 and began providing SSS retrievals over global oceans from August 2010 [Lagerloef et al., 2012].

While satellite SSS retrieval techniques are still improving, the most recent versions of SSS products from the SMOS and Aquarius missions exhibit the ability to capture the spatial patterns and temporal variations of ocean surface salinity. Bias and random error of relatively large magnitude, however, exist in the current versions of both SMOS and Aquarius satellite SSS retrievals due to the combined effects of imperfect algorithms and the less than desirable satellite sampling [Lagerloef et al., 2012].

In general, in situ sensors measure SSS with reasonable quantitative accuracy at their deployment sites but the in situ network density is insufficient to depict spatial variations in the SSS, particularly on fine spatial scales. In addition, sampling error exists in representing the grid box mean SSS using in situ measurements from a limited number of profiles. Satellite observations, on the other hand, are capable of depicting the spatial patterns of SSS, with random error introduced by relatively infrequent sampling and bias varying with the magnitude of the SSS. The complementary nature of the in situ measurements and the satellite observations for the SSS suggests potential improvements of the global SSS observations through blending information derived from both the in situ and space-based platforms. Similar efforts have been effective in constructing gridded fields of complete spatial coverage and improved quality for precipitation and sea surface temperature (SST) over global and regional domains [Reynolds, 1988; Reynolds and Smith, 1994; Xie and Arkin, 1996, 1997; Thiébaux et al., 2003; Wang and Xie, 2006; Reynolds et al., 2007; Xie and Xiong, 2011].

The objective of this work is to develop a prototype algorithm for constructing monthly SSS analyses on a $1^\circ \times 1^\circ$ latitude/longitude grid over the global ocean through blending in situ measurements and retrievals from both the SMOS and Aquarius satellites. The algorithm is designed under the assumptions that (a) the in situ measurements are spatially sparse but unbiased and (b) the satellite retrievals, when accumulated over a monthly period, provide quasi-complete spatial coverage but contain bias and random error of non-negligible magnitude. A two-step approach is adopted. First, the biases in the raw satellite retrievals are removed through calibration against the in situ measurements. The bias-corrected satellite retrievals are then combined with the in situ data through the optimal interpolation (OI) technique of Gandin [1965] to produce the blended analysis of SSS. The blended analysis is intended to represent the first ocean layer (0–5.25 m) bulk SSS as measured by in situ sensors.

This paper is composed of five sections. Section 2 provides a brief description of the in situ and satellite data used as inputs to the blended SSS analysis. Sections 3 and 4 document the development of the algorithms to remove satellite biases and to combine the bias-corrected satellite retrievals with the in situ data, respectively. Section 5 describes a comparison of our blended SSS analysis with the NODC in situ based analysis and a diagnostic study of the SSS-freshwater relationship using the new data set. Finally a summary concludes the paper.

2. Data

Three SSS data sets are used as inputs to the blending algorithm, the in situ SSS measurements aggregated and quality controlled by NOAA’s National Oceanographic Data Center (NODC); and the passive microwave (PMW) retrievals derived from the Aquarius/SAC-D and SMOS satellites.

NOAA/NODC collects and quality controls salinity observation data from various in situ sources, including Argo, TAO/TRITON/PIRATA/ RAMA moored arrays, and research vessels, and combines them into a single set of global salinity data. Quality control is performed for the raw salinity profiles at the source and this information is preserved and passed on to NODC. Further quality control is performed at NODC for the in situ salinity measurements following procedures outlined in Boyer and Levitus [1994] and Boyer et al. [2009].

Salinity data for the first oceanic layer (0–5.25 m) are extracted from the profile database. In situ measured SSS for each $1^\circ$ latitude/longitude grid box are computed as the arithmetic mean of the values from all profiles inside the grid box available during the target month. In addition to the monthly mean SSS, the
number of SSS profiles and the standard deviation among the measurements from different profiles are also calculated at NODC and used in this study to define the estimated sampling error for the in situ measured SSS.

The current in situ salinity measurement network covers most parts of the global ocean, from 60°S to 60°N, reasonably well (Figure 1, bottom). However, the in situ network density is relatively sparse, especially in the sub-tropics (e.g., South Pacific ~30°S, central North Pacific around 30°N). Overall, in recent years, only about 10% of the 1° latitude/longitude grid boxes of the global oceans are observed by one or more profiles over a monthly period (Figure 1, bottom). Analyses defined by interpolating these in situ measurements (Figure 1, top), therefore, are relatively poor in capturing salinity variations of fine spatial scales. To avoid confusion, in subsequent discussion we will refer to the in situ based grid box mean and the gridded analysis as the NODC SSS grid box mean and NODC SSS analysis, respectively.

Satellite SSS retrievals used in this study are from ESA’s SMOS mission and the U.S.-Argentina bilateral Aquarius/SAC-D mission. Launched in November 2009, the SMOS satellite observes the earth from a sun-synchronous orbit, passing over a geophysical location on the earth twice a day, at approximately 06LST (ascending node) and 18LST (descending node), respectively. Oceanic salinity information is derived from measurements of an L band (21 cm, 1.4 GHz) passive microwave (PMW) 2-D interferometric radiometer [Kerr et al., 2000]. Three sets of Level 2 retrievals of instantaneous SSS are derived from the PMW measurements using different oceanic surface roughness models [Reul et al., 2012a, 2012b]. Those based on model 1 are used in this study because preliminary comparisons against the NODC in situ measurements showed slightly better performance of the SMOS SSS retrievals based on surface roughness model 1 than those based on models 2 and 3. The SMOS data utilized as inputs to our blending procedures are the reprocessed Version 5.50 Level 2 retrievals, covering a nearly 4 year period from 12 January 2010 to December 2013.

The Aquarius/SAC-D satellite was launched into space in June 2011. Similar to the SMOS mission, the Aquarius/SAC-D measures the oceanic salinity by observing thermal emission from the surface at 1.413 GHz using passive microwave radiometers [Lagerloef et al., 2012]. The spacecraft is also on a sun-synchronous orbit, with its ascending/descending node at 06LST/18LST, respectively. The Aquarius SSS data used in this study are Version 3.0 Level 2 products, covering a 28 month period from 25 August 2011 to the end of 2013.

Preliminary tests showed that the current versions of the SMOS and Aquarius Level 3 gridded fields may be compromised by land contaminations and radio frequency interference (RFI) over coastal regions and by sea ice contaminations over high latitudes. Contamination by even a single satellite field of view (FOV) may cause error in the gridded monthly mean at a level comparable to the magnitude of monthly SSS anomalies (~0.2–0.3 psu). Therefore, Level 2 retrievals of instantaneous SSS at individual satellite observation FOV are collected and quality controlled for use in this study.

To this end, Level 2 retrievals at the raw satellite observation FOVs are binned into gridded fields of instantaneous SSS on a resolution of 1° latitude/longitude. A set of quality control (QC) procedures are developed and applied to remove the erroneous or suspicious SSS data. First, SSS retrievals over grid boxes of possible sea ice contaminations are eliminated. This is done by checking against the daily SST analysis of Reynolds et al. [2007] and the gridded fields of sea ice fraction included in the daily SST data set. Any SMOS/Aquarius Level 2 SSS retrievals over a grid box of 1° latitude/longitude with any sea ice concentration or covered by SST colder than 2°C are flagged and removed for possible contaminations.

The second QC step is then performed to eliminate retrievals with unrealistically large departures from their expected normal values. A map of median SSS is constructed for the SMOS and Aquarius retrievals, respectively, using the screened Level 2 data for the entire data periods. The SMOS and Aquarius retrievals are then compared to the median values at their respective geophysical locations. Retrieved SSS lower than the median by more than 2.0 psu are identified and eliminated from the inputs. Manual inspections revealed that most of these cases are caused by the RFI and land/sea ice contaminations not detected in the first QC step.

The quality controlled SMOS and Aquarius SSS retrievals are then integrated into gridded fields of monthly mean SSS on a 1° latitude/longitude grid over the global oceans. In addition to monthly mean SSS,
Figure 1. (top) NODC in situ based analysis (psu), (middle) NODC in situ grid box mean (psu) of the monthly SSS, and (bottom) the number of in situ profiles used to define the grid box mean for October 2011.
number of Level 2 retrievals at individual FOVs is also included to be used for the definition of random error statistics. As illustrated in Figures 2 and 3, SMOS and Aquarius retrievals provide estimates of monthly mean SSS with a quasi-complete spatial coverage over a broad latitude band from ~50°S to 60°N, presenting spatial patterns similar to those in the NODC in situ measurements and in situ based analysis (Figure 1). The number of sampling size for the SMOS SSS retrievals is larger than that for the Aquarius because the size of the FOVs for SMOS (approximately 50 km) is smaller than that for the Aquarius (94–156 km).

3. Bias Correction for the Satellite Retrievals

A comprehensive understanding of the performance of the individual input SSS data sets is the foundation to the successful development of the blending algorithm. An inter-comparison is first conducted between the SMOS/Aquarius satellite-based retrievals and the NODC in situ analysis of monthly mean SSS for a 28 month period from September 2011 to December 2013. As shown in Figure 4, SSS retrievals from both the SMOS and the Aquarius satellite observations agree very well in the large-scale spatial patterns, with maxima and minima SSS observed for the belts of subtropical subsidence and the Inter-Tropical Convergence Zone (ITCZ) freshening, respectively. Differences appear in high-latitude oceans and in coastal regions.

Figure 2. (top) Monthly mean SMOS SSS retrievals (psu) and (bottom) number of Level 2 orbital pixels used to define the monthly mean SSS for October 2011.
where decreased microwave measurement sensitivity and increased uncertainty in the current versions of
the satellite retrieval algorithms exist [Lagerloef et al., 2012].

Another noticeable difference between the in situ analysis and the satellite retrievals is in the SSS magni-
tude. It is clear from Figure 4 that the SMOS/Aquarius retrievals (left-top/left-bottom) have smaller/higher
SSS over the North Pacific and the Southern Oceans compared to the in situ data. An analysis of the proba-
bility density function (PDF) of the SSS magnitude for the three data sets shows that both the SMOS and
Aquarius SSS retrievals present a bias that appear as a function of the magnitude of SSS (Figure 5, top). For
example, the PDF differences between SMOS and the in situ data are relative large (small) with salinity in
range 34.5–35.5 psu (larger than 36 psu). Overall, Aquarius is closer to the in situ data than SMOS. The bias
observed in the satellite retrievals is attributable to multiple factors, including imperfect retrieval algorithms
for deriving skin SSS from the passive microwave observations, and the differences between the skin SSS
observed by satellites and the near-surface bulk SSS measured by in situ platforms.

This magnitude-dependent bias needs to be removed before the satellite SSS retrievals can be combined
with the in situ measurements. Following Wang and Xie [2006] and Xie and Xiong [2011], a technique was
developed to remove the bias in the satellite SSS retrievals through matching the cumulative probability
density functions (CPDF) of the satellite retrievals with the CPDF for the colocated in situ measurements. To
this end, CPDF tables are constructed for the retrievals from each satellite, for each 1° latitude/longitude

Figure 3. (top) Monthly mean Aquarius SSS retrievals (psu) and (bottom) number of Level 2 orbital pixels used to define the monthly
mean SSS for October 2011.
grid box, and for each time step (month) using colocated satellite and in situ data in a region centered on the target grid box and over a 5-month time period centered around the target month. The size of the domain is expanded until data pairs of sufficient number is secured to ensure the construction of stable CPDFs. In this study, the average size (radius) of data collection domain is nine 1° latitude/longitude grid boxes to collect 100 or more colocated in situ–satellite SSS data pairs.

The bias in the satellite SSS retrievals is then removed by matching the CPDF of the satellite SSS against that for the in situ measurements. As shown in Figure 4, the bias-corrected satellite SSS retrievals (right) present much improved agreements with the NODC in situ analysis compared to the raw data (left). The SSS magnitude of the bias-corrected satellite retrievals is very close to that of the NODC in situ analysis over the high-latitude oceans. In particular, displacement in the positions of the SSS maxima over the Atlantic Ocean in the raw SMOS retrievals has been corrected to match closely with the NODC in situ analysis.

Cross-validation tests are performed to quantitatively examine the performance of the bias-corrected satellite SSS retrievals. The in situ SSS measurements at 10% randomly selected grid boxes are withdrawn. Bias correction is then performed for the SMOS and Aquarius satellite SSS retrievals utilizing the CPDF matching technique described above using the in situ data at the remaining 90% grid boxes. This process is repeated 10 times so that the in situ data at each grid box is withdrawn once. The raw and bias-corrected satellite retrievals are then compared with the in situ measurements at the withdrawn grid boxes.

The cross-validation results show that the CPDF matching technique developed in this study is capable of removing SSS bias for both the SMOS and Aquarius satellite retrievals. Over the global domain, the bias is...
reduced from −0.267 (−0.073) psu to 0.014 (0.012) psu for the SMOS (Aquarius), respectively, while the correlation is improved from 0.902 (0.923) to 0.948 (0.967) (Table 1). More importantly, the magnitude-dependent bias in the satellite SSS retrievals (Figure 5, top) has been removed substantially, resulting in an almost perfect agreement in the PDFs with the in situ measurements (Figure 5, bottom). The CPDF matching technique developed in this work is capable of removing SSS bias for both the SMOS and Aquarius satellite retrievals. As described in section 2, the in situ measurements represent the bulk mean SSS for the upper-ocean layer (0–5.25 m) while the satellite retrievals are for the SSS at the subskin level. Applying the CPDF matching technique developed in this work, the satellite retrievals are calibrated toward the bulk SSS measured by the in situ platforms, removing the bias in the satellite retrievals in representing subskin SSS and the systematic differences between the subskin SSS and bulk SSS at once. Since the CPDF matching is performed using colocated in situ and satellite data over a spatial domain centered at the target grid box and over a 5 month period centering at the target month, this procedure removes the mean satellite retrieval bias and bulk-subskin differences averaged over the combined time/space domain. Any bulk-subskin SSS differences within the combined time/space domain, especially those over coastal regions where relatively large changes in surface stratification are observed by a sparse in situ network, may remain as random error in the analysis.

Following the procedures described above, bias in the raw SMOS and Aquarius satellite retrievals are removed through the PDF matching against colocated in situ measurements for their respective data periods. Gridded fields of monthly SSS anomaly are then computed for the SMOS and Aquarius retrievals, respectively, by subtracting the monthly climatology of Boyer and Levitus [1994] from the bias-corrected satellite retrievals. An Empirical Orthogonal Function (EOF) analysis is then implemented to the resulting satellite-based monthly SSS anomaly fields to identify erroneous satellite SSS anomaly data. Unrealistically large SSS departures caused by the RFI and land/sea ice contaminations not detected and removed in the QC procedures described in section 2 yield leading modes EOF patterns of extremely large spatial loadings.

Table 1. Cross-Validation Statistics for the Raw and Bias-Adjusted Monthly Mean Sea-Surface Salinity (SSS) from the SMOS and Aquarius Satellite Retrievals

<table>
<thead>
<tr>
<th>Statistics</th>
<th>SMOS Raw</th>
<th>SMOS Adjusted</th>
<th>Aquarius Raw</th>
<th>Aquarius Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias (psu)</td>
<td>−0.267</td>
<td>0.014</td>
<td>−0.073</td>
<td>0.012</td>
</tr>
<tr>
<td>RMS error (psu)</td>
<td>0.503</td>
<td>0.299</td>
<td>0.402</td>
<td>0.248</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.902</td>
<td>0.948</td>
<td>0.923</td>
<td>0.967</td>
</tr>
</tbody>
</table>

Figure 5. Probability density function (PDF, %) of monthly mean SSS (psu) at a 1° latitude/longitude grid box over the global ocean for the NODC in situ measurements (black), SMOS (red), and Aquarius (green) satellite retrievals. Results for the raw and bias-corrected satellite retrievals are plotted in the top and bottom plots, respectively.
over high-latitude regions and along the coasts. Grid boxes with a spatial loading of 4.0 or higher are identified as locations of bad satellite data and the anomaly data over the grid boxes are set as missing. These “cleaned-up” SSS anomaly fields for the SMOS and Aquarius retrievals are used as inputs to the OI-based blending procedures described in the following section (Figure 6, for example, for October 2011). Corresponding SSS anomaly over a grid box with spatial loading of 4.0 or greater differs with the EOF pattern but is usually larger than 2.0 psu. While this practice eliminates virtually all anomalous SSS data, it may also remove data of real signal over regions (e.g., river mouth) where SSS exhibit substantial changes. In this preliminary work, a rather strict approach is taken to ensure most wrong reports from either the in situ or satellite retrievals are removed. Further work needs to improve the quality control procedures to retain as much SSS information as possible from the raw inputs especially over the coastal regions.

4. Combining the Bias-Corrected Satellite Retrievals With In Situ Measurements

4.1. Basic Framework

The blending procedures described in this section aim to construct an analysis of monthly SSS anomaly on a 1° latitude/longitude grid over the global oceans. The bias-corrected SMOS and Aquarius satellite SSS retrievals are combined with the in situ measurements through the optimal interpolation (OI) technique of Gandin [1965] to produce an in situ-satellite merged analysis of monthly SSS with further improved quality. Under the OI framework, the final analysis value ($A_{k,t}$) at a target grid box ($k$) for time step ($t$) is defined through modifying the first guess ($F_{k,t}$) at the grid box using observations at and near the target grid box [Japan Meteorological Agency (JMA) Prediction Division, 1990]:

$$A_{k,t} = F_{k,t} + \sum_{i=1}^{n} W_i (O_{i,t} - F_{i,t})$$

where $O_{i,t}$, $F_{i,t}$, and $W_i$ are the observations, first guess, and weighting for time step ($t$) at grid box ($i$) where the observations are available, while $n$ is the number of observation grid boxes used in the interpolation. In our application here, the monthly mean analysis for the previous time step ($t-1$) is used as the first guess, while combined in situ measurements and bias-corrected SMOS/Aquarius retrievals, called super observations, for the target analysis time step ($t$) are defined and used as the observations to modify the first guess field.

Under the assumptions that there are, (a) no bias in the observations and the first guess; (b) no correlation between error in the first guess and error in the observations; and (c) no correlation between errors of the observations at two different grid boxes, the weighting coefficients $W_i$ in equation (1) can be calculated by solving the following linear equations:

$$\sum_{j=1}^{n} (\lambda_{ij} \phi_i + \lambda_{ij} \phi_j) W_j = \lambda_{ik} \phi_k$$

where $\lambda_{ij} \phi_i$ is the first guess/observation error correlation at two grid boxes $i$ and $j$, $\lambda_{ik}$ is the first guess error correlation between the target grid box $k$ and the observation grid box $i$, respectively. Under the assumption (c), $\lambda_{ij} \phi_j$ is set to 1 for $i = j$, and to 0 for $i \neq j$. $\lambda_{ik} = \frac{\sigma_k^2}{\sigma_i^2}$ is the ratio between the standard deviation of the observation error and that of the first guess error at grid box $i$.

Error variance $(E_{k,t})^2$ for the resulting analysis $A_{k,t}$ may be estimated as:

$$(E_{k,t})^2 = (\sigma_k^2)^2 \left(1 - \sum_{i=1}^{n} W_i \lambda_{ik} \phi_k\right)$$

Key to the development of an OI-based in situ-satellite combined algorithm is the quantification of error structures for the input first guess and the observations. Under the assumptions described above, three
error parameters need to be quantified to calculate the OI-based analysis. These include: (a) error estimation for the in situ SSS measurements ($\sigma^2$), (b) error estimation for the super observations ($\sigma^2$), and (c) error correlation for the first guess at two different grid boxes ($\omega_{ij}$).

Figure 6. Monthly SSS anomaly (psu) for October 2011, derived from (top) the NODC in situ measurements, as well as (middle) the SMOS and (bottom) Aquarius satellite retrievals.
4.2. Super Observations and Their Error Definition
Following the practices of Reynolds and Smith [1994] and Wang and Xie [2006], observation-based inputs from individual sources are combined into a single field before they are fed into the OI processing. Called super observation (\(O_{it}\)), the combined observation at a grid box \((i)\) for time step \((t)\), is defined as the weighted mean of the individual input observation data \((S_{itm})\) from all sources available at the grid box:

\[
O_{it} = \sum_{m=1}^{M} V_{im} \cdot S_{itm}
\]

Here, the individual inputs \((S_{itm})\) used to define the super observations at grid box \((i)\) include the NODC grid box means of in situ measurement, and the bias-corrected satellite retrievals from the SMOS and Aquarius. The weighting coefficient \((V_{im})\) at a grid box \((i)\) for individual input \((m)\) is defined through maximum likelihood estimation (MLE):

\[
V_{im} = \left(\frac{r_{im}}{\sum_{l=1}^{M} (r_{il})^{-2}}\right)^{-1/2}
\]

where \(r_{im}\) is the estimated error for individual observation \((m)\) at grid box \((i)\). Estimated error \((\sigma^2o_i)\) for the super observation \((O_i)\) at a grid box \((i)\) is given by:

\[
\sigma^2o_i = \left[\sum_{m=1}^{M} \left(\frac{\sigma^2o_{km}}{(\sigma^2r_{m})^{-2} \cdot n_{ref}/n_S}\right)^{-1/2}\right]^{-2}
\]

where \(M\) is the number of individual observations used to define the super observation \((M\) equals to three in this case). In this study, observation error for the bias-corrected satellite estimates is defined as a function of season, latitude, and sampling size through comparison with colocated NODC box mean of in situ measurements.

Following Xie and Xiong [2011], error variance \((\sigma^2r_m)^2\) for the bias-corrected satellite retrievals \((m)\) at grid box \((i)\) is assumed to be inversely proportional to the sampling size \((n_i)\):

\[
(\sigma^2r_m)^2 = (\sigma^2r_{m})^2 \cdot n_{ref}/n_S
\]

Here, \(n_{ref}\) is the average number of samples (number of satellite FOVs), while \((\sigma^2r_{m})^2\) is the error variance expected for the monthly mean SSS retrievals defined from \(n_{ref}\) individual satellite SSS FOVs, defined for each type of satellite retrievals as a function of season and latitude. This is done through comparison of the target bias-corrected satellite SSS retrievals against colocated NODC grid box means. Colocated satellite and in situ data over a 5 month period centering at the target calendar month and over a latitude band of 7° centering at the target latitude are used in the calculation of the reference error variance.

The error variance for the NODC grid box mean of the in situ SSS is also assumed to be inversely proportional to the sampling size \((n_b)\), i.e., the number of profiles used to define the grid box mean:

\[
(\sigma^2b)^2 = (\sigma^2r_{b})^2 / n_b
\]

where \((\sigma^2r_{b})^2\) is the error variance for the in situ measurement from a single profile to represent the mean SSS over a grid box of 1° latitude/longitude. As described in section 2, in addition to the grid box mean SSS and the number of profiles used to define the grid box mean, the NODC monthly in situ SSS data set also include information on the standard deviation of the in situ SSS measurements among all profiles used to define the grid box mean. This standard deviation is used here to define \((\sigma^2r_{b})^2\). To this end, mean difference
variance (the square of the standard deviation) is computed using the monthly in situ data for the 48 month period from January 2010 to December 2013 over all grid boxes where/when four or more profiles were used to define the grid box mean and the standard deviation. The resulting statistic ($\sigma_b = 0.158$) shows that average error for the measurement from a single profile to represent monthly mean SSS over a $1^\circ$ latitude/longitude grid is approximately 0.158 psu.

### 4.3. First Guess, Its Error, and Error Correlation

Anomaly fields SSS constructed through the OI system for the previous month are used in this study as the first guess to define the gridded analysis for the target month. Since oceanic systems evolve at a relatively slow pace, state of ocean for the last time step often is a good approximation of that for the current time step [Reynolds and Smith, 1994]. The total error for the anomaly SSS for the previous month to represent that for the current month ($\sigma_{i,t}$) are composed of two independent components, i.e., error in the analyzed anomaly SSS field ($\sigma_{i,t-1}^a$) for the previous time step ($t-1$), and error caused by assuming no changes in the anomaly SSS from the previous month to the target month, $\sigma_{i,t}^f$.

$$\left(\sigma_{i,t}^a\right)^2 = \left(\sigma_{i,t-1}^a\right)^2 + \left(\sigma_{i,t}^f\right)^2$$ (9)

The first component of the first guess error, the analysis error for the previous month, is defined by equation (3) as part of the OI analysis procedures. The second component, the tendency error, is estimated for each grid box as the standard deviation of the SSS tendency calculated using the NODC gridded monthly analysis for the 48 month period from January 2010 to December 2013. No seasonal variations in the tendency error are considered in this first round estimation.

Construction of the OI-based gridded analyses also requires the definition of correlation between the first guess error at grid boxes separated by various distance. In the two components of the first guess error described in the last paragraph, the analysis error at a grid box for the previous time step is random in nature and bears no correlation with that at a different location. The first guess error correlation associated with the tendency error is calculated using the tendency data derived from the NODC SSS analysis over the entire global oceans for the entire data period from January 2010 to December 2013. Due to the lack of sufficient data, the spatial error correlation is modeled only as an exponential function of the separation distance through a least square fitting. The results showed an e-folding distance of $\sim$600 km, indicating a correlation of approximately 0.4 for the first guess errors at two grid boxes of approximately 600 km apart.

Once the error statistics for the first guess and observations are defined as described above, the SSS analysis and the associated error estimations are then computed through equations (1) and (3) using the first guess and the super observations at grid boxes over a domain around the target grid box. This interpolation data domain is expanded isotropically until the super observation data are available over at least five grid boxes. Over the tropical oceans where most grid boxes are covered by in situ measurements and/or satellite observations, the interpolation data domain is often a circular region of three $1^\circ$ latitude/longitude grid boxes in diameters, ensuring the delineation of SSS variations of relatively small spatial scales.

Preliminary monthly SSS analyses are produced through the OI-based algorithm using error statistics defined above for the 48 month period from January 2010 to December 2013. The error statistics for the first guess field are refined using the preliminary SSS analyses. The first tendency error is now defined for each time step and for each oceanic grid box as the standard deviation of the SSS anomaly differences between the current and previous months computed using data over a domain of $7^\circ$ latitude by $7^\circ$ longitude centering at the target grid box. The first guess error correlation calculated using the preliminary SSS analyses yielded a shorter e-folding distance of approximately 400 km. The longer e-folding distance derived from the NODC gridded SSS analysis probably is caused by the fact that the analysis is defined by interpolation of in situ measurements over an extended area within a circle of 770 km radius.

### 4.4. The Blended SSS Analysis and Cross-Validation Tests

Following the procedures described above, the in situ-satellite blended analysis of monthly SSS is constructed on a $1^\circ$ latitude/longitude grid over the global ocean for a 48 month period from January 2010 to December 2013. As illustrated in Figure 7, the blended analysis has a quasi-complete coverage of monthly
SSS anomaly over the global oceans. Its spatial patterns are similar to those observed in the individual input data sources (Figure 6), with reduced noise observed in the satellite retrievals. In addition to the analyzed value of the monthly SSS anomaly, error estimation for the SSS analysis is also included as part of the output package (Figure 7, bottom). Overall, the estimated analysis error is smaller over tropical and subtropical oceans but higher over high latitudes, a reflection of the quality and sampling size of the satellite SSS retrievals used in this study as the inputs. In particular, estimated analysis error may increase to more than 1 psu over high-latitude oceans covered by cold SST and sea ice. At the meantime, the error is much smaller (0.15 psu or less) for grid boxes over tropical and subtropical oceans where in situ measurements are available.

To quantitatively examine the performance of the OI-based blending algorithm in producing the gridded analysis of SSS, cross-validation tests are executed following the same strategy applied for the bias correction procedures. In situ measurements at 10% randomly selected grid boxes are dropped. In situ measurements at the remaining 90% of the grid boxes are then blended with the bias-corrected satellite retrievals generated by the cross-validation experiments for the bias correction procedures described in section 3. The blended SSS analysis is then compared against the withdrawn in situ measurement at the 10% grid boxes.

Figure 8 shows the serial correlation between the monthly SSS anomaly derived from the cross-validation tests and that from the withdrawn in situ measurements for the 48 month data period. Only results over
grid boxes with colocated data for 10 or more months are plotted. Anomaly correlation is higher than 0.5 over most of the grid boxes over tropical and subtropical oceans. Table 2 displays bias, root-mean-square (RMS) error, and correlation for the blended analysis monthly SSS anomaly over the tropical (Table 2), and the entire global ocean (Table 2), respectively. Bias of the blended SSS analysis is very small, less than 0.005 psu, over the global ocean, indicating success of the bias correction in the previous processing step. Correlation for the blended SSS analysis is also improved compared to that for the individual input satellite retrievals. In particular, for tropics, the correlation/RMS error for blended SSS analysis is 0.704/0.203 psu, compared to 0.548/0.285 and 0.665/0.228 psu, respectively, for the bias-corrected SMOS and Aquarius retrievals. Relatively high correlation between the blended analysis and the in situ measurements may be also partially attributable to the fact that majority of the moored buoy observations have their shallowest measurements made at 1.0 m depth, much closer to the surface. The RMS error for the blended analysis should be slightly smaller than those shown in Table 2 for two reasons. The statistics are derived through comparisons against in situ measurements which also contain random error that are not excluded in calculating the RMS error for the blended analysis. In addition, the cross-validation experiments, by nature, only included the tests over grid boxes where in situ data are withheld and only inputs from satellite retrievals are used. The quality of the blended analysis should be better over grid boxes where data from both satellite and in situ platforms are available.

5. Comparisons and Applications

5.1. Comparison With the NODC In Situ-Based Analysis

Gridded monthly SSS anomaly fields derived from our blended algorithm are compared to those from the NODC in situ based analysis for a 48 month period from January 2010 to December 2013. The two sets of analyses present good overall agreements in both the large-scale distribution patterns and the overall anomaly magnitude over majority of the
tropical and subtropical oceans (see example for July 2012 shown in Figure 9). The NODC in situ analysis depicts a very smooth spatial distribution patterns, caused at least partially by the interpolation of in situ data from a relatively broad area (approximately 770 km in radius). The blended analysis, meanwhile, is defined using both in situ and satellite data collected over a domain of approximately three 1° latitude/longitude grid boxes in diameter, resulting in a much shorter spatial scale. While more details in the SSS spatial distribution are revealed in the blended analysis, without a high quality reference observation network of extensive coverage, it is not an easy task to demonstrate and quantify to what extent the blended analysis presents improved skills in capturing spatial variations of SSS of differing scales. Since the oceanic system is evolving over the combined time-space domain, one thing we may explore in the future is to compare the time series of the blended and the in situ based SSS analyses of various time resolutions against independent observations (e.g., withdrawn in situ measurements) over selected grid boxes and examine how the temporal variations of SSS are captured.

Figure 9. SSS anomaly fields for July 2012, derived from (top) the NOAA blended analysis of sea salinity (BASS), and (bottom) the NODC in situ based analysis.
Overall, close agreement is observed between the two sets of SSS analyses over most of the tropical and subtropical oceans (Figure 10). Anomaly correlation between the two analyses is higher than 0.5 over most grid boxes there, with the maximum correlation (greater than 0.8) observed over the tropical Indian Ocean and the tropical Pacific (Figure 10, top). The RMS differences between the two analyses are at a level of approximately 0.2 psu, much smaller than the magnitude of the signal as shown by the standard deviation of the blended analysis (Figure 10, bottom).

Figure 10. (top) Correlation and (middle) root mean square (RMS) differences (psu) between the monthly anomaly of the BASS and that of the NODC in situ-based analysis computed for a 48 month period from January 2010 to December 2013. (bottom) Standard deviation of the BASS monthly anomaly is also plotted (psu) as a proxy indicator of the SSS variability intensity.
Substantial differences between the two analyses are observed over high latitudes and in coastal regions (Figure 10, middle). For those areas, in situ measurements are extremely sparse and the satellite retrievals are either of nonexistence or of poor quality. Both the in situ based and the in situ-satellite blended analyses are therefore largely defined through extrapolation of the observations from distant locations. Quality of the analyses there is substantially degraded as shown in Figure 7. The differences between the two analyses over these regions are attributable to the different input data used and the different weighting coefficients applied in the analysis procedures. Since the in situ-based analysis itself presents limited quality over the coastal regions, the large differences there should not be regarded simply as the “error” in the blended analysis and thereby the satellite retrievals. Blended analysis showed clear freshwater paths from some major rivers (e.g., Amazon) that are missing in the in situ analysis. More work is needed to reduce the uncertainty in the blended analysis over coastal regions.

5.2. Application Examples

Oceanic freshwater flux plays an important role in determining the oceanic salinity and their variations [Delcroix and Picaut, 1998; Delcroix and McPhaden, 2002; Liu and Xie, 2008; Large and Yeager, 2009; Singh et al., 2011]. As an illustration of the applications of the newly developed blended SSS analysis, relationship between the anomaly of the monthly SSS and that of the freshwater flux is examined for the 4 year data period from 2010 to 2013. The precipitation data used in this study are from the satellite-based estimates defined using the CPC Morphing technique (CMORPH) [Joyce et al., 2004]. The evaporation data are those generated by OAFlux of Yu and Weller [2007] and Yu et al. [2008].

As shown in Figure 10 (bottom), variations of SSS are large in the tropical western Pacific, eastern Indian Oceans, tropical northwestern, and southeastern Atlantic Oceans, where intensive precipitation associated with deep convection is observed, implying a connection between SSS and precipitation. Correlation between the monthly anomalies of SSS and precipitation reaches −0.4 and stronger over most oceanic areas with active heavy precipitation associated with the ITCZ, the South Pacific Convergence Zone (SPCZ), and storm tracks (Figure 11, top). In particular, SSS-precipitation correlation is the strongest (approximately −0.8) over the equatorial western Pacific, suggesting the dominant influence of the freshwater in determining the SSS in this oceanic region of heavy precipitation. Evaporation exhibits much weaker correlation with the SSS anomaly, compared to that for the precipitation (Figure 11, middle). Anomaly correlation of 0.4 or stronger is observed only over part of the equatorial tropics and along the storm tracks where surface wind is strong. Between the two components of oceanic freshwater flux, precipitation plays more important role in determining the SSS variations over most of the oceanic areas. The relatively lower correlation between evaporation and salinity probably results from the evaporation being a slower process with smaller magnitude anomalies than precipitation, as well as precipitation reducing the density of near-surface water, which tends to remain at the surface, while evaporation increases near-surface density, leading to convective overturning that masks the process. Overall, the net freshwater flux (E-P) shows slightly better correlation with the SSS variations, with wider area covered by significant correlation especially over the tropical Pacific (Figure 11, bottom).

A combined EOF analysis is performed for the standardized anomaly fields of SSS, precipitation, and evaporation for the 48 month period from January 2010 to December 2013. Despite the relatively short time period, the first mode of the EOF reveals a physically coherent pattern of SSS and freshwater flux variations in association with the evolution of ENSO (Figure 12). A decrease in precipitation, coupled with enhanced evaporation, increases the SSS over the equatorial western Pacific and tropical Indian Ocean. In the equatorial Atlantic Ocean, excessive precipitation overwhelms increased evaporation, resulting in a net decrease in the SSS. Reduced evaporation over the southeast Pacific dry zone, however, does not yield freshened seawater, suggesting importance of other oceanic processes in determining the surface salinity over the region.

6. Summary and Conclusions

As a joint effort among NOAA Climate Prediction Center (CPC), NOAA/NESDIS National Oceanographic Data Center (NODC), and NOAA/NESDIS Center for Satellite Applications and Research (STAR), an analysis of monthly sea surface salinity (SSS) has been developed on a 1° latitude/longitude grid over the global ocean for a 4 year period from 2010 to 2013 by blending information from in situ measurements and satellite retrievals. The in situ measurements used as inputs to the blended analysis are the grid box mean of SSS...
produced by NOAA/NESDIS/NODC from quality controlled measurements from Argo and tropical moored buoys (TAO/TRITON, PIRATA, RAMA), CTDs, and gliders. The satellite data used in this study are the Level 2 products of SSS retrievals from both the Soil Moisture-Oceanic Salinity (SMOS) of European Space Agency (ESA) and the US-Argentina joint venture Aquarius missions.

A two-step algorithm is developed to blend the in situ and satellite data. First, the biases in the satellite retrievals are removed through PDF matching against temporally-spatially colocated in situ measurements. The blended analysis is then achieved through OI, where the analysis for the previous time step is used as the first guess while the in situ measurements and bias-corrected satellite retrievals are employed as the observations to update the first guess. Cross validations are conducted by randomly removing in situ SSS measurements and validating the blended analysis against the withheld in situ data. Results show improved accuracy of the blended analysis compared to the individual inputs, with reduction in bias and random errors over most of the global oceans. However, uncertainty of large magnitude in the blended analysis exists for high-salinity oceans and coastal regions where the in situ networks are sparse and current-generation satellite retrievals have limitations.
The blended monthly SSS analysis, called the NOAA "Blended Analysis of Surface Salinity" (BASS), is constructed for a 4 year period from 2010 to 2012. Our in situ-satellite blended SSS analysis shows good agreements with the NODC in situ-based analysis over most of the tropical and subtropical oceans, with large differences observed for high-latitude oceans and along coasts. Correlation analysis for monthly anomaly fields suggests the dominate role of the precipitation played in SSS variation over deep convection regions, such as the tropical western pacific. Combined empirical orthogonal function (EOF) analysis of the SSS from

Figure 12. (first row) Time series of the principal component (PC) and the spatial loadings of the (second row) SSS, (third row) precipitation, and (fourth row) evaporation variability associated with the first mode of the combined EOF performed for a 36 month period from January 2010 to December 2012. Monthly anomaly fields of SSS, precipitation, and evaporation data are, respectively, from the BASS, CMORPH, and OAFlux.
the BASS analysis, precipitation (P) from CMORPH, and evaporation (E) from OAFlux reveals global patterns of the covariability between the SSS and oceanic freshwater flux (E-P) in association with the evolution of the El Niño-Southern Oscillation (ENSO).

One great challenge in constructing the blended SSS analysis is the quality control of the satellite retrievals. Despite the extensive efforts by the satellite products developers, current generations of the satellite-based SSS retrievals from both the SMOS and Aquarius missions still present unrealistic values caused by radio frequency interference (RFI) and land/sea ice contaminations. In addition, uncertainties exist in the SSS retrievals over oceanic regions of high wind. As a result, random error in the monthly anomaly fields of the raw satellite retrievals is often of magnitude comparable to or even greater than that of the signals over coastal regions and high latitudes. A set of QC procedures were developed to identify and remove anomalous SSS retrievals through both cross checking among different sources of inputs and comparisons against climatology. Additional examinations, however, are needed to clean up retrieved SSS fields through EOF analysis of monthly anomaly fields and manual inspections of individual monthly fields.

Another area for future improvements is the inputs from the in situ measurements of sea surface salinity, especially over coastal regions where no in situ data are available from regular sources. A quick survey of all coastal moored buoy SSS measurements may provide insights on how we may better capture the temporal/spatial variations of SSS there. In addition, SSS measurements made by gliders, not included in the current version of the NOCD in situ data set, may enhance the coverage of boundary currents and other regions where Argo floats do not go or do not stay long.

The current version of the NOAA BASS sea-surface salinity analysis is constructed for a time space resolution of monthly and 1° latitude/longitude. Further work is underway to explore the feasibility and strategy for producing the blended SSS analysis at a refined resolution (e.g., 10 day), with analysis updates on a quasi real-time basis for monitoring applications.

Acknowledgments
This work is a joint effort among NOAA/NWS/CPC, NOAA/NESDIS/NODC, and NOAA/NESDIS/STAR. Research and development work at CPC is supported by the NOAA Climate Program Office (CPO) Climate Monitoring and Observation (CMO) program as a part of the "Climate Prediction Center Analyses and Monitoring in Support of the Ocean Observing System for Climate" project. Salinity data from the Soil Moisture-Ocean Salinity (SMOS) mission was provided courtesy of ESA in conjunction with ESA Category-1 Project 7118, "Applications of SMOS sea-surface salinity data to improved operational modeling." Salinity data from the Aquarius Mission were provided courtesy of NASA and the Jet Propulsion Laboratory (JPL) Physical Oceanography Distributed Active Archive Center (PO.DAAC). Comments and suggestions by Zeng-Zhen Hu and Calhong Wen of NOAA Climate Prediction Center and two anonymous reviewers helped improve the quality of our work and the manuscript.

References
Lagerloef, G., et al. (2009), Resolving the global surface salinity field and variations by blending satellite and in situ observations, paper presented at OceanObs’09 Conference, Venice, Italy, 21–25 Sep.