Improved global net surface heat flux

James A. Carton, Gennady A. Chepurin, Ligan Chen, and Semyon A. Grodsky

Department of Atmospheric and Oceanic Science
University of Maryland
College Park, MD 20742

Corresponding Author: James Carton (carton@atmos.umd.edu)

Key Points:

- Current atmospheric reanalysis heat flux estimates differ by 10-30 Wm^{-2}.
- We identify the errors and reduce them to $< \pm 5$ Wm^{-2} at most locations.
- The Southern Ocean is shown to be a source of heat to the atmosphere.

Abstract

Surface heat flux estimates from widely used atmospheric reanalyses differ locally by 10-30 Wm^{-2} even in time mean. Here a method is presented to help identify the errors causing these differences and to reduce these errors by exploiting hydrographic observations and the resulting temperature increments produced by an ocean reanalysis. The method is applied to improve the climatological monthly net surface heat fluxes from three atmospheric reanalyses: MERRA-2, ERA-Interim, and JRA-55, during an eight year test period 2007-2014. The results show that the time mean error, as evaluated by consistency with the ocean heat budget, is reduced to less than ±5 Wm^{-2} over much of the subtropical and midlatitude ocean. For the global ocean, after all the corrections have been made the eight-year mean global net surface heat imbalance has been reduced to 3.4 Wm^{-2}. A method is also presented to quantify the uncertainty in the heat flux estimates by repeating the procedure with many different atmospheric reanalyses and then examining the resulting spread in estimates. This reevaluation of net surface flux reveals, among other results, that the Southern Ocean is a source of heat to the atmosphere.

1 Introduction

Heat exchange between the atmosphere and ocean is the primary way the ocean gains or loses heat. Even though the seasonal exchange in some regions may be hundreds of Watts per meter squared (Wm^{-2}), a multi-year imbalance of only a few Wm^{-2} is sufficient to explain decadal mixed layer temperature variability as well as the secular trend of ocean warming. Yet while planning documents call for reducing uncertainties in seasonal net surface heat flux estimates to below ±5Wm^{-2} (Taylor et al., 2000; Curry et al., 2004), comparisons among current estimates show differences remaining stubbornly in the range of 10-30Wm^{-2} (Yu and Weller 2007; Smith et al., 2010; Cerovecki, et al., 2011; Brunke et al., 2011). Here we describe a method to ingest the recent rich set of global ocean hydrographic observations along with atmospheric reanalysis estimates in a sequential ocean data assimilation reanalysis to improve estimates of seasonal net surface heat flux and to quantify the remaining error. We demonstrate this approach with a series of experiments during the eight year period 2007-2014.

Annually averaged, net surface heat flux varies from a maximum of $+150Wm^{-2}$ flowing into the cool eastern equatorial Pacific and Atlantic to a maximum outflow from the north subtropical western boundary current regions of the Kuroshio and Gulf Stream (**Fig. 1**). Along the equator and to a lesser extent over the cool subtropical coastal upwelling zones the high value of net downward flux is due to high rates of shortwave radiation reaching the surface ocean. At low latitude the net cooling effect of outgoing long wave radiation is reduced due to elevated atmospheric humidity and the presence of warm low level clouds.

The most widely used estimates of surface heat flux are produced by atmospheric reanalyses such as the NASA Modern-Era Retrospective Analysis For Research And Applications version 2 (MERRA-2), the European Center for Medium Range Weather Forecasts (ECMWF) atmospheric reanalysis interim (ERA-Interim), and the Japan Meteorological Agency 55 year reanalysis (JMA-55). These reanalyses compute the turbulent components of net surface heat flux using bulk formulae unique to each reanalysis driven by instantaneous state variables at the lowest model level. The downwelling part of the radiative components are determined by the full column radiative properties of the atmosphere while the upwelling components of radiation are determined by specified surface albedo, surface emissivity, and SST.

Because reanalysis surface heat flux is a residual of atmospheric reanalysis rather than being directly observed, estimates from different atmospheric reanalysis products easily differ by tens of Wm^{-2}, obscuring potentially interesting climate signals. Seasonal averages of MERRA-2, ERA-Interim, and JRA-55 net surface heat flux products show differences of ±20 W/m^2, in line with the spread of estimates shown in previous intercomparisons (individual components and net flux are shown in *Supporting Information* **Fig. S1**). In contrast most ocean reanalyses compute turbulent and upwelling radiative components of net surface heat flux by applying a bulk formula to atmospheric variables such as winds, air temperature, and specific humidity at fixed elevation (we'll refer to this as bulk flux forcing) (Large and Yeager, 2009; Valdivieso et al., 2017). The two types of flux estimates can also produce values differing by tens of W/m^2 even when based on the same reanalysis (*Supporting Information* **Fig. S3**).

One approach to providing an absolute error estimate is to compare the reanalysis fluxes to fluxes calculated at various ocean mooring sites. For example, one of the most carefully calibrated flux time series, those from the STRATUS mooring in the subtropical southeastern Pacific (20°S, 85°W), has a 2004-2009 mean net surface heating rate of +32 W/m^2, with a measurement error which is the equivalent of just a few W/m^2 and an additional larger, but rarely quantified error associated with the use of a bulk formula (Zeng and Dickinson, 1998; Holte et al., 2014). At this location, of the three reanalyses we consider MERRA-2 has the lowest net surface heating rate of +3 W/m^2, ERA-Interim has a net surface heating rate of +19 W/m^2, while JRA-55 has the highest net surface heating rate of +28 W/m^2. In another example, at the subtropical northwest (32°N, 145°E) Kuroshio Outflow Experiment (KOE) mooring the 2004-2014 mean net surface rate of cooling is -57W/m^2 (Zhang et al. 2016). At this location MERRA-

2 has the lowest mean net surface cooling rate of the three of -74 W/m^2 while JRA55 is losing heat at a rate of -115 W/m^2. In contrast, at PAPA in the Alaskan gyre (50°N, 145°W) all three multi-year reanalysis estimates land within the year-to-year variations in observed net surface bulk heat flux estimates (whose average is ~ +26 W/m^2). Such comparisons are sometimes used to quantify the reanalysis heat flux error. Further comparisons are provided in *Supplementary Information*. However, the use of such comparisons to verify the reanalyses is likely circular reasoning since observations from many mooring sites and their service ships have been previously assimilated into the reanalysis data sets. As an alternative in this paper we use the compatibility of the heat flux with the ocean heat budget to quantify our flux error thus bringing the full oceanographic data set to bear on flux evaluation.

Systematic errors in reanalysis surface fluxes are related to systematic biases in reanalysis surface radiation and surface variables entering the bulk formulas. Recognition of this has led to efforts to de-bias reanalysis variables a posteriori (e.g. Dussin et al. 2016). Later we will present results of an experiment with such a bias-corrected forcing set.

Ocean-based corrections to net surface heat flux estimates have been derived from ocean previously, either calculated directly (Grist and Josey, 2003), or derived from observation-constrained numerical solutions to the ocean heat budget (Stammer et al., 2004). Unfortunately, prior to the 2000's the ocean observation set was limited, handicapping these efforts. More recent efforts such as that of Cerovecki et al. (2011) and Forget et al. (2015) have been carried out in the context of 4DVar. The calculation of net surface heat flux becomes much easier if averaged over the whole ocean so that advective effects are eliminated and net flux equals ocean heat storage. von Schuckmann and LeTroan (2011) and Levitus et al. (2012) provide objective analysis-based global average warming rates for spans of years since 2000 of > 1/2 Wm^{-2} while Trenberth et al. (2016) report an ocean reanalysis-based estimate for years 2005-2013 of 0.8 Wm^{-2}. These warming rates are widely viewed as reflecting the alteration of atmospheric radiative properties due to the increase in greenhouse gases. The uncertainties reported with these, which are based on assumptions of unbiased Gaussian statistics, are all approximately ±0.2 Wm^{-2}.

In the current study our goal is to improve atmospheric reanalysis estimates of the spatially varying climatological seasonal net surface heat flux. Our approach follows the theme of Stammer et al. (2004) and Cerovecki et al. (2011) in using ocean data assimilation to constrain the meteorological estimates. However, instead of using 4DVar we use the more common sequential data assimilation, in which temperature and salinity increments are calculated for each assimilation cycle. The misfit between the meteorological forcing and the ocean observations is reflected in these increments and so we examine their statistics to infer errors in surface heat flux. The computational efficiency of our data assimilation algorithm allows us to present multiple experiments in the global domain using a full eddy-permitting ocean general circulation model.

2 Data and Methods

The experiments described here are carried out with the global Simple Ocean Data Assimilation v3 (SODA3) ocean reanalysis system for the period 2007-2014. The ocean/sea ice component models are built on Geophysical Fluid Dynamics Laboratory MOM5/SIS1 numerics with a quasi-isotropic $28km \times \cos(lat)$ eddy-permitting tripolar grid (Delworth et al., 2012). The model has 50 vertical levels with 10m resolution near-surface. This model uses a third order advection scheme on an Arakawa B-grid with no explicit horizontal diffusion and a scale-selective Smagorinsky viscosity, enhanced in the region of the western boundary currents. In addition, enhanced nearshore tidal mixing is parameterized. Vertical mixing is given by the K-profile scheme of Large et al. (1994). Climatological monthly river discharge is provided by Dai et al. (2009). The initial conditions in January, 2007, common to all experiments, are provided by a preliminary one-year long integration beginning from World Ocean Atlas 2013 potential temperature and salinity (Boyer et al., 2013).

The SODA3 multivariate sequential data assimilation filter operates on a 10 day update cycle using the incremental analysis updates digital filter of Bloom et al. (1996) and pre-specified error covariances following Carton and Giese (2008). Each 10 days the filter ingests all available temperature and salinity profile observations from the World Ocean Database (Locarnini et al., 2013; Zweng et al., 2013) updated as of fall, 2016. During our eight year target period 2007-2014 this profile data set consists of 2.7×10^6 fairly uniformly distributed profiles, of which 32% are from the ARGO system. The uncertainty for ARGO temperature and pressure observations is likely dominated by representativeness error rather than the small ±0.002K and ±2.4dbar measurement error. In situ and satellite SST observations (Xu and Ignatov, 2010) are also assimilated. The gridded increments of temperature ($\delta\theta = \theta^a - \theta^f$) and salinity ($\delta S = S^a - S^f$) are computed each 10 days.

Atmospheric reanalyses make available not only near-surface variables at fixed height, but also calculated surface momentum, mass, and heat fluxes. The reanalyses calculate the latter from instantaneous model state variables on the lowest model level, whose height varies in time and is different for each model. Each reanalysis uses their own flux parameterizations and releases fluxes averaged over different forecast intervals (Smith et al., 2010; Brunke et al., 2011; Brodeau et al., 2017). Here we present experiments using two different ways of specifying surface heat flux. For most experiments we use bulk flux forcing based on the same set of generally daily near-surface variables ($\mathbf{U}_{10m}$, T_{2m}, q_{2m}, SLP, and radiative terms) obtained for each of three reanalyses: MERRA-2 (Gelero et al., 2017), ERA-Interim (Dee et al., 2011), and JRA-55 (Kobayashi et al., 2015) and apply the same Coupled Ocean Atmosphere Response Experiment version 4.0 (COARE4) (Edson et al., 2013 describes version 3.5) flux parameterizations to each set of atmospheric state variables. To explore the sensitivity of the results to the choice of bulk formula we describe repeating one set of experiments using an alternative parameterization of Large and Yeager (2009) in *Supporting Information* **Fig. S4**.

For MERRA-2 and ERA-Interim we carried out experiments for the full eight year period 2007-2014. For JRA-55 we were only able to obtain seven years of data spanning 2007-2013. As a result of some preliminary experiments it was realized that the MERRA-2 ten meter height winds are weaker than the corresponding ERA-Interim, and JRA-55 winds (*Supporting Information* **Fig. S4**). This weakness is compensated for by MERRA-2 having a surface stress formula that has a larger drag coefficient than COARE4 for moderate winds. To bring the stresses we compute from MERRA-2 winds closer to the released version of ERA-Interim stress we have increased MERRA-2 $\mathbf{U}_{10m}$ by 10%. We recognize that this is an overly simplistic, minimalist, correction which deserves a justification more rigorous than that provided in the Supporting Information, although we also note that the more rigorous efforts of Large and Yeager (2009) and Dussin et al. (2016) both end up applying a similar inflation factor to surface winds.

For each reanalysis surface net shortwave radiation is calculated from daily average downwelling radiation assuming a surface albedo of 6% over the open ocean, increasing to between 40-80% over sea ice (depending on the presence of snow or melt). Surface upwelling longwave radiation is calculated from the Stefan–Boltzmann law assuming a surface emissivity of 1.0. For each experiment a full suite of state and derived variables is saved on the original model grid at 5dy intervals and then remapped onto a uniform 1/2°×1/2° horizontal grid. Derived variables include the depth of the mixed layer calculated using a 0.003σ criterion from the five-day files. Most of the results presented here were computed using the remapped data set.

The time average net bulk heat fluxes for the three reanalyses during our period of interest are displayed in **Fig. 1,** left-hand panels. Averaged 70°S-60°N all three atmospheric reanalyses have net rates of excess global mean ocean heating, with JRA-55 showing a surplus of nearly 20 Wm^{-2}. As also shown in *Supporting Information* **Fig. S2** bulk heat flux differs significantly from the heat flux provided by the reanalysis centers. In experiment soda3.3.3a we compute the climatological seasonal difference between the MERRA-2 bulk heat flux for soda3.3.2a and the MERRA-2 reanalysis net surface flux 2007-2014. We then compute the mean, annual and semiannual harmonics of the difference between MERRA-2 bulk flux and reanalysis flux, add these harmonics as an additional forcing term to the MERRA-2 bulk flux forcing and then repeat the experiment with this augmented bulk heat flux forcing (experiment soda3.3.3a). This two-step procedure (first soda3.3.2a, then soda3.3.3a) has the property that the climatological seasonal augmented surface heat flux is within ±5W/m^2 of the reanalysis heat flux (**Fig. 2**, lower lefthand panel shows the difference), but at the ocean reanalysis retains feedbacks in which changes in SST or surface currents can alter surface fluxes.

To explore the error in net surface heat flux (as represented by the ocean heat budget imbalance) we construct a perturbation forecast model as is done in incremental 4DVar (e.g. Lorenc, 2003). We begin with the vertically integrated heat conservation equation (Moison and Niiler, 1998):

$$\rho C_p h \overline{\frac{D\theta^f}{Dt}} = Q^f \qquad (0.0)$$

where θ^f is the model forecast potential temperature, Q^f is the specified atmospheric reanalysis heat flux, ρ is water density, C_p is specific heat at constant pressure, and the overbar represents an average over depth h. For simplicity we have neglected diffusive processes. Subtracting (0.0) from the analysis heat conservation equation gives an equation for the analyzed surface heat flux the ocean heat budget expects, Q^a, as a function of Q^f :

$$Q^a = Q^f + \rho C_p h \left[\overline{\frac{D\theta^a}{Dt}} - \overline{\frac{D\theta^f}{Dt}} \right] \qquad (0.0)$$

The right-hand side is nonlinear as it includes components such as the advection of temperature increments by velocity increments. In incremental 4DVar these terms would be approximated by a Taylor Series expansion. Here we exploit the slow advective and seasonal mixed layer entrainment timescales of the ocean relative to the length of an assimilation cycle to carry out even greater simplification.

The length of a forecast is the length of a data assimilation cycle, $\Delta t = 10$ days. Over this time interval the ratio of the advective terms to local storage is: $O(U\Delta t/\Delta x)$ where Δx is typically the horizontal scale of the temperature increments. For $U \sim 0.1$ m/s and a horizontal scale of a few hundred kilometers or greater the right-hand side of (0.0) is dominated by local storage over depth h. We choose h to be the depth of the mixed layer, evaluated every 5 days (we require h to lie in a depth range 20m < h < 200m) on the assumption that during a single forecast cycle any excess heat will not have time to penetrate below the mixed layer. Thus we have an approximate relation for the analyzed flux:

$$Q^a \cong Q^f + \frac{\rho C_p h}{\Delta t}\overline{\delta\theta} \qquad (0.0)$$

Equation (0.0) provides a simple formula for improving atmospheric reanalysis surface heat flux Q^f based on the imbalance of the ocean mixed layer heat budget and the constraints imposed by the archive of ocean observations (including the hydrographic sections). Here we limit our correction of net surface heat flux to the climatological seasonal components of Q.

One of the terms neglected in deriving (0.0) is the heat stored in the melting-freezing cycle of sea ice, which in turn is reflected in increments of ice thickness δh_{ice}. In ice-covered regions we must add a term of the form $\rho L \delta h_{ice} / \Delta t$ to account for seasonal heat storage in sea ice, which in the Arctic is similar in size to heat storage in the liquid ocean (Serreze et al., 2007; Carton et al., 2015). To avoid this additional complication we have masked out ice-covered regions in this study.

3 Results

We begin by presenting the three seven- to eight-year long reanalysis experiments forced by MERRA-2, ERA-Interim, and JRA-55 bulk flux forcing, and then discuss three additional experiments in which these surface heat and freshwater fluxes are improved based on (1.3) and the equivalent relation for net freshwater flux (**Table 1**). In the numbering system used to identify the experiments the first digit refers to the version of SODA (version 3), the second identifies the surface forcing, and the third refers to the bulk formulae used to calculate surface fluxes ('1' refers to Large and Yeager, 2009, while '2' refers to COARE4). The experiments described below (except in *Section 3.3*) use the COARE4 formulae.

3.1 Initial experiments

The time mean mixed layer temperature increments identify a deficit or surplus of downwelling heat flux that the observations are acting to correct. For example, positive increments of mixed layer temperature mean that there is a positive deficit of net surface heat flux. The three experiments show different patterns of deficit/surplus (**Fig. 3,** left-hand column). soda3.3.2a, driven by MERRA-2 bulk fluxes, shows a positive deficit of up to 15 Wm^{-2} in western tropical Pacific extending into the Indian Ocean and the subtropical Atlantic. At higher latitudes and in the eastern ocean the increments become negative indicating a surplus. In contrast, soda3.4.2a, driven by ERA-Interim bulk fluxes, has a distinctly positive deficit of 10-20 Wm^{-2} in the subtropics, but only a weak deficit close to the equator. In the eastern ocean and inshore of the north wall of the Gulf Stream all three have a surplus (negative deficit) of as much as 20 Wm^{-2} or more. JRA55 has a large 20 Wm^{-2} surplus almost everywhere except in areas in the western subtropical Pacific and Indian Oceans.

A comparison of the three reanalysis net versus bulk surface heat fluxes in *Supporting Information* (**Fig. S3**) shows that the bulk surface heat flux used to force soda3.3.2a is larger in time mean than the reanalysis flux used to drive soda3.3.3a by up to $25W/m^2$, with the largest differences occurring in the western subtropics (compare **Fig. 2** lefthand panels). When reanalysis flux forcing is used, instead of bulk flux forcing, the deficits generally shift to be more positive (**Fig. 2** righthand panels) but do not disappear.

By averaging the increments through the depth of the mixed layer we obscure some compensating vertical structures. We illustrate the presence of these structures for a region of the subtropical North Pacific (15°N-35°N, 150°E-130°W) in **Fig. 4** and the subtropical South Indian Ocean (40°S-20°S, 40°E-80°E) in **Fig. 5**. The time mean vertical profile of the temperature increments for these two regions are shown in **Fig. 6** (black curves).

In the North Pacific our subtropical region spans zones where the mixed layer increments of soda3.3.2a and soda3.7.2a are both positive and negative, while soda3.4.2a is only positive (**Fig. 3**). In the South Indian soda3.3.2a is near-neutral, soda3.4.2a is positive and soda3.7.2a is negative. These differences are reflected in differences in the sign of the temperature increments within the upper 50 m. At these depths the soda3.4.2a increments are positive and in the range of 0.01 to 0.05K/10dy while soda3.3.2 and soda3.7.2a are negative by a similar amount.

Despite the differences in the structures of time mean mixed layer temperature increments all three experiments (soda3.3.2a, soda3.4.2a, soda3.7.2a) have positive increments (up to 0.05 K/10dy) at the base of the spring and summer thermocline (approximately the depth of the 22°C contour in the North Pacific and the 20°C contour in the South Indian) (**Figs. 4, 5**). This pattern of positive summer, negative winter increments means the observations are strengthening the stratification of the summer mixed layer and weakening the stratification of the winter mixed layer, which we expect means the data assimilation is accelerating winter deepening. One possible explanation for why this same effect shows up in different regions and for different forcings is that we are seeing the impact on mixed layer properties of an ocean model deficiency.

The mixed layer temperature increments also show some interannual variability. For example, the temperature increments for both soda3.3.2a and soda.3.7.2a are strongly negative both within and below the mixed layer in North Pacific during the winter of 2009-10 (**Fig. 4**). An examination of the historical meteorology shows that the central North Pacific was unusually cold during that winter. Thus the negative temperature increments within and below the mixed layer may indicate that MERRA-2 and JRA-55 have produced insufficient surface cooling during that unusual winter.

Finally we note the presence of temperature increments that are apparent at depths well below the mixed layer. At 300m depth the time and geographic mean increments are smaller than $\pm 0.005\,K/10dy$. An increment this small even if it spanned a column 500m deep would only represent a storage error of 10 Wm^{-2}. The cause of these weak subthermocline temperature increments is still unclear.

3.1.1 ERA-Interim

In our previous discussion we identified several features of the time mean mixed layer temperature increments that are common among the three experiments, including a heat surplus in the eastern Equatorial Pacific and Atlantic and inshore of the north wall of the Gulf Stream. To look at those regions in more detail we focus on one experiment: soda3.4.2, driven by ERA-Interim forcing. A histogram of the temperature increments in the whole latitude band 60°S-60°N in **Fig. 7** shows a skewed distribution with a most likely imbalance of 10 Wm^{-2}, and with many regions where the imbalance is ±20 Wm^{-2} or greater.

Along the Pacific equator the vertical profile with longitude shows increments concentrated at thermocline depths (approximately the depth of the 20°C isotherm) (**Fig. 8** left-hand panel). The increments penetrate into the mixed layer only in the east. Thus the heat deficit in the eastern equatorial Pacific in **Fig. 8** is a manifestation of a dipole pattern of temperature increments that is concentrated within the thermocline. Such a pattern would result from a systematic error in the zonal tilt tendency of the equatorial thermocline. For a mean thermocline vertical stratification of 0.1 K/m, we estimate the tilt tendency to be 4m/10dy too deep in the east and 4 m/10dy too shallow in the west. Such a tilt tendency error represents development of a roughly 4%

reduction in the 200m mean west to east tilt of the equatorial thermocline during our 10 day assimilation cycle. Since the time mean zonal tilt of the thermocline is controlled by the time mean zonal wind stress, this in turn suggests that the zonal stress used in soda3.4.2 is slightly weaker along the equator than is compatible with the ocean observations, a result consistent with a direct comparison to scatterometer winds (Dussin et al., 2016).

We next consider a section of the eastern Pacific along 9°N. At that latitude a recent examination of ERA-Interim fluxes by Josey et al. (2014) identified a pattern of depressed specific humidity over the locations of the TAO/Triton moorings, which have been placed at a regular spacing of 15° longitude (170°W, 155°W, 140°W, 125°W, 110°W, 95°W). When used to drive an ocean model Josey et al. point to ocean circulation errors corresponding to errors in the specification of turbulent surface heat flux.

Here we explore what additional information we can extract from the ocean temperature increments regarding the impact of the TAO/Triton moorings on the ERA-Interim fluxes. We find large ±0.1 to ±0.2 K/10dy fluctuations associated with the locations of the moorings but concentrated at the depth of the thermocline (**Fig. 9** lower panel). When integrated over the upper 100m the heat storage increments are up to $60 Wm^{-2}$ (**Fig. 9** upper panel). Because they are large at thermocline depths and only weak near-surface we think it unlikely that they are caused by errors in surface heat flux (as suggested by Josey et al.) and much more likely that they are caused by the erroneous Ekman pumping. The presence of the latter is evident in the wind stress curl variations that also occur every 15° longitude, shown in **Fig. 9** (middle panel).

The final problem we consider is the band of mixed layer heat surplus that lies inshore of the north wall of the Gulf Stream in all three experiments. A detailed picture of the North Atlantic mixed layer heat deficit/surplus is shown in **Fig. 10** (upper left). The position of the Gulf Stream current lies just offshore of the position of maximum SST gradient (the grey lines in **Fig. 10**), and thus the Gulf Stream lies exactly between the inshore region of mixed layer heat surplus and the offshore region of mixed layer heat deficit. Interestingly, this pattern follows the shape of the continental shelf slope, curving around its eastward extension at 47°N, 45°W (a topographic feature known as Flemish Cap). This jog in the direction of the Gulf Stream is evident in the isolines of observed SST and also in the paths taken by surface drifters (e.g. McClean et al., 2002). Thus the temperature increments are acting to improve the path of the Gulf Stream, causing it to follow the shelf slope instead of heading directly eastward.

To explore the origin of the dipole pattern in mixed layer temperature increments crossing the Gulf Stream we next examine a vertical cross-section of the mean forecast temperature increments along a representative meridional section (65°W) (**Fig. 10** lower left). In this section the Gulf Stream has a mean latitude of 38°N, while the depth of the permanent thermocline is approximately indicated by the depth of the 10°C isotherm. Again the temperature increments are concentrated along the sloping thermocline and their sign is such as to indicate that the forecast model is acting to reduce that thermocline slope, in effect causing the north wall of the Gulf Stream to slump and thus the geostrophically related strength of the Gulf Stream current to weaken. For thermocline stratification of 0.02K/m the temperature increments indicate the inshore side of the north wall would sink at a rate of 10 m/10dy if not uplifted by the temperature corrections made by the data assimilation.

3.2 Flux correction

Based on the discussion above we propose to use (0.0) and the equivalent for freshwater to improve our estimate of climatological monthly net bulk heat flux. To test this proposition we augment the three sets of daily heat fluxes (with a ±50 Wm^{-2} limit on the size of the correction) and then repeat the three experiments with the improved fluxes (soda3.3.2, soda3.4.2, and soda3.7.2; **Table 1**). We then repeat the calculation of mixed layer heat flux deficit for the second set of experiments.

When averaged 70°S-60°N the net heating of the ocean prior to correction ranged from 5.2 Wm^{-2} (ERA-Interim) to 19.5 Wm^{-2} (JRA-55) (**Fig. 1** left-hand panels). After correction, the range of global heating estimates is reduced to between 1.5 Wm^{-2} (MERRA-2) and 7.1 Wm^{-2}(**Fig. 1** right-hand panels) If we go one step further and account for the mixed layer heat flux deficit associated with the second set of experiments (**Fig. 3** right-hand panels) we bring the range of heating estimates 70°S-60°N to between 1.0 Wm^{-2} (JRA-55) and 6.3 Wm^{-2} (ERA-Interim), which begins to approach the < 1 Wm^{-2} level of anthropogenic heating that we expect to be present.

Prior to flux correction the three time mean net bulk heat fluxes initially differed by 10-30 Wm^{-2} regionally. After correction the spatial maps of net surface heat flux come to resemble each other much more closely (**Fig. 1** compare left-hand and right-hand columns). Where the initial subtropical and midlatitude heat flux deficits were ±10-20 Wm^{-2} or more the deficits have been reduced, generally, below our target of ±5 Wm^{-2}. The spatial histogram of mean mixed layer heat flux deficits 60°S-60°N for the two experiments using ERA-Interim forcing shows the average heat flux deficit in a 1°x1° box has been reduced to -2.5 to +7.5 Wm^{-2} (**Fig. 7**).

In the initial set of experiments the time mean mixed layer temperature increments were surface trapped – largest in the mixed layer. We find that augmenting the surface fluxes using (0.0) mainly impacts the size of the temperature increments within the mixed layer and reduces these to a size similar to what we find at thermocline depths (**Fig. 6** compare black and red lines). Below the thermocline, at depths of 300m the temperature increments remain small.

In *Section 3.1* we examined the seasonal increments in the subtropical North Pacific and South Indian Oceans for the initial experiments (**Figs. 4** and **5**). In **Fig. 11** we show the corresponding figures for one of the experiments, soda3.4.2, after correction. Improving heat flux does reduce temperature increments within the upper 50 m, but does not alter the pattern of positive and negative increments at the base of the mixed layer. The lack of impact of changing surface heat flux on the temperature increments at the base of the mixed layer is consistent with the suggestion that those errors result from errors in mixed layer dynamics rather than errors in the rate at which surface heat is supplied.

Along the equator as well as along 9°N improving fluxes only reduces the increments near-surface in the eastern side of the Pacific, a result consistent with the idea that the equatorial temperature increments, large within the mixed layer, result from systematic errors in surface stress (**Figs. 8**, **9**). In the region of the Gulf Stream the impact is also only nearsurface, inshore of the north wall (**Fig. 10,** compare left-hand and right-hand panels), consistent with the idea that the temperature increments aligned with the Gulf Stream front show the effects of error in the model's maintenance of the cross-stream tilt of this front.

3.3 CORE2 Forcing

In order to provide daily surface forcing for the Coordinated Ocean Research Experiments (CORE) Large and Yeager (2009) carried out an exercise in which they used a variety of ancillary data sets to adjust downwelling short and longwave radiation and surface variables. This procedure increased wind speeds in the tropics by 10-40%, decreased specific humidity by 0.25 to to 1 g/kg and also adjusted radiative terms to match satellite (Zhang et al., 2004) and mooring measurements. Here we examine an experiment, soda3.6.1, driven by a slightly updated version of this daily bulk forcing (CORE2), produced using the Large and Yeager (2009) bulk formulas. Since this forcing was not available after 2009 we have carried out this experiment for the 29 year period 1980-2009, but limit our analysis to the last eight years (2001-2009). In all other aspects the experimental setup is the same as before. The resulting CORE2 bulk heat flux is rather similar to the JRA-55 bulk heat flux with a high net rate of heat entering the ocean (compare **Fig. 12** to **Fig. 1** lower left). The ocean heat budget reacts to this excess heating by showing a large heat surplus (negative deficit) throughout much of the ocean except the western Pacific and North Atlantic.

3.4 Improved net surface heat flux

To construct our final estimate of net surface heat flux displayed in **Fig. 13** we average the estimates coming from the three atmospheric reanalyses after correction and after adding their mixed layer heat flux deficits (**Fig. 3**, right-hand side). The broad features of the seasonal and annual mean maps are consistent with previously published estimates (e.g. Grist and Josey, 2003; Large and Yeager, 2009). Heat enters the ocean in the summer hemisphere and exits in the winter hemisphere. Averaging across many seasons, heat enters along the equator in all three ocean basins as well as in the eastern upwelling zones of the Atlantic and Pacific. Heat exits the ocean from warm subtropical western boundary currents and the Southern Ocean.

If we view the ocean basins separately, the Pacific, because of its vast size, is the place where most heat, $+1.3 \times 10^{15}$ W, enters the ocean (basin-integrated heating rates are given in *Supporting Information*, **Table S1**). Dividing this number by the area of the Pacific gives a basin-average heating rate of 8.4 Wm^{-2}. The net heating of the Atlantic Ocean is half as large as the Pacific (5.6×10^{14} W) as a result of its smaller width and the cooling occurring in the subpolar North Atlantic. Additional cooling occurs in the Barents Sea region of the Arctic Ocean and some other areas exposed at high latitudes. However, the contribution to total energy flux of these areas is small. The net contribution of the Indian Ocean to the Earth's heat budget is even weaker (8×10^{12} W), because heat uptake into the northern Indian Ocean is largely compensated for by heat loss from the southern Indian Ocean.

3.3.1 The Southern Ocean

From a time mean Eulerian prospective the meridional circulation in the upper 2 km of the Southern Ocean is dominated by the overturning Deacon Cell, with wind-driven equatorward transport near-surface, sinking along the northern flank of the Circumpolar Current, and upwelling on its southern flank near 60°S (Doos and Webb, 1994). Stability considerations require a source of buoyancy for the accompanying water mass transformation. Many previous studies have suggested that the necessary buoyancy is supplied by net surface heating (e.g. Speer et al., 2000). The idea that the Southern Ocean is being warmed by the atmosphere, however, is surprising given the severe winter weather conditions which occur there, and has provoked some complicated explanations (Czaja and Marshall, 2015). In this estimate we find that the Southern Ocean is actually losing heat to the circumpolar atmosphere at a rate of -6.3×10^{14} W. If we are are right about the sign of this flux then another buoyancy source such as net freshwater flux, as originally suggested by Doos and Webb (1994), needs to be invoked to maintain the Deacon Cell.

4 Summary

Net surface heat flux estimates produced by different atmospheric reanalyses differ by amounts that substantially exceed the size of interesting climate signals, as well as our target of ±5 Wm^{-2}. In many previous studies the accuracy of net surface heat flux estimates over the ocean has been evaluated by comparison to moored or shipboard measurements, but as we point out such evaluations are frequently based on circular reasoning. This study explores what can be learned about errors in net surface heat flux by examining temperature increment ($\theta^a - \theta^f$) statistics produced by a sequential ocean data assimilation reanalysis. We explore this mixed layer heat budget approach in a series of experiments carried out during the data-rich period 2007-2014 during which such an experiment produces nearly 300 short 10 day long forecasts. These short forecasts are less affected by slower processes such as advection and seasonal entrainment than an eight year simulation making it easier to use this mixed layer heat budget to improve estimates of net surface heat flux. An alternative approach to improving bulk surface fluxes by adjusting radiative fluxes and surface variables to match other observation sets (e.g. Large and Yeager, 2009; Dussin et al., 2016) is discussed in *Section 3.3*.

The temperature increments are produced by assimilating the full set of historical hydrographic observations of which, for example, the RAPID section along 26°N is a small subset (Cunningham, et al., 2007). To exploit the complete set of information contained in the temperature increment statistics we derive an approximate form of the incremental heat budget leaving us with a simple formula for improving net surface heat flux based on the temperature increments integrated through the oceanic mixed layer. We apply this formula to identify the seasonal error and improve the estimates of three widely used reanalysis heat fluxes: MERRA-2, ERA-Interim, and JRA-55. Over the subtropical gyres we find that on average ERA-Interim supplies the ocean with 10-20 Wm^{-2} too little heat, while JRA-55 has excess heating by a similar amount.

Our assumption that errors in net surface heat flux are accumulated only within the oceanic mixed layer helps us to reduce the impact of error sources whose impact is concentrated at deeper levels (e.g. **Fig. 9**). For example, along the Pacific Equator one such additional error source is the impact of zonal wind stress error on zonal tilt of the thermocline. In a second example, just north of the Pacific Equator a mismatch of reanalysis background winds and the observations from TAO/Triton moorings cause spurious variations in upwelling at 15° longitude, and that in turn causes a spurious series of shifts in thermocline depth. In a third example, in the North Atlantic a dipole pattern of thermocline temperature increments oriented perpendicular to the path of the Gulf Stream is the result of the inability of the ocean model to properly maintain the north wall of the . In each of these examples the impact of the thermocline temperature increments on our heat flux error estimates is mitigated by limiting our attention to the mixed layer and reducing the impact of errors in wind stress and model physics on our ocean heat budget.

In the second part of this study we test our increment-based approach to balancing the ocean heat budget by improving the seasonal components of net surface flux from the three atmospheric reanalyses: MERRA-2, ERA-Interim, and JRA-55; and then repeat the ocean reanalysis experiments with the improved fluxes. We find that improving the fluxes reduces the mean temperature increments in the mixed layer by as much as a factor of five. Repeating the calculation of seasonal heat flux deficit on the modified fluxes reduces the error to within ±5Wm^{-2} throughout much of the subtropical and midlatitude ocean. Averaged 70°S-60°N our estimates of net surface heating rate range from a minimum of 1.0 to a maximum of 6.3 Wm^{-2}. Further correction, accounting for the mixed layer heat deficit calculated for our second set of experiments and expanding to a global domain reduces our estimate of global surface flux imbalance to 3.4 Wm^{-2}. This global imbalance is still is larger than estimates of the decadal global ocean heating $[\partial(\iiint \rho C_p \theta^a dvol)/\partial t]/area$ of less than 1 Wm^{-2} reported in the *Introduction*, suggesting that we still are not at the point where we can track the excess heating of the ocean by inspection of analyses of surface flux.

An unanticipated consequence of this work is that we find the Southern Ocean to be a source of heat to the atmosphere, a result that makes intuitive sense, but differs from many previous studies. This result, if correct, puts minimum constraints on the amount of buoyancy contributed to the Southern Ocean by net freshwater flux.

In the *Introduction* it was pointed out that the approach taken in this study bears a close relationship to the 4DVar methodology adopted by the Estimating the Circulation and Climate of the Ocean (ECCO). In **Fig. 14** we compare the time mean corrections to ERA-Interim net surface flux produced by this study with the corresponding time mean corrections obtained by the recent ECCO4 release 1 (Forget et al., 2015). The patterns of the corrections obtained using these two approaches are indeed qualitatively similar, showing a reduction of heat going into the ocean at high latitudes and in the eastern equatorial regions and a weak increase in the heating of the central and western subtropical gyres.

In this preliminary study we have limited ourselves to considering only the climatological seasonal component of surface heat flux. We think the same approach can used to improve net surface heat flux at monthly resolution. We have avoided improving heat flux over regions with seasonal sea ice. When sea ice is present the increment temperature conservation equation, (0.0), must be augmented to account for the heat stored in freezing and melting. Finally, we note that the same approach can be used to improve estimates of net surface freshwater flux.

Acknowledgments, Samples, and Data

Computing resources were provided by the Computational and Information Systems Laboratory of NCAR. The complete hydrographic data set was obtained from the NOAA National Center for Environmental Information World Ocean Database (www.nodc.noaa.gov). In situ SST observations were obtained from the ICOADS archive which is hosted by the U.S. National

Oceanic and Atmospheric Administration (icoads.noaa.gov/e-doc). Satellite AVHRR Pathfinder Version 5.2 (PFV5.2) SSTs were obtained from the US National Oceanographic Data Center and GHRSST (pathfinder.nodc.noaa.gov). PFV5.2 is an updated version of the Pathfinder Version 5.0 and 5.1 collection described in Casey et al. (2010). ACSPO MODIS SST data are provided by NOAA STAR. We strongly recommend contacting NOAA SST team led by A. Ignatov before the data are used for any publication or presentation. We are indebted to our data providers. The MERRA-2 atmospheric reanalysis is available through NASA's Science Mission Directorate (gmao.gsfc.nasa.gov/reanalysis/MERRA-2/data_access/). The ERA-Interim reanalysis was obtained from the European Center for Medium Range Weather Forecasts (apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc). The JRA-55, produced by the Japanese Meteorological Agency, was obtained from the NCAR Research Data Archive (rda.ucar.edu/datasets/ds628.0/). The SODA3 reanalysis data sets are all available through www.soda.umd.edu. We gratefully acknowledge the Physical Oceanography Program of the National Science Foundation (OCE1233942) for providing financial support to the authors for this work.

References

Bloom, S.C., L.L. Takacs, A.M. Da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates, *Mon. Wea. Rev.*, **124**, 1256-1271 DOI: http://dx.doi.org/10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2

Brodeau, L., B. Barnier, A.-M. Treguier, T. Penduff, and S. Gulev, 2010: An ERA40-based atmospheric forcing for global ocean circulation models, *Ocean Mod.*, **31**, 88-104. DOI: 10.1016/j.ocemod.2009.10.005

Boyer, T.P., J. I. Antonov, O. K. Baranova, C. Coleman, H. E. Garcia, A. Grodsky, D. R. Johnson, R. A. Locarnini, A. V. Mishonov, T.D. O'Brien, C.R. Paver, J.R. Reagan, D. Seidov, I. V. Smolyar, and M. M. Zweng, 2013: World Ocean Database 2013, NOAA Atlas, NESDIS, **72**, S. Levitus, Ed., A. Mishonov, Technical Ed.; Silver Spring, MD, 209 pp., http://doi.org/10.7289/V5NZ85MT

Brodeau, L., B. Barnier, S.K. Gulev, and C. Woods, 2017: Climatologically Significant Effects of Some Approximations in the Bulk Parameterizations of Turbulent Air–Sea Fluxes. J. Phys. Oceanogr., 47, 5–28, doi: 10.1175/JPO-D-16-0169.1.

Brunke, M. A., Z. Wang, X. B. Zeng, M. Bosilovich and C. L. Shie, 2011: An assessment of the uncertainties in ocean surface turbulent fluxes in 11 reanalysis, satellite-derived, and combined global datasets, *J. Clim.*, **24**, 5469-5493.

Carton, J.A., and B.S. Giese, 2008: A reanalysis of ocean climate using Simple Ocean Data Assimilation (SODA), *Mon. Wea. Rev.*, **136**, 2999–3017, http://dx.doi.org/10.1175/2007MWR1978.1.

Carton, J. A., Y. Ding, and K. R. Arrigo, 2015: The seasonal cycle of the Arctic Ocean under climate change, *Geophys. Res. Lett.*, **42**, 7681–7686, doi:10.1002/ 2015GL064514.

Cerovecki, I., L. D. Talley and M. R. Mazloff 2011: A Comparison of Southern Ocean Air-Sea Buoyancy Flux from an Ocean State Estimate with Five Other Products, *J. Clim.*, **24**, 6283-6306.

Cunningham, S.A., T. Kanzow, D. Rayner, M.O. Baringer, W.E. Johns, J. Marotzke, H.R. Longworth, E.M. Grant, J. J.-M. Hirschi,. L.M. Beal,. C.S. Meinen, and H.L. Bryden, 2007: Temporal Variability of the Atlantic Meridional Overturning Circulation at 26.5N, *Science*, **317**, 935. DOI 10.1126/science.1141304.

Curry, J. A., et al. ,2004: SEAFLUX, *Bull. Am. Meteorol. Soc.*, **85**, 409–424, doi:10.1175/BAMS-85-3-409.

Czaja, A., and J. Marshall, 2015: Why is there net surface heating over the Antarctic Circumpolar Current, *Ocean Dynam.*, **65**, 751–760. DOI 10.1007/s10236-015-0830-1

Dai, A., T. Qian, K. E. Trenberth, and J. D. Milliman, 2009: Changes in continental freshwater discharge from 1948-2004, *J. Clim.*, **22**, 2773-2791.

Dee, D.P., with 35 co-authors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. R. Meteorol. Soc.*, **137**, 553-597, DOI: 10.1002/qj.828.

Delworth, T.L., A. Rosati, W. Anderson, Alistair J. Adcroft, V. Balaji, Rusty Benson, Keith Dixon, Stephen M. Griffies, Hyun-Chul Lee, Ronald C. Pacanowski, Gabriel A. Vecchi, Andrew T. Wittenberg, Fanrong Zeng, and Rong Zhang, 2012: Simulated Climate and Climate Change in the GFDL CM2.5 High-Resolution Coupled Climate Model. *J. Clim.*, **25**, 2755–2781. doi: http://dx.doi.org/10.1175/JCLI-D-11-00316.

Doos, K., and D.J. Webb, 1994: The Deacon Cell and other meridional cells of the Southern Ocean, *J. Phys. Oceanogr.*, **24**, 429-442.

Dussin, R., B. Barneir, L. Brodeau, and J.M. Molines, 2016: The making of the DRAKKAR forcing set DFS5 DRAKKAR/MyOcean Report 01-04-16. https://www.drakkar-ocean.eu/publications/reports/report_DFS5v3_April2016.pdf. Accessed August, 2017.

Edson, J. B., and Coauthors, 2013: On the exchange of momentum over the open ocean. *J. Phys. Oceanogr.*, **43**, 1589–1610, doi:10.1175/JPO-D-12-0173.1.

Forget, G., J. M. Campin, P. Heimbach, C. Hill, R. M. Ponte, and C. Wunsch, 2015: ECCO version 4: An integrated framework for non-linear inverse modeling and global ocean state estimation. *Geosci. Model Dev*., **8**, 3071–3104, doi:https://doi.org/10.5194/gmd-8-3071-2015.

Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30,** 5419–5454, doi:https://doi.org/10.1175/JCLI-D-16-0758.1.

Grist, J.P., and S.A. Josey, 2003: Inverse analysis adjustment of the SOC air-sea flux climatology using ocean heat transport constraints, *J. Clim.*, **16**, 3274-3295. DOI: 10.1175/1520-0442(2003)016<3274:IAAOTS>2.0.CO;2

Holte, J., F. Straneo, J. T. Farrar, and R. A. Weller, 2014: Heat and salinity budgets at the Stratus mooring in the southeast Pacific, *J. Geophys. Res. Oceans*, **119**, 8162–8176, doi:10.1002/2014JC010256.Josey, S. A., L. Yu, S. Gulev, X. Jin, N. Tilinina, B. Barnier, and L. Brodeau, 2014: Unexpected impacts of the Tropical Pacific array on reanalysis surface meteorology and heat fluxes, *Geophys. Res. Letts.*, **41**, 6213–6220, doi:10.1002/2014GL061302.

Kobayashi, S., Y. Ota, Y. Harada, A. Ebita, M. Moriya, H. Onoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, K. Miyaoka, and K. Takahashi, 2015: The JRA-55 Reanalysis: general specifications and basic characteristics, *J. Meteorol. Soc. Japan*, **93**, 5−48 DOI:10.2151/jmsj.2015-001

Large, W.G., J.C. McWilliams, and S.C. Doney, 1994: Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. *Rev. Geophys.*, **32**, 363-403.

Large, W. G. and S. G. Yeager, 2009: The global climatology of an interannually varying air-sea flux data set, *Clim. Dynam.*, **33**, 341-364.

Levitus, S., et al., 2012: World ocean heat content and thermosteric sea level change (0–2000 m) 1955–2010, *Geophys. Res. Lett*., **39**, L10603, doi:10.1029/2012GL051106.

Locarnini, R. A., A. V. Mishonov, J. I. Antonov, T. P. Boyer, H. E. Garcia, O. K. Baranova, M. M. Zweng, C. R. Paver, J. R. Reagan, D. R. Johnson, M. Hamilton, and D. Seidov, 2013. World Ocean Atlas 2013, Volume 1: Temperature. S. Levitus, Ed., A. Mishonov Technical Ed.; NOAA Atlas NESDIS 73, 40 pp.

Lorenc, A.C. 2003. Modelling of error covariances by 4D-Var. *Q. J. R. Meteorol. Soc.,* **129**, 3167–3182.

McClean, J.L., P. Poulain, J.W. Pelton, and M.E. Maltrud, 2002: Eulerian and Lagrangian Statistics from Surface Drifters and a High-Resolution POP Simulation in the North Atlantic. *J. Phys. Oceanogr*., **32**, 2472–2491, doi: 10.1175/1520-0485-32.9.2472.

Moisan, J. R., and P. P. Niiler, 1998: The seasonal heat budget of the North Pacific: net heat flux and heat storage rates (1950–1990), *J. Phys. Oceanogr.*, **28**, 401–421.

von Schuckmann, K., and P.-Y. Le Traon, 2011: How well can we derive global ocean indicators from Argo data?, Ocean Sci., 7, 783–791, doi:10.5194/osd-8-999-2011.

Serreze, M. C., A. P. Barrett, A. G. Slater, M. Steele, J. L. Zhang, and K. E. Trenberth, 2007: The large-scale energy budget of the Arctic, *J. Geophys. Res*., **112**, D11122, doi:10.1029/2006JD008230.

Smith, S. R., P. J. Hughes and M. A. Bourassa, 2010: A comparison of nine monthly air-sea flux products, *Int. J. Climatol.*, **31**, 1002–1027, doi:10.1002/joc.2225.

Speer, K., S. R. Rintoul, B. Sloyan, 2000. The diabatic Deacon cell, *J. Phys. Oceanogr*., **30**, 3212-3222.

Stammer, D., K. Ueyoshi, A. Kohl, W. G. Large, S. A. Josey and C. Wunsch, 2004: Estimating air-sea fluxes of heat, freshwater, and momentum through global ocean data assimilation, *J. Geophys. Res.*, **109**, 16.

Taylor, P. K., Ed., 2000: Intercomparison and validation of ocean-atmosphere energy flux fields—Final report of the JointWCRP/SCOR Working Group on Air–Sea Fluxes. WMO Rep. WCRP-112, WMO/TD-1036, 306 pp. (www.noc.soton.ac.uk/ooc/WGASF/)

Trenberth, K.E. J.T. Fasullo, K. von Schuckmann, and L. Cheng, 2016: Insights into Earth's energy imbalance from multiple sources, *J. Clim.*, **29**, 7495-7505. DOI: 10.1175/JCLI-D-16-0339.1

Valdivieso, K. Haines, M. Balmaseda, Y.-S. Chang, M. Drevillon, N. Ferry, Y. Fujii, A. Köhl, A. Storto, T. Toyoda, X. Wang, J. Waters, Y. Xue, Y. Yin, B. Barnier, F. Hernandez, A. Kumar, T. Lee, S. Masina, and K. A. Peterson, 2017: An assessment of air–sea heat fluxes from ocean and coupled reanalyses, *Clim. Dyn.*, doi:10.1007/s00382-015-2843-3.

Xu, F. and A. Ignatov, 2010: Evaluation of in situ sea surface temperatures for use in the calibration and validation of satellite retrievals, *J. Geophys. Res.*, **115**, C09022, doi:10.1029/2010JC006129.

Yu, L. S., and R. A. Weller, 2007: Objectively analyzed air-sea heat fluxes for the global ice-free oceans (1981-2005), *Bull. Amer. Meteorol. Soc.*, **88**, 527-539.doi: 10.1175/BAMS-88-4-527.

Zhang Y, W. Rossow, A. Lacis, V. Oinas, and M. Mishchenko, 2004: Calculation of radiative flux profiles from the surface to top-of atmosphere based on ISCCP and other global data sets: refinements of the radiative transfer model and input data. *J. Geophys Res.*, **109**, D19105, doi:10.1029/2003JD004457

Zhang, D., M. F. Cronin, C. Wen, Y. Xue, A. Kumar, and D. McClurg, 2016: Assessing surface heat fluxes in atmospheric reanalyses with a decade of data from the NOAA Kuroshio Extension Observatory, *J. Geophys. Res. Oceans*, **121**, 6874–6890, doi:10.1002/ 2016JC011905.

Zeng, X., M. Zhao, and R.E. Dickinson, 1998: Intercomparison of bulk aerodynamic algorithms for the computation of sea surface fluxes using TOGA COARE and TAO data. *J. Climate*, **11**, 2628–2644, https://doi.org/10.1175/1520-0442(1998)011<2628:IOBAAF>2.0.CO;2

Zweng, M.M, J.R. Reagan, J.I. Antonov, R.A. Locarnini, A.V. Mishonov, T.P. Boyer, H.E. Garcia, O.K. Baranova, D.R. Johnson, D. Seidov, M.M. Biddle, 2013: World Ocean Atlas 2013, Volume 2: Salinity. S. Levitus, Ed., A. Mishonov Technical Ed.; NOAA Atlas NESDIS, **74**, 39 pp.

Figure Legends

Fig. 1 Time mean net surface heat flux into the ocean for the six experiments listed in **Table 1**. Units are Wm^{-2}. Panels on the left (BEFORE) show fluxes based on atmospheric reanalysis near-surface variables (soda3.3.2a, soda3.4.2a, soda3.7.2a). The spatial averages 70°S-60°N are shown in the upper right of each panel. Panels on the right (AFTER) show fluxes after application of (1.3).

Fig. 2 Comparison of heat flux and heat flux deficits when MERRA-2 reanalysis heat flux versus bulk heat flux forcing is used. Upper left panel shows MERRA-2 net surface heat flux averaged 2007-2014. Middle and bottom left panels show the difference in the net surface heat flux forcing for soda3.3.2a (bulk) and soda 3.3.3a (reanalysis), also averaged 2007-2014. Righthand panels show corresponding mixed layer heat flux deficits from (1.3). Units are W/m^2.

Fig. 3 Time mean mixed layer heat flux deficit from (1.3) evaluated for six experiments listed in **Table 1**. Left-hand panel shows results for the initial experiments: soda3.3.2a, soda3.4.2a, and soda3.7.2a. Right-hand panel shows the mixed layer heat flux deficit for the experiments using the improved net surface heat fluxes: soda3.3.2, soda3.4.2, and soda3.7.2. The spatial averages 70°S-60°N are shown in the upper right of each panel. Units are Wm^{-2}.

Fig. 4 Monthly forecast temperature increments in the subtropical North Pacific (15°N-35°N, 150°E-130°W) with depth and time for the years 2007-2011. The mixed layer depth (approximately the 18°C isotherm depth), varies from 20-30 m in summer to 90 m or more in late winter. Units are K/10dy. Grey contours show isotherms in this domain at 2K intervals.

Fig. 5 Monthly forecast temperature increments in the subtropical South Indian Ocean (40°S-20°S, 40°E-80°E) with depth and time for the years 2007-2011. The mixed layer depth approximately the 20°C isotherm depth), varies from 10 m in summer to 100 m or more in late winter. Units are K/10dy. Grey contours show isotherms at 2K intervals.

Fig. 6 Mean forecast temperature increments (upper panels) in the subtropical North Pacific (15°N-35°N, 150°E-130°W), and (lower panels) the subtropical South Indian Ocean (40°S-20°S, 40°E-80°E) with depth. Units are K/10dy. Black shows experiments before correction (soda3.3.2a, soda3.4.2a, and soda3.7.2a), red shows experiments after correction (soda3.3.2, soda3.4.2, and soda3.7.2).

Fig. 7 Histogram of time mean soda3.4.2a (black) and soda3.4.2 (red) mixed layer heat flux deficits for the 1°×1° squares in the latitude range between 60°S-60°N. The spread of the deficits declines from 13.2 Wm^{-2} before heat flux correction to 6.2 Wm^{-2} after heat flux correction.

Fig. 8 Mean temperature increments for experiments using ERA-Interim fluxes, with and longitude at 0N. (Left) before flux correction (soda3.4.2a). (Right) After flux correction (soda3.4.2). Units are K/10dy. Mean positions of isotherms are shown in grey.

Fig. 9 Forecast temperature increments along 9°N for experiments using ERA-Interim forcing. Upper panel shows the mean mixed layer heat deficit before (soda3.4.2a, black) and after (soda3.4.2, red) flux correction (Wm^{-2}). Middle panel shows wind stress curl (10^{-8} N/m). Longitudes of TAO/Triton mooring locations are indicated by vertical lines. Bottom panel shows mean temperature increments with depth. Mean depths of isotherms are shown in grey.

Fig. 10 Upper panels: mean heat imbalance (0-75m) in the North Atlantic when forced by ERA-Interim before and after flux correction (soda3.4.2a and soda3.4.2). Units are W/m^2. Contours of mean SST are superimposed. Lower panels: meridional sections of mean forecast temperature increments along 65°W (location of section indicated by a line in upper right-hand panel) for the same two experiments. Depths of mean temperature isolines are superimposed in grey.

Fig. 11 Monthly soda3.4.2 forecast temperature increments (upper panel) in the subtropical North Pacific (15°N-35°N, 150°E-130°W) , (lower panel) in the South Indian Ocean (40°S-20°S, 40°E-80°E) with depth and time for the years 2007-2011. Units are K/10dy. Grey contours show lines of constant temperature in this domain at 2K. These panels can be compared to Figs. 3 and 4.

Fig. 12 Results from experiment soda3.3.3a with CORE2 bulk flux forcing averaged 2001-2009. (left) Net bulk heat flux. Units and contour interval are the same as in **Fig. 1**. (right) Mixed layer heat flux deficit with contour interval and units similar to **Fig. 3**.

Fig. 13 Net surface heat flux obtained by averaging the three flux estimates (2007-2014) after flux correction (**Fig. 1**, right-hand panels). Upper panels show seasonal fluxes, lower panel shows the annual mean.

Fig. 14 Comparison of mean net surface heat flux correction for this study with that obtained using Time mean net surface heat flux difference from the Estimating the Circulation and Climate of the Ocean (ECCO), version 4 release 1 4DVar (Forget et al., 2015). Left-hand panel shows the difference between SODA3.4.2 and ERA-Interim net surface heat flux while right-hand panel shows the difference between ECCO4 and ERA-Interim net surface heat flux. In both figures the averaging period is 2003-2010 (experiment SODA3.4.2 was repeated over an extended period for this comparison). Units are Wm^{-2}.

Table 1 Eight SODA3 data assimilation experiments discussed in this paper. Each begins with the same initial conditions on January 1, 2007 and all assimilate the full suite of observational data, but differ only in prescribed surface fluxes. Three initial experiments use atmospheric reanalysis bulk flux forcing provided by the atmospheric reanalysis centers (soda3.3.2a, soda3.4.2a, and soda3.7.2a). A fourth experiment, soda3.3.3a uses MERRA-2 augmented reanalysis forcing. A fifth experiment uses the CORE2 forcing of *Large and Yeager* (2009). Three additional experiments (soda3.3.2, soda3.4.2, and soda3.7.2) are carried out with surface heat and freshwater flux modified based on (0.0) and an equivalent relation for net freshwater flux. Most experiments span eight years (2007-2014). Since CORE2 forcing was only available through 2009 that experiment begins in 1980.

Experiment	Surface forcing	Time period
3.3.2a	MERRA-2 bulk flux forcing	2007-2014
3.3.2	MERRA-2 bulk flux forcing MERRA-2with modified heat and freshwater flux	2007-2014
3.3.3a	MERRA-2 augmented reanalysis forcing	2007-2014
3.6.1	CORE2 forcing	1980-2009
3.4.2a	ERA-I MERRA-2 bulk flux forcing	2007-2014
3.4.2	ERA-I MERRA-2 bulk flux forcing with modified heat and freshwater flux	2007-2014
3.7.2a	JRA55 MERRA-2 bulk flux forcing	2007-2013
3.7.2	JRA55 MERRA-2 bulk flux forcing with modified heat and freshwater flux	2007-2013