A New Sea Surface Temperature and Sea Ice Boundary Dataset for the Community Atmosphere Model

JAMES W. HURRELL, JAMES J. HACK, DENNIS SHEA, JULIE M. CARON, AND JAMES ROSINSKI

National Center for Atmospheric Research,* Boulder, Colorado

(Manuscript received 18 October 2007, in final form 26 February 2008)

ABSTRACT

A new surface boundary forcing dataset for uncoupled simulations with the Community Atmosphere Model is described. It is a merged product based on the monthly mean Hadley Centre sea ice and SST dataset version 1 (HadISST1) and version 2 of the National Oceanic and Atmospheric Administration (NOAA) weekly optimum interpolation (OI) SST analysis. These two source datasets were also used to supply ocean surface information to the 40-yr European Centre for Medium-Range Weather Forecasts Re-Analysis (ERA-40). The merged product provides monthly mean sea surface temperature and sea ice concentration data from 1870 to the present: it is updated monthly, and it is freely available for community use. The merging procedure was designed to take full advantage of the higher-resolution SST information inherent in the NOAA OI.v2 analysis.

1. Introduction

The Community Atmosphere Model (CAM), a three-dimensional atmospheric general circulation model (AGCM), is utilized by many scientists for climate research (Collins et al. 2006a). The CAM serves as the atmospheric component of the Community Climate System Model (CCSM), a fully coupled global climate model (Collins et al. 2006b), and it is the latest in a series of AGCMs previously known as the Community Climate Model (CCM). The most recent version of the CCSM (version 3.0) was released in June 2004, and the release included complete collections of component model source code, documentation, and input data, as well as model output from several experiments. The purpose of this note is to document the global sea surface temperature (SST) and sea ice concentration (SIC) boundary dataset that has been developed specifically for uncoupled simulations with present and future versions of CAM.

Perhaps the most important field in climate system modeling is SST. A significant advantage of fully coupled models is that the ocean not only exerts an influence on the overlying atmosphere, but it also responds to fluctuations in surface heat fluxes driven by atmospheric variability, as in the real world. The one-way forcing in uncoupled AGCM experiments is, therefore, incorrect, and the implied infinite oceanic heat content can significantly limit the utility of such experiments (e.g., Barsugli and Battisti 1998; Kushnir et al. 2002). Yet, flaws in simulated SSTs in coupled models can also seriously distort the modeled climate. Errors in tropical SSTs, for instance, can greatly impact moist convection and the hydrological cycle, thereby affecting the water vapor feedback and global teleconnections such as those observed during El Niño events. Simulation errors in high-latitude SSTs can significantly affect the sea ice extent, with resultant ice-albedo feedbacks potentially exacerbating the original SST errors. Uncoupled AGCM simulations thus remain useful, especially for the detection of atmospheric climate signals related to variations in ocean surface conditions, such as the role of SST in anomalous flood and drought episodes and in decadal climate variability.

An experimental protocol that has become a standard is one where an AGCM is forced with the known global evolution of SST and SIC. Such integrations form the basis of the Atmospheric Model Intercompari-
son Project (AMIP; Gates 1992; Gates et al. 1999), and therefore are commonly referred to as AMIP experiments. Knowledge of SSTs as well as sea ice is also required in analyses of atmospheric fields; consequently, the integrity of global reanalyses, for example by the National Centers for Environmental Prediction (NCEP) and the European Centre for Medium-Range Weather Forecasts (ECMW), depends critically on the specified lower boundary conditions.

The 40-yr ECMWF Re-Analysis (ERA-40; Uppala et al. 2005) utilizes two new surface boundary datasets, developed partly in response to the needs of the project. These source datasets are 1) the monthly mean Hadley Centre sea ice and SST dataset version 1.1, hereafter referred to as HadISST1 following Rayner et al. (2003); and 2) version 2 of the National Oceanic and Atmospheric Administration (NOAA) weekly optimum interpolation (OI.v2) SST analysis (Reynolds et al. 2002). HadISST1 is used in ERA-40 to supply ocean surface information for the period 1958 through 1981, while OI.v2 is used thereafter. The only special processing of these data for ERA-40 was the application of the AMIP-II midmonth calculation (Taylor et al. 2000), which ensures that the monthly mean of the daily interpolated data is identical to the input monthly mean. These datasets were selected for ERA-40 because they not only provide globally complete coverage from analyzed in situ and satellite-derived SST, but also because they share a common SIC dataset and they employ the same SIC–SST relationship in sea ice margins (Fiorino 2004).

Noting these advantages, we have also used the HadISST1 and OI.v2 data to construct a lower-boundary forcing dataset for studies with CAM, but with some additional processing to support a smoother transition between the products as well as to eliminate some SIC data that were judged to be spurious. Below, we provide concise descriptions of the HadISST1 and OI.v2 analyses, and we explain our processing methodology. The need for this documentation lies not only in the fact that CAM is a widely used community tool, but that other major modeling centers and research groups are also employing our processed SST and SIC analyses, which are updated monthly and are freely available.

2. The SST and SIC datasets

a. HadISST1

We employ the Hadley Centre SST and SIC dataset developed at the Met Office Hadley Centre for climate prediction and research. The HadISST1 dataset is fully described and assessed by Rayner et al. (2003). The explicit purpose of HadISST1, which improves upon previous Global Sea Ice and Sea Surface Temperature datasets (GISST), is to force AGCMs in the simulation of recent climate, although it is also routinely used for climate monitoring (e.g., Folland et al. 2001). It is a combination of monthly, globally complete SST and SIC fields on a 1° latitude–longitude grid since 1871, although intermediate processing of anomalies is done on coarser grids.

For SST through 1981, HadISST1 is derived from gridded, bias-adjusted (Folland and Parker 1995) in situ observations. Data-sparse regions are infilled using reduced space OI (RSOI; Kaplan et al. 1997) on a 2° grid from 1949 onward and a 4° grid for earlier data. The RSOI technique utilizes a set of fixed empirical orthogonal functions (EOFs) from a generally well-observed period to describe the characteristic spatial patterns of SST anomaly variations. The advantage of this technique is a more reliable projection of the SST anomaly patterns that exist based on limited observations, but it depends critically on the assumption of stationarity of the statistics (Kaplan et al. 1997, 1998). In particular, the presence of trends, such as those expected with climate change, seriously violates this assumption (Hurrell and Trenberth 1999). Thus, the trend is analyzed separately in HadISST1 (see Fig. 4 of Rayner et al. 2003), and RSOI is used to reconstruct the residual “interannual” variations over 1871–1981. For the latter purpose, EOFs were computed from seasonal, detrended in situ and bias-adjusted satellite data for 1958–97.

While such techniques exploit teleconnection patterns within and between ocean basins, they reduce local variance. Thus, the reconstructed SSTs were blended with noninterpolated gridded in situ SST data to restore some of the local variance. This procedure has also contributed to the improved intermonthly persistence of anomalies in HadISST1 (not shown), which was a major shortcoming of the earlier GISST analyses (Hurrell and Trenberth 1999; Rayner et al. 2003). For the more recent period since 1981, the RSOI technique was applied to the combined in situ and satellite data, and an additional analysis was performed for the Southern Ocean (see Rayner et al. 2003 for details). We use, however, the OI.v2 data for reasons explained later.

Some of the largest discrepancies among earlier SST products occurred at high latitudes where, for instance, annual mean differences between the adjusted OI climatology of Smith and Reynolds (1988) and the GISST climatology of Parker et al. (1995) exceeded 2°C (e.g., Fig. 1 of Hurrell and Trenberth 1999). These differences stem from the extremely low number of in situ (due to navigation hazards) and satellite (due to cloud
cover) observations in these regions and the very different methodologies employed by the two groups to estimate SST near sea ice.

To address these deficiencies, under the coordination of ERA-40, a common and “best” SIC was developed from various data sources, including substantial efforts to remove inhomogeneities to the extent possible (Rayner et al. 2003). Moreover, both groups now employ the same quadratic algorithm to determine SST in sea ice regions. The constants of this algorithm are derived from analyses of collocated SST and SIC data (Rayner et al. 1996), with the constraint that SST is set to $-1.8^\circ C$ (0°C) when sea ice coverage exceeds 90% over the ocean (freshwater lakes). SSTs with SIC values between 0.5 and 0.8 are usually considerably warmer than the $-1.8^\circ C$ value assigned in earlier versions of the NOAA product. Because of differences in the in situ SST data utilized by the two groups, HadISST and OI.v2 still exhibit high-latitude SST differences in marginal sea ice zones (Fig. 1), but these differences are much reduced relative to earlier SST products (see Fig. 1 in Hurrell and Trenberth 1999).

b. OI.v2

The OI SST analysis technique described by Reynolds and Smith (1994) was developed for operational purposes at NCEP. It follows on the analysis methods of Reynolds (1988) and Reynolds and Marsico (1993), which combine in situ and satellite-derived SST data using Poisson’s equation to produce “blended” products. The in situ SST data used consist of quality-controlled ship and buoy observations from the Comprehensive Ocean–Atmosphere Data Set (COADS; Slutz et al. 1985; Woodruff et al. 1998) and the Global Telecommunication System (GTS). Satellite data are obtained from the Advanced Very High Resolution Radiometer (AVHRR) on NOAA polar orbiting satellites beginning in November 1981.

The global coverage provided by satellite estimates of SST is a considerable advantage over the sparse coverage of in situ data, and satellites also provide useful information about patterns and gradients of SSTs. The absolute accuracy of satellite-derived SST, however, is uncertain; substantial corrections are necessary where in situ data are available to provide calibration (Reynolds 1988), yet biases and other uncertainties in the in situ data complicate any correction of the satellite data (e.g., Reynolds et al. 2002). However, without real-time bias corrections, SST analyses using operational AVHRR retrievals are not useful for climate monitoring.

A disadvantage of the blending technique to correct biases in the satellite data relative to the in situ data is the considerable degradation of the spatial resolution of the SST analysis. With the OI product, however, the high resolution of the satellite data is better preserved and the analysis is done weekly (and daily for opera-

---

Fig. 1. Differences between the 30-yr 1971–2000 SST climatologies (OI.v2 – HadISST1 in °C). The color contours are 0.5°C up to absolute 3°C.
tions). The first step is to use the blending technique to provide a preliminary large-scale time-dependent correction of satellite biases. The in situ and bias-corrected satellite SST data are then analyzed using OI on a 1° latitude and longitude grid. Optimal interpolation produces an interpolated value from a weighted sum of the data. Weights are computed using estimates of local spatial covariance and data error variance. The first guess is the previous analysis of the anomalies, which therefore means the anomalies persist in the absence of new information. The technique does not otherwise utilize information from earlier or later times.

Reynolds et al. (2002) discuss the OI.v2 product in detail, including a comparison to HadISST1. The major updates to the OI.v1 product involve adoption of the sea ice to SST conversion algorithm employed in the production of HadISST1 (as discussed above), and a further reduction of satellite bias. Regarding the latter, Reynolds et al. (2002) show that, averaged over 60°S–60°N, OI.v1 data are biased negative by roughly 0.05°C relative COADS, with a slightly larger bias in the 1990s. At the time the OI.v1 product was developed, the in situ data came from a preliminary version of COADS prior to 1990 and from the GTS from January 1990 onward. For OI.v2, the in situ input prior to 1998 comes from COADS, with the GTS providing data in real time afterward. Since COADS contains significantly more ship data than does the GTS, switching from the preliminary to the current version of COADS for in situ data from 1981 to 1997 increased the ship SST coverage, and the impact of this change was positive. Over the 60°S–60°N average, the bias of the OI.v2 relative to COADS was reduced, although a residual of approximately −0.03°C remains (see Fig. 8 of Reynolds et al. 2002).

3. Methodology and results

a. Why merge?

A reasonable option for uncoupled simulations with the CAM or any other AGCM would be to employ the HadISST1 product through time. However, we chose to merge HadISST1 with OI.v2 for several reasons. First, the OI.v2 product is regularly updated and is easily accessible. The merged SST and SIC product is, therefore, easily kept up to date for community use. This allows, for instance, existing AGCM simulations to be easily extended to examine the role of the ocean in forcing near-real-time climate events. Second, in part because it better resolves small-scale features of significance to climate, such as the Gulf Stream, the OI.v2 product is arguably the best global SST analysis currently available (Hurrell and Trenberth 1999; Reynolds et al. 2002). Finally, a merged SST and SIC product was used in the ERA-40 project (largely for the two aforementioned reasons), and these reanalysis data are routinely used to monitor climate and evaluate the simulated climates of AGCMs such as CAM (e.g., Hurrell et al. 2006).

b. Merging procedure

The simplest merge is, of course, to transition from HadISST1 to the OI.v2 SST and SIC data in November 1981. Doing so, however, creates spurious climate signals. The leading EOF of boreal winter (January–March) North Atlantic SSTs and its associated principal component time series reveal a clearly false signal (Fig. 2) that arises from a straightforward merging of the two SST products. In particular, the spatial dipole pattern and sharp discontinuity in 1982 reflect the better resolution of the narrow Gulf Stream in the OI.v2 SST data, which is also apparent in the difference map of the climatological SST values (Fig. 1). Similarly, a simple merging produces large temporal discontinuities in the Kuroshio Extension in the North Pacific, the equatorial Pacific upwelling region, the retroflection region south of Africa, near the Peru, Falkland, and Benguela Currents, and in other areas where real small-scale structures and sharp SST gradients are better resolved by the OI.v2 SST product (Fig. 1).

Instead of simply merging total SST fields, these discontinuities can be eliminated by first producing HadISST1 anomalies relative to the mean HadISST climatology, then adding the monthly anomaly fields onto the OI.v2 SST climatology for the same base period (1971–2000). Over the North Atlantic, for instance, this procedure yields a leading EOF (Fig. 3) that is the well-known observed tripolar pattern of SST variability (e.g., Cayan 1992a,b; Visbeck et al. 2003). This pattern is marked, in one phase, by a warm anomaly in the subpolar North Atlantic, a cold anomaly in the midlatitudes centered off Cape Hatteras, and a warm subtropical anomaly between the equator and 30°N. Moreover, it is known (e.g., Deser and Timlin 1997) that this structure is driven by changes in the surface wind and air–sea heat exchanges associated with the leading pattern of regional atmospheric variability, the North Atlantic Oscillation (NAO; Visbeck et al. 2003): the correlation between the mean winter NAO index of Hurrell (1995) and the principal component in Fig. 3 is ~0.7.

This merging procedure, therefore, was used to create a more homogenous, blended product covering the period since 1870. Adding HadISST1 SST anomalies to the OI.v2 climatology, however, also modified the SST in marginal sea ice zones so that, for instance, the SST
no longer was set to $-1.8^\circ$C over the ocean when ice coverage exceeded 90% (Fig. 4, middle panel). Moreover, in regions with less ice coverage, a small number of our adjusted SSTs were warmer than would be expected for a given SIC value. Therefore, some additional adjustments were made to the collocated SIC and SST data to eliminate outliers and create values that were more physically realistic and consistent with the criteria discussed in section 2a.

In particular, based on scatterplots of collocated SST and SIC for each year of the original HadISST1 analysis, we found that more than 99.8% of data points since...
1870 fell within the following empirical distribution function:

\[ \text{SST}_{\text{max}} = 9.328(0.729 - \text{SIC}^3) - 1.8, \]

where \( \text{SST}_{\text{max}} \) is the maximum SST (°C) for a given SIC (%) less than or equal to 90%. This empirical function is given by the thick black curve in Fig. 4, where the scatter for a typical year (1968) in the original HadISST1 analysis is also shown (top panel).

We made the following adjustments: SSTs were set back to 1.8 °C 1) if the adjusted values were colder and 2) for all ocean points in regions where collocated

**Fig. 3.** As in Fig. 2, but the HadISST1 anomalies are relative to the 30-yr climatology (1971–2000) from the OI.v2 SST analysis (see text for details).
FIG. 4. Scatterplots of global, collocated SST and SIC data over 1968 from (top) the original HadISST1 data; (middle) the HadISST1 data after adjustment of SST to the OI.v2 SST climatology; and (bottom) after final adjustment. The latter includes adjusting SIC for points (diamonds in the middle panel) beyond the range given by the empirical function (thick black curve) and setting SST to $-1.8^\circ$C where 1) the adjusted values were colder and 2) for all ocean points in regions where collocated SIC $\geq 0.9$ (see text for details).
SIC \geq 0.9.\textsuperscript{1} Otherwise, in regions of less ice coverage, the SST data were considered to be more reliable than the SIC data, especially in the presatellite era (Rayner et al. 2003). Therefore, the adjusted SST values were retained and corrections to the SIC data (at points represented by the diamonds in Fig. 4, middle panel) were made. In particular, if the SST for any SIC exceeded SSTmax (Fig. 4, bottom panel) were otherwise retained and corrections to the SIC data (at points represented by the diamonds in Fig. 4, middle panel) were made. In particular, if the SST for any SIC exceeded SSTmax (this condition most typically was found for inland sea regions). Otherwise, if SST exceeded SSTmax, the ice concentration was reduced to the SIC value that corresponded to SST = SSTmax (Fig. 4, bottom panel). This adjustment procedure affected only a very small number of points, and the magnitude of the adjustments to the SIC data were usually 10% or less.

4. Summary

The objective of this short paper is to document the global SST and SIC dataset that has been developed specifically for uncoupled simulations with the CAM. This AGCM is widely used by investigators throughout the climate research community, and AMIP integrations performed with it are available for analysis by the community as well. Therefore, the lower boundary forcing dataset used in these simulations needs to be documented. Moreover, other modeling groups are using this merged dataset to perform AMIP simulations with their atmospheric models, and the Program for Climate Model Diagnosis and Intercomparison (PCMDI) is relying on this dataset in responding to requests for AMIP boundary conditions (K. Taylor 2007, personal communication).

Our methodology for constructing a merged HadISST1 and OI.v2 product was straightforward: one objective was to minimally process the source data. To achieve a smoother transition from HadISST1 to OI.v2 in November 1981, and to take advantage of the better resolution of small-scale structures and sharp SST gradients by the OI.v2 SST product (Fig. 1), we first produced HadISST1 SST anomalies relative to their own 1971–2000 mean. Those anomalies were then added to the OI.v2 climatology for the same base period. Doing so, however, modified the SST and SIC relationships present in the original HadISST1 and OI.v2 data. We therefore constructed an adjustment procedure that eliminated a small number of outliers and created collocated values that were physically realistic, based on a scatterplot analysis of the original HadISST1 data, and the assumption that the newly adjusted SST values were more reliable than the SIC analysis. This assumption is defensible given the multiple uncertainties in the original SIC analysis (see Rayner et al. 2003).

The dataset is updated regularly, and it is freely available (after registration) in network common data format (netCDF) through the National Center for Atmospheric Research’s (NCAR) Community Data Portal (http://cdp.ucar.edu/MergedHadleyOI).

Acknowledgments. The authors wish to thank the Editor, Dr. David Straus, and two anonymous reviewers for their helpful reviews of an earlier version of the manuscript. We also wish to thank Dr. Richard Reynolds and Dr. Nick Rayner for discussions that helped to motivate this project.

REFERENCES


