UCRL-JC-125597

# PCMDI Report No. 60

# THE SEA SURFACE TEMPERATURE AND SEA-ICE CONCENTRATION BOUNDARY CONDITIONS FOR AMIP II SIMULATIONS

**by**

**Karl E. Taylor[1], David Williamson[2], and Francis Zwiers[3]**

[1]Program for Climate Model Diagnosis and Intercomparison
Lawrence Livermore National Laboratory Livermore, CA

[2]Climate and Global Dynamics Division
National Center for Atmospheric Research
Boulder, CO, USA

[3]Canadian Centre for Climate Modelling and Analysis
Atmospheric Environment Service, Victoria, Canada

**Setember 2000**

**PROGRAM FOR CLIMATE MODEL DIAGNOSIS AND INTERCOMPARISON**
**UNIVERSITY OF CALIFORNIA, LAWRENCE LIVERMORE NATIONAL LABORATORY,**
**LIVERMORE, CA 94550**

UCRL-MI-125597

PCMDI Report No. 60

# The Sea Surface Temperature and Sea-Ice Concentration Boundary Conditions for AMIP II Simulations

Karl E. Taylor[1], David Williamson[2] and Francis Zwiers[3]

[1]Program for Climate Model Diagnosis and Intercomparison
Lawrence Livermore National Laboratory
Livermore, CA

Climate and Global Dynamics Division
[2]National Center for Atmospheric Research
Boulder, CO

[3]Canadian Centre for Climate Modelling and Analysis
Atmospheric Environment Service
Victoria, Canada

September 2000

**ABSTRACT**

In many atmospheric general circulation model (AGCM) simulations, the sea surface temperature (SST) and sea-ice concentration (i.e., percentage of area covered by sea ice) are both prescribed, based on observations. In particular this experimental design is fundamental to the Atmospheric Model Intercomparison Project simulations. For AMIP, observed monthly mean SSTs and sea-ice concentration have been compiled for the period January 1979 through February 1996 and also for the month preceding and for several months following the designated AMIP II period. These observed monthly means do not constitute the official AMIP II "boundary conditions," but the boundary conditions are based on them, as explained in this brief report.

In AMIP II the SST and sea-ice concentration boundary conditions should be specified such that the monthly means computed from the model output precisely agree with the observed monthly means. There are several ways to assure compliance with this AMIP II requirement, but for most models which rely on linear interpolation between monthly values, appropriate AMIP II boundary condition data sets can be obtained from the Program for Climate Model Diagnosis and Intercomparison (PCMDI).

This report contains a description of and justification for the procedure recommended here for preparing and applying the AMIP II boundary condition data sets.

## 1. Introduction

In AMIP I simulations (Gates, 1992; Gates *et al.*, 1999) the observed monthly mean sea surface temperatures (SSTs) were specified in most models as applying to the middle of each month, and daily SSTs were obtained by linearly interpolating between these monthly mean values. This is the traditional method for prescribing SSTs in GCM simulations. Under this procedure much of the SST variance on sub-monthly time-scales is filtered out. Perhaps less obvious is the fact that the amplitudes of seasonal and interannual variations in SSTs are also damped under this "traditional" procedure. In fact the monthly mean SST computed from the linearly interpolated time-series will differ in general from the monthly mean series from which it was derived. (Consider, for example, the warmest month of the year. In order to recover the monthly mean value, the maximum temperature in the month must exceed the mean, but this is clearly impossible if temperatures are linearly interpolated between monthly mean values.)

In AMIP II the SST and sea-ice concentration should be prescribed in such a way that the correct (observed) monthly means are recovered. Clearly, a number of different procedures can be devised to satisfy this constraint. Harzallah and Sadourny (1995), for example, have interpolated monthly mean data to daily values by an iterative cubic spline method that preserves the monthly means. Many models, however, are formulated such that monthly, not daily, data are specified, and daily values are then obtained by linear interpolation. For these models the simplest procedure is to construct a data set of (artificial) mid-month values that, upon interpolation to daily values, yields the observed monthly means. Here we describe how such a data set can be generated and used in atmospheric simulations such as AMIP II. We refer to this procedure of specifying mid-month values that preserve the observed monthly means as the "new" method, although variations on it are already in use by some groups. In particular, Sheng and Zwiers (1998) describe a similar approach to specifying sea surface temperature, but with a different treatment of sea ice.

In Section 2 of this report we provide evidence that under the "traditional" method followed in earlier AMIP simulations, the damping of the seasonal and interannual variations of SSTs is not negligible. This is followed in Section 3 by a description of the "new" method for generating mid-month values for use in AMIP II (i.e., the AMIP II boundary conditions). We then show in Section 4 that compared with the "traditional" method, the time series of daily values interpolated from the AMIP II boundary condition data is in better agreement with observations. Section 5 provides instructions on how to obtain and apply this procedure for specifying AMIP II SST and sea-ice boundary condition data. In Section 6 we recommend a spin-up procedure that reduces initial transients, and we describe the SST and sea-ice data that can be used in the spin-up period. This is followed by some brief concluding remarks and three appendices containing mathematical and practical details concerning application of the procedure described here.

## 2. Shortcomings of the traditional method

As noted above, linear interpolation of monthly mean values (under the "traditional" procedure) will damp seasonal and interannual variations. In Figure 1 we show, for example, that at a location in the North Pacific, the maximum temperature observed in 1988 was 295.6 K, but the monthly mean temperature for the warmest month was only 294.3 K. In this case if the temperatures were linearly interpolated between monthly mean values in a GCM simulation, then the peak temperatures would be missed by more than a degree. The particular grid cell and year shown in Figure 1 was chosen to illustrate the problem of concern here, and although the differences there are particularly large, this grid cell is not exceptional. We have analyzed the monthly mean SSTs globally for a 14 year period from 1982 through 1995. We find that under the traditional procedure the climatological mean seasonal cycle over much of the tropics is damped by a few percent (i.e., typically, no more than a few tenths of a degree), but in some regions of the mid-latitudes, it is damped by a half degree or more. See Sheng and Zwiers (1998) for further evidence of the reduction in amplitude of the seasonal cycle.

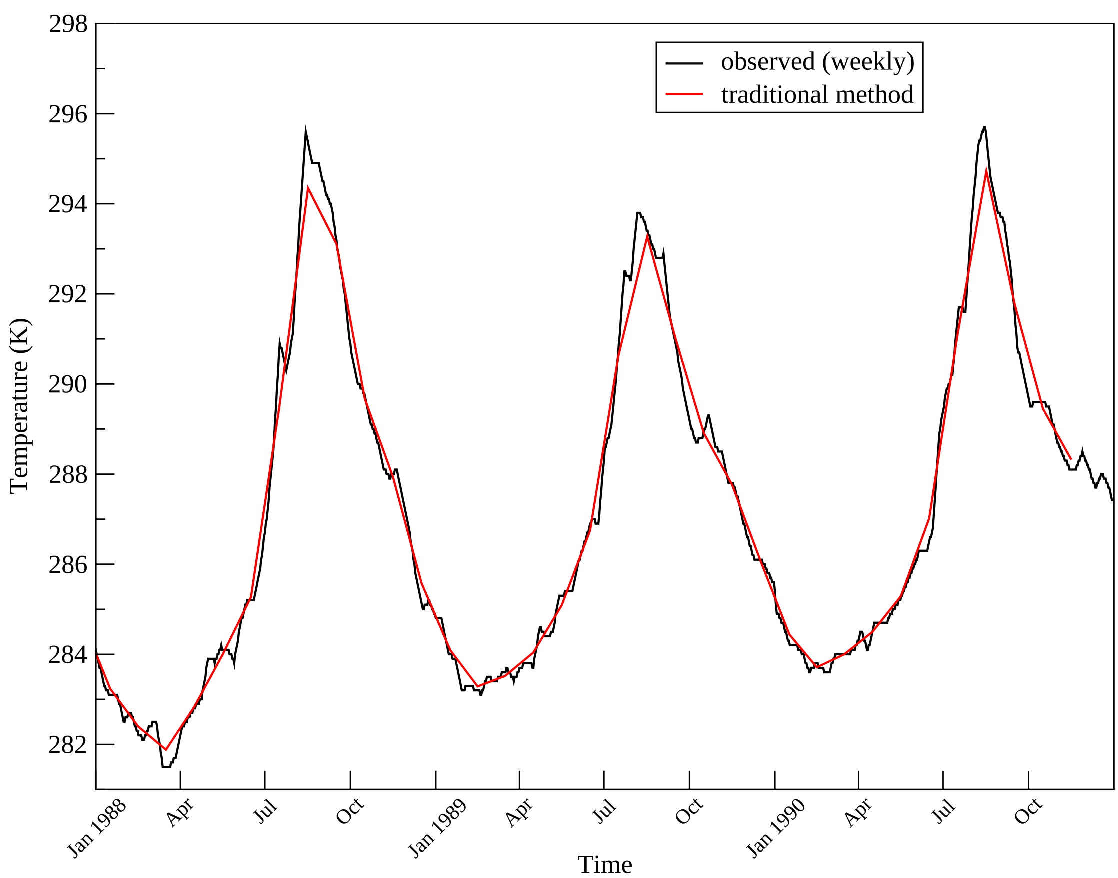


**Fig. 1:** Observed weekly temperature near 180 W 40 N for the years 1988, 1989 and 1990. Also shown is a temperature time-series generated by the traditional method of interpolating linearly between monthly mean values.

Perhaps more troubling than the underestimation of temperature extremes under the traditional procedure, the monthly means are, in general, not preserved. It can be shown analytically that the monthly mean temperature for a given month, $T_i^*$, is related to the monthly boundary condition data $T_i$ by the following formula (assuming here for simplicity that all months have the same number of days):

$$T_i^* = (T_{i-1} + 6T_i + T_{i+1})/8 .$$

Thus for month *i*, the monthly mean resulting from this procedure is a weighted average of the temperature for three months centered on month i. The monthly mean, $T_i^*$, will therefore generally differ from $T_i$. The process of linear interpolation to daily values in effect filters the data in such a way as to decrease the variance of the monthly means on all time-scales, but especially at the highest frequencies. Under the traditional procedure generally followed in AMIP I simulations, $T_i$ was taken as the observed monthly mean, so these monthly means were clearly not preserved.

How large are the differences between $T_i^*$ and $T_i$? For ice-free regions of the oceans, Figure 2a shows the maximum positive difference between $T_i$ and $T_i^*$ (when $T_i$ is set to the observed monthly mean values), considering all months of the AMIP II period

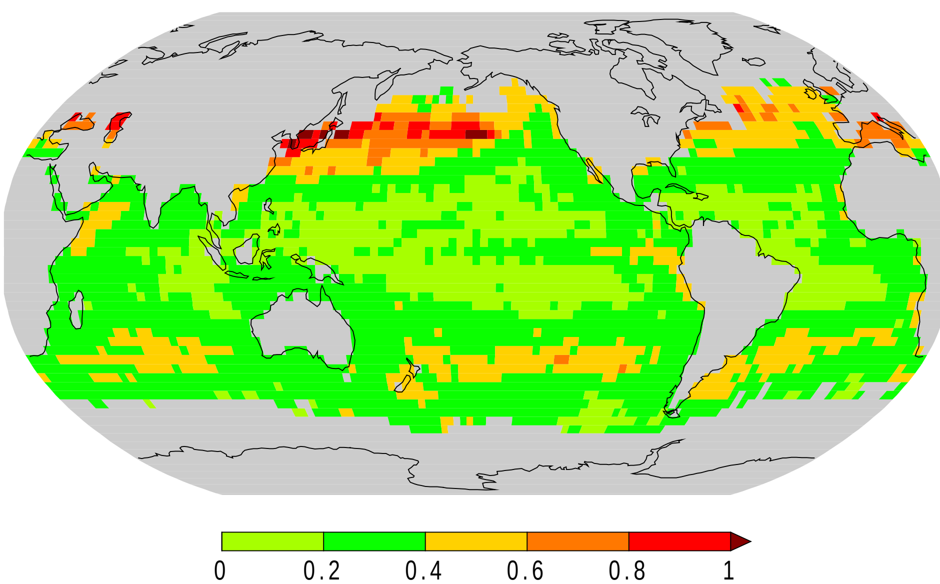


**Fig. 2a**: Maximum positive difference found between *T* and *T** (i.e. the difference between the observed monthly mean and the monthly mean calculated from daily values obtained by linearly interpolating between observed monthly means) for the AMIP II period (January 1979 through February 1996). (Units: K)

(January 1979 through February 1996). In most places this maximum difference is less than a half degree, but in the North Pacific and North Atlantic it reaches a degree or more. If the observed monthly mean values were specified in a GCM simulation, the resulting monthly mean temperature (after daily interpolation) would therefore differ from the observed by a few tenths of a degree or more almost everywhere during at least one month of the AMIP period. The difference shown almost invariably occurs in the warmest month of the year, so in fact the warmest month under the traditional method of prescribing SSTs is always too cool. Similarly, the coolest month of the year is always too warm, but typically the discrepancy is smaller in this case, as shown in Figure 2b.

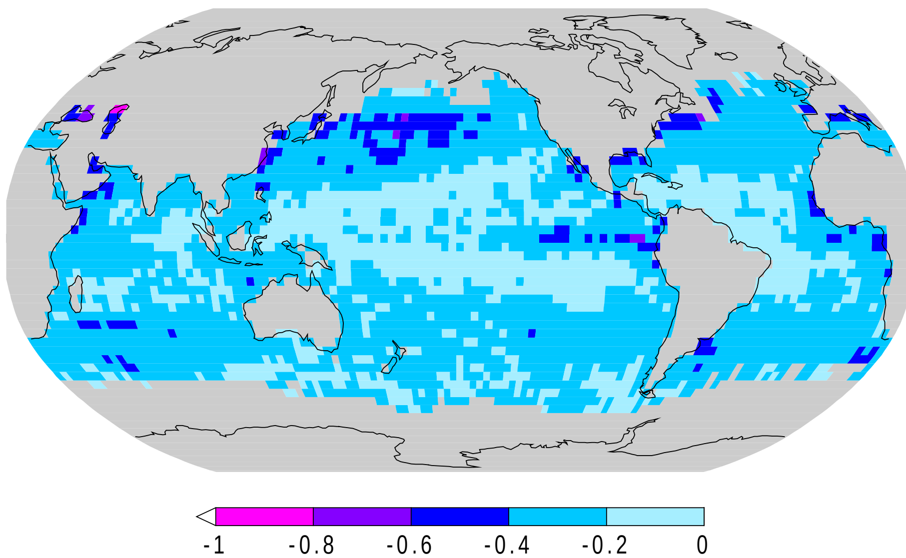


**Fig. 2b**: As in Figure 2a, but the maximum negative difference found.

The size of monthly temperature anomalies (i.e, the variations in temperature once the climatological mean seasonal cycle has been removed) is also reduced under the traditional procedure of interpolating to daily values from monthly mean SSTs. The fractional reduction in the variance of monthly anomalies can be shown to be

$$\frac{\overline{T'^2} - \overline{T^{*\prime 2}}}{\overline{T'^2}} = \frac{13}{32} - \frac{3}{8}R_1 - \frac{1}{32}R_2,$$

where $T'$ is the SST anomaly, the overbar indicates an average over all months, and $R_1$ and $R_2$ are the 1-month and 2-month lag correlations, respectively, for temperature.

An analysis of the SST anomalies during the AMIP II period gives the following global mean values for the lag correlations: $R_1 = 0.67$ and $R_2 = 0.47$. According to the above formula, this implies that under the traditional method of specifying monthly mean temperatures in GCM simulations, the variance of the monthly mean temperature anomalies is typically reduced below the observed variance of monthly means by 14%. Clearly, under this procedure the true strength of the SST anomalies will not be preserved in the actual forcing of the GCM. If the GCM response to this forcing scales linearly, then obviously the response will also be reduced typically by 14%. (For sea-ice concentration, the corresponding number is 18%.)

## 3. Construction of the AMIP II boundary conditions

Because of the shortcomings described above, a new approach to specifying the boundary condition data sets has been devised for AMIP II. The boundary conditions are based on the monthly mean observed data (Fiorino, 1997), but these data have been modified such that their use in GCMs leads to monthly means that are identical to the observed. Moreover, the artificially constructed mid-month temperatures and sea-ice concentrations, which constitute the boundary conditions, are found on average to correspond more closely to the actual mid-month temperatures than the observed monthly means.

To create the AMIP II boundary conditions satisfying the constraints discussed above, we have:

1. Calculated a *climatological* monthly mean for each month of the year, based on observations (Fiorino, 1997) available on a 1° by 1° latitude-longitude grid for all the months of the AMIP II time period (i.e., eighteen years of January and February and seventeen years of March, April, May, etc).

2. Generated twelve mid-month values that (when interpolated to daily values) exactly reproduce the observed *climatological* monthly mean values. These values can be used during most of the spin-up period prior to the AMIP II simulation period.

3. Generated mid-month values (for the 206 AMIP II months plus December of 1978 and March of 1996) that (when interpolated to daily values) exactly reproduce the observed monthly means. (In order to do this, *observed* monthly means were required for a few months before and a few months after the AMIP II period.)

4. Generated a data set of mid-month values for 1978 that can be used in the last year of the model spin-up, prior to the AMIP II simulation period.

5. Written computer code to generate mid-month values appropriate for any model, given its grid-structure and an appropriate land-sea mask. An algorithm that produces an objectively constructed land-sea mask on any model grid has also been developed (Taylor and Doutriaux, 2000).

Appendix 1 contains the mathematical details concerning steps 2, 3, and 4. For sea-ice concentration ($C$), some of the mid-month values in the boundary condition data set are by design negative or greater than 100%. These values are clearly unphysical, but have been deliberately defined in this way so that the correct monthly means will be preserved if, after a daily value is calculated based on linear interpolation, it is "clipped", if necessary, by applying the following filter:

$$C = \max[\ \min(C, 100.0), 0.0].$$

Similarly for sea surface temperature ($T$), the following filter will ensure that the observed monthly mean is preserved and the temperature is held at or above the freezing point of sea water (taken to be 271.38 K):

$$T = \max(T, 271.38).$$

Accurate preservation of the monthly means requires use of a realistic (Julian or Gregorian) calendar and requires that daily (or higher frequency) data be generated through linear interpolation between the appropriate mid-month values, assuming the mid-month values apply precisely at the middle of each month (i.e, 12Z January 16, 0Z February 15 (for non-leap years), 12Z February 15 (for leap years), etc.). We note, however, that for models with unrealistic calendars (e.g., twelve 30-day months), shifting these times by a day or so probably makes little practical difference. (See Appendix 3 for a complete table of mid-month dates and times.)

Temperature anomaly data for the period preceding and following the AMIP II period have been used to minimize the influence of a somewhat arbitrary mathematical boundary condition that must be applied to solve for mid-month values. (See Appendix 1.)

## 4. Comparison of the "traditional" and "new" methods

Figure 3 reproduces the data shown in Figure 1, along with the time-series generated by linear interpolation between mid-month values contained in the AMIP II boundary condition data set (identified in the figure as "new"). For 1988 and 1990 the new method clearly better reproduces the observed annual cycle. For 1989 the observed maximum temperature occurs near September 1, whereas both the traditional method and new method can produce maxima only at the center of a month. Still, the new method produces a maximum value closer to the observed than the traditional method, and, more importantly, it yields the correct monthly mean.

In order to determine whether in general the new method, compared to the traditional method, yields daily values in closer agreement with observations, we have considered the daily time series at each grid cell for the period January 1982 through December 1995 (during which weekly observed SSTs are available for computing an approximate daily time series). At each grid cell we have computed two root-mean square (RMS) differences: the RMS difference between the observed daily values and the

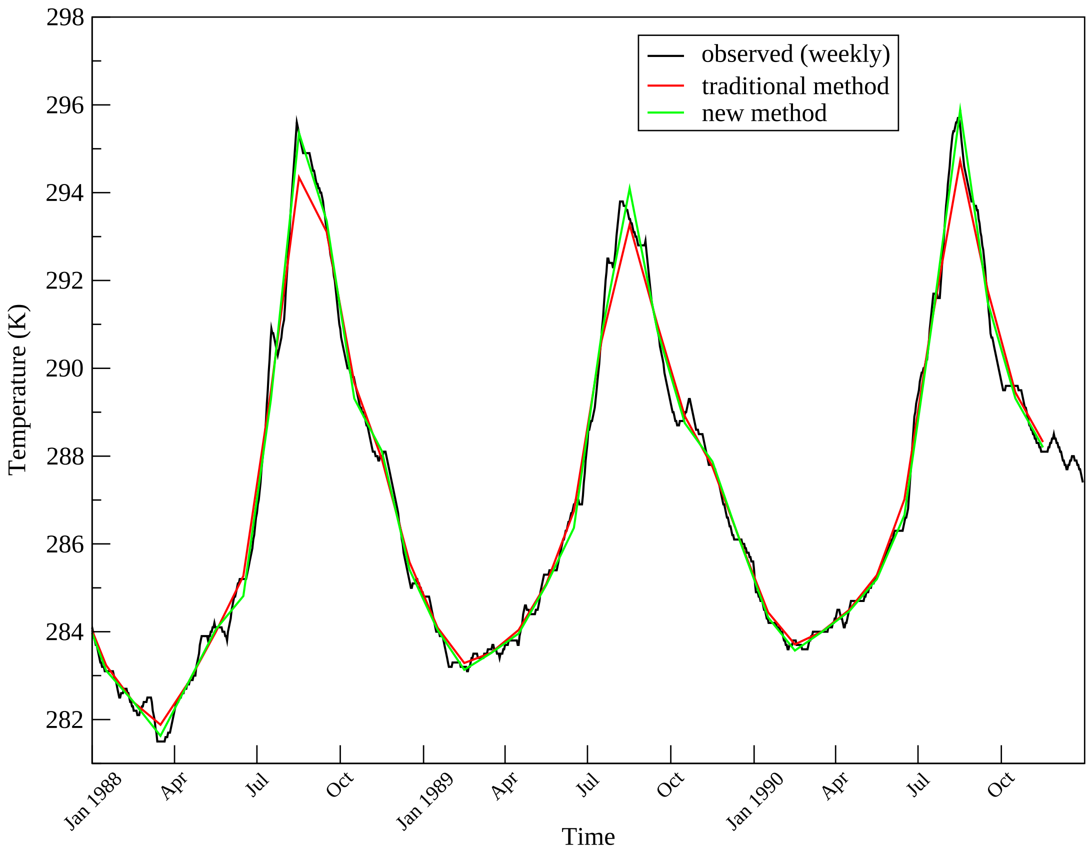


**Fig. 3**: Observed daily temperature near 180 W 40 N for the years 1988, 1989 and 1990, based on weekly data. Also shown is a temperature time-series generated by interpolating linearly between monthly mean values (the "traditional" method) and a temperature time series generated by interpolating between mid-month values constructed according to the "new" method.

daily values computed from the "traditional" method ($E_t$), and, similarly, the RMS difference between the observations and the time series resulting from the "new" method ($E_n$). The RMS difference represents the average error in the daily temperatures that are used to force GCMs under each of the methods. Next we computed the fractional difference in these errors:

$$(E_t - E_n)/E_n \ .$$

This difference, expressed as a percent, indicates by what percent the error in daily temperatures is reduced by the "new" method. Figure 4a shows that the reduction in error is positive everywhere, indicating that the "new" method is everywhere better than the "traditional" method. In the tropics the "new" method reduces the error typically by less than 10%, but in middle latitudes the error is reduced in some areas by more than 20%. Since the RMS errors are of the order of a degree or so, the actual reduction in error is quite modest, as shown in Figure 4b. Still it is remarkable that the "new" method, which

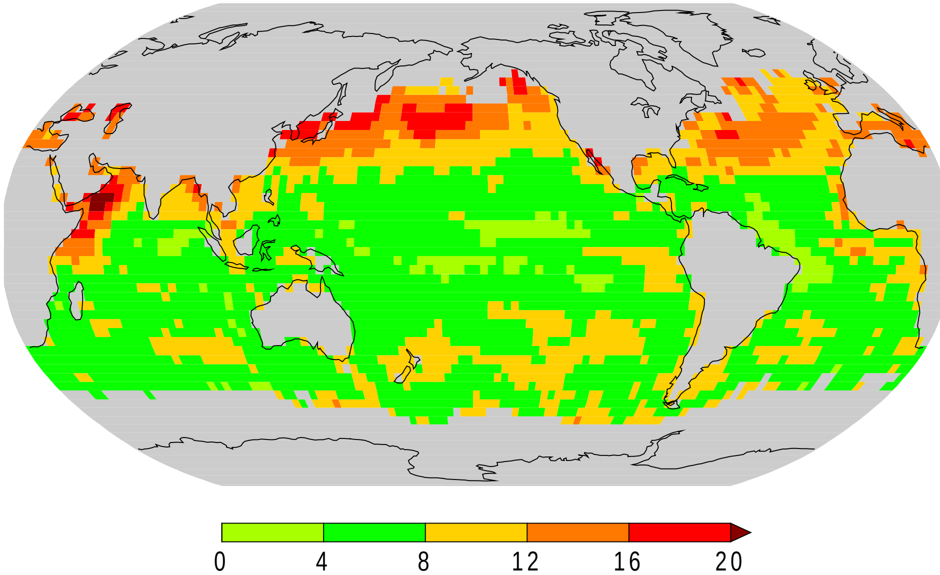


**Fig. 4a:** The percent decrease in RMS error in daily temperature, which is achieved using the "new" method in comparison to the "traditional" method. All values are positive, which demonstrates improvement everywhere. See text for further explanation.

gives perfect agreement with monthly mean observations, also yields daily temperature series that are everywhere in better agreement with observations than the traditional method.

As noted earlier, the variance of monthly mean temperature anomalies is reduced by 14% under the "traditional" method, whereas the new method again forces exact agreement with observations. Thus, the new method is also superior in this regard.

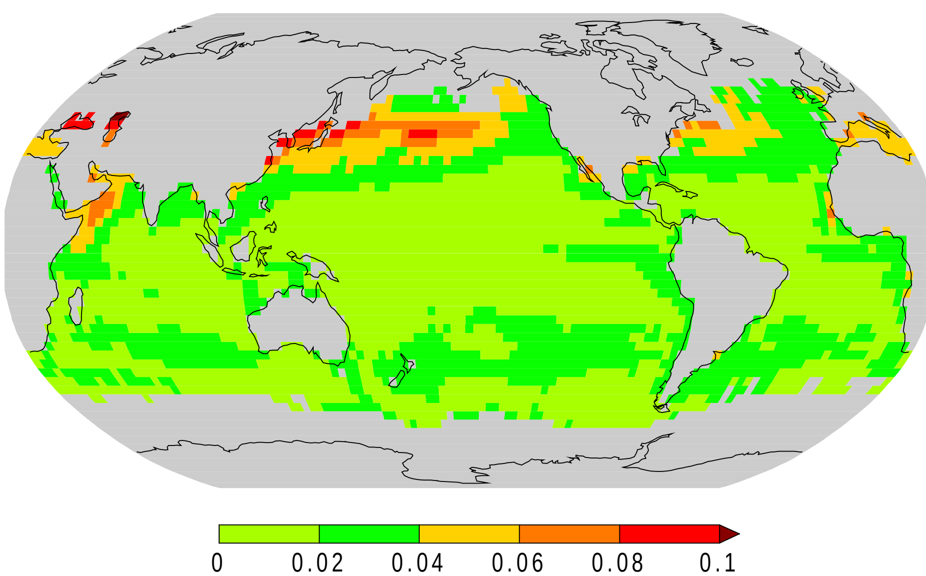


**Fig. 4b:** The actual reduction in the RMS error in daily temperature, which is achieved using the "new" method in comparison to the "traditional" method. See text for further explanation. (Units: K)

## 5. Use of boundary condition data sets in AMIP II simulations

We have outlined how mid-month values of SST and sea-ice concentration can be generated from observed monthly mean data (and further details are provided in Appendix 1). Here we describe how to prepare the boundary condition data for use in climate model simulations. Most of the work can be avoided by submitting a request to PCMDI (http://www-pcmdi.llnl.gov/amip/AMIP2EXPDSN/BCS/amip_grid.html), which will prepare the data on whatever grid is appropriate for a particular model. Appendix 2 provides a step-by-step procedure for obtaining the data and using it in AMIP II simulations. Here we summarize the PCMDI method for generating boundary condition data on any model's grid.

The monthly mean AMIP II observed data sets are available on a 1° by 1° latitude-longitude grid, as described by Fiorino (1997). Before this data can be used in models it must be mapped to a model's (generally coarser) grid and it must be processed, so that interpolation to daily values preserves the monthly means. There are two ways to proceed: map the monthly mean data to the model grid and then produce the mid-month values, or *vice versa*. The boundary condition data sets resulting from these alternative procedures will not be identical in regions of sea ice, where, because of the maximum and minimum limits imposed (100% and 0%, respectively), the problem is nonlinear. Under our procedure, we mapped the data to the model's grid first. This ensures that when the boundary condition data are generated on the model's grid, temporal interpolation will preserve these observed mean values, as required under AMIP II.

It is interesting to see how the monthly mean temperatures calculated from the boundary condition data depend on the order of interpolation/calculation. Figure 5a shows for sea-ice concentration the root-mean-square (RMS) difference between monthly means resulting from interpolating the observed monthly means first, calculating the mid-month boundary condition data second (i.e., the recommended procedure) and *vice versa*. The differences are small (less than 4%), but of course the RMS difference is an average over all months. Figure 5b, on the other hand, shows that for an individual month during the AMIP period the order of interpolation/calculation can change the monthly mean by 10% and more (up to 35% at one grid cell for one month).

For SST the dependence of monthly mean values on the order of the operations (interpolation/calculation) is also quite modest. Figure 6a shows that the RMS difference is almost everywhere less than 0.1 K (and never exceeds 0.25 K). Figure 6b shows that at a few grid cells the maximum difference exceeds 1 K for at least one month of the AMIP period. In spite of these rather small differences, if the objective is to run all AMIP models under the same boundary conditions (to the extent that this is possible at different resolutions), we recommend that all groups follow the recommended procedure and obtain boundary condition data from PCMDI at the resolution needed for their models.

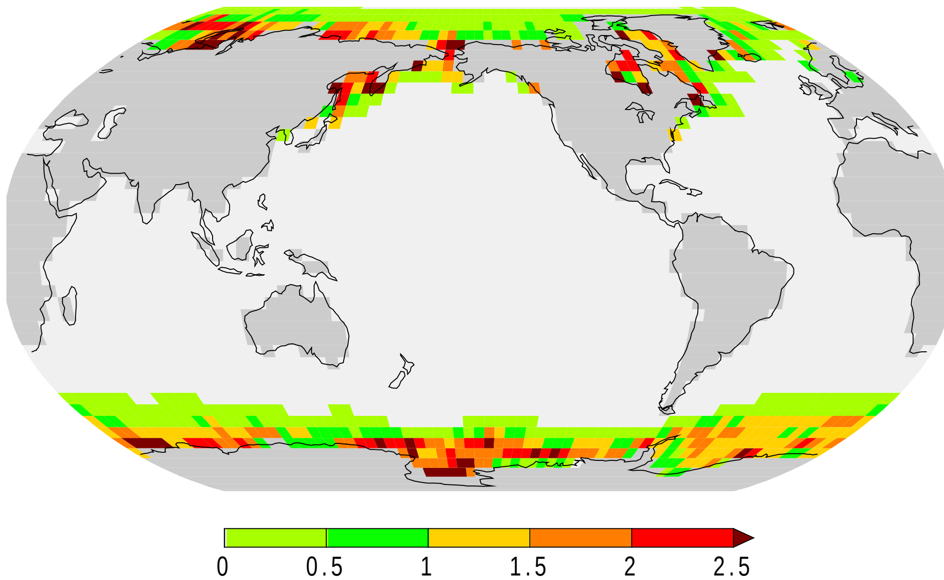


**Fig. 5a:** The RMS difference in monthly mean sea-ice concentration resulting from mid-month boundary condition data that have been generated in two different ways: 1) area-weighted mapping of 1° by 1° observed monthly mean data to a 4° by 5° grid, followed by calculation of mid-month values, and 2) calculation of mid-month values from the 1° by 1° observed monthly mean data, followed by area-weighted mapping to a 4° by 5° grid. The first procedure is the one recommended in this report. (Units: %)

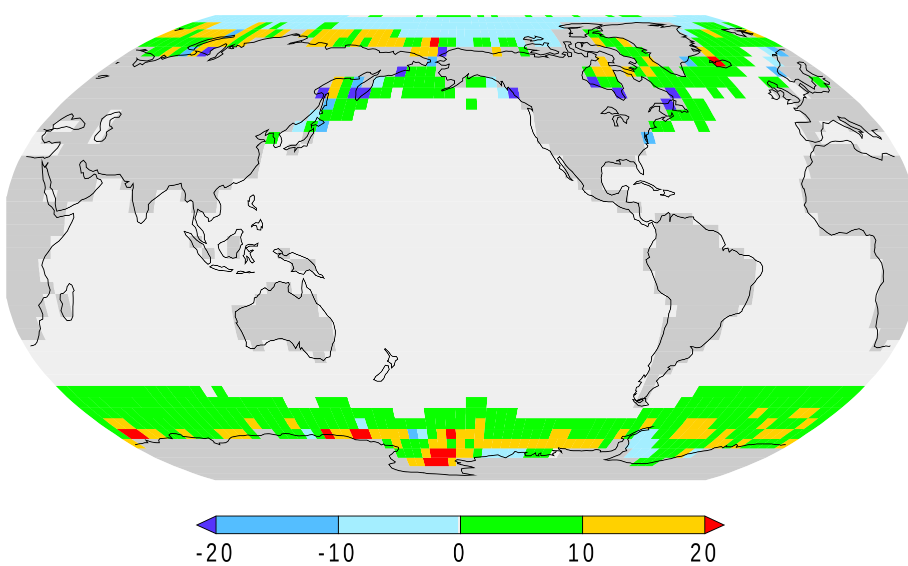


**Fig. 5b:** The maximum difference in monthly mean sea-ice concentration resulting from mid-month boundary condition data that have been generated in two different ways: 1) area-weighted mapping of 1° by 1° observed monthly mean data to a 4° by 5° grid, followed by calculation of mid-month values, and 2) calculation of mid-month values from the 1° by 1° observed monthly mean data, followed by area-weighted mapping to a 4° by 5° grid. The difference shown [(1) - (2)] is for the month with the largest absolute difference. The first procedure is the one recommended in this report. (Units: %)

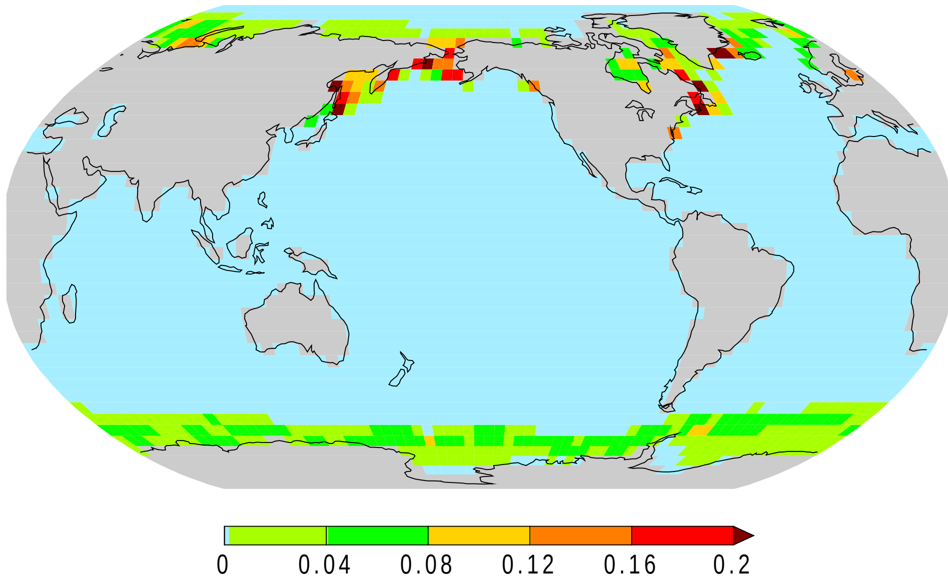


**Fig. 6a:** The RMS difference in monthly mean sea surface temperature resulting from mid-month boundary condition data that have been generated in two different ways: 1) area-weighted mapping of 1° by 1° observed monthly mean data to a 4° by 5° grid, followed by calculation of mid-month values, and 2) calculation of mid-month values from the 1° by 1° observed monthly mean data, followed by area-weighted mapping to a 4° by 5° grid. The first procedure is the one recommended in this report. (Units: K)

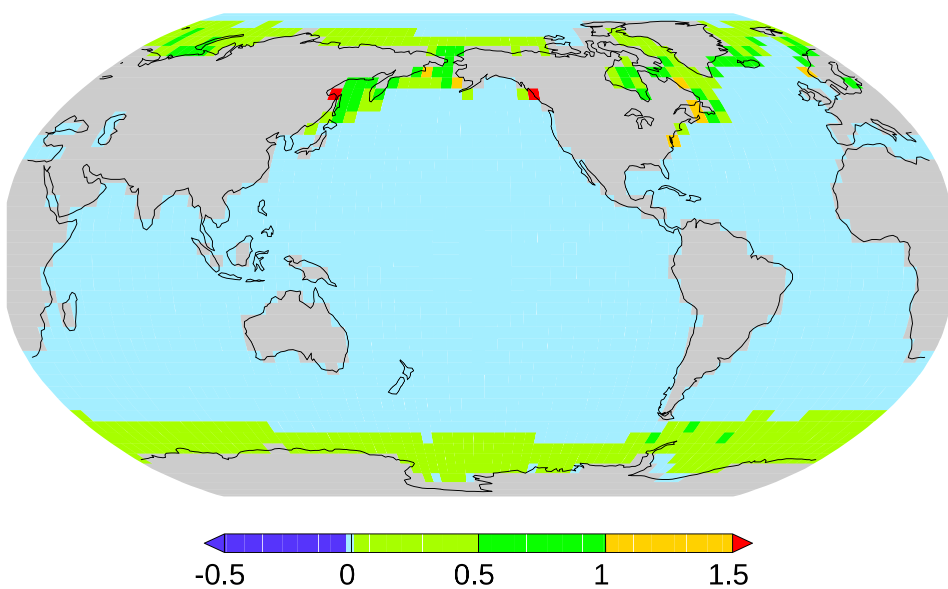


**Fig. 6b:** The maximum difference in monthly mean sea-ice concentration resulting from mid-month boundary condition data that have been generated in two different ways: 1) area-weighted mapping of 1° by 1° observed monthly mean data to a 4° by 5° grid, followed by calculation of mid-month values, and 2) calculation of mid-month values from the 1° by 1° observed monthly mean data, followed by area-weighted mapping to a 4° by 5° grid. The difference shown [(1) - (2)] is for the month with the largest absolute difference. The first procedure is the one recommended in this report. (Units: K)

In interpolating data to a model’s grid, we use a mapping algorithm (area-weighted averaging) that preserves area means. Near ocean boundaries the artificial SST and sea-ice data for land grid cells found in the 1° by 1° monthly mean data are generally ignored. (Note that in the observational data set, the values over land may be reasonable, but in fact they are extrapolated from nearby ocean regions using a Cressman scan analysis procedure dubbed the "weaver" at NCEP. With our method, then, if a model’s ocean grid cell overlaps a region that according to the 1° by 1° data set is partly land and partly ocean, only values from the 1° by 1° ocean grid cells contribute to the mean value assigned to the model’s grid cell. In this way only real ocean data are used to produce a data set on the model’s grid. If for some grid cell a model’s land/sea "mask" is completely incompatible with the 1° by 1° data (i.e., a model’s ocean grid cell contains only land grid cells according to the 1° by 1° grid), then we are forced to use the extrapolated land data to estimate the SST and sea-ice concentration for the model. The 1° by 1° land/sea mask used in creating the monthly mean data is unrealistic in the sense that if only a small fraction of a cell is ocean, the cell is designated ocean. This tends to exaggerate the ocean area, but it means that if a model’s land/sea mask is reasonably realistic, then there should be no need to use the artificial land data. This is also true of models with fractional ocean area in individual grid cells.

In interpolating data to a model’s grid, we use a mapping algorithm (area-weighted averaging) that preserves area means. Near ocean boundaries the artificial SST and sea-ice data for land grid cells found in the 1° by 1° monthly mean data are generally ignored. (Note that in the observational data set, the values over land may be reasonable, but in fact they are extrapolated from nearby ocean regions using a Cressman scan analysis procedure dubbed the "weaver" at NCEP. With our method, then, if a model’s ocean grid cell overlaps a region that, according to the 1° by 1° data set, is partly land and partly ocean, only values from the 1° by 1° *ocean* grid cells contribute to the mean value assigned to the model’s grid cell. In this way only real ocean data are used to produce a data set on the model’s grid. If for some grid cell a model’s land/sea "mask" is completely incompatible with the 1° by 1° data (i.e., a model’s ocean grid cell contains only land grid cells according to the 1° by 1° grid), then we are forced to use the extrapolated land data to estimate the SST and sea-ice concentration for the model. The 1° by 1° land/sea mask used in creating the monthly mean data is unrealistic in the sense that if only a small fraction of a cell is ocean, the cell is designated ocean. This tends to exaggerate the ocean area, but it means that if a model’s land/sea mask is reasonably realistic, then there should be no need to use the artificial land data. This is also true of models with fractional ocean area in individual grid cells.

It should be noted that the for most inland lakes, the 1° by 1° observational data set used in AMIP II is inaccurate; the values for these inland regions may in fact be extrapolations from the nearest ocean regions. Thus, it is *not* recommended that the lake temperatures and sea-ice cover be prescribed according to the AMIP II boundary condition data set. An alternative is to include a simple lake model (e.g., a well mixed layer of prescribed depth) or to set the lake temperature equal to the local annual mean surface temperature.

Objectively constructed land/sea masks for any model resolution are available from PCMDI (Taylor and Doutriaux, 2000), based on a 1/6° by 1/6° U.S. Navy data set (Cuming and Hawkins, 1981; also see http://www.scd.ucar.edu/dss/datasets/ds754.0.html), as explained in Appendix 2. For models with fractional land area in individual grid cells, the land fraction will be preserved; for models with traditional land-only or ocean-only grid cells, the method optimizes the agreement between the model's mask and the original high resolution data.

Once the mid-month boundary condition data set has been obtained, it will not be difficult to use it in AMIP II simulations. There are, however, certain details that should be attended to which are discussed in Appendix 2.

## 6. Boundary conditions for suggested spin-up procedure

Before beginning the simulation of the AMIP II time period, the models should be "spun-up" in such a way as to minimize initial transients. PCMDI has created a climatological SST and sea-ice data set (the mid-month values created in step 2 of Section 3) that can be used to specify the boundary conditions during most of the spin-up period. In order to avoid an initial shock when the observed SST anomalies are first imposed at the beginning of the AMIP period, a special transition boundary condition data set has been prepared (see step 4 of Section 3) for use during the last year of the spin-up period (i.e., a representation of the year 1978). This data set of mid-month values has been constructed based on climatology and on artificially generated monthly SST anomalies that are initially 0, but gradually approach the monthly anomalies observed in December 1978 (the first month where adequate global data are available). The artificially constructed anomalies have an autocorrelation structure that agrees approximately with the observed (i.e., for SST on a 3° grid, a 1-month lag correlation of 0.69, 2-month lag correlation of 0.47, and similarly for 3, 4, 5, ... 12 month lags, correlations of 0.34, 0.27, 0.21, 0.17, 0.14, 0.12, 0.09, 0.06, 0.03, 0.00, respectively; similarly for sea-ice concentrations, the corresponding lag correlations are 0.58, 0.27, 0.15, 0.09, 0.06, 0.03, 0.0, 0.0, 0.0, 0.0, 0.0, and 0.0). The temporal autocorrelation structure depends somewhat on location and on spatial resolution, but not enough to significantly affect the boundary conditions in the months needed for AMIP II.

## 7. Concluding remarks

For intercomparison of models participating in AMIP II it is critical that all simulations are run under the same conditions, notably that the SST and sea-ice concentration be identical for all models. Here we have described how these boundary conditions can be specified to meet the AMIP II requirement that the observed monthly means be preserved when the data are interpolated to daily values. Alternative procedures that meet this AMIP II requirement are permitted, but it may be more difficult to adapt those procedures to the architecture of many climate models that rely on linear interpolation of monthly observations.

Because of the considerable care and effort required to correctly prepare the AMIP II boundary conditions (and the potential for subtle errors), we recommend that each modeling group provide information to PCMDI concerning their model's grid structure and land-sea distribution and rely on PCMDI to produce the boundary condition data sets appropriate to their model. If PCMDI is asked to generate boundary condition data, the mapping from the original resolution to the model resolution will preserve area-average temperatures and sea-ice concentration. On request, PCMDI will also produce an objectively constructed land/sea mask. See Appendix 2 for a step-by-step procedure for obtaining the boundary condition data sets.

If a modeling group elects to create the boundary condition data sets itself, then it is suggested that rather than proceeding as described above (i.e., interpolating the monthly mean data to the model grid and then generating the mid-month values), it is better to obtain the 1° by 1° mid-month boundary condition data set available from PCMDI and interpolate this data to the model's grid. Although in regions of partial sea-ice coverage this procedure will not exactly recover the observed monthly means (obtained by interpolating the 1° by 1° monthly mean data to the model grid), it is considerably easier than solving the (nonlinear) problem of creating a mid-month data set from an observed monthly mean data set.

**Acknowledgments**. We thank Peter Gleckler for useful discussions concerning model spin-up and AMIP II experimental design, as well as useful comments on draft versions of this paper. Charles Doutriaux has been very helpful by providing web-based forms that enable efficient processing of boundary condition requests and assisting in many aspects of data processing. Larry Gates provided helpful editorial corrections and suggestions. Work at LLNL was performed under the auspices of the U.S. Department of Energy, Environmental Sciences Division, by the University of California under contract No. W-7405-Eng-48.

## Appendix 1: Mathematical Details

Here we outline the mathematical procedure for calculating mid-month SSTs that, when linearly interpolated to generate a continuous time-series, will yield monthly mean temperatures that are in exact agreement with the observed monthly mean SSTs. For simplicity we describe the procedure under the approximation that all months are of equal length (say 30 days), but in the actual creation of the AMIP II boundary condition data set, months were assigned the correct number of days, and the equations given here were appropriately modified. In addition, the complications of physical limits on the SSTs (i.e. the freezing point) or sea-ice concentration (i.e. non-negative values no greater than 100%) are not included in this summary, but again were correctly accounted for in constructing the AMIP II boundary conditions. (Note that placing lower and/or upper limits on the data introduces a nonlinearity that requires an iterative method for solution.)

As will be shown below, solution for the AMIP boundary condition data requires specification of monthly mean data for several months preceding and several months following the AMIP II period (January 1979 - February 1996). Since good data are not available prior to December 1978, artificial data were generated for those months, based on the monthly mean climatological data (as described further below). In addition, the climatological data are needed for model spin-up. The first step is then to generate a *climatological* monthly mean boundary condition data set. This data set, comprising twelve mid-month values for each grid cell, is generated from the observed climatological monthly mean data as follows (described here with specific reference to sea surface temperature, but applicable also to sea ice).

Let $S_i$ be the observed mean SST for month $i$ ($i$=1, $I$, where $I$=206 is the number of months in the AMIP II period: January 1979 through February 1996). Let $T_i$ be the mid-month SST for month $i$, specified such that the monthly mean temperatures calculated from the time-series resulting from linear interpolation of these mid-month values are identical to the observed monthly mean SSTs. $T_i$ will constitute the AMIP II SST boundary condition data set.

Let an overbar represent a climatological monthly mean value and a prime an anomaly (relative to the climatological monthly mean value) such that

$$S_i = \overline{S}_j + S_i' \quad \text{and} \quad T_i = \overline{T}_j + T_i',$$

where $i$=1 corresponds to January and $j-1$ is the remainder resulting from division of ($i$-1) by 12 (i.e., $j$-1 = ($i$-1) modulo 12).

The observed climatological monthly mean is defined as

$$\overline{S}_j = \tfrac{1}{K_j} \sum_{k=0}^{K_j-1} S_{j+12k} \qquad j = 1, 12$$

and similarly for $\overline{T}_j$, where $K_j$ is the number of years in the AMIP II period for month *j*. For January and February (*j*=1 and *j*=2), , there are 18 years in the AMIP period (*K*=18), and for all other months there are 17 years (*K*=17). Note that if a grid cell is completely covered by sea ice, the SST is assumed to be the temperature of the ocean just below the ice and is set (following Fiorino, 1997) to 271.38 K. This temperature will contribute to the climatological monthly mean temperature.

Setting the observed climatological monthly mean temperatures $\overline{S}_j$ to the monthly means obtained after linearly interpolating from the climatological mid-month temperatures, we obtain

$$\tfrac{1}{8}\overline{T}_{j-1} + \tfrac{3}{4}\overline{T}_j + \tfrac{1}{8}\overline{T}_{j+1} = \overline{S}_j \qquad j = 1, 12$$

and a cyclic boundary condition:

$$\overline{T}_0 = \overline{T}_{12} \quad \text{and} \quad \overline{T}_{13} = \overline{T}_1.$$

Given the observed climatological mean SSTs, these equations can be solved for the mid-month temperatures that constitute the climatological boundary conditions: $\overline{T}_j \quad (j = 1, 12)$.

The above climatological boundary condition data should be used in the initial spin-up period, as described in Section 6. For AMIP II simulations, however, SSTs include the interannual variability, and should be specified consistent with the observed monthly mean data for the period January 1979 through February 1996. The mid-month temperatures for this period are generated similarly to the climatological boundary condition data, but simple application of a cyclic boundary condition is no longer appropriate. The set of equations to be solved is:

$$\tfrac{1}{8}T_{i-1} + \tfrac{3}{4}T_i + \tfrac{1}{8}T_{i+1} = S_i \qquad i = 1, I$$

where *I*=206. We have *I* equations and *I*+2 unknowns (noting that $T_0$ and $T_{I+1}$ appear in the equation for *i*=1 and *i*=*I*, respectively). Two additional constraints are needed to close the problem mathematically. We note that however $T_0$ and $T_{I+1}$ are determined, they primarily influence the temperatures generated for months 1 and *I*, respectively. They have less influence on the temperatures generated for months 2 and *I*-1, and their influence decreases exponentially for months further away from the beginning and end of the period. If observed monthly mean data were available for many months prior to the AMIP period and for many months following the AMIP period, then one could simply extend the period for which the above equation applied, and one could then prescribe the temperature prior to the beginning and following the end of this extended period to be equal to the climatological monthly mean, knowing this would not significantly affect the mid-month boundary condition data for the AMIP period. The problem is that good global monthly mean data are not available prior to December 1978. We have therefore created an artificial set of monthly mean data for the months from January through December 1978, and have used observed monthly mean data for several months beyond the end of the AMIP II period.

Prior to December 1978, artificial "observed" monthly mean values were prescribed. The artificial data were generated such that temperature anomalies smoothly approach the actual anomalies at the beginning of the AMIP period, but approach zero (i.e, the monthly means approach climatology) within 1 year preceeding the AMIP period. Thus for January 1978 through November 1978, the monthly mean data were specified as:

$$S_{-i} = \overline{S}_{-i} + r_i S_0' \qquad i = 1, 11$$

where $S_{-i}$ is the monthly mean temperature for month $i$ preceding December 1978, $\overline{S}_{-i}$ is the climatological mean temperature for month $i$, $S_0'$ is the observed temperature anomaly (with climatological mean removed) for December 1978, and $r_i$ is the global mean of the auto correlation coefficient (with a lag of $i$ months) calculated from the time series for the AMIP period at each grid cell. Empirically derived values for $r_i$ are given Section 6.

Following the end of the AMIP period, observed monthly mean data were available through August of 1996, and these were supplemented by a full year of data generated artificially using a procedure similar to that described above. Finally, four more months of data (corresponding to September 1996 through December 1996) were prescribed to be equal to the climatological monthly mean values.

Through this procedure, a 20-year hybrid data set of observed and artificial monthly mean data was obtained, covering the period January 1978 through December 1997. These data, along with a cyclic boundary condition (in effect setting the month prior to January 1978 equal to the value for December 1997 and setting the month following December 1997 equal to January 1978) were used to generate the AMIP II boundary condition data sets.

There is one final detail concerning the calculation of the artificial monthly mean for a month in which the climatological value is near one of its limits (e.g., sea-ice concentration near 100%). In this case the equation given above involving the lag correlation coefficient is applied, but the resulting value is "clipped," if necessary, to avoid exceeding the limits.

We reiterate that the artificial data used to generate the AMIP II boundary conditions have only a small effect during the AMIP period itself. Analysis indicates that in general an error in the monthly mean data has a large effect on the mid-month value for the month where the error occurs, but the effect of this error on the adjacent months is diminished by a factor of about 6. The effect on the next nearest neighbors is diminished even further, by another factor of about 6, so the error does not propagate significantly to later or earlier times. We might therefore expect that the use of artificial data for months preceding the AMIP period will only affect the first few months of the simulation, and the effect on those months will be small. Following the end of the AMIP II period, we used observed data through August of 1996 (six months after the last AMIP month), so we do not expect the artificial data following these months to have any noticeable effect on any of the AMIP months.

In the above derivations, we assumed that all months were of equal length. In creating the actual AMIP II boundary condition data sets, we did not make this assumption; months were defined according to the Gregorian calendar. This required minor adjustments to the coefficients in the equations being solved (i.e., changes in the coefficients, 1/8 and 3/4, which depend on the month).

Also in the above derivation, the values for temperature (or sea ice) were not constrained to lie within certain limits. In fact the ocean temperature should never drop below freezing and sea-ice concentration should never be negative or greater than 100%. If these constraints are imposed, then the linear equations given above no longer apply everywhere. In the case of temperature, if linear interpolation from mid-month values yields a value less than 271.38 K, the value is set to 271.38 K, so that when a monthly mean is computed, the simple 1/8, 3/4, 1/8 weighting of mid-month temperatures is no longer correct. In general the coefficients depend on the mid-month temperatures and the equations become nonlinear. This leads to a number of complications, but the equations can be solved using an iterative technique as long as the following additional constraint is imposed (which eliminates the possibility of more than one solution for a given set of observed monthly mean data): If for some month the observed monthly mean is near its upper limit (e.g., a grid cell is nearly completely ice-covered for an entire month), then the mid-month value is set to the minimum value that (when interpolated in time and "clipped" as described in Section 5 and Appendix 2) will yield the correct (i.e., observed) monthly means. A similar constraint is imposed when the observed monthly mean is equal to its lower limit.

In the case of sea ice, an additional complication arises when a grid cell is completely ice-covered in one month but becomes completely ice free the next month. In this case the mid-month value would have to approach negative infinity in the first month and positive infinity in the next month in order for linear interpolation followed by the "clipping" filter to yield the correct monthly mean. In this case unacceptably large "ringing" can also occur in neighboring months. It is of course highly unusual for sea ice to completely cover a grid cell through an entire month and then completely melt at the very end of that month and become completely ice free for the entire next month. Nevertheless, the observed monthly mean data set does contain occasional inland grid cells (which have been assigned artificial data through an extrapolation technique described in Fiorino, 1997) with nearly complete ice cover in one month followed by nearly complete melting at or just following the end of the month. To treat this case the data are temporally filtered in such a way as to preserve the time-mean over the two months, but preventing abrupt jumps from one limit to the other. Specifically, if the change in the sea-ice concentration from one month to the next exceeds 96%, then a small amount of sea ice is moved from one month to the other to reduce the difference to less than 96%. This is done in such a way as to preserve the mean over the two months. These smoothed monthly means are then used in calculating the mid-month AMIP boundary condition values.

## Appendix 2:

### Recipe for Obtaining and Applying the AMIP II Boundary Condition Data Sets

The AMIP II SST and sea-ice concentration boundary conditions should be specified such that the monthly means computed from the model output precisely agree with the observations as given by Fiorino (1997). There are several ways to assure agreement, but for models relying on linear interpolation between monthly values, we strongly recommend that PCMDI be asked to prepare the data sets *at the resolution appropriate for each model*, not the original 1° by 1° resolution. The following procedure should be followed (but for those electing to ignore this recommendation, the next best alternative is to obtain the boundary condition mid-month data sets on the 1° by 1° grid: see http://www-pcmdi.llnl.gov/ AMIP2EXPDSN/BCS/amipbc_dwnld.html and proceed as explained in Section 5).

To request the mid-month boundary condition data sets, please fill out the form found at http://www-pcmdi.llnl.gov/AMIP2EXPDSN/BCS/amip_grid.html (or send an e-mail message to taylor13@llnl.gov), answering the following questions:

- What is your model's grid structure? (sample answer: T42 spherical harmonic Gaussian grid with a longitude/latitude resolution of 128x64 and with the first longitude grid point, i.e., the center of the first grid box) located at 0° east, and with the first latitude grid point at -87.8638°.)

- Does your model accommodate fractional land coverage in individual grid cells, or does it use a "binary" land/sea mask?

- Do you want PCMDI to create an objectively constructed land/sea mask for your grid or do you prefer to use your own land/sea mask? If you plan to rely on your own land/sea mask, please send either an ascii file containing the land fraction (expressed as a percentage) for each grid cell or a file containing the same information written by a PCMDI output subroutine called LATS (see http://www-pcmdi.llnl.gov/software/ lats/). If you send an ascii file with the land percentage data, please put it in the following format:

  ```
  write(*,'(a80)') name
  write(*,'(2i5)') ii, jj
  write(*,'(10f8.3)') ((sftl(i,j), i=1,ii), j=1,jj)
  ```

  where "`name`" should be a string containing information that identifies the modeling group and model (e.g., "NCAR ccm3"), "`ii`" and "`jj`" are the longitude and latitude dimensions, respectively, and "`sftl`" is an array containing the percent land in each grid cell (0.000 or 100.000 for models without fractional coverage). The first longitude in the array, `sftl`, should coincide with the Greenwich meridian (or the first grid point to the east of the Greenwich meridian if there if no grid point coincides with 0°), and longitudes should be stored from west to east. Latitudes should be stored from south to north. A note should be sent to PCMDI (taylor13@llnl.gov) indicating how your land/sea mask data can be retrieved (e.g., provide an anonymous ftp address).

- Which of the following file types do you prefer for the data sets that you will receive from PCMDI: netcdf (COARDS compatible), grib, drs, or ascii? If netcdf or grib files are requested, they will be written with the LATS output subroutine (see http://www-pcmdi.llnl.gov/software/lats/). In ascii output files, grid cells that are entirely land will have a sea-ice concentration of 0.0 and an SST value of 0.0. In all other types of output files, the values in these cells will be set to 1.0e20.

You will be notified when the data sets appropriate for your model's grid have been prepared and placed on anonymous ftp. (This should take only a few days.) The following files will be found:

- SST and sea-ice concentration mid-month boundary condition data (for the AMIP period and the preceeding year: January 1978 - March 1996).

- SST and sea-ice concentration mid-month boundary condition *climatology* based on the period January 1979 - February 1996.

- SST and sea-ice concentration observed monthly-mean data (for model verification from January 1979 - February 1996).

- SST and sea-ice concentration observed *climatological* monthly-mean data based on the period January 1979 - February 1996.

- Objectively constructed land/sea mask or percentage land in each grid cell (if requested).

A "readme" file will accompany the data with a full explanation of how to read the data, along with a sample FORTRAN code segment for reading the data.

In applying the boundary condition data in AMIP II simulations, you should:

- Read in monthly SST and sea-ice concentration mid-month boundary condition data.

- Interpolate linearly in time every model time-step (or at least once each day) between the appropriate mid-month values. (Note that we have found that interpolating once each day, rather than continuously, introduces errors in the monthly mean that are quite small: for SST, maximum errors of less than 0.01 K, and for sea-ice concentration, maximum errors of less than 0.1%).

- For SSTs "clip" the data, if necessary, by applying the filter $T = \max(T, 271.38)$ to prevent temperatures below freezing. For sea-ice concentration, "clip" the data, if necessary, by applying the filter $C = \max[\min(C, 100.0), 0.0]$ to prevent concentrations less than 0% or greater than 100%. All "clipping" should be done after interpolating in time. Note that we have generated the mid-month values such that if this algorithm is followed, the correct monthly means will result.

- For models with grid cells that are either completely sea-ice covered or completely ice free (i.e., models without fractional sea-ice coverage) use a threshold value of 50% to determine whether or not sea ice is present. Do this test after interpolating in time. It is *not* recommended that a flag in the SST data set (such as the value -1.8 °C) be used to determine whether or not sea ice is present (as has been historically the practice in some models). This may therefore require some groups to modify their computer codes to read in the monthly sea-ice concentration data set in addition to the SSTs.

A suggested spin-up procedure is as follows:

- Simulate several years under the *climatological* boundary condition data set (supplied by PCMDI) until initial transients become acceptably small.

- Simulate one year (1978) under an artificially generated boundary condition data set (supplied by PCMDI), with anomalies that are initially zero, but gradually approach the monthly anomalies observed in December 1978 (the first month with adequate global data).

## Appendix 3: Table of Mid-Month Dates and Times

| | Non-Leap Year | | Leap Year | |
|---|---|---|---|---|
| Month | Time/Date | Day* | Time/Date | Day* |
| January | 12Z/16 | 15.5 | 12Z/16 | 15.5 |
| February | 00Z/15 | 45.0 | 12Z/15 | 45.5 |
| March | 12Z/16 | 74.5 | 12Z/16 | 75.5 |
| April | 00Z/15 | 105.0 | 00Z/15 | 106.0 |
| May | 12Z/16 | 135.5 | 12Z/16 | 136.5 |
| June | 00Z/15 | 166.0 | 00Z/15 | 167.0 |
| July | 12Z/16 | 196.5 | 12Z/16 | 197.5 |
| August | 12Z/16 | 227.5 | 12Z/16 | 228.5 |
| September | 00Z/15 | 258.0 | 00Z/15 | 259.0 |
| October | 12Z/16 | 288.5 | 12Z/16 | 289.5 |
| November | 00Z/15 | 319.0 | 00Z/15 | 320.0 |
| December | 12Z/16 | 349.5 | 12Z/16 | 350.5 |

*Number of days following 00Z January 1.

# REFERENCES


Cuming, M.J. and B.A. Hawkins, 1981: TERDAT: The FNOC system for terrain data extraction and processing. Technical Report MII Project M-254 (Second Edition). Prepared for the Fleet Numerical Oceanography Center (Monterey, CA). Published by Meteorology International Incorporated.

Gates, W.L., 1992: The Atmospheric Model Intercomparison Project. *Bull. Amer. Meteor. Soc.*, **73**, 1962-1970.

Gates, W.L., J.S. Boyle, C. Covey, C.G. Dease, C.M. Doutriaux, R.S. Drach, M. Fiorino, P.J. Gleckler, J.J. Hnilo, S.M. Marlais, T.J. Phillips, G.L. Potter, B.D. Santer, K.R. Sperber, K.E. Taylor, and D.N. Williams, 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Amer. Meteor. Soc.*, **80**, 29-55.

Fiorino, M, 1997: AMIP II sea surface temperature and sea ice concentration observations. http://www-pcmdi.llnl.gov/amip/AMIP2EXPDSN/BCS_OBS/amip2_bcs.htm.

Harzallah, A., and R. Sadourny, 1995: Internal versus SST-forced atmospheric variability as simulated by an atmospheric general circulation model. *J. Clim.*, **8**, 474-495.

Sheng, J., and F. Zwiers, 1998: An improved scheme for time-dependent boundary conditions in atmospheric general circulation models. *Clim. Dynam.*, **14**, 609-613.

Taylor, K.E., and C. Doutriaux, 2000: An objective method for generating land/sea masks for use in GCM simulations. PCMDI Report No 58, 16 pp.