

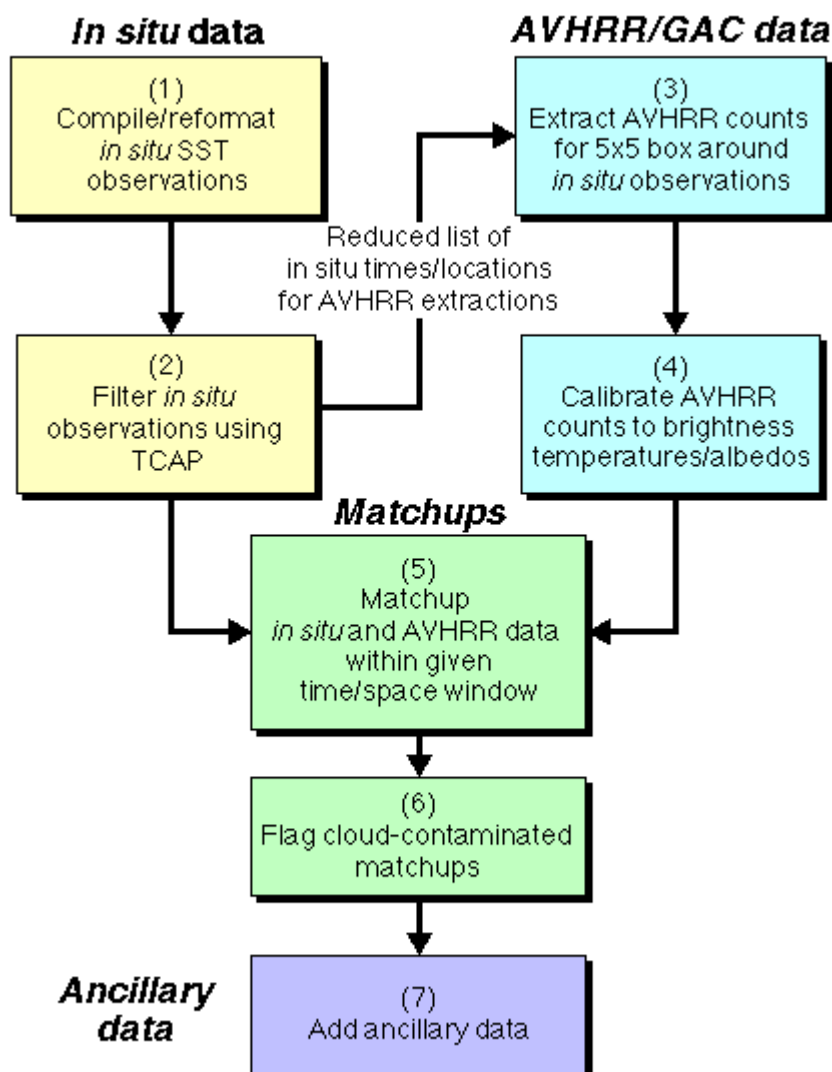


Compilation of the AVHRR Pathfinder Matchup Databases

There are five main steps involved in the compilation of the Pathfinder Matchup Database (Figure 1).
These steps, described in detail below are:

- Compilation and reformatting of in situ data
- Extraction of AVHRR/GAC data
- Matchup process
- Record filtering and flagging
- Addition of ancillary data

Figure 1. Steps involved in the compilation of the Pathfinder Matchup Database. The number in each box indicates the sequence in which the steps take place.



Compilation and Reformatting of *In Situ* Data

The first step in the development of the PFMDB is to assemble and reformat the in situ SST data from various sources (Figure 1, box 1). A matchup database must include observations with a wide geographic and temporal distribution to encompass a wide range of atmospheric and oceanic conditions, as well as observational conditions (e.g., many viewing geometries). Furthermore, the in situ SST measurements should be of good quality. The SST values included in the Pathfinder MDB were obtained from buoys deployed and operated by various meteorological and oceanographic agencies or programs in the U.S. and other countries. The rationale was to use reasonably well calibrated measurements which had undergone some degree of quality control (although some concerns on data quality are discussed later). For the same reason, we avoided the use of SST ship reports, as their quality may be quite variable.

The PFMDB includes in situ SST data from two main sources: moored buoys and drifting buoys. The moored buoy data include platforms operated by the [National Data Buoy Center](#)* (NDBC), located off the east and west coasts of the U.S., in the gulfs of Mexico and Alaska, and around Hawaii; platforms elsewhere in the Pacific (Guam and southeastern Pacific) are also present. Note that the distributed matchup data set includes data for NDBC buoys deployed in the Great Lakes, in case there is interest in estimating algorithms for lake temperature; instructions are given below for excluding data from these buoys from analyses. Brief descriptions of the NDBC moored buoys are presented by Hamilton (1986) and Meindl and Hamilton (1992). Further information on these platforms can be obtained from NDBC's World Wide Web (WWW) server; Universal Resource Locators (URLs) for

these and other in situ data sources are listed in Table 1. The archived NDBC data were provided by the National Oceanographic Data Office (NODC).

The second source of moored buoy data is the [Tropical Ocean-Global Atmosphere \(TOGA\) Tropical Atmosphere Ocean \(TAO\) array](#)* in the equatorial Pacific Ocean. This array is described in Hayes et al. (1991) and McPhaden (1993). The TOGA-TAO data were provided by the TOGA-TAO Project Office at NOAA's Pacific Meteorological and Environmental Laboratory (PMEL). Information on the TOGA-TAO program, as well as data display software and near real-time data may be accessed through the WWW. The third set of moored buoy data includes platforms around Japan, operated by [Japan's Meteorological Agency](#)*. Starting with 1994 matchups, moored buoy data in the northeast Atlantic were obtained from the [United Kingdom Met. Office](#)*. The water temperature sensor in the moored buoys from the various data sources is generally located at a depth of about 1 m.

The drifting buoy data included in the PFMDB were provided by various sources. The first drifter data set was compiled by the [Drifting Buoy Data Assembly Center \(DAC\) at NOAA's Atlantic Oceanographic and Meteorological Laboratory \(AOML\)](#)* as part of the Global Drifter Program. The geographic focus of the early AOML data is the tropical Pacific, as drifters were mostly deployed as part of the TOGA program. However, in recent years the geographic distribution of the data has widened. Maps of drifter tracks and other useful information (e.g., drifter diagrams and photographs) can be obtained through the WWW. A second source of drifting buoy data was [Canada's Marine Environmental Data Service \(MEDS\)](#)*. The MEDS drifter database included all drifter positions and SSTs reported through the Global Telecommunications System (GTS). Therefore, the MEDS data provided useful information for high latitude regions, or in areas other than the tropical Pacific.

The MEDS data included some drifter records also reported in the AOML set. To avoid duplicates, records present in both drifter data sets had to be identified. This could not be done simply by using the drifter ID, as the naming convention was different in the two data sets: the IDs listed by MEDS are the WMO IDs number, whereas the AOML data set lists the Argos IDs (the PTT numbers). A particular combination of time, latitude and longitude was considered to define a unique drifter record. Whenever a pair of records which shared that combination of values was identified, the record in the AOML database was kept and the corresponding MEDS record was excluded. (Please note that use of the MEDS data is for non-commercial use only)

In an effort to enhance the number of matchups in higher latitudes, a data set including drifters deployed in the NW Atlantic and assembled by NATO's SACLANT Undersea Research Center was kindly provided by Dr. Peter Minnett (University of Miami). Although these data encompassed the period 1991–1995, they were only incorporated to a couple of years of the NOAA-11 matchups, because the extractions for all other AVHRR data sets had been completed. If AVHRR data are reextracted in the future, the remainder of the NATO drifter data will be added to the matchup database. The NATO drifter data have been recently made available to NOAA-AOML's Drifting Buoy Data Assembly Center, and can be obtained through AOML. A summary of the sources of in situ SST data included in the PFMDB is shown in Table 1.

Table 1. Sources of in situ SST values included in the AVHRR Pathfinder Oceans global matchup data base.

Platform Type	Source	Platform ID	WWW URL
Moored Buoys			
NDBC	U.S. National Data Buoy Center	Original + 300,000	http://seaboard.ndbc.noaa.gov *
Japanese	Japan Meteorological Agency (English)	Original + 100,000	http://www.jma.go.jp/jma/indexe.html *
TOGA-TAO	TOGA-TAO Project Office (NOAA Pacific Marine and Environmental Laboratory)	Original + 200,000	http://www.pmel.noaa.gov/toga-tao *
UK	United Kingdom Meteorological Office	Original + 400,000	http://www.metoffice.gov.uk *

Drifting Buoys

AOML	Drifting Buoy Data Assembly Center (NOAA Atlantic Oceanographic and Meteorological Laboratory)	Original + 600,000	http://www.aoml.noaa.gov /phod/dac/ *
MEDS	Canadian Marine Environmental Data Service (Fisheries and Oceans Canada)	Original	http://www.dfo-mpo.gc.ca/index- eng.htm *
NATO	NATO SACLANT Undersea Research Centre	Original + 700,000	http://www.nurc.nato.int/ *

Because most *in situ* data are obtained from archive sites, they have been subject to various stages of quality control. Therefore, we assumed that further quality control was not required for most of the data sets. An exception was the early MEDS drifter data, which had undergone only limited quality control. In this case, we implemented a set of tests that essentially excluded erroneous buoy locations. The tests were adapted from routines kindly provided by Dr. Richard Reynolds (Climate Analysis Center). For consistency, these tests were applied to the entire MEDS data set. Preliminary analyses of satellite and drifter data suggested the possibility of problems such as miscalibration of drifter SSTs and wrong values reported towards the end of the lifetime of the drifter's battery. The possibility of biases in SST measurements from drifting buoys was suggested by Bitterman and Hansen (1993). In the future, we expect to investigate further the issue of variable biases among the different sources of SST data in the PFMDB.

We warn users that despite the fact that *in situ* data came from archives, there may still be some quality problems. We identified, for instance, SST reports of 0°C in the tropics. Users can eliminate these problems by checking for large differences between buoy SST and either Pathfinder or Reynolds SST. Most of the initial *in situ* data compilation and quality control was done in collaboration with Dr. Charles McClain and his research group at NASA's Goddard Space Flight Center (NASA/GSFC).

The original platform IDs (usually a 4- or 5-digit code representing the WMO or Argos ID) were modified to facilitate the identification of the various *in situ* SST sources once all the matchups were assembled. A different constant for each data source was added to the original ID (Table 1). For instance, after adding 300,000 to their original IDs, the NDBC buoys' IDs will have values between 300,000 and 399,999. As mentioned above, the PFMDB includes data for NDBC buoys in the Great Lakes. These buoys can be easily identified (and excluded), as their IDs range between 345,000 and 345,999.

The main reformatting of the *in situ* data was the computation of a continuous time coordinate to facilitate the matchup procedure. The dates and times of the *in situ* SST reports data were converted to a continuous time coordinate, "seconds since January 1, 1981", here referred to as "Pathfinder time". The calculation of Pathfinder time took into account leap days, but not any leap seconds added from the beginning of the time coordinate to the present. The Gregorian dates and UTC times of the *in situ* data are included in the PFMDB for the convenience of the users. Because the dates and times were recalculated from the continuous coordinates, there might be roundoff errors in the original *in situ* time; this should cause differences between the original and listed time of no more than one second.

Extraction of AVHRR/GAC Data

AVHRR data had to be extracted at the times and locations of the *in situ* observations. The AVHRR data are provided by the Global Area Coverage (GAC) data stream. The original GAC data tapes were transcribed into optical disks by the AVHRR Land Pathfinder component, and the disks were made available to us in order to perform the extractions.

Due to the time-consuming nature of the extraction process, we developed a procedure (Figure 1, step 2) to exclude those times/locations for which there was not an AVHRR pass within ± 30 minutes from the *in situ* measurement; 30 minutes was the maximum temporal separation allowed between satellite and *in situ* data for inclusion in the PFMDB (more on this below). The procedure involves the computation of the Time of Closest Approach (TCAP) of the NOAA polar platforms with respect to a given point and time (that of an *in situ* SST report). This procedure (termed the "TCAP filter") relies on

orbital routines kindly provided by D. Baldwin (Univ. of Colorado). The TCAP filter significantly reduces the time required for the extraction of GAC data.

AVHRR data were extracted for 5x5-pixel boxes centered at each in situ SST location that passed the TCAP filter (Figure 1, step 3). The initial extraction data set included only the basic sensor count information and was not corrected for any instrument behavior. A navigation correction (time and attitude) was applied to ensure correct geolocation of the satellite data. The initial extraction files were then processed to (a) convert the sensor counts for visible and near-infrared channels (1 and 2) first to Rayleigh-corrected radiances and then to aerosol optical thickness (t_A) values, which removed effects of viewing and illumination geometry, and (b) convert the counts for infrared channels (3-5) into brightness temperatures (Figure 1, step 4). The brightness temperature calculation incorporated the consensus Pathfinder correction for sensor non-linearity described in Rao (1993). Data for the visible channels have been corrected for degradation using calibration correction tables supplied by Dr. C.J. Tucker (NASA/GSFC).

The AVHRR portion of the PFMDB records includes sensor geometry information (solar zenith angle, scanner zenith angle). Using the geometric information, a sun glint index was calculated to assist in the identification of pixels contaminated by glint. Other geometric quantities are provided to identify whether a matchup is located on the side of the AVHRR scan that may be contaminated by sun glint in daytime data. There is also information relevant to sensor calibration, such as the temperature of the baseplate on which the internal calibration blackbody is mounted, and the slopes and intercepts of the counts-to-radiance conversions for the IR channels.

A new feature of Version 19 of the PFMDB is that it includes, for each of the five AVHRR channels, the values of all 25 pixels inside a 5x5 extraction box. This new information may be used for various purposes: for instance, textural quantities may be computed in order to aid in cloud flagging. The PFMDB includes the value of the central pixel in the extraction box for each channel, as well as summary statistics (mean, median, maximum, minimum) for all values inside an extraction box. We note that, in earlier versions of the PFMDB, AVHRR extractions were made for 3x3-pixel boxes, and thus summary statistics were computed for a box of that size. Because the cloud flagging step (see below) includes spatial coherence tests derived from the earlier 3x3 boxes, the summary statistics presented in this version of the PFMDB are computed for 3x3-pixel boxes, for the sake of consistency. Furthermore, the Pathfinder global SST fields also are processed using a 3x3 window for cloud tests. All the satellite quantities are described in further detail below.

Matchup Process

The next step in the generation of the PFMDB was to temporally match in situ records against the AVHRR extractions (Figure 1, step 5). To limit the variability introduced by the time separation between the two data sources (Minnett 1991), the absolute difference between the time of the in situ SST report and the time at which that location was viewed by the AVHRR (i.e., the matchup time window) was restricted in most cases to a maximum of 30 minutes (more on this below). In situ records that did not fall within the stipulated time window were rejected. In situ records that passed the temporal matchup subsequently had to pass a spatial test. A maximum distance of 0.1° in latitude and longitude (approximately 10 km) was allowed between the in situ SST location and the location of the central pixel in the AVHRR extraction box.

For all platforms except the TOGA/TAO buoys, the matchup process is based on a ± 30 minute window centered at the time of the SST report. Although, in some cases, the actual SST measurement may be made a little earlier, the reporting time is the one listed in the PFMDB. For instance, most of the NDBC moored buoys report measurements every hour on the hour, and the reported SST value is averaged over a relatively short period (8–10 minutes) ending about 10 minutes before the reporting time. The drifting buoys report SST measurements taken during the short interval (about 13–15 minutes) in which the polar satellite which receives the Argos data is above the buoy's horizon.

The matchup procedures for the TOGA/TAO buoys were slightly different from those used for other platforms. These buoys differ from the other platforms because reported SST values represent averages over longer times. The SSTs reported by the early TOGA/TAO buoys (prior to about 1991) were averaged over periods ranging between 4 and 24 hours. To limit temporal variability, we only included SSTs from those TOGA buoys with averaging times of 8 hours. The more recent TOGA-TAO data include more frequent (hourly) SST observations. However, the reported SSTs are the average of six measurements taken every 10 minutes; the reporting time is the end of the averaging period.

For both early and recent data from the TOGA-TAO buoys, therefore, the center of the matchup temporal window was the center of the averaging period. The in situ time listed in the PFMDB for the TOGA/TAO buoys also corresponds to the center of the averaging period. For instance, if a TOGA/TAO buoy reported an SST at 6:00 UTC and the averaging period was 4 hours (from 2:00 to 6:00 UTC), the time used as the center of the matchup temporal window (and reported in the PFMDB) was 4:00 UTC. Furthermore, because of the longer averaging period of the TOGA/TAO buoys, the temporal matchup window for early data (prior to 1992) was wider than for all other platforms. Regardless of the actual averaging period (4–8 hours), the matchup temporal window for TOGA/TAO observations for 1985–1991 was set to ± 4 hours from the central time (the center of the averaging window). The wider matchup window coincided with the maximum averaging period accepted. The output of the matchup process is a series of records which contain both satellite-derived and in situ data. Recently it came to our attention that the geographical position of TOGA/TAO buoys listed in the original archive file was inaccurate. The original files at the NOAA/PMEL archive listed the position to the nearest degree. We compared the actual positions obtained via the ARGOS satellite to those listed in the archive for 1996; 58% of the TOGA/TAO buoys were within 10 Km and 90% are within 30Km of the nominal archive position. Beginning in 1997, the TOGA/TAO buoys present in the matchups will show the ARGOS-recorded geographic position.

The last step in the matchup process is to check whether a given matchup record is unique. Duplicate matchups may occur in the unfiltered matchups because consecutive GAC files have a small temporal overlap. Consecutive passes are recorded on alternate tape recorders. Typically, the second recorder is started before the first one is stopped, thus creating the overlap. This means that GAC data may be extracted twice for an in situ observation located in the overlap area. The in situ part of the records will coincide for the duplicate matchups (i.e., the same buoy ID, latitude, longitude and SST will appear in both records), but the satellite time and locations may differ slightly because the automatic navigation and clock corrections are estimated separately for each GAC file. Because of this small difference in satellite times or locations, two records that should be considered as duplicate will be considered "unique" by sorting routines. To eliminate these duplicate records, a test was implemented that excludes one of two records with the same in situ part and satellite times that are less than 5 minutes apart (although differences in satellite times for overlapping records are usually much smaller, of the order of a few seconds). This test also eliminates multiple in situ reports from a drifting buoy corresponding to the same AVHRR pass.

In a small number of cases, the satellite part of the matchup record is the same for two in situ records collected a short time apart. For instance a few NDBC buoys report data every 30 minutes. In these cases, therefore, the same satellite extraction may satisfy the matchup criteria for two consecutive in situ observations (if the satellite time falls between the two in situ measurements). In contrast, when the matchup temporal window is ± 30 minutes, no in situ record can match two satellite passes (for the same spacecraft) because the orbital period of the NOAA spacecraft is of the order of 100 minutes. However, when a wider (± 4 hours) matchup temporal window is used for the early TOGA/TAO data it is possible for one in situ report to match more than one satellite pass.

The end result of the three steps described so far is a set of matchup records with both in situ and AVHRR quantities. The next step is the identification of cloud-contaminated matchups. The description of the cloud-flagging tests, however, makes reference to a number of the fields included in the matchups which have not been discussed in detail so far. Therefore, in the following section we will describe all the fields in the matchup database. The discussion of the remaining processing steps (cloud-flagging and ancillary data addition) is resumed in subsequent sections.

 [Matchups Home](#)  [Previous Section](#)  [Next Section](#)

Page last Updated: Saturday, June 30, 2001 at 3:47 PM

Contact: Guillermo Podestá (gpodesta@rsmas.miami.edu),
Telephone: +1.305.361.4142