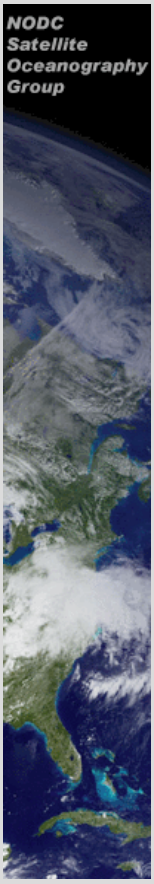


4 km Pathfinder Version 5.0 User Guide

Last Updated: 17 November 2004

Quick Links

- **Recent Changes:** See the Current Status page for a listing of changes to the dataset.
- **Data Access:** See the Available Data page for quick access to the dataset.
- **Contacts:** RSMAS: Edward J. Kearns; PO.DAAC: Jorge Vasquez; NODC: Kenneth S. Casey
- **PDF Version:** PFV50_UserGuide.pdf

Abstract

The 4 km AVHRR Pathfinder Version 5.0 SST Project (Pathfinder V5) is a new reanalysis of the AVHRR data stream developed by the University of Miami's Rosenstiel School of Marine and Atmospheric Science (RSMAS) and the NOAA National Oceanographic Data Center (NODC). In partnership with NODC and RSMAS is NASA's Physical Oceanography Distributed Active Archive Center (PO.DAAC), which has years of experience serving and developing earlier versions of the Pathfinder dataset. Currently in the third year of a three-year demonstration effort, it is hoped that this system can be implemented as an ongoing effort as part of a broader SST climate data record system. This reprocessing uses an improved version of the Pathfinder algorithm and processing steps to produce twice-daily global SST and related parameters back to 1985, at a resolution of approximately 4 km, the highest possible for a global AVHRR data set. Temporal averages for 5-day, 7-day, 8-day, Monthly, and Yearly periods are also produced. Current key improvements over the original 9 km Pathfinder SST data set include a more accurate, consistent land mask, higher spatial resolution, and inclusion of sea ice information. Additional improvements including better flagging of aerosol-contaminated retrievals and the provision of wind and aerosol ancillary data are being implemented in the Version 6 reprocessing to take place during FY2005. See http://www.nodc.noaa.gov/sog/pathfinder4km for more information.

Background

The NOAA/NASA AVHRR Oceans Pathfinder Program has been developing global SSTs at 9.28 km resolution since the early 1990's and presently releases a Version 4.x (V4) dataset. The V4 dataset, distributed through the JPL's PO.DAAC (http://podaac.jpl.nasa.gov/sst/), has found widespread use and is among their most popular products. For more information regarding the V4 Pathfinder data, see:

Kilpatrick, K.A., G.P. Podesta and R. Evanso (2001). Overview of the NOAA/NASA Advanced Very High Resolution Radiometer Pathfinder algorithm for sea surface temperature and associated matchup database, *Journal of Geophysical Research-Oceans*, 106 (C5): 9179-9197 MAY 15 2001.

Despite its success, some problems with the V4 data set limit its use. For example, biases are known to exist in areas strongly impacted by atmospheric aerosols. Additionally, the relatively crude landmask creates significant problems for coastal applications and lack of sea ice information limits the use of the data in the ice-prone high latitudes. This new reprocessing by RSMAS/NODC seeks to correct these deficiencies and provide a higher resolution and ultimately

more accurate SST product by the end of the three-year effort. This global 4 km resolution dataset is termed "Pathfinder Version 5.0 (V5)" since it builds directly on the V4 dataset, including the use of the same set of matchup coefficients.

Satellites Used

The following table describes which NOAA polar orbiting satellites were used in the Pathfinder Version 5.0 processing for SST. The table gives dates in both YYYYDDD format (4 digit year followed by 3 digit day of year) and YYMMDD format (2 digit year, 2 digit month, 2 digit day of month).

Satellite	YYYYDDD	YYMMDD
NOAA-9	1985004 - 1988312	850104 - 881107
NOAA-11	1988313 - 1994256	881108 - 940913
NOAA-9	1994257 - 1995021	940914 - 950121
NOAA-14	1995022 - 2000285	950122 - 001011
NOAA-16	2000286 - present	001012 - present

Note that there are differences between the satellites chosen during 2000 and beyond between the old V4 Pathfinder and the new V5 data. Changes were required due to the detection of a digitizer error in NOAA-14 data. Please see the following article for complete details on this problem:

Podesta, G.P., M. Arbelo, R. Evans, K. Kilpatrick, V. Halliwell, and J. Brown (2003). Errors in high-latitude SSTs and other geophysical products linked to NOAA-14 AVHRR channel 4 problems, *Geophysical Research Letters* , vol 30, no 11., 1548, doi: 10.1029/2003GL017178.

Up to date information on the status of the NOAA polar orbiting satellites can be found at the NOAA Office of Satellite Operations POES Status web site and from the Weekly NOAA Polar Orbiting Spacecraft Status Reports.

Times of Observation

A common question about the Pathfinder SSTs concerns the time of observation for each pixel. This question is a complicated one because there are several factors to consider, some having to do with the satellite orbits and some having to do with the manner in which the data are processed. Each satellite in the NOAA polar orbiting satellite series was launched in a specific orbit, with specified ascending and descending node times (the ascending node time is the time when the satellite passes from south to north over the equator and the descending node time is when it passes from north to south over the equator). Each orbit requires about 102 minutes to complete, and the swaths observed by the satellite overlap, resulting in possibly multiple observations from multiple orbits at any given spot on Earth. This overlap is more pronounced at higher latitdes. Since the Pathfinder algorithm combines the multiple observations (as long as they have they same quality level), the actual time of observation of any given pixel is not known precisely. While an average, or median, time could be determined during the processing, it would require some significant changes to the processing code making it difficult to carry out under the current funding situation. To complicate matters, the actual orbits of the satellites drift over time, introducing a temporal component as well. The table below (assembled from information in Table 1.2-1 of the NOAA Polar Orbiter Data User's Guide and Weekly NOAA Polar Orbiting Spacecraft Status Reports) shows the Local Solar Times (LST) at launch and the values they had drifted to by later dates for the satellites used in the Version 5.0 Pathfinder algorithm (NOAA-17 is also given for reference). Drift rates are also given if known.

Satellite	Ascending and Descending Node Times (LST)	Drift Rate
NOAA-9	Ascending: 1420 (Jan 1985) to 2116 (Mar 1995)	
	Descending: 0220 (Jan 1985) to 0916 (Mar 1995)	
NOAA-11	Ascending: 1330 (Sep 1988) to 1723 (Mar 1995)	
	Descending: 0130 (Sep 1988) to 0523 (Mar 1995)	
NOAA-14	Ascending: 1330 (Dec 1994) to 2006 (Nov 2004)	+5.0 min/month
	Descending: 0130 (Dec 1994) to 0806 (Nov 2004)	
NOAA-16	Ascending: 1400 (Sep 2000) to 1424 (Nov 2004)	+1.0 min/month
	Descending: 0200 (Sep 2000) to 0224 (Nov 2004)	
NOAA-17	Ascending: 2200 (Jun 2002) to 2224 (Nov 2004)	
	Descending: 1000 (Jun 2002) to 1024 (Nov 2004)	+1.0 min/month

Parameters Produced

In addition to the SST values, the Pathfinder V5 Project is making several other useful parameters available. Among other uses, these parameters can be used to calculate higher-order statistics, examine in detail the quality tests passed and failed, and to individually determine what consitutes a "good" SST value by selecting SSTs based on their overall quality assignment. The seven parameters currently available are listed below.

1. "All-pixel" SST
   The all-pixel SST files contain values for each pixel location, including those contaminated with clouds or other sources of error. The Overall Quality Flag values may be used to filter out these unwanted values. The SST value in each pixel location is an average of the highest quality AVHRR Global Area Coverage (GAC) observations available in each roughly 4 km bin.

2. First-guess SST
   The Pathfinder algorithm uses a first guess SST provided by the Reynolds Optimally Interpolated SST (OISST), Version 2 product. The OISST V2 is also used in the quality control procedures.

3. Number of Observations
   This parameter indicates the number of AVHRR GAC observations falling in each approximately 4 km bin.

4. Standard Deviation
   This is the standard deviation of the observations in each 4 km bin.

5. Overall Quality Flag
   The overall quality flag is a relative assignment of SST quality based on a hierarchical suite of tests. The Quality Flag varies from 0 to 7, with 0 being the lowest quality and 7 the highest. For more information regarding the suite of tests, see the Kilpatrick et al. (2001) paper cited above.

6. Mask 1
   These files contains a mask code, which along with Mask 2, can be to determine the tests in the hierarchical suite that were passed or failed, resulting in the Overall Quality Flag.

7. Mask 2
   These files contains a mask code, which along with Mask 1, can be to determine the tests in the hierarchical suite that were passed or failed, resulting in the Overall Quality Flag.

Additionally, in the Version 6.0 data to be produced during FY2005, wind speeds and aerosol

content will also be provided.

Data Format

The data are all stored in Hierarchical Data Format Verison 4 (HDF 4), using the Scientific Data Set (SDS) model. HDF-SDS is a self-describing format capable of storing multiple layers of data as well as metadata describing the file contents. As of July 2003, the metadata do not yet conform to FGDC standards, though work is underway to meet that goal. HDF's capacity to chunk, or tile, the data is taken advantage of, with 512 by 512 element tiles. This internal compression scheme enables both efficient access to subsets of the data as well as enormous savings in disk volume required compared to uncompressed requirements.

Each of the seven parameter files listed above contains a mapped array with 8192 elements in longitude and 4096 in latitude, plus a vector of length 8192 identifying the longitudes and a vector with 4096 values indicating the latitudes. There are also global tags describing the entire contents as well as tags describing each of the 2 vectors and 1 array. The seven parameters are stored either as 8-bit or 16-bit unsigned integers which may be converted linearly (y = mx + b) to geophysical units using a scale (i.e., slope=m) and offset (i.e., intercept=b) according to the following table:

PARAMETER	# BITS	SCALE	OFFSET	UNITS
"All-pixel" SST	16 bit	0.075	-3.0	Deg C
First-guess SST	16 bit	0.075	-3.0	Deg C
Standard Deviation	16 bit	0.150	0.0	Deg C
Number of Observations	8 bit	1.000	0.0	Unitless
Overall Quality Flag	8 bit	1.000	0.0*	Unitless
Mask 1	8 bit	1.000	0.0	Unitless
Mask 2	8 bit	1.000	0.0	Unitless

* - Note that the offset parameter in the Overall Quality Flag files is incorrectly set to a value of 1. The value of 0 listed here is correct.

Information for GIS Users

While not directly readable in most GIS systems, mechanisms are being developed to dynamically convert HDF files into suitable formats. Upon request, NODC will convert specific files into GeoTIFF format (see http://remotesensing.org/geotiff/geotiff.html) which is capable of being of being read by most GIS software.

Map Projection Information

The following combination of parameters works with software, such as GIS systems, needing this level of detail (the four digit codes are from the GeoTIFF web site http://remotesensing.org/geotiff/geotiff.html). These pieces of information comprise the related tags in a GeoTIFF file:

```
Geotiff_Information:
   Version: 1
   Key_Revision: 1.0
   Tagged_Information:
      ModelTiepointTag (2,3):
         0               0               0
         -180            90              0
      ModelPixelScaleTag (1,3):
```

```
         0.0439453      0.0439453      0
      End_Of_Tags.
   Keyed_Information:
      GTModelTypeGeoKey (Short,1): ModelTypeGeographic
      GTRasterTypeGeoKey (Short,1): RasterPixelIsArea
      GTCitationGeoKey (Ascii,17): "LONG/LAT    E005"
      GeographicTypeGeoKey (Short,1): GCS_WGS_84
      GeogAngularUnitsGeoKey (Short,1): Angular_Degree
      ProjLinearUnitsGeoKey (Short,1): Linear_Meter
      End_Of_Keys.
   End_Of_Geotiff.

GCS: 4326/WGS 84
Datum: 6326/World Geodetic System 1984
Ellipsoid: 7030/WGS 84 (6378137.00,6356752.31)
Prime Meridian: 8901/Greenwich (0.000000/  0d 0' 0.00"E)
Projection Linear Units: 9001/metre (1.000000m)

Corner Coordinates:
Upper Left    (180d 0' 0.00"W, 90d 0' 0.00"N)
Lower Left    (180d 0' 0.00"W, 90d 0' 0.00"S)
Upper Right   (180d 0' 0.00"E, 90d 0' 0.00"N)
Lower Right   (180d 0' 0.00"E, 90d 0' 0.00"S)
Center        (  0d 0' 0.00"E,  0d 0' 0.00"N)
```

Other systems, such as MATLAB, can read the HDF-SDS format directly and require less detailed information to re-project the data. If using the MATLAB Mapping Toolbox, for example, use a map legend of:

ml = [8192/360 90 -180];

Key Differences Between V5 Pathfinder and V4 Pathfinder

In addition to the provision of the seven parameters listed above (V4 only provides SST, number of observations, and quality flag), there are three key enhancements which distinguish the new V5 data from the older V4 data: improved resolution, ice mask identification, and improved land mask. These are are discussed below.

- Improved Resolution

  While High-Resolution Picture Transmission (HRPT) or Local Area Coverage (LAC) data exist for some parts of the world at roughly 1.1 km resolution, the GAC data at roughly 4.6 km are the highest resolution AVHRR data available globally. When mapped to an equal angle grid, this resolution results in an array of 8196 by 4096 pixels. This improved spatial resolution greatly enhances the utility of the Pathfinder V5 dataset in coastal zones and areas of strong SST gradients.

- Ice Mask Identification

  A tiered approach to flagging pixels compromised by sea-ice is being taken. First, weekly ice-masks derived from the Special Sensor Microwave Imager (SSMI) and processed by Remote Sensing Systems, Inc. (RSS), are used to identify regions with ice concentrations exceeding 1% (in other words, if any ice is present the region is flagged as ice covered). Currently, RSS version 5 data are being used at 25 km resolution. These data are available back to July 1987. Prior to that time, the sea ice information from the Reynolds OISST V2 is used. Because SSMI sensor data are not valid near land, ice concentrations in coastal regions are not properly identified. These gaps are filled using the Reynolds OISST V2 data

at 1-degree resolution. Finally, because some areas appear inconsistent (such as a gap right at the poles in the SSMI data), an automated approach to "backfilling" the ice mask into interior regions has been developed. Therefore, the first tier uses the SSMI data, the second tier uses the OISST V2 data, and the third tier fills gaps using an automated algorithm. SSTs are not generated for pixels identified as ice covered. These areas are assigned an SST pixel value of 0 (SST=-3.0) and a quality flag of 0.

- Improved Land Mask

  A new land mask developed from MODIS data is employed in the V5 Pathfinder Project as well. This mask represents an enormous improvement over the old 9 km land mask used in Pathfinder V4 and earlier data. The new mask is based on a 1 km resolution dataset derived by the USGS Land Processes Distributed Active Archive Center (see http://edcdaac.usgs.gov/modis/mod12q1.html for more info.) The classifications in that USGS dataset are as follows:

```
SHALLOW_WATER          0
LAND                   1
SHORELINE              2
SHALLOW_INLAND_WATER   3
EPHEMERAL_WATER        4
DEEP_INLAND_WATER      5
MODERATE_OCEAN         6
DEEP_OCEAN             7
```

The USGS dataset is at 1 km resolution so must be degraded to 4 km before use in the Pathfinder V5 processing. A liberal approach is taken so if any of the 16 USGS 1 km values within a given 4 km bin are identified as water, then the entire 4 km bin is classified as water in Pathfinder V5. Values of 1, 2, and 4 are selected to represent land; the remaining values (0, 3, 5-7) are classified as water. The result is a global array (8192 x 4096 pixels) of land/water classifications. This file, in which land points have a value of 2, and water a value of 0, may be accessed by following this link: pfv50_land.m04.hdf. In the Pathfinder V5 processing, SSTs and quality levels are not calculated for pixels falling on land. These areas are assigned an SST pixel value of 0 (SST=-3.0) and quality flag of 0.

File Naming Conventions

The file names in the Pathfinder V5 dataset contain a great deal of information. Therefore, they tend to be rather long and somewhat complicated at first examination. Each set of temporally-averaged data uses the following convention:

OBSDATE.[BITCODE][RESO][AVGPERIOD][DAYNIGHT]pfv50-TYPE(-BITS).hdf

Where:

- OBSDATE = The date(s) the observations were made. OBSDATE may be in one of the following forms:
  - Daily Files: YYYYDDD, with:
    - YYYY = 4 digit year of observation (1985-200x)
    - DDD = 3 digit day of year (001-365, or 001-366 for leap years) For help converting between calendar dates and day of year, please see our Calendar Date Conversion Chart
  - 5-Day Files: SYYYYDDD-EYYYYDDD, with:

        - SYYYYDDD = starting 4-digit year and 3-digit day
        - EYYYYDDD = ending 4-digit year and 3-digit day
    - 7-Day Files: YYYYWW, with:
        - YYYY = 4 digit year of observation
        - WW = 2-digit week number between 01 and 52
    - 8-Day Files: SYYYYDDD-EYYYYDDD, with:
        - SYYYYDDD = starting 4-digit year and 3-digit day
        - EYYYYDDD = ending 4-digit year and 3-digit day
    - Monthly Files: YYYYMM, with:
        - YYYY = 4 digit year of observation
        - MM = 2-digit month between 01 and 12
    - Yearly Files: YYYY, with:
        - YYYY = 4 digit year of observation
- BITCODE = Indicates the bit length of the pixel values in the file. "s" is for 16 bit files, "m" is for 8 bit files (see BITS below)
- RESO = Approximate resolution in km. Set to "04" for 4 km files
- AVGPERIOD = Indicates the averaging period used to create the file. AVGPERIOD may be in one of the following forms:
    - Daily Files = "d"
    - 5-Day Files = "5"
    - 7-Day Files = "w"
    - 8-Day Files = "8"
    - Monthly Files = "m"
    - Yearly Files = "y"
- DAYNIGHT = Indicates nighttime, descending pass (1) or daytime, ascending pass (3) observations
- pfv50 = Shorthand for "Pathfinder Version 5.0"
- TYPE = Indicates the type of data stored in the file. TYPE may be one of the following:
    - sst: Pathfinder all-pixel SST
    - bsst: OISST V2 first-guess SST field
    - sdev: Standard deviation
    - num: Number of observations
    - qual: Overall quality value
    - msk1: Quality mask 1
    - msk2: Quality mask 2
- BITS = Number of bits in each pixel. Only present for 16-bit files (-16b)
- hdf = Indicates HDF-SDS Version 4 file format

Examples:

- 1990182.s04d1pfv50-sst-16b.hdf
  A daily, nighttime all-pixel SST file with 16-bit pixel values at 4 km resolution from 1990, day 182.

- 1994181-1994185.m0451pfv50-num.hdf
  The number of observations in a 5-day, nighttime mean at 4 km resolution using observations from 1994, days 181-185.

- 199035.m04w3pfv50-qual.hdf
  The overall quality flag values for the weekly, daytime mean SST for 1994, week 35, at 4 km resolution.

- 1991169-1991176.s0483pfv50-bsst-16b.hdf
  The first-guess SST field used to create the 8-day mean SST for year 1991, days 169-176. The data have a spatial resolution of 4 km and are stored as 16 bit unsigned

integers.

- 199004.s04m1pfv50-sdev-16b.hdf
  The standard deviation of the monthly mean nighttime SST for April 1990, at 4 km resolution, stored in 16-bit format.

How to Access the Data

Work is underway to make these Pathfinder V5 data available through a wide variety of mechanisms, in an attempt to serve the broadest possible array of users. These mechanisms at NODC include:

- **Direct FTP Access**: This is a good approach if you know what you want and have a large number of files you want to download. Simply point your favorite ftp client to ftp://data.nodc.noaa.gov/pub/data.nodc/pathfinder/Version5.0/ and begin downloading files.
- **Simple Web Access**: Simply click on http://data.nodc.noaa.gov/pathfinder/Version5.0 and begin browsing through the file hierarchy. Clicking on any of the files will prompt you to download that file or will launch any application you might have associated with HDF files. If you want to just take a look around and see what is there, or if you know which file or small number of files you want to download, then this is a good approach.
- **OPeNDAP Access**: These data are also served using OPeNDAP (formerly DODS) server so you may access all of these data using that distributed data access system. The base URL is http://data.nodc.noaa.gov/cgi-bin/nph-dods/pathfinder/Version5.0 In addition, you may simply follow the links provided in the table on the Available Data page to access the OPeNDAP directory listings. For a listing of OPeNDAP clients which you can use to access the 4 km Pathfinder data (as well as any other OPeNDAP-enabled data sets), please see the OPeNDAP web site and follow the link called "Download". Examine the list of "Clients" for various mechanisms to access and display the data directly using OPeNDAP. For example, the OPeNDAP Data Connector is an excellent standalone client which will allow you to quickly search, access, browse, and even generate images of the Pathfinder SST data.
- **Live Access Server**: A Live Access Server (LAS) is being developed to provide a web interface to the data. The LAS will provide on-line display, subsetting, and the capability to download the data in a wide variety of formats.
- **Low Resolution Browse Images**: Simply click on http://data.nodc.noaa.gov/pathfinder/Version5.0/browse_images and begin browsing through low resolution (only one out of 32 pixels in latitude and one out of 32 pixels in longitude is displayed) JPEG images for every data file on the system.

To see a table listing the most current information on the data presently available at NODC , please see the Available Data page. Much of the data for Pathfinder V5 are also available through the PO.DAAC's data access systems.

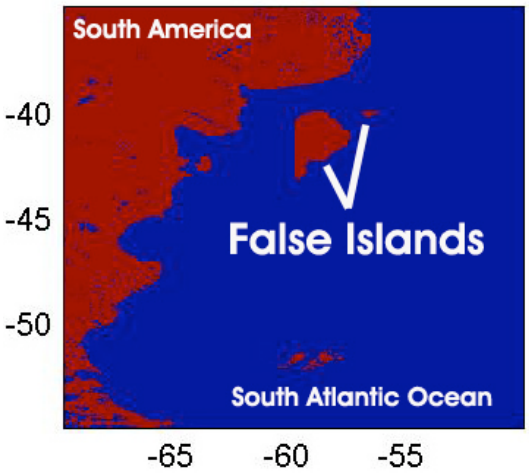
- **PO.DAAC FTP Server**: ftp://podaac.jpl.nasa.gov/sea_surface_temperature/avhrr/pathfinder/data_v5 contains daily all-pixel SST data with quality flags from 1985-current.
- **PO.DAAC POET** : http://podaac.jpl.nasa.gov/poet provided sophisticated online display and subsetting capabilities.

Known Problems

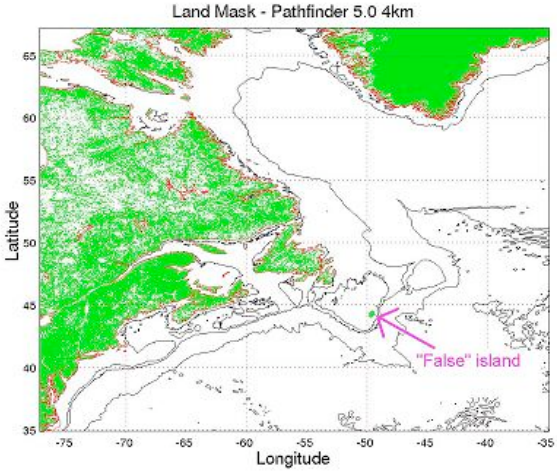
- A problem involving spurious SST gradients has been detected by Dave Ullman and Peter Cornillon at the University of Rhode Island. This problem manifests itself as a jump in SST value every 18 lines and is not clearly visible in a single image. Instead, it was detected

using edge detection algorithms which examine sequences of data. This problem is also known to exist in the older 9 km Pathfinder SST data and has also been seen in MODIS SST data produced at the University of Miami. The problem is related to the mapping procedures by which the equal-area space in which the data are generated are mapped into the equal-angle, uniform grid space in which the data are distributed. Future versions of the Pathfinder dataset will correct this problem.

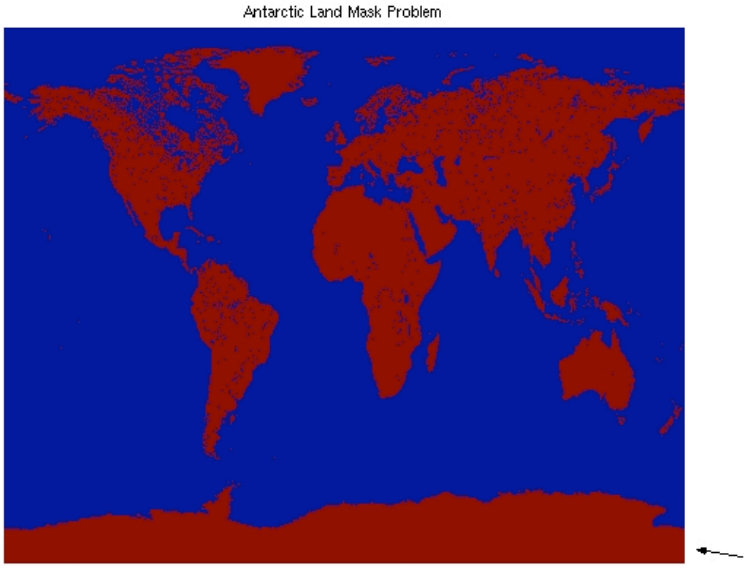
- Biases related to aerosol contamination from Saharan dust storms have been identified in both 9 km and 4 km Pathfinder fields. Techniques to better mask these contaminated pixels and possibly correct them are being investigated.
- A problem with the landmask was detected and reported by Dohyung Koo at the Southampton Oceanography Centre in the UK. He found two areas which are incorrectly classified as land off the east coast of South America, around the Patagonian Shelf (around 35-50S and 70-50W). An image illustrating these "false islands" is shown below.



Another problem has also been identified by Clive Mason (Bedford Institute of Oceanography) near the South East Shoal, as illustrated in the image (with thanks to Dr. Mason) below:



A third land mask problem was identified in the Antarctic. The last column (column = 8192) in the land data file is incorrectly set and indicates water not land. Pixels in rows 3827 through the last row (4096) are affected. The arrow on the figure below points to the area affected, but the area is too small to notice on a global plot.



Improvements to the land mask are being developed and will be implemented in the Pathfinder Verison 5.1 reprocessing. However, these problem are evident in all of the Pathfinder Version 5.0 data files now available.

Advice on Using Pathfinder V5 Data

Under development.

Example Read and Display Software

All of the Pathfinder SST data and ancillary fields are stored in HDF (version 4) Scientific Data Set format, or HDF-SDS. Many tools, libraries, and data analysis and visualization environments are available which work directly with this powerful, self-describing format:

- **General Tools**: HDF is developed and maintained at the National Center for Supercomputing Applications. For access to many tools and libraries, see their HDF4 web site. Remember as you are working with these files that they are HDF-SDS, version 4. They are not HDF-raster and they are not HDF Version 5.
- **Matlab**: Matlab has many tools for working with HDF files and a complete implementation of the HDF library, which enable efficient and direct access to these data. For a GUI interface, examine the help information on "hdftool", or simply do an "hdftool *filename*" to start working with a file. You can also do a "help hdf" for general help in using HDF files in Matlab, and doing a "help hdfsd" will show you a list of functions for working with HDF-SDS files. There is also a simplified function called "hdfread" so you can examine the help information on that as well.
- More to come...

Acknowledgements

Under development.

References

Under development.

4 km Home	Version 5 User Guide	Schedule
Current Status	Available Data	Quality Assurance

Kenneth.Casey@noaa.gov

SOG NODC NOAA CLASS AVHRR SST GODAE MPMC GAC RSMAS GHRSST-PP MCSST NLSST SeaWiFS GOSTA NPOESS VIIRS OPeNDAP DODS LAS HRPT LAC GAC HDF-SDS DMAC PO.DAAC LTSRF