README Document for

Reprocessing and Goddard Satellite-based Surface Turbulent Fluxes (GSSTF) Data Set for Global Water and Energy Cycle Research

Goddard Earth Sciences Data and Information Services Center (GES DISC)
http://disc.gsfc.nasa.gov
NASA Goddard Space Flight Center
Code 610.2
Greenbelt, MD 20771 USA
## Revision History

<table>
<thead>
<tr>
<th>Revision Date</th>
<th>Changes</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>11/19/2012</td>
<td>Updated time coverage</td>
<td>Andrey Savtchenko</td>
</tr>
<tr>
<td>6/25/2012</td>
<td>Document Created</td>
<td>Andrey Savtchenko</td>
</tr>
<tr>
<td>5/6/2015</td>
<td>Tair_2m Corrected</td>
<td>Chung-Lin Shie</td>
</tr>
</tbody>
</table>
# Table of Contents

1 INTRODUCTION ....................................................................................................................... 5
   1.1 DATASET DESCRIPTION ................................................................................................... 5
   1.2 ALGORITHM BACKGROUND ............................................................................................ 6
   1.3 DATA DISCLAIMER ........................................................................................................... 6

2 DATA ORGANIZATION ........................................................................................................... 7
   2.1 HDF-EOS5 PRODUCTS RESULTING FROM GSSTF3 ....................................................... 7
   2.2 FILE NAMING CONVENTION .......................................................................................... 9
   2.3 FILE FORMAT AND STRUCTURE .................................................................................... 10
   2.4 KEY SCIENCE DATA FIELDS .......................................................................................... 11
   2.5 SCIENCE AREA .............................................................................................................. 11

3 DATA CONTENTS ...................................................................................................................... 13
   3.1 DIMENSIONS ................................................................................................................. 13
   3.2 GLOBAL ATTRIBUTES .................................................................................................... 13

4 OPTIONS FOR READING THE DATA .................................................................................... 14
   4.1 COMMAND LINE UTILITIES .......................................................................................... 14
   4.2 TOOLS FOR SIMPLE VISUALIZATION AND FILE CONTENT VIEW ............................. 16
      4.2.1 Panoply .................................................................................................................... 16
      4.2.2 Commercial .............................................................................................................. 16

5 DATA SERVICES ..................................................................................................................... 17
   5.1 MIRADOR ....................................................................................................................... 17
   5.2 OPENDAP ...................................................................................................................... 17

6 MORE INFORMATION ............................................................................................................ 17

7 ACKNOWLEDGEMENTS .......................................................................................................... 19

8 REFERENCES ........................................................................................................................... 19
1 Introduction

This document provides basic information for using Goddard Satellite-based Surface Turbulent Fluxes Version 3 (a.k.a. GSSTF3) Data Set products.

The GSSTF3 consists of products generated for the focus on Global Water and Energy Cycle Research. The global water cycle’s provision of water to terrestrial storage, reservoirs, and rivers rests upon the global excess of evaporation to precipitation over the oceans. Variations in the magnitude of this ocean evaporation excess will ultimately lead to variations in the amount of freshwater that is transported (by the atmosphere) and precipitated over continental regions. The air-sea fluxes of momentum, radiation, and freshwater (precipitation – evaporation) play a very essential role in a wide variety of atmospheric and oceanic problems (e.g., oceanic evaporation contributes to the net fresh water input to the oceans and drives the upper ocean density structure and consequently the circulation of the oceans). Information on these fluxes is crucial in understanding the interactions between the atmosphere and oceans, global energy, and water cycle variability, and in improving model simulations of climate variations.

1.1 Dataset Description

In contrast to previous versions, the GSSTF Version 3 is on a 0.25x0.25 degree grid. It contains only one set of combined data, "Set1", and discontinues the variable “WB”, the lowest 500-m precipitable water (see details addressed as follows). The Version 3 series now covers the period from 1987.07.09 to 2008.12.31.

The GSSTF, Version 3, daily fluxes have first been produced for each individual available SSM/I (Special Sensor Microwave Imager) satellite tapes (e.g., F08, F10, F11, F13, F14 and F15). Then, the Combined daily fluxes are produced by averaging (equally weighted) over available flux data/files from various satellites. These Combined daily flux data are considered as the “final” GSSTF, Version 3, and are stored in this HDF-EOS5 collection.

Similar to the preceding GSSTF2c production (Shie et al. 2011; Shie 2011), a corrected/improved set of SSM/I Tb's (brightness temperature) has been used for producing GSSTF3. Basically, the artificial trends found in the original SSM/I V6 Tb's (used for producing GSSTF2b previously) due mainly to the variations/trends of Earth incidence angle (EIA) in the SSM/I satellites were genuinely removed. Details about the Tb correction can be found in Hilburn and Shie (2011).
1.2 Algorithm Background

The bulk flux model (based on the surface layer similarity theory, Chou, 1993) used for performing the newly produced GSSTF3 is essentially the same as that for GSSTF2 (Chou et al., 2003) and GSSTF2c (Shie 2011). Similar to GSSTF2c, GSSTF3 requires the same methodology and mostly the same kinds of input data such as the surface/10-m wind speeds (U), SST, 2-m air temperature (T_a), and sea level pressure (SLP). In the currently distributed GSSTF2c datasets, the surface air humidity (Qair) retrieval involved applications of the WB-Tb formulation (Schulz et al. 1993) and the EOF method (Chou et al. 1995; 1997). This newly produced GSSTF3 dataset involves an updated Qair retrieval method, i.e., using the Qair-Tb formulation in Bentamy et al. 2003. As such, WB is no longer distributed in GSSTF3. Also note that the SSM/I total precipitable water (W) is no longer used in the GSSTF3 production, but still distributed as one of the output data for providing extra information.

The daily 0.25° GSSTF3 turbulent fluxes were derived from the SSM/I surface winds and surface air humidity, as well as the 2-m air and sea surface temperatures of the NCEP-NCAR reanalysis, using a bulk aerodynamic algorithm based on the surface layer similarity theory. Transfer coefficients, which reflect efficiency of the vertical transportation of momentum, heat, and moisture flux, are a non-linear function of the vertical gradient in wind speed, temperature, and water vapor near the surface, and hence are affected by the stability of the surface air.

The coefficients increase with decreasing wind when the 10-m winds are less than 3 m s^{-1}: The GSSTF2 transfer coefficients (for 10-m winds up to < 18 m s^{-1}) are in close agreement with those of Zeng et al. (1998), Fairall et al. (1996), and Renfrew et al. (2002).

1.3 Data Disclaimer

Data should be used with care and proper citations, i.e., properly indicating your applications with, e.g., "using the combined 2001 data file of Version 3" or "using the 2001 F13 data file".
2 Data Organization

2.1 HDF-EOS5 products resulting from GSSTF3

The original GSSTF3 collections are in binary format, and their organization is described in:


The MEaSUREs project at GES DISC, however, converted all data into HDF-EOS5 format and reorganized it into separate daily and monthly files, in a manner consistent with the on-going Earth Observing System (EOS) missions such as Aura, Aqua, and Terra. The monthly, seasonal and yearly climatologies are also in HDSF-EOS5 format, in separate files. I.e. the daily files are now easily searchable and downloadable by data day (section 5).

The essential meaning of HDF-EOS5 is that the data are now in a standard “Grid” format. The seven major variables are grouped into one separate grid, named “SET1”. Further, the “minor” variable - total precipitable water - is grouped together with the eight major variables in the corresponding SET1. Thus the grid contains 8 variables. This organization is identical for all daily and monthly files, apart from the model reanalysis. The “common”, NCEP/DOE Reanalysis II, data are stored in separate files with one Grid containing the four “common” variables. The climatologies (monthly, seasonal, and yearly) also contain one SET1 grids, but in addition are also containing the four NCEP variables in a separate grid.

The “individual tapes” representing the individual SSM/I satellites are stored in separate collections and daily files, one day per file. They contain one grid that is named on the satellite name (F08, F10, F11, F13, F14, and F15). The grid contains the eight variables (7 major + 1 minor) that go into the computation of the final eight “combined” retrieved variables.

All data within these HDF-EOS5 are easily recognizable as 1440x720 arrays, representing global map, at 0.25x0.25 deg grid cell size. The endian-ess of the remote storage and users’ local machines become irrelevant. The concept of headers and offsets, typical and critical for the binary format, disappears and all that matters are Grid and data field names that are very easy to list out using command line utilities. Examples are listed in section 4.1.
This data organization results in 11 data types, with Short Names listed below. The Short Name is the first string in all filenames, and it is also stored inside the files as a global file attribute “ShortName”.

A. Daily: 
There are total of six (8) daily data types.

GSSTF
GSSTF_NCEP
GSSTF_F08
GSSTF_F10
GSSTF_F11
GSSTF_F13
GSSTF_F14
GSSTF_F15

1) GSSTF  
It has 1 grid, "SET1". Every grid has 8 parameters, the 7 "major" plus the "minor" total precipitable water.

2) GSSTF_NCEP
It has one grid, "NCEP", with 4 parameters. The original were the "common" binary files, that is the NCEP/DOE Reanalysis II.

3) GSSTF_Fxx
These are the individual satellites, where Fxx is one of the following (F08, F10, F11, F13, F14, and F15). The HDF-EOS5 files have only one grid which takes the name of the individual satellite. Every grid has 8 parameters, the 7 "major" plus the "minor" total precipitable water.

B. Monthly:
There are two Monthly data types in he5.
GSSTFM
GSSTFM_NCEP

1) GSSTFM
It has two (2) grids: SET1 and SET1_INT, where "INT" stands for interpolated. Each of these has 8 parameters: the 7 "major" plus the "minor" total precipitable water.

2) GSSTFM_NCEP
This is the "common" monthly NCEP/DOE Reanalysis II. It has one grid with four parameters.
C. Climatology
There are three (3) climatological data types: Monthly, Seasonal, and Yearly in HDF-EOS5:
GSSTFMC
GSSTFSC
GSSTFYC

The climatologies (monthly, seasonal, and yearly) also contain the SET1 grid, but in addition are also containing the four NCEP/DOE Reanalysis II variables in a separate grid, “NCEP”.

Summary of All Data Types:
GSSTF
GSSTF_NCEP
GSSTF_F08
GSSTF_F10
GSSTF_F11
GSSTF_F13
GSSTF_F14
GSSTF_F15
GSSTFM
GSSTFM_NCEP
GSSTFMC
GSSTFSC
GSSTFYC

2.2 File Naming Convention

The file naming convention for the non-climatological HDF-EOS5 files produced at GES DISC for the GSSTF3 project is as follows:

ShortName.vv.yyyy.mm.dd.he5

Where:
• ShortName = one of the following Data Types:
  GSSTF
  GSSTF_NCEP
  GSSTF_F08
  GSSTF_F10
  GSSTF_F11
GSSTF_F13
GSSTF_F14
GSSTF_F15
GSSTFM
GSSTFM_NCEP
GSSTFMC
GSSTFSC
GSSTFYC

- vv = 3 for this release
- yyyy = data year
- mm= data month
- dd = start date for the data
- he5 = commonly accepted extension for HDF-EOS5 files.

Filename example for the daily “combined” turbulent fluxes retrieval, for November 1, 2000:
GSSTF.3.2000.11.01.he5

Climatologies have slightly different file names that reflect the Month (for monthlies), and the range of Months (for seasonal), and the range of years for which the climatology is built. The yearly climatology has one file only.

Example file name for Monthly climatology for November:

Example file name for seasonal climatology for September-November:

2.3 File Format and Structure

GSSTF3 files are in HDF-EOS5 format, which is an extension of the Hierarchical Data Format Version 5 (HDF-5), developed at the National Center for Supercomputing Applications http://www.hdfgroup.org/. This extension facilitates the creation of a standard Grid data structures that are relevant to this project. All data arrays represent global map of size 1440x720 (columns x rows) grid cells.

Thanks to the HDF-EOS5 format, the path to the data fields inside the file is always standardized. For the Grid format, it is always of the form:

/HDFEOS/GRIDS/<grid_name>/Data Fields/<data_field_short_name>.
For example, the sensible heat flux in all relevant GSSTF3 files in HDF-EOS5 will be found under this path:

/HDFEOS/GRIDS/SET1/Data Fields/H

Thus, users need to know only the grid_name and the data_field_short_name. The rest of the path is standard.

2.4 Key Science Data Fields

In the tables below, the data_field_short_name is the one that is used in the HDF-EOS5 path inside the files to uniquely identify a data field. The long name is an attribute to the particular data field that serves as a short description of the field.

Each SET1 grid (including interpolated, SET1_INT) contains the following 8 science data fields:

<table>
<thead>
<tr>
<th>data_field_short_name</th>
<th>Data field long_name (units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;E&quot;</td>
<td>'latent heat flux' (W/m**2)</td>
</tr>
<tr>
<td>&quot;STu&quot;</td>
<td>'zonal wind stress' (N/m**2)</td>
</tr>
<tr>
<td>&quot;STv&quot;</td>
<td>'meridional wind stress' (N/m**2)</td>
</tr>
<tr>
<td>&quot;H&quot;</td>
<td>'sensible heat flux' (W/m**2)</td>
</tr>
<tr>
<td>&quot;Qair&quot;</td>
<td>'surface air (~10-m) specific humidity' (g/kg)</td>
</tr>
<tr>
<td>&quot;U&quot;</td>
<td>'10-m wind speed' (m/s)</td>
</tr>
<tr>
<td>&quot;DQ&quot;</td>
<td>'sea-air humidity difference' (g/kg)</td>
</tr>
<tr>
<td>&quot;Tot_Precip_Water&quot;</td>
<td>'total precipitable water' (g/cm**2)</td>
</tr>
</tbody>
</table>

Each NCEP Grid contains the following 4 data fields:

<table>
<thead>
<tr>
<th>Data field short name</th>
<th>Data field long_name (units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;SST&quot;</td>
<td>'sea surface skin temperature' (C)</td>
</tr>
<tr>
<td>&quot;Psea_level&quot;</td>
<td>'sea level pressure' (hPa)</td>
</tr>
<tr>
<td>&quot;Tair_2m&quot;</td>
<td>'2m air temperature' (C)</td>
</tr>
<tr>
<td>&quot;Qsat&quot;</td>
<td>'sea surface saturation humidity' (g/kg)</td>
</tr>
</tbody>
</table>

2.5 Science Area

The global water cycle’s provision of water to terrestrial storage, reservoirs, and rivers rests upon the global excess of evaporation to precipitation over the oceans. Variations in the magnitude of this ocean evaporation excess will ultimately lead to variations in the amount of freshwater that is transported (by
the atmosphere) and precipitated over continental regions. The air-sea fluxes of momentum, radiation, and freshwater (precipitation – evaporation) play a very essential role in a wide variety of atmospheric and oceanic problems (e.g., oceanic evaporation contributes to the net fresh water input to the oceans and drives the upper ocean density structure and consequently the circulation of the oceans). Information on these fluxes is crucial in understanding the interactions between the atmosphere and oceans, global energy, and water cycle variability, and in improving model simulations of climate variations.
3 Data Contents

3.1 Dimensions

All data fields in the GSSTF3 data types represent global maps of size 1440x720 (columns x rows) grid cells. This corresponds to 0.25x0.25 degree cylindrical map representation. There only two dimension names in the grid structure, Xdim and Ydim, that address the longitude, and latitude, respectively.

3.2 Global Attributes

In addition to SDS arrays containing variables and dimension scales, global metadata is also stored in the files. Some metadata are required by standard conventions, some are present to meet data provenance requirements and others as a convenience to users of GSSTF3 products. Under the HDF-EOS5 standards, the global (file) attributes are grouped under the following path:

/HDFEOS/ADDITIONAL/FILE_ATTRIBUTES

<table>
<thead>
<tr>
<th>Global Attribute</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BeginDate</td>
<td>Character string</td>
<td>Begin time of data</td>
</tr>
<tr>
<td>EndDate</td>
<td>Character string</td>
<td>End time of data</td>
</tr>
<tr>
<td>ShortName</td>
<td>Character string</td>
<td>Short name of the data type (collection) from section 2.1</td>
</tr>
<tr>
<td>LongName</td>
<td>Character string</td>
<td>Long name of the collection</td>
</tr>
<tr>
<td>CollectionDescription</td>
<td>Character string</td>
<td>Description of the collection</td>
</tr>
<tr>
<td>DOI</td>
<td>Character string</td>
<td>Digital Object Identifier</td>
</tr>
</tbody>
</table>

Apart from the global (file) attributes, every data field has its own “local” attributes. Presently there are only three local attributes:

_FillValue
long_name
units

_FillValue=-999 is preserved from the binary originals for the entire GSSTF3 collection. The long_name and the Units were described above and can be found in the tables of sec. 2.4.
4 Options for Reading the Data

4.1 Command Line Utilities

The HDF Group lists a number of HDF5 command line tools on their website:

http://www.hdfgroup.org/products/hdf5_tools/index.html#h5dist

Among them, the one that is the most useful for previewing HDF5 content is h5dump. Since the HDF-EOS5 is just a standard-structured HDF5, the entire functionality of h5dump is fully preserved. H5dump is the tool to use to have a very first look at any HDF5-formatted file, even if no documentation describing data is available. The tool by itself does not need extensive documentation, it is very straightforward to use and its all options can be conveniently listed at the command line using the help option “-h”:

h5dump -h

As of the current naming setup of the data fields, a simple list request “-n” will produce the following output, where full paths to groups and individual data fields (dataset) are obvious:

h5dump -n GSSTF.3.1998.12.19.he5

HDF5 "GSSTF.3.1998.12.19.he5" {
FILE_CONTENTS {
  group / 
  group /HDFEOS 
  group /HDFEOS/ADDITIONAL 
  group /HDFEOS/ADDITIONAL/FILE_ATTRIBUTES 
  group /HDFEOS/GRIDS 
  group /HDFEOS/GRIDS/SET1 
  group /HDFEOS/GRIDS/SET1/Data Fields 
  dataset /HDFEOS/GRIDS/SET1/Data Fields/DQ 
  dataset /HDFEOS/GRIDS/SET1/Data Fields/E 
  dataset /HDFEOS/GRIDS/SET1/Data Fields/H 
  dataset /HDFEOS/GRIDS/SET1/Data Fields/Qair
It is easy to distinguish the Set1 group, and the associated data fields that are always under the group “Data Fields”. The dataset “StructMetadata.0” is written by the HDF-EOS5 libraries, during the process of the Grid structures creation. It can also be inspected by more advanced users, to validate dimensions and data organization (note, some outputs can be very long, so they are abbreviated here for clarity):

```
```

The option -d can be used to list numerical datasets, too:

```
```

(75,165): 196.857, 191.538, 191.316, 184.728, 184.373, 198.597, 188.133,
(75,172): 178.673, 177.775, 181.975, 169.355, 146.068, 124.323, 112.992,
where numbers in parentheses indicate (row, column).

4.2 Tools for simple visualization and file content view

Among these, HDFView is one of the simplest to use and install on your local desktop. A strength of HDFView is that it is good for both, HDF4 and HDF5, formats and it will open files in both formats seamlessly for the user. Another strength is that file content is presented graphically in a very friendly fashion – all groups are presented as folders where user can easily drill down the file hierarchy, which is especially useful for HDF5. Numerical data sets can be viewed as spreadsheets, and as images (multi-dimensional datasets). Simple data manipulations are possible. For download and more information, follow this link:


There are many other free tools for HDF viewing.

4.2.1 Panoply
Developed at the Goddard Institute for Space Studies (GISS), the tool is compliant with NetCDF Climate and Forecast (CF) Metadata Convention that is gaining popularity. A strength of the tool is that data can be previewed “remotely” over the network – i.e. user can preview file content of HDF files stored on a remote site, without downloading them. Panoply is available from
GISS:http://www.giss.nasa.gov/tools/panoply/

With Panoply, GSSTF data types are best viewed through OPeNDAP (more on this service below), by entering the following in the “file”->“open remote catalog” dialog:

http://measures.gsfc.nasa.gov/opendap/GSSTF/catalog.xml

4.2.2 Commercial
IDL and MatLAB, are two major commercial data languages, coming with their libraries that fully support all HDF formats. The HDFEOS group has added simple IDL and MatLAB recipes on their web site in particular for GSSTF data in the HDF-EOS5:

http://hdfeos.org/zoo/index_openGESDISC_Examples.php
5 Data Services

5.1 Mirador

GES DISC provides basic temporal and advanced (event) searches through its search and download engine, Mirador:

http://mirador.gsfc.nasa.gov/

Mirador offers various download options that suit users with different preferences and different levels of technical abilities. Users can start from a point where they don’t know anything about these particular data, its location, size, format, etc., and quickly find what they need by just providing relevant keywords, like “GSSTF3”, or “flux”, or “turbulence”, or just “Heat”.

5.2 OPeNDAP

OPeNDAP stands for “Open-source Project for a Network Data Access Protocol”. OPeNDAP is a framework that simplifies all aspects of scientific data networking. It provides simple means for parameter and spatial subset. The spatial subset is actually a subset by array index - array being the grid data. In the most simplistic case, OPeNDAP can be used to convert data from HDF-EOS5 to ASCII. The data directory hierarchy, as served by OPeNDAP, can be viewed in any browser:

http://measures.gsfc.nasa.gov/opendap/GSSTF/contents.html

6 More Information

All GSSTF3 data types are published to the Global Change Master Directory:

http://gcmd.nasa.gov

This is a centralized depository of climate data information that is catalogued by popular keywords, which facilitate data discovery. Since the Directory is visited by a large number of people working on a broad range of research topics, it is an excellent forum to popularize data collections. It also serves as source of data type information for Mirador.

A special notice issued by Dr. Chung-Lin Shie (PI of this GSSTF Project) on May 6, 2015:
A glitch has recently been found by a careful and thoughtful user on the Daily Tair_2m (originally the NCEP/DOE Reanalysis II) for the period of May 12-31, 1998. This glitch has been further confirmed by this appreciative data provider as due to an unexpected error occurred during the data “regrid” process, i.e., converting data from its original Gaussian format into the global 0.25° x 0.25° grids. The affected Tair_2m and the hence produced Fluxes (E, H, and STu/STv) of 1) the Daily product of May 12-31, 1998; 2) the Monthly-mean product of May 1998; 3) the May Climatology product; 4) the Spring (March/April/May) Climatology product; and 5) the Yearly Climatology product have all been reprocessed and redistributed accordingly. Fortunately, the E, H and STu/STv of the reprocessed product for Daily of May 1998 have generally retained similar patterns qualitatively. The quantitative changes in the fluxes of the reprocessed 1998 May Monthly-Mean are relatively small, while the quantitative changes in the fluxes for all the affected Climatology products are generally negligible, especially for the Spring Climatology and the Yearly Climatology. The quantitative changes, i.e., domain-averaged, in LHF, SHF, STu and STv are 0.8%, 1.5%, 0.65%, and 9.0%, respectively, for the 1998 May Monthly-mean, while the associated changes for the May Climatology are significantly reduced to 0.03%, 0.6%, 0.2%, and 0.2%, respectively. We sincerely apologize for the inconvenience that may have caused for those earlier users (prior to this finding and revision).
7 Acknowledgements

The creators for GSSTF3 are:


8 References


