GPCP Satellite-Gauge (SG) Monthly Precipitation

1. Intent of This Document and POC

1a) This document is intended for users who wish to compare satellite-derived precipitation estimates with climate model output in the context of the CMIP5/IPCC historical experiments. Users are not expected to be experts in satellite-derived Earth system observational data. This document summarizes essential information needed for comparing this dataset to climate model output. References are provided at the end of this document to additional information.

This dataset, which is computed by NASA as a contribution to the Global Water and Energy Exchange (GEWEX) is provided as part of an experimental activity to increase the usability of NASA and related satellite observational data for the modeling and model analysis communities. This particular archive of data is not a standard NASA satellite instrument product, but does represent an effort on behalf of data experts to repackage a standard product that is appropriate for routine model evaluation. The data may have been reprocessed, reformatted, or created solely for comparisons with climate model output. Community feedback to improve and validate the dataset for modeling usage is appreciated. Email comments to HQ-CLIMATE-OBS@mail.nasa.gov .

Dataset File Names (as they appear on the ESG):

pr_GPCP-SG_L3_v2.2_*YYYYM1-YYYYM2*.nc
prStderr_GPCP-SG_L3_v2.2_*YYYYM1-YYYYM2*.nc

where *YYYY* = year
M1 = first month (of the year)
M2 = last month (of the year)

1b) Technical point of contact for this dataset:

George J. Huffman, george.j.huffman@nasa.gov

2. Data Field Description

CF variable name, units:	pr (precipitation_flux), units of kg / m^2 / s
CF variable name, units:	prSterr (precipitation_flux_standard_error), units of kg / m^2 / s
Spatial resolution:	2.5°x2.5° latitude/longitude
Temporal resolution and extent:	Monthly averages, January 1979 – July 2011 in yearly files
Coverage:	latitudes 90°N – 90°S

3. Data Origin

The Satellite-Gauge data set has the identifier "SG" within the Global Precipitation Climatology Project (GPCP). Within the ESG these datasets are posted with file names of the form

pr_GPCP-SG_L3_v2.2_*YYYYM1*-*YYYYM2*.nc
prStderr_GPCP-SG_L3_v2.2_*YYYYM1*-*YYYYM2*.nc

The SG is designed to provide fully global, consistently processed precipitation estimates based on a relatively homogeneous set of input data. This design meets Climate Data Record (CDR) standards, although the SG pre-dates the formulation of the CDR concept. Data are drawn from four sources, namely passive microwave (PMW) radiances at multiple frequencies and polarizations observed from the DMSP sensors at 6 a.m./p.m. during the DMSP epoch (F08, F11, and F13 SSMI, and F17 SSMIS), thermal infrared brightness temperatures (IR Tb; observed by the international constellation of low-Earth-orbit [leo] and geosynchronous-Earth-orbit [geo] satellites), atmospheric soundings computed from National Oceanic and Atmospheric Administration (NOAA)-series Television-Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS) and NASA Aqua Advanced Infrared Sounder (AIRS) data, and surface precipitation gauge measurements. The monthly estimates contained in this data set are an optimal combination of the monthly satellite precipitation estimates and the precipitation gauge analysis (see below).

Each of the PMW data streams is processed into precipitation estimates separately over land and ocean. Over ocean the microwave emission technique infers the quantity of liquid water in a column from the increased low-frequency observed microwave brightness temperatures. Greater amounts of liquid water in the column tend to correlate with greater surface precipitation. The algorithm applied is the Wilheit et al. (1991) iterative histogram approach to retrieving precipitation from emission signals in the 19-GHz SSMI channel. It assumes a log-normal precipitation histogram and estimates the freezing level from the 19 and 22 GHz channels. The fit is applied to the full month of data. Individual estimates on the 2.5ºx2.5º grid occasionally fail to converge. In that case the estimate is set to the separately computed 5ºx5° precipitation estimates available in the box for the month. Over land and coastal surface areas the algorithm reduces to a scattering-type procedure using only the higher-frequency channels. This loss of information arises from the physics of the emission signal in the lower frequencies when the underlying surface is other than all water. The microwave scattering technique infers the quantity of hydrometeor ice in a column from the depressions in the 85 GHz channel brightness temperatures. More ice aloft typically implies more surface precipitation. This relationship is physically less direct than in the emission technique, but it works equally well over land and ocean whenever deep convection is important. However, icy surfaces also exhibit scattering, so no scattering estimates are possible in areas with an icy, snowy, or frozen surface. The algorithm applied is based on the Grody (1991) Scattering Index (SI), supplemented by the Weng and Grody (1994) emission technique in oceanic areas. A similar fall-back approach was used during the period June 1990 - December 1991 when the 85.5-GHz channels were unusable. The scheme showed anomalously high coastal values in many locations, and lacked snow screening. Pixel-by-pixel retrievals are

accumulated onto separate daily ascending and descending 0.333ºx0.333º lat/lon grids, then all the grids are accumulated for the month on the 2.5º grid. Because SSMIS observes at 91 GHz, while the SI expects 85 GHz data, Vila et al. (2012) developed a 91 GHz-based 85 GHz proxy channel.

Through most of the period of record the geo-IR Tb (with leo-IR Tb fill in) data are processed as part of the SG algorithm (see below). However, for the pre-DMSP epoch (January 1979 – July 1987) the leo-IR data are used as Outgoing Longwave Radiation (OLR) Precipitation Index (OPI; Xie and Arkin 1998). Colder OLR radiances are directly related to higher cloud tops, which are related to increased precipitation rates. It is necessary to define "cold" locally, so OLR and precipitation climatologies are computed and a regression relationship is developed for anomalies in OLR and precipitation. In use, the total precipitation inferred is the estimated anomaly plus the local climatological value. A backup direct OLR-precipitation regression is used when the anomaly approach yields unphysical values. In this analysis, the precipitation climatology used to develop the OLR-derived precipitation estimates was based on the GPCP Version 2.2 satellite-gauge estimates over the time period 1988-2007. The resulting spatially and temporally varying climatological calibration is then applied to the independent OPI data covering the span 1979-1987 to fill all months lacking SSMI(SSMIS) data. The OPI data for the first two satellites (covering January 1979 through August 1981) were given additional adjustments, described in Huffman and Bolvin (2012). This adjusted OPI data provides a globally complete proxy for the SSMI(SSMIS) data.

The TOVS(AIRS) precipitation estimate is based on Susskind and Pfaendtner (1989) and Susskind et al. (1997). The TOVS(AIRS) precipitation estimates infer precipitation from deep, extensive clouds. The technique uses a multiple regression relationship between collocated rain gauge measurements and several TOVS(AIRS)-based parameters that relate to cloud volume: cloud-top pressure, fractional cloud cover, and relative humidity profile. This relationship is allowed to vary seasonally and latitudinally. Furthermore, separate relationships are developed for ocean and land. The TOVS data are used for the period August 1987 – April 2005 and are provided at the 1º spatial resolution and at the daily temporal resolution. The data covering the span up to February 1999 are based on information from two satellites. For the period March 1999 – April 2005, the TOVS estimates are based on information from one satellite. In addition, the date span 1-17 February 2004 experienced partial (1st and 17th) or total (2-16) loss of TOVS data, so AIRS data are used for February 2004 and after April 2005.

Summarizing Adler et al. (2003) and Huffman et al. (2009), during the DMSP era TOVS(AIRS) is merged in with SSMI(SSMIS) where the SSMI(SSMIS) is suspect (outside about 45ºN-S) or missing (snowy/icy surfaces). Then SSMI(SSMIS) and geo-IR are approximately time-matched to compute local coefficients to adjust the full geo-IR Geosynchronous Orbit Environmental Satellite (GOES) Precipitation Index (GPI; Arkin and Meisner 1987) to the bias of the SSMI(SSMIS) in the 40ºN-S band. As well, leo-IR GPI is approximately scaled to the SSMI(SSMIS). This Adjusted GPI (AGPI) is built from geo-IR AGPI where possible and leo-IR AGPI elsewhere. The Multi-Satellite (MS) intermediate product is composed of AGPI in the band 40ºN-S and the merged SSMI(SSMIS)–TOVS(AIRS) elsewhere.

During the late pre-DMSP period (January 1986 - July 1987 and December 1987), the OPI data, as calibrated by the GPCP satellite-gauge estimates for part of the DMSP period (1988-2007), are used as a proxy for the merged SSMI(SSMIS)/TOVS(AIRS) field in the AGPI procedure described for the DMSP period. During the early pre-DMSP period (January 1979 – December 1985) there is no geo-IR GPI, and therefore no AGPI. The OPI data, calibrated by the GPCP satellite-gauge estimates for the same part of the DMSP period (1988-2007), are used "as is" for the multi-satellite estimates

Throughout, the gauge analysis is used to remove large-area biases in the MS, then combined with the (debiased) MS data using optimal weighting by the inverse (estimated) error variances to form the GPCP SG combination (ESG datasets of the form pr_GPCP-SG_L3_v2.2_*YYYYM1*-*YYYYM2*.nc). The error variance calculation (following Huffman 1997) results in fields of estimated random error (ESG datasets of the form prStderr_GPCP-SG_L3_v2.2_*YYYYM1*-*YYYYM2*.nc). In the CMIP5 collection the precipitation and random error are referred to as fields pr (precipitation_flux) and prStderr (precipitation_standard_error). The detailed SG technical documentation (Huffman and Bolvin 2012) is posted at *ftp://precip.gsfc.nasa.gov/pub/gpcp-v2.2/doc/V2.2_doc.pdf*. The GPCP version number for this series of SG is Version 2.2. A summary of the upgrades from Version 2.1 to Version 2.2 are provided in the technical document. Updates are planned to the CMIP collection of GPCP SG after each additional year of the data is computed.

For most of the period of record essentially every grid box has a value, so sampling is not typically an issue. The primary sampling issue is that the Indian Ocean sector lacked geo-IR data before July 1998. As a partial offset, we employed GPI data computed from LEO-IR data, but even in combination with the PMW data the sampling is reduced. Otherwise, the sampling inside the latitude band 40°N-S during the DMSP era is higher than the zones outside and than the entire globe in the early pre-DMSP era (1979-1985).

The precipitation research group in the NASA/GSFC Mesoscale Atmospheric Processes Laboratory is responsible for technical development and maintenance for the SG. Gerald L. Potter developed the conversion routines to CMIP-standard files.

4. Validation and Uncertainty Estimate

Combinations are difficult to validate as they tend to include data that would otherwise be independent. An early validation of the old Version 1a data set against the Surface Reference Data Center analysis yields the statistics in Table 1. Revised statistics are being developed for Version 2.2. Overall, the combination appears to be working as expected.

Analysis against dense gauge data in Finland also shows reasonable behavior for the Version 2 monthly SG, with better results in the summer than the winter (Bolvin et al. 2009). Overall, the original SG appears to have worked as expected in both the AGPI and TOVS(AIRS) data, and this should continue to be true in Version 2.2.

Krajewski et al. (2000) develop and apply a methodology for assessing the expected random error in a gridded precipitation field. Their estimates of expected error agree rather closely with the errors estimated for the multi-satellite and satellite-gauge combinations.

Table 1. Summary statistics for all cells and months comparing the Version 1a SSMI composite, Multi-Satellite, Gauge, and Satellite-Gauge products to the SRDC analysis for July 1987 – December 1991.

Product	*Bias (mm/mo)*	*Avg. Diff. (mm/mo)*	*RMS Error (mm/mo)*
SSMI composite	4.03	60.10	88.05
Multi-satellite	-5.80	44.20	62.47
Gauge (GPCC)	6.77	18.85	35.11
Satellite-Gauge	3.70	20.29	32.98

An initial comparison between Versions 2 and 2.1 was developed as part of the transition to Version 2.1 and appeared in Huffman et al. (2009). It is summarized in the technical document, with an example in Table 2.

Table 2. Global- and tropical-average land, ocean, and total precipitation for Versions 2.1 and 2 in mm/d. The percentage increase of Version 2.1 over Version 2 is given in parentheses. "Ocean" and "land" regions are defined by 100% and <100% coverage by water.

1979-2007	Global 90°N-90°S		Tropical 25°N-25°S	
Version	2	2.1	2	2.1
Land and Ocean	2.62	2.68 (+2%)	3.12	3.22 (+3%)
Land	2.39	2.53 (+6%)	3.49	3.73 (+7%)
Ocean	2.78	2.78 (+0%)	2.88	2.88 (+0%)

An initial comparison between Versions 2.1 and 2.2 is being developed. In general the changes are much smaller than for the transition from 2 to 2.1. The only additional data boundary to note is the transition from SSMI to SSMIS starting with January 2009. The total, land, coast, and ocean averages for each of the Versions are given in Table 3. See "accuracy in space/time averages" for comments on the importance of defining the masks for each surface type.

Table 3. Global-average land, coast, ocean, and total precipitation for Versions 2.1 and 2.2 in mm/d. The percentage increase of Version 2.1 over Version 2 is given in parentheses. "Ocean" and "land" regions are defined by 100% and <5% coverage by water.

August 1987 – September 2009	Global 90°N-90°S	
Version	2.1	2.2
Total	2.675	2.681 (+0.2%)
Land	1.918	1.924 (-0.3%)
Coast	3.347	3.341 (-0.2%)
Ocean	2.788	2.800 (+0.4%)

5. Considerations for Model-Observation Comparisons

Collecting the issues raised in other parts of this document:

- There is a boundary between IR- and SSMI(SSMIS)/TOVS(AIRS)-based data at 40°N and S. Typically, discontinuities are not an issue.
- TOVS data are used outside the latitude band 40°N-S in the early part of the record, while AIRS data are used for February 2004 and after April 2005. There is good, but not perfect agreement between the two, with the SG typically having somewhat higher values at high latitudes during the AIRS epoch.
- Likewise, in the latitude belt 40°N-S there is a transition from SSMI to SSMIS starting January 2009. The intercalibration of these two sensor series is still in progress.

As well, a few additional factors should be noted:

- Coastal zones present special challenges for retrievals due to the heterogeneity of the surface scene. In a few cases where the land/ocean contrast in precipitation is strong (such as Jamaica), the gauge values tend to bleed into the surrounding coastal waters on 2.5°x2.5° blocks related to the SG resolution.
- Orographic enhancement of precipitation is sometimes a challenge for the satellite schemes. The issue arises when the enhancement takes place (mostly) in the liquid phase, which current PMW algorithms cannot "see" over land, causing underestimation. On the other hand, in a few places the orography provokes very inefficient storms that create large amounts of ice near cloud top relative to the precipitation reaching the ground. The satellites consequently overestimate the rainfall in these cases.
- Current PMW schemes cannot make retrievals over snowy or frozen surfaces, which yield signals similar to frozen precipitation. The SG substitutes TOVS(AIRS) estimates, but these are presumed to be of lower quality. As a result, statistics over cold-season land situations should be examined for possible degradation by these snow effects.

6. Instrument Overview

The instruments contributing to the SG are drawn from a wide variety of sources. The goal of the SG dataset is to use a relatively homogeneous set of quasi-global precipitation estimates from the international constellation of precipitation-relevant satellites to create a CDR-like product with complete coverage over the chosen domain and period of record (global, 1979-present). Fig. 1 summarizes the periods of record for the various inputs:

- PMW radiometers in the form of selected SSMI and SSMIS conical-scan imagers that fly on the DMSP series, which feature multiple channels and dual polarization well-suited to estimating precipitation; provide constant footprint sizes, although these sizes differ for different channels; and are processed with sensor-specific algorithms for land and ocean separately.
- IR imagers provide Tb's that are converted to GPI before transmission to the GMDC. For the period 1986-March 1998 the geo-GPI data are accumulated on a 2.5ºx2.5º lat/lon grid for pentads (5-day periods). Starting with October 1996 the geo-GPI data are accumulated on a 1ºx1º lat/lon grid for individual 3-hrly images. In both data sets gaps in geo-IR are filled with leo-IR data from the NOAA series of polar orbiting meteorological satellites. However, the 2.5ºx2.5º data only contain the leo-IR used for fill-in, while the 1ºx1º data contain the full leo-IR. The GPI product is based on the 2.5ºx2.5º data for the period 1987-1996, and the 1ºx1º beginning in 1997. The boundary is set at January 1997 to avoid placing the boundary during the 1997-1998 ENSO event.

  In the pre-DMSP epoch leo-IR data are processed into OLR estimates of broadband outgoing longwave radiation based on an algorithm applied to the narrow-band IR channels on the Advanced Very High Resolution Radiometer (AVHRR) aboard the polar-orbiting NOAA series of satellites. Typically two satellites are available, but occasionally the OLR is based on only one satellite. Then, these OLR estimates are converted to OPI values before transmission to the GMDC.
- TOVS(AIRS) soundings are retrieved from a series of sensors. Up through April 2005, the TOVS dataset of surface and atmospheric parameters is derived from analysis of High-Resolution Infrared Sounder 2 (HIRS2) and Microwave Sounding Unit (MSU) data aboard the NOAA series of polar-orbiting operational meteorological satellites. Thereafter, the AIRS dataset of surface and atmospheric parameters is derived from analysis of High-Resolution Infrared Sounder data aboard the Aqua polar-orbiting satellite. In both cases the retrieved fields include land and ocean surface skin temperature, atmospheric temperature and water vapor profiles, total atmospheric ozone burden, cloud-top pressure and radiatively effective fractional cloud cover, outgoing longwave radiation and longwave cloud radiative forcing, and precipitation estimate.

The SG technical document (Huffman and Bolvin 2012) provides expanded summaries for each sensor and references to relevant documentation.

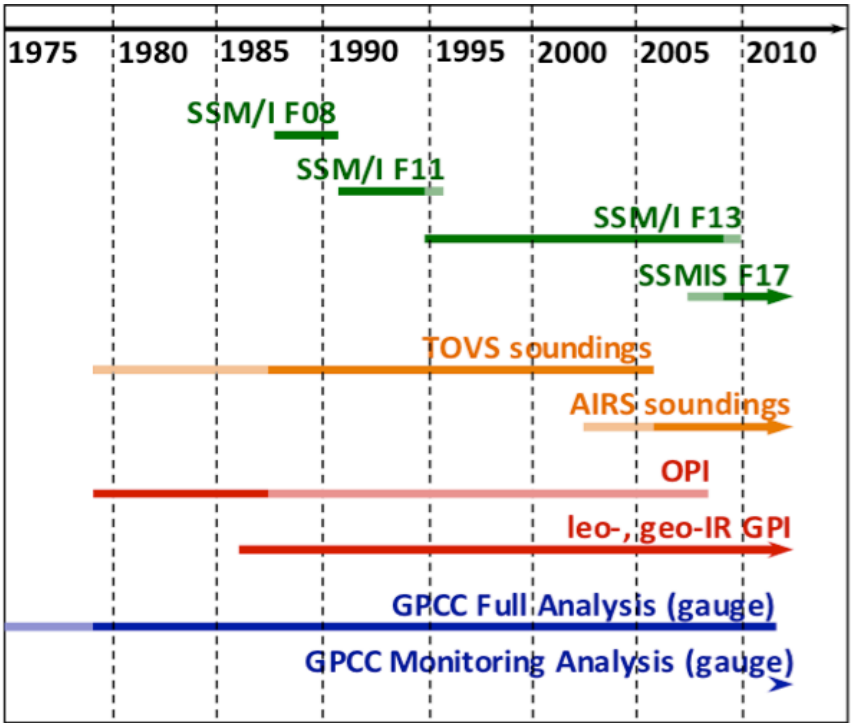

Fig. 1. Periods of record for the various data sets used in computing the SG (solid lines). Some of these sensors' periods of record extend beyond the periods of use, shown in light colors.

7. References

The International Polar Year (IPY) Data policy guidelines (http://.ipydis.org/data/citations.html) suggest a formal reference for data sets of the form

> Huffman, G.J., D.T. Bolvin, R.F. Adler, 2012, last updated 2012: *GPCP Version 2.2 SG Combined Precipitation Data Set*. WDC-A, NCDC, Asheville, NC. Data set accessed <date> at http://www.ncdc.noaa.gov/oa/wmo/wdcamet-ncdc.html.

As an "Acknowledgment", one possible wording is: "The GPCP SG combined precipitation data were developed and computed at the NASA/Goddard Space Flight Center's Mesoscale Atmospheric Processes Laboratory – Atmospheres as a contribution to the GEWEX Global Precipitation Climatology Project."

Additional details: At frequencies below about 37 GHz the radiative transfer signal in PMW sensor channels is primarily a combination of emission from the surface and then from the overlying atmosphere, including cloud and precipitation liquid water. At higher frequencies the useful signal results from scattering of the upwelling radiant energy out of the line of sight. Unfortunately, the land surface is radiometrically emissive and heterogeneous, so current-generation algorithms can only use the emission channels over ocean. The restriction to frozen hydrometeors alone over land is an issue because they only represent the upper reaches of clouds, while the liquid phase tells about precipitation

nearer the surface. Thus, conical-scan radiometers, which span both radiometric regimes, provide better answers over ocean than land. This is also the basis for the issues with retrievals over snowy/frozen surfaces and when orographic enhancement is in the liquid phase.

Data source:

http://www.ncdc.noaa.gov/oa/wmo/wdcamet-ncdc.html

_____, G.J. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, E.J. Nelkin, 2003: The Version 2.1 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979 - Present). *J. Hydrometeor.*, **4**(6), 1147-1167.

Arkin, P.A., and B. N. Meisner, 1987: The relationship between large-scale convective rainfall and cold cloud over the Western Hemisphere during 1982-1984. *Mon. Wea. Rev.*, **115**, 51-74.

Bolvin, D.T., R.F. Adler, G.J. Huffman, E.J. Nelkin, J.P. Poutiainen, 2009: Comparison of GPCP Monthly and Daily Precipitation Estimates with High-Latitude Gauge Observations. *J. Appl. Meteor. Climatol.,* **48**(9), 1843–1857Grody, N.C., 1991: Classification of snow cover and precipitation using the Special Sensor Microwave/Imager (SSM/I). *J. Geophys. Res.*, **96**, 7423-7435.

Grody, N.C., 1991: Classification of snow cover and precipitation using the Special Sensor Microwave/Imager (SSM/I). *J. Geophys. Res.*, **96**, 7423-7435.

Huffman, G.J., 1997: Estimates of root-mean-square random error contained in finite sets of estimated precipitation. *J. Appl. Meteor.*, **36**, 1191-1201.

_____, R.F. Adler, D.T. Bolvin, G. Gu, 2009: Improving the global precipitation record: GPCP Version 2.1. *Geophys. Res. Lett.*, **36**, L17808, doi:10.1029/2009GL040000.

_____, D.T. Bolvin, 2012: GPCP Version 2.2 SG Combined Precipitation Data Set Documentation. *ftp://precip.gsfc.nasa.gov/pub/gpcp-v2.2/doc/V2.2_doc.pdf*, 46 pp.

Krajewski, W.F., G.J. Ciach, J.R. McCollum, C. Bacotiu, 2000: Initial validation of the Global Precipitation Climatology Project over the United States. *J. Appl. Meteor.*, **39**, 1071-1087.

Susskind, J., J. Pfaendtner, 1989: Impact of interactive physical retrievals on NWP. *Report on the Joint ECMWF/EUMETSAT Workshop on the Use of Satellite Data in Operational Weather Prediction: 1989-1993, Vol. 1*, T. Hollingsworth, Ed., ECMWF, Shinfield Park, Reading RG2 9AV, U.K., 245-270.

_____, P. Piraino, L. Rokke, L. Iredell, A. Mehta, 1997: Characteristics of the TOVS Pathfinder Path A Dataset. *Bull. Amer. Meteor. Soc.*, **78**, 1449-1472.

Vila, D., C. Hernandez,R. Ferraro, H. Semunegus, 2012: The Performance of Hydrological Monthly Products Using SSM/I – SSMI/S Sensors. *J. Hydrometeor.*, doi:10.1175/JHM-D-12-056.1, to appear.

Weng, F., N.C. Grody, 1994: Retrieval of cloud liquid water using the Special Sensor Microwave Imager (SSM/I). *J. Geophys. Res.*, **99**, 25535-25551.

Wilheit, T., A. Chang, L. Chiu, 1991: Retrieval of monthly rainfall indices from microwave radiometric measurements using probability distribution function. *J. Atmos. Ocean. Tech*., **8**, 118-136.

Xie, P., P.A. Arkin, 1998: Global monthly precipitation estimates from satellite-observed outgoing longwave radiation. *J. Climate*, **11**, 137-164.

8. Revision History

Rev 0 – 9/30/2012 – This is a new document/dataset [G.J. Huffman]