Reviews of Observational Data Available over Africa for Monitoring, Attribution and Forecast Evaluation

David Parker, Elizabeth Good and Rob Chadwick
Met Office Hadley Centre, Exeter
February 2011; link revised February 2012

1. Introduction

This review was developed in 2010 in support of the Met Office Hadley Centre and Department for International Development Climate Science Research Programme (CSRP). The objective was to record the status of observational dataset availability for climate monitoring, for future attribution, and for forecast evaluation both on seasonal timescales and at the much higher temporal resolution required for onset, duration and dry spell assessments.

Consultations with African and other stakeholders showed the need for rainfall data on 10-day as well as daily and monthly timescales, so all these are included. Basic requirements are in Section 2. Our application-dependent recommendations on the choice of datasets for particular purposes are in Section 3. Annex A is a catalogue of information on temperature and rainfall datasets to guide users in more detail. References, and a list of acronyms and abbreviations, are at the end of Annex A. Soil moisture and aerosol datasets are catalogued likewise in Annex B.

Unsurprisingly, the sparsity of *in situ* data over Africa remains a major theme of this update: satellite data provide very valuable coverage of rainfall estimates, but more *in situ* data are essential for interpretation of the satellite data and for on-the-ground monitoring and evaluation of forecasts.

2. Dataset Requirements

An initial list of requirements for observational data was identified through consultation within the Met Office. African stakeholders then prioritised air temperature and precipitation: datasets are detailed in Annex A. Soil moisture and aerosol were accorded lower priority but preliminary information, gathered early in 2010, is in Annex B.

2.1. Spatial Resolution

Seasonal forecasting is performed at approximately 120 km resolution, with some runs at 60 km, and the decadal model is run at 240 km resolution. Satellite data for validation as well as monitoring therefore become more useful as resolution is improved from 2.5° (~280km at the equator) to 0.5° (<60km at the equator). But these data require much finer resolution *in situ* data for calibration and validation. For example it may take >20 gauges to represent a 2.5° x 2.5° area's monthly rainfall (Section 3.4).

2.2. Temporal Resolution

In addition to daily and monthly temperature and rainfall data, African stakeholders requested 10-day rainfall data. These capture much of the intra-seasonal variability of rainfall and are often more accurate than the available daily datasets.

2.3. Length and Timescale

Based on hindcast evaluation periods, the requirements for data for forecast evaluation are from 1989 and 1960 for seasonal and decadal forecast evaluation respectively. For attribution studies, datasets would ideally span many decades.

2.4. Real Time Data Availability Requirements

At least weekly updates will be required for many applications.

3. Recommendations

3.1. Daily and monthly temperature data.

Available gridded daily temperature data over Africa are limited to HadGHCND, and this has very poor coverage (Section A1.1.4 in the Annex). The underlying GHCN station density is too low for comprehensive monitoring of extremes.

The University of Cape Town is blending daily station temperature data with the ERA-interim Reanalysis for the CORDEX regional model downscaling project. The analysis relaxes back to ERA-interim where there are no nearby data, so it will be affected by any biases in ERA-interim (Bruce Hewitson, pers. comm.) and should be used with caution.

We stress the need for satellite observations to supplement *in situ* surface air temperature analyses especially on daily time-scales. By analogy with the needs for daily rainfall gauge data to develop and validate satellite-based rainfall algorithms (Section 3.4), we recommend the acquisition of national maximum, minimum and fixed-hour surface air temperature data for development and validation of satellite-based land surface temperature products. This would in particular benefit drought-monitoring and could be done bilaterally in the context of CSRP study fellowships.

3.2. Daily precipitation data.

There are significant differences between available daily precipitation datasets. Care should be taken when using gauge-only datasets since they are based on so few *in situ* measurements. The authors of the CPC dataset, for example, specifically point out that the quality of their gauge-analysis (Section A2.1.2) is poor over Africa. They recommend interpreting their precipitation analysis in conjunction with the gauge density information that is supplied in the product to indicate where confidence is particularly low/high. Roca *et al*. (2010) (see Section A2.1.4) discuss the sampling uncertainties of both rain-gauge and satellite estimates of precipitation. The unrealistic 'bulls-eye' appearance of the CPC analysis (Figure A11) is a typical artefact of optimal interpolation where data are sparse. See also Section 3.4 for recommendations on the acquisition of daily as well as monthly gauge data.

So we recommend at present satellite-only or combined gauge-satellite analyses for forecast evaluation and attribution. Based on intercomparison studies at 10-day and monthly scales (Sections A2.2.2 and A2.3.17), a good choice of daily satellite-only product would appear to be TRMM 3B 42 (Section A2.1.3). Nonetheless, even this has biases (Section A2.3.17)

Of the combined satellite-gauge products, the operational RFE 2.0 (Section A2.1.7) is recommended for forecast validation, while the climatological RFE is probably more appropriate for attribution studies, due to its longer record, though it may have biases owing to the varying number of gauges included in the analysis.

The poor quantitative performance of all current satellite rainfall products at high spatial resolution on daily scales should be kept in mind. Averaging these products to larger spatial scales is expected to improve their accuracy.

3.3. 10-day precipitation data.

10-day rainfall estimates are key for forecasting and attribution where sub-monthly information is required, and where daily datasets do not provide sufficient accuracy or reliability. The sparsity of reporting gauges in most of Africa means that satellite-only and combined gauge-satellite datasets are preferable to gauge-only datasets. For forecast verification, intercomparison studies indicate that of the satellite-only products, TAMSAT (Section A2.2.1) and TRMM 3B 42 (Section A2.1.3) may be the best, while the operational RFE 2.0 (Section A2.1.7) is the most suitable combined gauge-satellite product. For attribution studies, the TAMSAT and climatological RFE datasets are most appropriate; TAMSAT may be more consistent over time as RFE includes a gauge component of varying spatial coverage.

Validation studies (Section A2.2.2) show the value of local calibration or constraint of satellite rainfall products, either with gauges (TAMSAT and RFE) or precipitation radar (TRMM 3B 42). This is also expected to apply to other timescales and is discussed further in Section 3.4.

3.4. Monthly precipitation data.

Over the southern Sahel, a rain-gauge density of at least 10 gauges per 2.5° grid box is required to give a monthly precipitation error of less than 10%. The requirement exceeds 20 gauges per 2.5° grid box in the northern Sahel where rainfall is less coherent (Section A2.3.18). The gauge density needed for a given % monthly precipitation error elsewhere in Africa depends on the spatial coherence of monthly rainfall. Mountainous areas (e.g. most of East Africa) and all zones on the edge of monsoon-penetration (e.g. parts of the Sudan, Namibia etc.) are likely to have highly intermittent, patchy rainfall so it is likely that >20 gauges will be needed per 2.5° grid box, and even more for 10-day rainfall which suffers greater, less coherent variations. These requirements are very rarely met in available data for Africa, so in general monthly gauge errors are likely to be much larger than 10%. So improvement of *in situ* data availability is an important objective. This applies for monthly and longer timescales because attribution and seasonal hindcast studies need *in situ* rainfall data predating the satellite era. But daily satellite-era gauge data are also needed, because the validation studies cited in Section A2.2.3 show the value of local calibration or constraint of satellite rainfall products. This is supported by GCOS (2006). Release of monthly data will benefit stakeholders in that long-term trends and their causes can be better assessed; release of daily data will benefit them in that nationally-applicable satellite-based algorithms can be developed and applied to monitoring, studies of extremes, and seasonal forecast verification. This may be best achieved by bilateral agreements in relation to the proposed CSRP study fellowships.

Regarding currently available monthly gauge data, the University of Delaware dataset benefits from the inclusion of Sharon Nicholson's African gauge data for 1950-1996. She has extensive links with African Met services and it is likely that her gauge dataset contains a large number of gauges that have not been included in other datasets. For future development, the University of Cape Town data base includes a dense network for South Africa. At present, the GPCC first guess is recommended for near real time, and the University of Delaware, GPCC reanalysis or GPCC-VASClimO for historical analysis.

Owing to sparsity of available gauge data, we recommend that, at present, monthly datasets with a satellite component should usually be used in preference to those based on gauge-only observations. However we note that even blended satellite-*in situ* datasets may suffer biases owing to changes of *in situ* gauge coverage (Section A2.1.7). Of the satellite-gauge datasets, CMAP and GPCP are long enough (Table A5) and may be sufficiently stable for historical analysis. CMAP is reported to agree well with the GPCP merged analysis over land, and tropical and subtropical oceanic areas, but with some differences over extratropical oceans (Section A2.3.18). TAMSAT appears to be the best option for a historical satellite-only dataset (Section A2.2.3).

For near-real time applications, the CAMS-OPI analysis, operational RFE, TAMSAT or TRMM 3B 43 are recommended

3.5. Collaboration

We emphasise the benefits of collaboration with University of Reading (TAMSAT group), Université Paris 1 (J.C. Bergès: EPSAT-SG), GPCC (Andreas Becker), University of Cape Town (Bruce Hewitson), and Florida State University (Sharon Nicholson), in order to take advantage of the latest improvements and developments in observational datasets for Africa.

Annex A. Catalogue of Temperature and Precipitation Datasets for Africa.

In this Annex, we provide basic information and access details for temperature and precipitation datasets for Africa. Daily, 10-day (precipitation only) and monthly data are presented in turn. References and a list of acronyms and abbreviations are at the end.

A1. Land Surface Air Temperature

A1.1. Daily Observational Datasets

Table A1 provides a summary of available daily temperature datasets. Further details are provided in the following sections.

Table A1: Summary of daily air temperature datasets

Dataset	**Spatial Res.**	**Spatial Coverage**	**Temporal Coverage**	**Variables**
GHCN-D	Station	Global	19th C.-now	Tmin, Tmax
ISD	Station	Global	1901-now	Tmin, Tmax, Tsub-daily
Synops	Station	Global	19th C.-now	Tmin, Tmax, Tsub-daily
HadGHCND	3.75x2.5º	Global	1950-now	Tmin, Tmax anomalies

A1.1.1. GHCN-D

The Global Historical Climate Network-Daily (GHCN-D) dataset, maintained by the National Climatic Data Center (NCDC), contains more than 40,000 global station data records. Variables include maximum and minimum daily temperature, temperature at time of observation, daily precipitation totals, snowfall and snow depth. The data undergo rigorous quality control, the results of which are provided in the ascii data files available via FTP from NCDC.

Coverage of study region:

- Sparse, with the exception of parts of Southern Africa (Figure A1).

Dataset update information:

- Updated daily in real time.

Further information:

- http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/
- Durre et al. (2008).

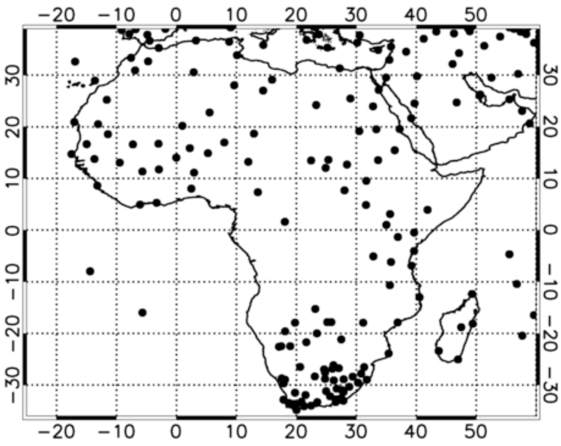

Figure A1: Stations reporting maximum daily temperature in GHCN-D in 2009.

A1.1.2. Integrated Surface Daily (ISD) dataset

The Integrated Surface Daily (ISD) dataset, maintained by NCDC, contains more than 20,000 global station data records. The station records contain numerous variables at daily and/or sub-daily intervals, including pressure, air temperature and precipitation. The data undergo quality-control implemented at NCDC. A static version of the ISD dataset between 1973 and 2010 that has undergone additional quality control is also available internally within the Met Office.

Coverage of study region:

- Sparse, with the exception of parts of Southern Africa.

Dataset update information:

- Updated daily in real time.

Further information:

- http://www.ncdc.noaa.gov/oa/climate/isd/index.php
- Lott and Baldwin (2002); Lott (2004).

A1.1.3. Synops

'Synops' stations are land-based meteorological stations that report surface synoptic observations via the Global Telecommunication System (GTS) in near real time. The data include many types of observation beside temperature, and are received at the Met Office and archived for 10 years. In their raw form, the data undergo no quality control. However, many of the Synops station reports, including historical observations beyond the past 10 years, can also be found in the quality-controlled GHCN-D (Section A1.1.1) and ISD (Section A1.1.2) datasets.

Coverage of study region:

- Fairly sparse (Figure A2)

Dataset update information:

- Reported in near-real time (e.g. six-hourly).

Further information:

- http://www-hc.metoffice.com/~hadobs/Monitoring.operational/crutem3/docs/mean_dailies/ (internal link)

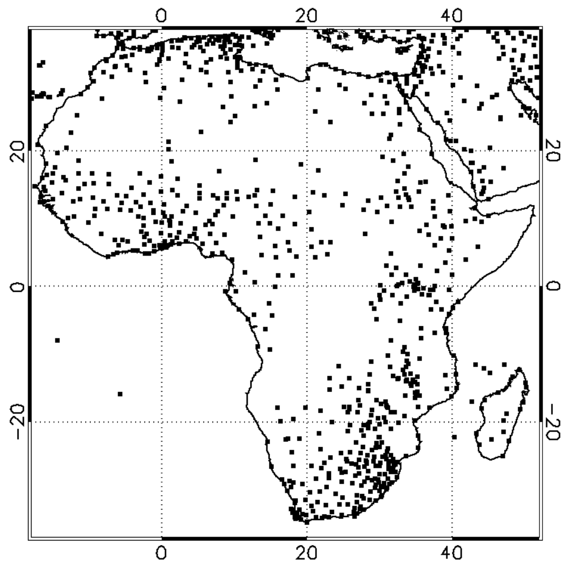

Figure A2: SYNOP stations reporting daily precipitation totals in July 2009.

A1.1.4. HadGHCND

The HadGHCND has been produced through a combined effort between the Met Office Hadley Centre and NCDC. It is based on selected stations from the GHCN-D dataset (Section A1.1.1), which have been interpolated onto a 2.5° latitude x 3.75° longitude grid through angular-distance weighting. The data are anomalies, referenced to the baseline period 1961-1990. At present, HadGHCND does not include any uncertainty estimates.

Coverage of study region:

- Poor. Limited coverage in northern and southern parts of Africa (Figure A3).

Dataset update information:

- Updated daily in real time.

Further information:

- http://hadobs.metoffice.com/hadghcnd/
- Caesar et al. (2006).

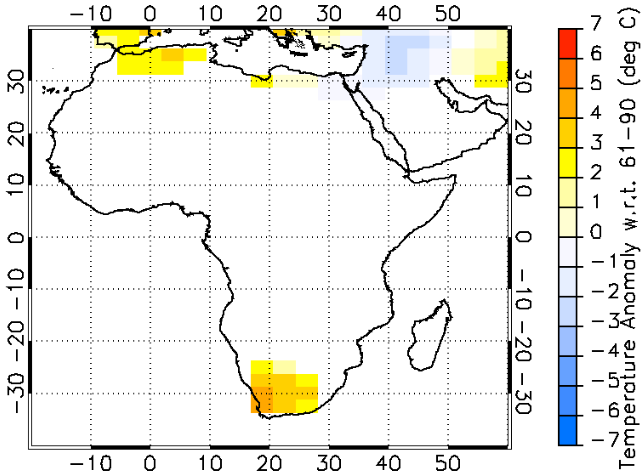

Figure A3: HadGHCND maximum temperature for 1 October 2009.

A1.1.5. Discussion of daily temperature datasets

Available gridded daily temperature data over Africa are limited to HadGHCND, and this has very poor coverage.

The University of Cape Town is blending station temperature data with the ERA-interim Reanalysis for the CORDEX regional model downscaling project. The analysis relaxes back to ERA-interim where there are no nearby data, so it will be affected by any biases in ERA-interim (Bruce Hewitson, pers. comm.) and should be used with caution.

We stress the need for scientific advance in using satellite data to supplement *in situ* observations.

A1.2. Monthly Datasets

Table A2 provides a summary of available monthly land temperature datasets. Further details are provided in the following sections.

Table A2: Summary of monthly air temperature datasets. Variables here reflect mean daily values over the course of one month.

Dataset	Spatial Res.	Spatial Coverage	Temporal Coverage	Variables
GHCN	Station	Global	19^{th} C -now	Tmax, Tmin, Tmean
CRUTEM3	5º	Global	1850-now	Tmean anomalies
GISS	2º	Global	1880-now	Tmean anomalies
NCDC	5º	Global	1880-now	Tmean anomalies
CPC	0. 5º	Global	1948-now	Tmean

University of Delaware	0. 5°	Global	1900-2008	Tmean
CRU	0.5°	Global	1901-2006	Tmean, Tmax, Tmin, Diurnal temp range

A1.2.1. GHCN

The Global Historical Climate Network-Monthly is a database of monthly station observations of temperature and precipitation, maintained by NCDC. The full network consists of around 7000 stations for mean temperature and 5000 stations for minimum/maximum temperature. A homogeneity-adjusted dataset is available for a subset of these stations (approximately 5000 for mean temperature and 3500 for minimum/maximum temperature). The data undergo rigorous quality control.

Coverage of study region:

- Sparse for mean temperature and very poor for minimum/maximum temperature (Figure A4).

Dataset update information:

- Updated monthly.

Further information:

- http://www.ncdc.noaa.gov/ghcnm/
- Peterson and Vose (1997); Peterson et al. (1998b).

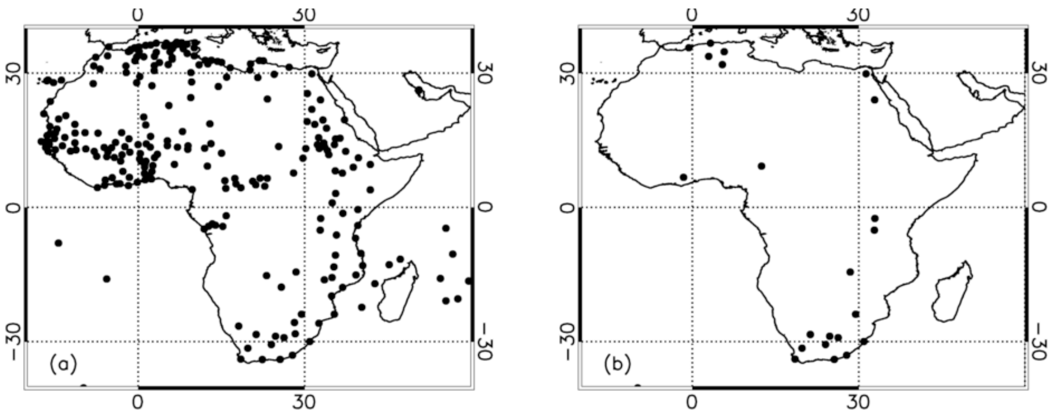

Figure A4: GHCN stations reporting (a) mean temperature and (b) maximum temperature in 2009.

A1.2.2. CRUTEM3

The CRUTEM3 dataset has been produced through a collaborative effort between Met Office Hadley Centre and the Climatic Research Unit at the

University of East Anglia. It is based on monthly mean temperatures from more than 4000 meteorological stations, which are converted to anomalies with respect to 1961-1990, before being gridded to 5º resolution. Prior to gridding, the station records undergo rigorous quality control. CRUTEM3 includes grid box uncertainty estimates.

Coverage of study region:

- Reasonably poor (Figure A5).

Dataset update information:

- Updated monthly.

Further information:

- http://hadobs.metoffice.com/crutem3/
- Brohan et al. (2006).

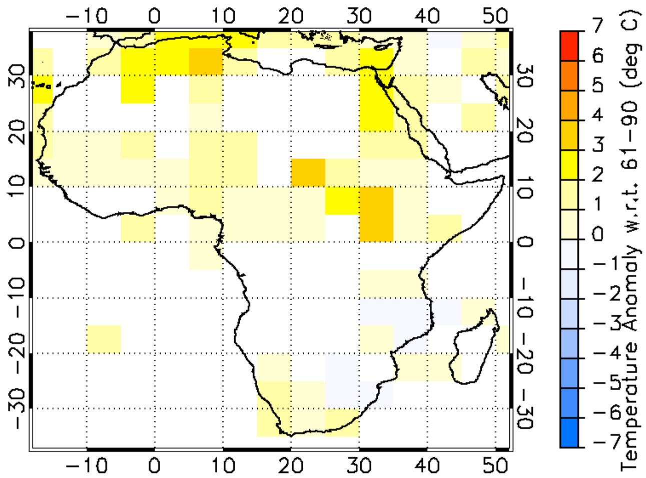

Figure A5: CRUTEM3 temperature anomalies for July 2009.

A1.2.3. GISS

The Goddard Institute for Space Studies (GISS) land surface temperature analysis is based on station data from the GHCN (unadjusted, Section A1.2.1), United States Historical Climatology Network (USHCN) data, and SCAR (Scientific Committee on Antarctic Research). The data are anomalies, referenced to the baseline period 1951-1980. The current analysis includes adjustment for urbanisation by using neighbouring rural stations; urban stations without at least 3 nearby rural stations are excluded from the analysis. In their native format, the data are available on an equal-area grid, with code provided by NASA to read in the data into a 2º latitude/longitude grid. Two levels of smoothing are available: 250 km and 1200 km.

Coverage of study region:

- Poor for 250 km smoothed version (Figure A6). Near-complete coverage for 1200 km smoothed version.

Dataset update information:

- Updated monthly.

Further information:

- http://data.giss.nasa.gov/gistemp/
- http://www.esrl.noaa.gov/psd/data/gridded/data.gisstemp.html
- Hansen et al., (1999); Hansen et al. (2001). Hansen et al. (2010).

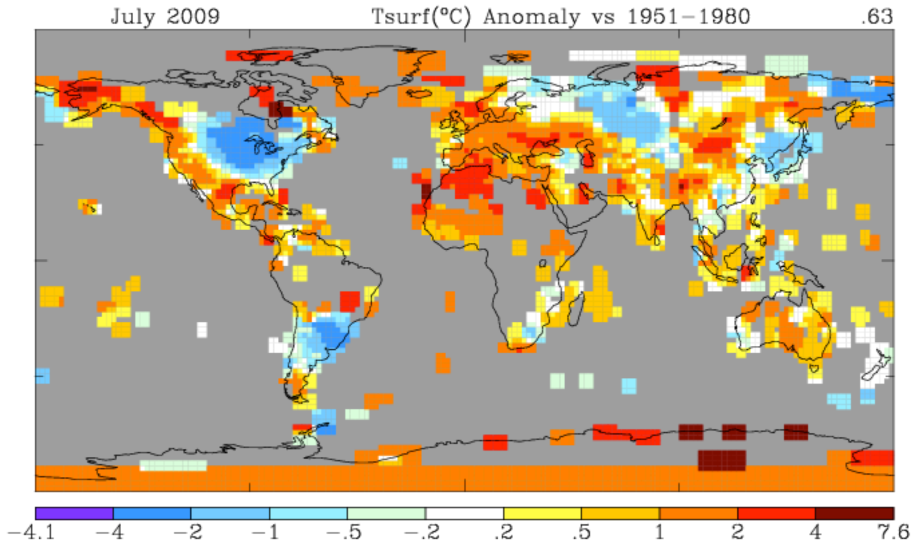

Figure A6: GISS Surface Temperature Analysis for July 2009, smoothed to 250 km. From NASA (http://data.giss.nasa.gov/gistemp/)

A1.2.4. NCDC

The NCDC data global land surface temperature data consists of gridded anomalies with respect to the baseline period 1961-1990. The dataset is based on land surface temperatures from the GHCN dataset (Section A1.2.1). The primary source of data is the GHCN homogenised station records, but for grid boxes without homogenised data, the raw station data are used.

Coverage of study region:

- Poor. Some coverage in western and southeastern parts (Figure A7).

Dataset update information:

- Updated monthly.

Further information:

- http://www.ncdc.noaa.gov/temp-and-precip/ghcn-gridded-products.php
- Peterson and Vose (1997); Peterson et al (1998a).

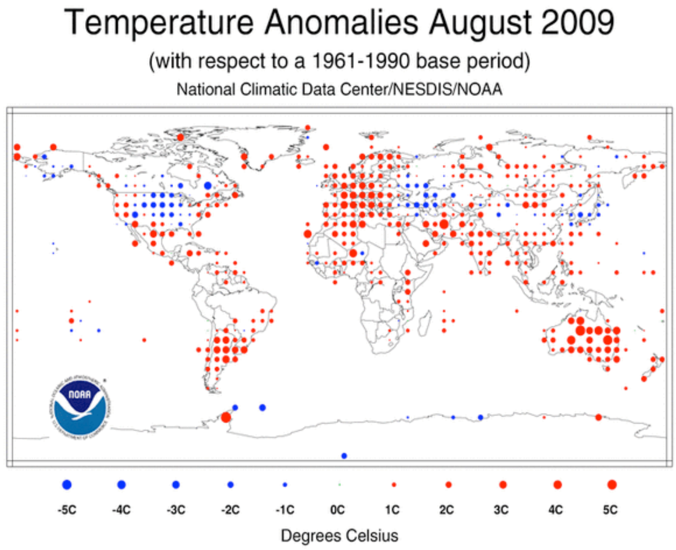

Figure A7: NCDC gridded land temperature dataset for August 2009. Source: NCDC.

A1.2.5. CPC GHCN-CAMS

The Climate Prediction Center (CPC) global surface air temperature dataset is derived from station data provided in the GHCN (Section A1.2.1) and Climate Anomaly Monitoring System (CAMS) datasets. The incorporation of the CAMS dataset improves spatial coverage, but the CAMS station data are not as rigorously quality-controlled as the GHCN. The CPC 0.5º analysis is produced through the interpolation of station anomalies relative to a gridded climatology. Gridded actual temperatures are then obtained by adding back this climatology, then reanalysis-based temperature lapse rates are used to compensate for the difference between the station elevation and the elevation associated with the climatology.

Coverage of study region:

- 100% coverage of global land (Figure A8). However, note, where station density is low the data are heavily interpolated and may be unreliable (e.g. circular artefacts in data at high latitudes are visible in global plot).

Dataset update information:

- Updated monthly.

Further information:

- ftp://ftp.cpc.ncep.noaa.gov/wd51yf/GHCN_CAMS/
- Fan and van den Dool (2008).

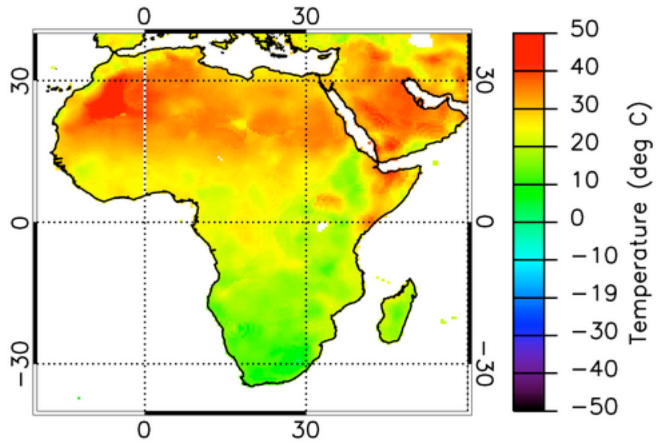

Figure A8: CPC GHCN-CAMS gridded air temperatures for July 2008.

A1.2.6. University of Delaware (v2.01)

The University of Delaware gridded land temperature data is primarily based on station observations from GHCN (Section A1.2.1). Additional station data from a number of other sources are used to infill gaps in the analysis (e.g. the Atmospheric Environment Service/Environment Canada, the State Hydrometeorological Institute, St. Petersburg, Russia, and the Global Surface Summary of Day (NCDC)). The gridded dataset is produced through interpolation of station anomalies, which are then combined with an equivalent gridded climatology to produce estimates of the absolute mean monthly temperature. The dataset includes interpolation error estimates derived from the results of cross validation experiments.

Coverage of study region:

- 100% coverage of global land (Figure A9). However, note, where station density is low the data are heavily interpolated and may be unreliable.

Dataset update information:

- Infrequently updated – essentially static.

Further information:

- http://climate.geog.udel.edu/~climate/html_pages/Global2_Ts_2009/README.global_t_ts_2009.html

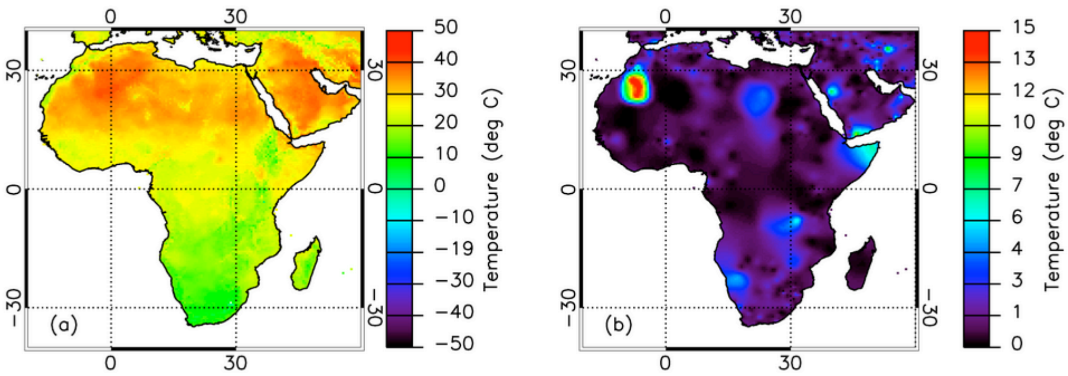

Figure A9: University of Delaware gridded air temperatures for July 2008 showing (a) absolute temperature and (b) corresponding cross validation errors.

A1.2.7. CRU TS2.1 and 3.0

The CRU TS2.1 temperature dataset is based on records provided by more than 4000 stations from around the world. The station records were sourced from several sources; quality control of the station records was carried out by the station data providers. Before interpolating to 0.5º spatial resolution, the data are converted to anomalies with respect to the 1961-1990 baseline period. After interpolation, the temperatures are adjusted to actual values by adding on the climatological normals. Interpolation is performed by considering a 'sphere of influence', which is the correlation decay distance. The correlation decay distance is 750 km for diurnal temperature range, and 1200 km for mean temperature. CRU TS2.1 includes an adjustment of inhomogeneities using an iterative procedure based on the GHCN homogeneity adjustment procedure (Section A1.2.1). CRU TS3.0, without homogenization, is in preparation.

Coverage of study region:

- 100% coverage of global land (Figure A10). However, note that the interpolation relaxes to baseline period means in regions with few stations. It is recommended that accompanying station density files are considered in conjunction with data files.

Dataset update information:

- Essentially static, with a new version released every few years.

Further information:

- http://badc.nerc.ac.uk/data/cru/
- http://www.cru.uea.ac.uk/cru/data/hrg/
- Mitchell and Jones (2005); New et al. (2000).

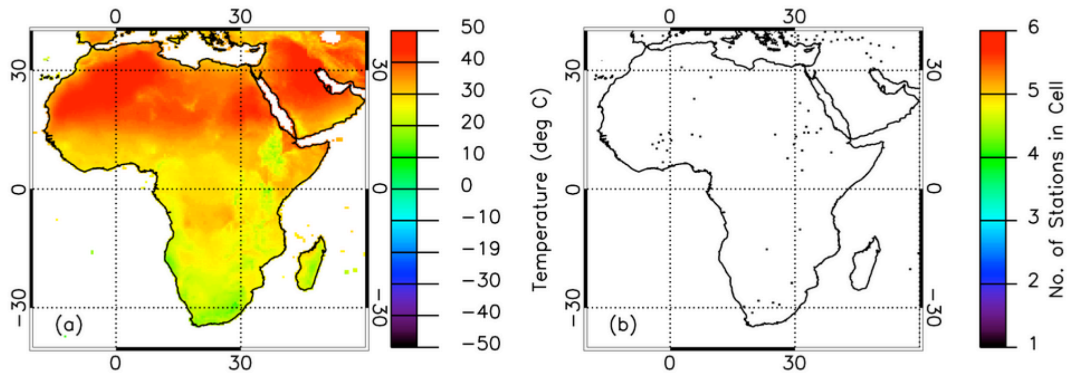

Figure A10: CRU TS3.0 data for July 2005 showing (a) Tmax and (b) number of stations in each grid cell for the diurnal temperature range product.

A1.2.8. Discussion of monthly temperature datasets

All these datasets, except for CRUTEM3 and CRU TS, use GHCN data as their main input; even CRUTEM3 and CRU TS include some GHCN data. However specific analysis techniques differ.

The underlying station density is too low for comprehensive monitoring of extremes.

Again we stress the need for incorporating satellite observations to supplement the *in situ* data.

A2. Precipitation

A2.1. Daily Datasets

Table A3 provides a summary of available daily precipitation datasets. Further details are provided in the following sections.

Table A3: Summary of daily precipitation datasets

Dataset	Spatial Res.	Spatial Coverage	Temporal Coverage	Variables	Data
GHCN-D	Station	Global	19^{th} C.-now	Total precipitation	Gauge
CPC gauge	0.5°	Global	1979-now	Total precipitation	Gauge
TRMM	0.25°	50N-S.	1997-now	Rain rate (mm/day)	Satellite
GPCP	1°	Global	Oct 1996-now	Rain rate (mm/day)	Satellite
GPCP GPI	1°	40N-40S	1996-now	Rain rate (mm/day)	Satellite
CMORPH	0.25°	60N-60S	2002-now	Rain rate	Satellite

				(mm/day)	
CPC-RFE operational	0.1°	40S-40N, 20W-55E	1995-now	Rain rate (mm/day)	Gauge & satellite
CPC-RFE climatological	0.1 °	40S-40N, 20W-55E	1983-now	Rain rate (mm/day)	Gauge & satellite
EPSAT-SG	0.05 °	West Africa	Not yet operational	Rain rate (mm/day)	Satellite
University of Cape Town	Various	African continent	In development	Rain rate (mm/day)	Various

A2.1.1. GHCN-D

See Section A1.1.1 for details.

Coverage of study region:

- Sparse, with the exception of parts of Southern Africa (precipitation coverage almost identical to that shown in Figure A1)

Dataset update information:

- Updated daily in real time.

Further information:

- http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/
- Durre et al. (2008).

A2.1.2. CPC Gauge-Based Analysis of Global Daily Precipitation

The CPC daily precipitation product is a recent initiative, and is the first stage of long-term plans at the CPC to develop, unify and improve their existing precipitation products. The product is based on gauge measurements, which are quality-controlled and then interpolated onto a grid using optimal interpolation that includes adjustments for topographic effects. Between 1979 and 2005, a retrospective version of the product is available that is based on more than 30,000 gauges. From 2006, a real time version is available that is based on approximately 17,000 gauges. The majority of stations are based in the US, Mexico, South America and Australia, with just ~5000 stations originating from outside of these regions. Station network density information is provided with the data products and users are encouraged to consult this when using the data. In addition, a quality check report is provided that lists known problems with the data that have been realised through manual checks.

Coverage of study region:

- Complete. However, note that the dataset is based on very few stations in Africa, particularly in tropical regions, and therefore has high uncertainty (Figure A11). Note significant differences from equivalent GPCP data in Figure A11.

Dataset update information:
- Updated in near real time – lag of 1-2 days.

Further information:
- ftp://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP
- Chen et al. (2008); Xie et al. (2007)

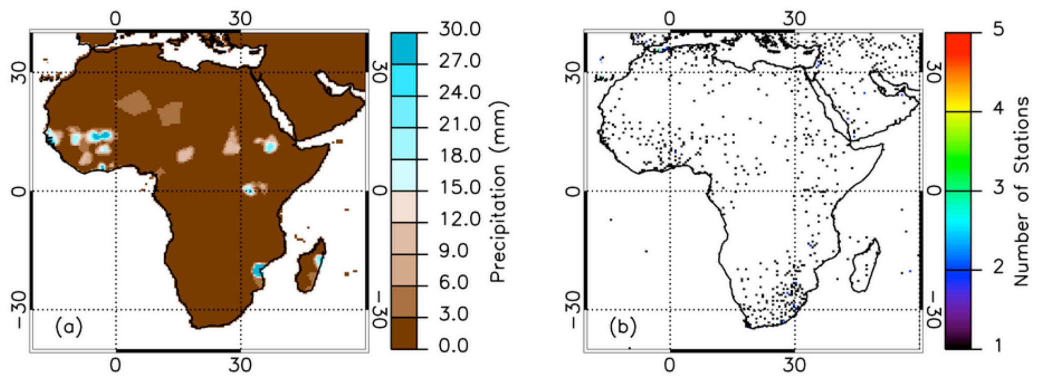

Figure A11: CPC daily gauge analysis for 01 July 2009 showing (a) precipitation total and (b) gauge density.

A2.1.3. TRMM Combination (3B 42 Version 6)

The TRMM is a joint satellite project between NASA and JAXA designed to improve observations of precipitation over the tropics. The mission incorporates several instruments, including a microwave imager, radar and visible-infrared scanner. A number of different products are made available to users, including eight monthly products, one daily and one 3-hourly. The daily product is derived from a combination of TRMM data and other satellite infrared observations (e.g. Meteosat, GMS, GOES and NOAA-12). The TRMM data are used to produce monthly infrared calibration parameters, which are then applied to 3-hourly precipitation estimates from the other satellite infrared datasets. The daily totals are estimated from the 3-hourly precipitation data between 00Z and 21Z. Finally the daily totals are scaled so that the monthly total matches that of the satellite-gauge TRMM Combination (3B 43 Version 6: Section A2.3.13). The final dataset consists of a daily precipitation rate and root-mean-square precipitation-error estimates.

Coverage of study region:
- Complete (Figure A12).

Dataset update information:
- Updated with a few days to weeks lag.

Further information:.

- http://precip.gsfc.nasa.gov/ and Huffman et al. (2007).

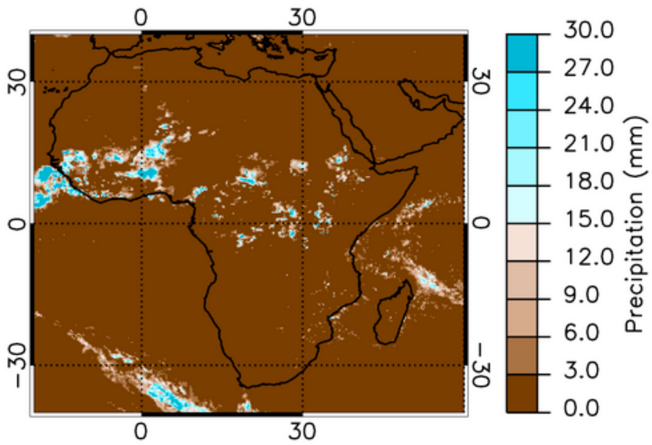

Figure A12: TRMM daily analysis for 01 July 2009.

A2.1.4. GPCP 1DD v1.1

This analysis is largely based on the monthly combined satellite-gauge product (Section A2.3.10) which is used to calibrate daily estimates derived from geostationary and polar-orbiting infrared sensors. Microwave and gauge estimates are not used explicitly owing to sampling limitations. The accuracy of GPCP 1DD is expected to be lower than that of the monthly dataset.

Coverage of study region:

- Complete (Figure A13).

Dataset update information:

- Updated with a few months delay.

Further information:

- http://precip.gsfc.nasa.gov/gpcp_daily_comb.html.
- Huffman et al. (2001); Roca et al. (2010).

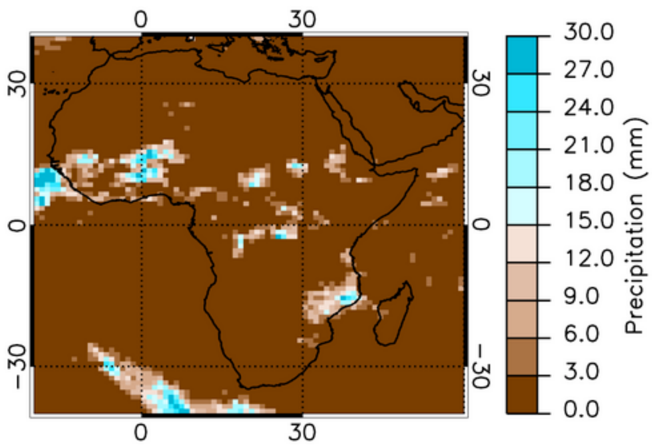

Figure A13: GPCP daily satellite-gauge analysis for 01 July 2009.

A2.1.5. GPCP GPI

The GPCP GOES Precipitation Index (GPI) is one of the elements that is used in the GPCP daily precipitation estimates. This dataset is based solely on infrared observations from geostationary satellite platforms. The approach used to produce the GPI is extremely simple, and is expressed as:

Precipitation (mm) = FRAC x RATE x TIME

Where 'FRAC' is the fractional coverage of infrared pixels with top of atmosphere brightness temperature of less than 235K over a reasonably large domain (50 km x 50 km and larger), 'RATE' is 3 mm/hour, and 'TIME' is the number of hours over which "FRAC" was compiled (equation and explanatory text reproduced from CPC GPI webpage cited below).

Coverage of study region:

- Complete (Figure A14).

Dataset update information:

- Updated monthly.

Further information:

- http://www.cpc.ncep.noaa.gov/products/global_precip/html/wpage.gpi.shtml.
- Arkin and Meisner (1987).

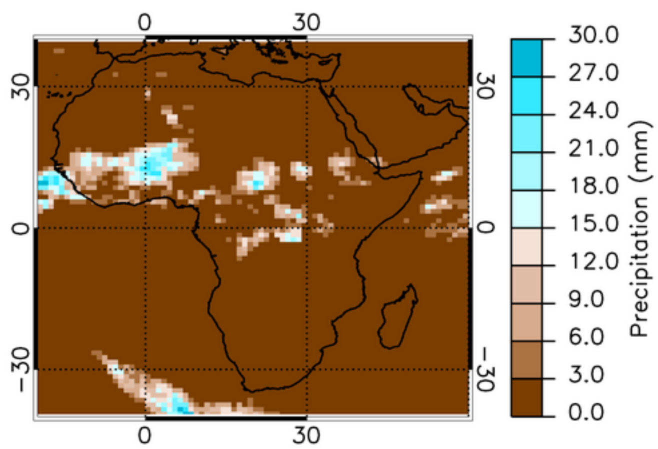

Figure A14: GPCP GPI daily analysis for 01 July 2009

A2.1.6. CMORPH

The CPC MORPHing technique (CMORPH) precipitation estimates are derived principally from passive microwave observations of precipitation from polar-orbiting platforms. Geostationary infrared data are then used to provide supplementary high-resolution information, 'propagating' the microwave estimates through both space and time, allowing precipitation rates to be estimated at a frequency of 30 minutes and at 8 km spatial resolution. The 30-minute data are then aggregated to provide 3-hourly estimates, which in turn are used to derive the daily estimates. At present, microwave data from SSM/I, AMSU-B, AMSR-E and TMI are used in the analysis. The objective of CMORPH is to provide a method of combining different microwave retrievals from different sources; CMORPH is not explicitly a precipitation retrieval algorithm in its own right.

Coverage of study region:

- Complete (Figure A15).

Dataset update information:

- Updated in near real time with a couple of days lag. Note: only the last few weeks of daily data are archived by CPC. For older data, the user must retrieve the three-hourly data and aggregate.

Further information:

- http://www.cpc.ncep.noaa.gov/products/janowiak/cmorph_description.html.
- Joyce et al. (2004).

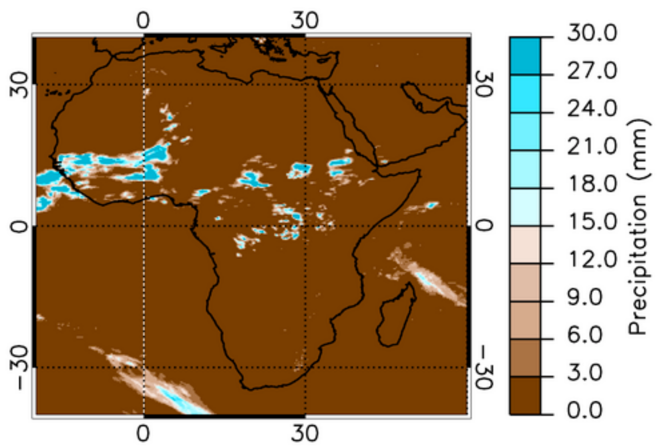

Figure A15: CMORPH precipitation analysis for 01 July 2009.

A2.1.7. CPC Africa RFE (FEWS-NET)

The CPC African Rainfall Estimate (RFE) is a combined gauge-satellite precipitation dataset and has been developed as part of the Famine Early Warning System Network (FEWS-NET) project. Operational and climatological versions of the estimates are produced using different (though similar) methodologies. The operational product went through a major change in 2001, from V1.0 to V2.0.

V1.0 of the operational product covers the period 1995-2000, while V2.0 is from 2001 to the present. The V2.0 dataset shares many similarities with the CMAP product, described in Section A2.3.12. In essence, a combined satellite dataset is generated from microwave AMSU and SSM/I, and GPI (Section A2.1.5) precipitation estimates using a maximum likelihood method. These are then merged with daily rain gauge data from up to 1000 stations, although typically, the number of stations is closer to 500 owing to erroneous station data and/or poor station maintenance. The merging process allows the final rainfall estimates to have the magnitude of the station data, with the shape of the precipitation field determined by the satellite estimates.

There are significant differences between RFE 2.0 and 1.0, leading to possible biases in the combined operational time series. RFE 1.0 does not include microwave data, but instead has a separate component for estimating orographic rainfall based on model humidity and wind fields combined with orography data.

The climatological version of RFE covers 1983 to the present, and is continually updated. It is similar to the operational RFE 1.0 product, combining GPI estimates with available gauge data, but does not include an orographic rainfall element. Microwave estimates are not included due to the shorter time period of available microwave observations compared to infrared data. However the number and position of gauges included in the analysis varies over time, which may introduce biases into the analysis (Yin and Gruber (2010)).

RFE operational and climatological estimates are available at various temporal resolutions, including daily, 10-day, monthly and seasonal composites.

Coverage of study region:

- Complete (Figure A16).

Dataset update information:

- Updated daily in near real time.

Further information:

- http://www.cpc.noaa.gov/products/fews/rfe.shtml

- http://www.cpc.noaa.gov/products/fews/rfe2.0_tech.pdf
- http://www.cpc.noaa.gov/products/fews/AFR_CLIM/afr_clim.shtml
- Herman et al. (1997); Xie and Arkin (1996: Yin and Gruber (2010).

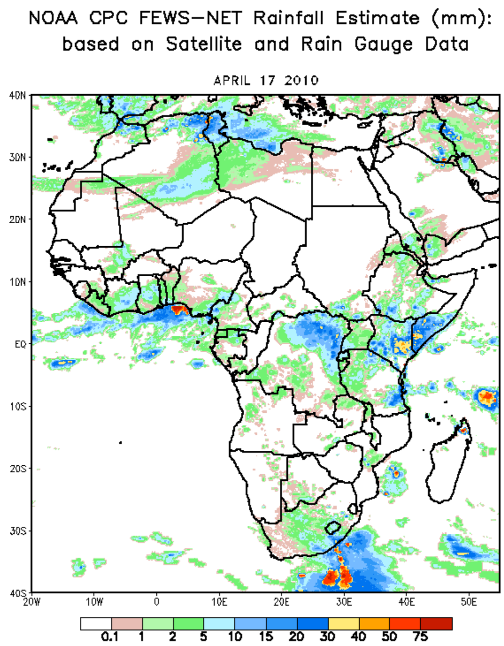

Figure A16: FEWS-NET operational RFE for 17 April 2010 (source: http://www.cpc.ncep.noaa.gov/products/fews/rfe.shtm).

A2.1.8. Future products

A recently developed algorithm designed to produce rainfall estimates for the West African region for the African Monsoon Multidisciplinary Analysis (AMMA) campaign is EPSAT-SG (Bergès *et al*. 2010). This uses an artificial neural network to calibrate multi-spectral infrared radiances with TRMM PR data to create a 'Rainfall Probability' field. This probability field is combined with downscaled GPCP 1 degree day (1DD) data to produce a field of 'Potential Rainfall Intensity'. Finally, the probability and potential intensity fields are multiplied together to form a final rainfall intensity estimate with Meteosat resolution of 3km and 15 min. Operational rainfall estimates are not currently produced using EPSAT-SG, though this may change in the near future.

The University of Cape Town is developing African rainfall datasets, using station data acquired from a variety of sources (Bruce Hewitson, pers. comm.). A high resolution gridded rainfall product will be available for South Africa based on the methodology of Hewitson and Crane (2005). A blended satellite-gauge rainfall product combining gauge data with TRMM (Section A2.1.3) and/or RFE 2.0 (FEWS) (Section A2.1.7) will also be available.

A2.1.9. Intercomparison studies

Laws *et al.* (2004) compared CPC-RFE 2.0 (combined gauge and satellite: Section A2.1.7), TRMM 3B 42 (Section A2.1.3) and CMORPH (Section A2.1.6) (both satellite only) daily rainfall estimates against GTS gauge data from across Africa. They found that the performances of all three algorithms were similar, with underestimation of low gauge rainfall and significant overestimation of high gauge rainfall.

Dinku *et al.* (2008) compared satellite rainfall products including RFE 2.0, TRMM 3B 42 and CMORPH against gridded 0.25º daily gauge estimates over Ethiopia and Zimbabwe. All products detected rainfall with some success, but quantitative satellite-rainfall estimates at this scale were poor, with CMORPH and 3B 42 better than RFE 2.0 over the complex Ethiopian terrain and all three products performing similarly over the flatter terrain of Zimbabwe.

Hirpa and Gebremichael (2010) compared, at 3-hourly 0.25° resolution, the *real-time* TRMM analysis 3B42RT (which is among the products alluded to in Section A2.1.3), CMORPH (Section A2.1.6), and the infra-red based Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) analysis, over Ethiopia. TRMM 3B42RT and CMORPH gave similar rainfall bias, spatial structure, elevation-dependent trend, and distribution function. These differed from PERSIANN which considerably underestimated rainfall in high-elevation areas.

A2.1.10. Discussion of daily precipitation datasets

There are significant differences between available daily precipitation datasets. Care should be taken when using gauge only datasets since they are based on so few *in situ* measurements. The authors of the CPC dataset, for example, specifically point out that the quality of their gauge-analysis (Section A2.1.2) is poor over Africa. They recommend interpreting their precipitation analysis in conjunction with the gauge density information that is supplied in the product to indicate where confidence is particularly low/high. Roca *et al*. (2010) (see Section A2.1.4) discuss the sampling uncertainties of both rain-gauge and satellite estimates of precipitation. The unrealistic 'bulls-eye' appearance of the CPC analysis (Figure A11) is a typical artefact of the optimal interpolation approach adopted by the authors, where data are sparse.

So we recommend satellite-only or combined gauge-satellite analyses for forecast evaluation and attribution. Based on intercomparison studies at 10-day and monthly scales (Sections A2.2.2 and A2.3.17), a good choice of satellite-only product would appear to be TRMM 3B 42. Of the combined satellite-gauge products, the operational RFE 2.0 (Section A2.1.7) is recommended for forecast validation, while the climatological RFE is probably more appropriate for attribution studies, due to its longer record, though it may have biases owing to the varying number of gauges included in the analysis.

The poor quantitative performance of all current satellite rainfall products at high spatial resolution on daily scales should be kept in mind. Averaging these products to larger spatial scales is expected to improve their accuracy.

A2.2. 10-day datasets

10-day rainfall estimates are often used for food security and drought monitoring applications in Africa. This interval is considered short enough to represent intra-seasonal variations in rainfall, while satellite rainfall products are more accurate at 10-day than daily scale due to averaging. The TAMSAT rainfall product is designed specifically to provide 10-day rainfall estimates for Africa, and is described below. 10-day estimates can also be produced by accumulation of any of the daily products described in Section A2.1. In particular, the CPC Africa-RFE product (see Section A2.1.7) produces operational 10-day estimates. Table A4 gives a summary of available 10-day rainfall products for the region.

Table A4: Summary of 10-day precipitation datasets

Dataset	Spatial Res.	Spatial Coverage	Temporal Coverage	Variables	Data
GHCN-D	Station	Global	19^{th} C.-now	Total precipitation	Gauge
CPC gauge	0.5º	Global	1979-now	Total precipitation	Gauge
TRMM	0.25º	50N-S.	1997-now	Rain rate (mm/day)	Satellite
GPCP	1º	Global	1996-now	Rain rate (mm/day)	Satellite
GPCP GPI	1º	40N-40S	1996-now	Rain rate (mm/day)	Satellite
CMORPH	0.25º	60N-60S	2002-now	Rain rate (mm/day)	Satellite
CPC-RFE operational	0.1º	40S-40N, 20W-55E	1995-now	Rain rate (mm/day)	Gauge & satellite
CPC-RFE climatological	0.1º	40S-40N, 20W-55E	1983-now	Rain rate (mm/day)	Gauge & satellite
TAMSAT	0.04 º	All Africa Land only	1982-now	Total precipitation & anomaly	Satellite

A2.2.1 TAMSAT

The TAMSAT (Tropical Applications of Meteorology using SATellite data) product provides operational 10-day rainfall estimates for the whole African continent from geostationary infrared data (Figure A17). TAMSAT uses a simple linear relationship between cold cloud amount and rainfall similar to that used by the GPI (see Section A2.1.5). but is calibrated specifically for Africa with different fixed calibrations used regionally and seasonally. Historical rain-gauge data are used to calibrate the product but real-time gauge data are not needed to produce the estimates.

10-day and monthly rainfall estimates, together with anomalies against the 2000-2009 TAMSAT rainfall climatology, are available from the TAMSAT website and are also propagated via the GEONETcast service of EUMETcast. These are produced at a basic spatial resolution of 0.0375° (4 km).

A TAMSAT all-Africa rainfall climatology of 10-day estimates from 1982-present is due to be completed by December 2010. In the future this climatology will be continually updated with operational estimates. TAMSAT uses only the Meteosat series of satellites, and is therefore likely to be more consistent over long periods than datasets compiled from multiple sources and satellites.

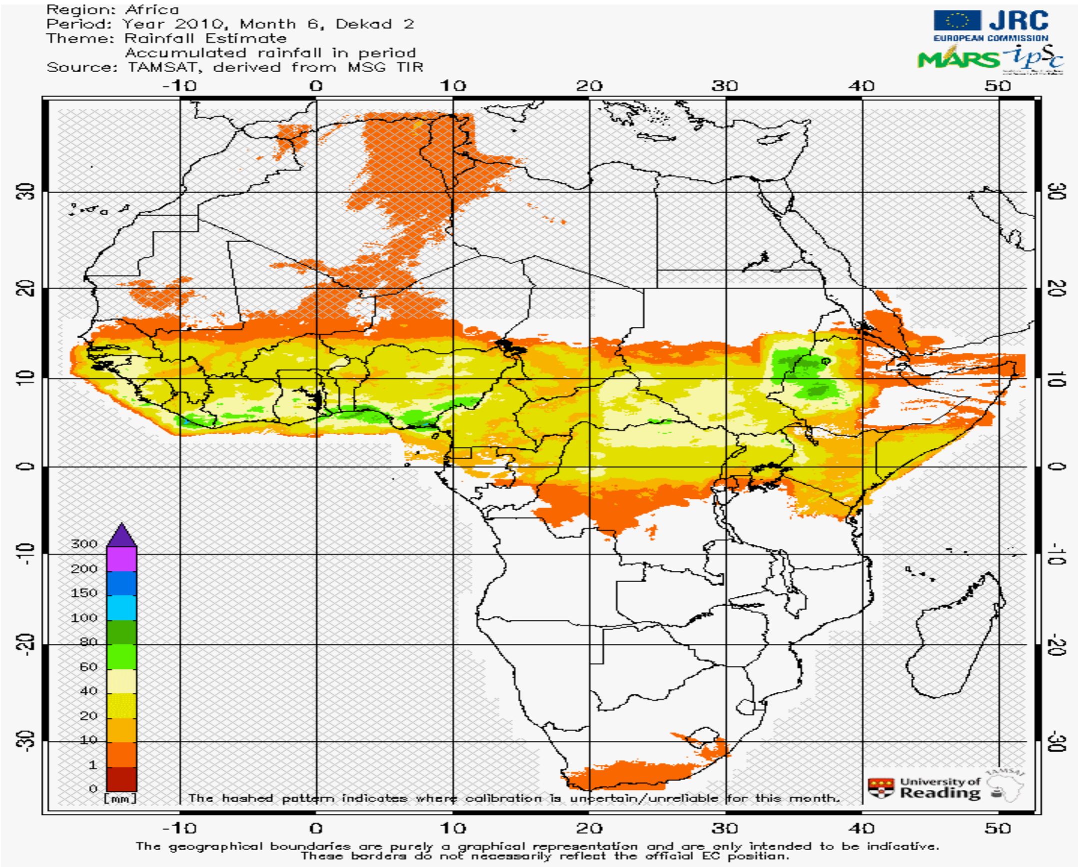

Figure A17: TAMSAT 10-day rainfall estimate for June 11th-20th 2010

Coverage of study region:

- Complete.

Dataset update information:

- Updated every 10 days in near real time.

Further information:

- http://www.met.reading.ac.uk/tamsat/
- Grimes et al. (1999); Dugdale et al. (1991).

A2.2.2. Intercomparison studies

Jobard *et al.* (2007) compared TRMM 3B 42, TAMSAT, RFE 2.0, GPCP 1DD, CMORPH and GPI 10-day estimates against gridded gauge data over the Sahel region of West Africa. The products designed specifically for Africa (RFE 2.0 and TAMSAT) gave the most accurate estimates, with CMORPH performing poorly for this region.

Dinku *et al.* (2007) examined 10-day estimates from TRMM 3B 42, TAMSAT, RFE 1.0 and 2.0, CMORPH and GPCP 1DD over the mountainous terrain of Ethiopia. CMORPH, TAMSAT and TRMM 3B 42 performed well, with the best estimates depending on which validation statistic is considered most important for a particular application.

A2.2.3. Discussion of 10-day rainfall datasets

10-day rainfall estimates are key for forecasting and attribution where sub-monthly information is required, and where daily datasets do not provide sufficient accuracy or reliability.

The sparsity of reporting gauges in most of Africa means that satellite-only and combined gauge-satellite datasets are preferable to gauge-only datasets. For forecast verification, intercomparison studies indicate that of the satellite-only products, TAMSAT and TRMM 3B 42 may be the best, while the operational RFE 2.0 is the most suitable combined gauge-satellite product. For attribution studies, the TAMSAT and climatological RFE datasets are most appropriate; TAMSAT may be more consistent over time as RFE includes a gauge component of varying spatial coverage.

Validation studies (Dinku *et al.* 2007; Jobard *et al.* 2007) show the value of local calibration or constraint of satellite rainfall products, either with gauges (TAMSAT and RFE) or precipitation radar (TRMM 3B 42). This is also likely to apply to other timescales.

A2.3. Monthly Datasets

Table A5 provides a summary of available monthly precipitation datasets. Further details are provided in the following sections.

Table A5: Summary of monthly precipitation datasets

Dataset	**Spatial Res.**	**Spatial Coverage**	**Temporal Coverage**	**Variables**	**Data**
GHCN	Station	Global	19th C-now	Precip total.	Gauge
GPCC first guess	1.0° & 2.5°	Global	Oct 2003 - now	Precip total	Gauge
GPCC monitoring	1.0° & 2.5°	Global	1986-now	Precip total	Gauge
GPCC reanalysis	0.5°, 1.0° & 2.5°	Global	1901 - 2007	Precip total & anomaly.	Gauge
GPCC VASClimO	0.5°, 1.0° & 2.5°	Global	1951 - 2000	Precip total	Gauge
GPCC gridded climatology	0.25°, 0.5°, 1.0° & 2.5°	Global	1951 - 2000	Precip total	Gauge
CRU	0.5º	Global	1901-2006	Precip total, wet-day frequency	Gauge
University of Delaware	0. 5º	Global	1900-2008	Precip total.	Gauge
NCDC	5º	Global	1900-now	Precip anomalies	Gauge
GPCP	2.5º	Global	1979-now	Rain rate (mm/day)	Gauge & satellite
GPCP-Int	2.5º	Global	1979-now	Rain rate (mm/day)	Gauge & satellite
CMAP	2.5º	Global	1979-2009	Rain rate (mm/day)	Gauge & satellite
TRMM	0.25º	50N-S	1998-now	Rain rate (mm/hour)	Gauge & satellite
CAMS-OPI	2.5º	Global	1979-now	Rain rate (mm/day)	Gauge & satellite
CPC-RFE operational	0.1º	40S-40N, 20W-55E	1995-now	Rain rate (mm/day)	Gauge & satellite
CPC-RFE climatological	0.1º	40S-40N, 20W-55E	1983-now	Rain rate (mm/day)	Gauge & satellite
TAMSAT	0.04º	All Africa land-only	1982-now	Precip total & anomaly	Satellite

A2.3.1. GHCN

GHCN contains more than 20,000 stations across the globe that report monthly precipitation. The data undergo rigorous quality control.

Coverage of study region:

- Patchy: varies from reasonable to very poor (Figure A18).

Dataset update information:

- Updated monthly in real time.

Further information:

- http://www.ncdc.noaa.gov/ghcnm/ and Section A1.2.1.

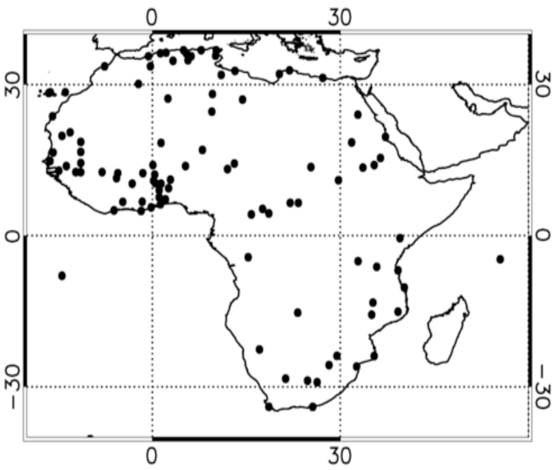

Figure A18: GHCN stations reporting precipitation during 2009 (filled circles).

A2.3.2. GPCC – First Guess

The GPCC produces several global precipitation products. The first guess product is a provisional product based on SYNOP reports from about 6000 stations worldwide and is available in near-real time. The data are subject to automated quality control before being included in the analysis. This dataset is derived from station anomalies, which are calculated with respect to the GPCC global normals (Section A2.3.6). Once gridded, the anomalies are superimposed on the gridded climatology before being released to users. The first guess product is considered to be the least accurate of the GPCC products, as the other products undergo more extensive pre-processing and quality control. Users are encouraged to consult the gauge density information supplied with the gridded product.

Coverage of study region:

- Complete, but derived from very few gauges (Figure A19).

Dataset update information:

- Available within 5 days of the end of the month.

Further information:

- http://gpcc.dwd.de and Schneider et al. (2010).

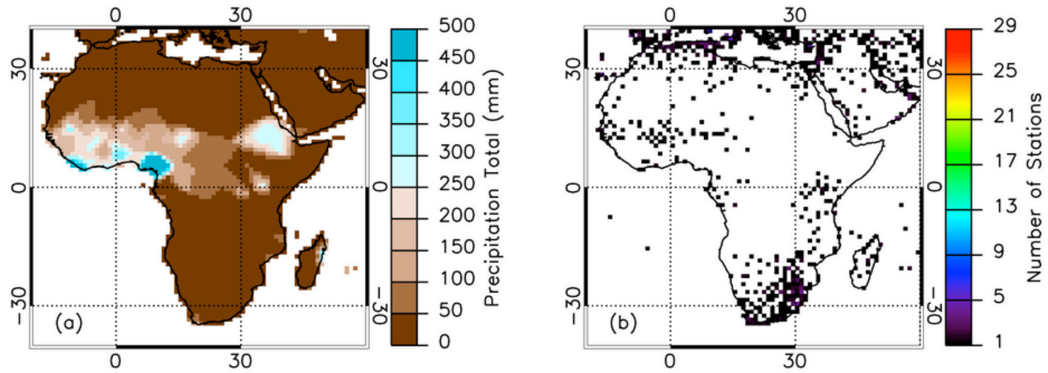

Figure A19: GPCC first guess monthly precipitation for July 2007 and the gauges used.

A2.3.3. GPCC – Monitoring

The GPCC monitoring product is based on SYNOP and CLIMAT reports from 7000-8000 stations. It is similar to the first guess product (Section A2.3.2), but uses additional stations and more rigorous quality control that involves both automatic and manual methods. Like the first guess product, the dataset is based on station anomalies that are gridded and then superimposed on the climatology to produce gridded total precipitation. Gauge density information is supplied with the precipitation data. From 2007, both absolute and relative (%) error fields are also available in the data files. The GPCC monitoring product constitutes the *in situ* component of both the GPCP (Section A2.3.10) and CMAP (Section A2.3.12) analyses.

Coverage of study region:

- Complete, although gauge network is sparse in some regions so product may be unreliable in places (Figure A20).

Dataset update information:

- Available within about two months after the end of the month.

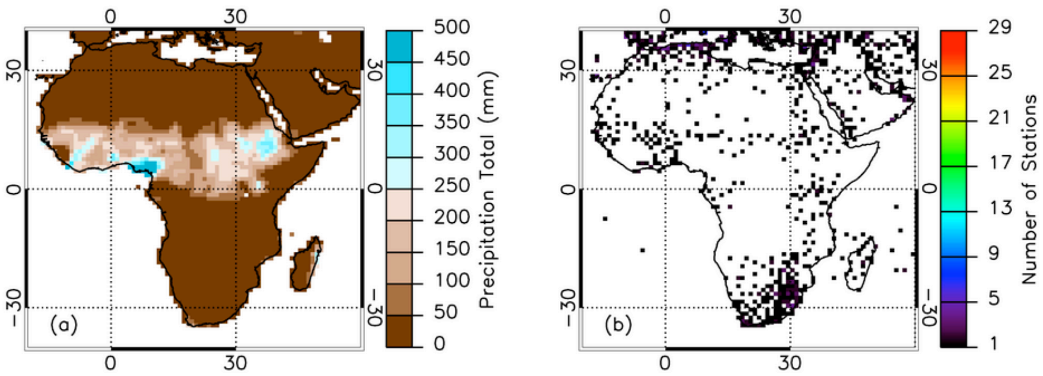

Figure A20: GPCC monitoring monthly precipitation for July 2007 and the gauges used.

A2.3.4. GPCC – Reanalysis (v4)

The GPCC reanalysis product is more accurate than the first guess (Section A2.3.2) and monitoring (Section A2.3.3) products. It is optimised to achieve maximum spatial coverage and is based on quality-controlled data from all available station records in the GPCC data base, including both real-time and non real-time data. Coverage ranges between less than 10,000 and about 45,000 stations. The analysis method is essentially the same as for the first guess and monitoring products and does not use a physical model. The data files include both anomalies and actuals, along with the number of gauges per grid cell. GPCC recommends that the gauge density information is consulted in conjunction with the precipitation data.

Coverage of study region:

- Complete: may be unreliable where gauges are sparse (Figure A21).

Dataset update information:

- Essentially static, updated occasionally.

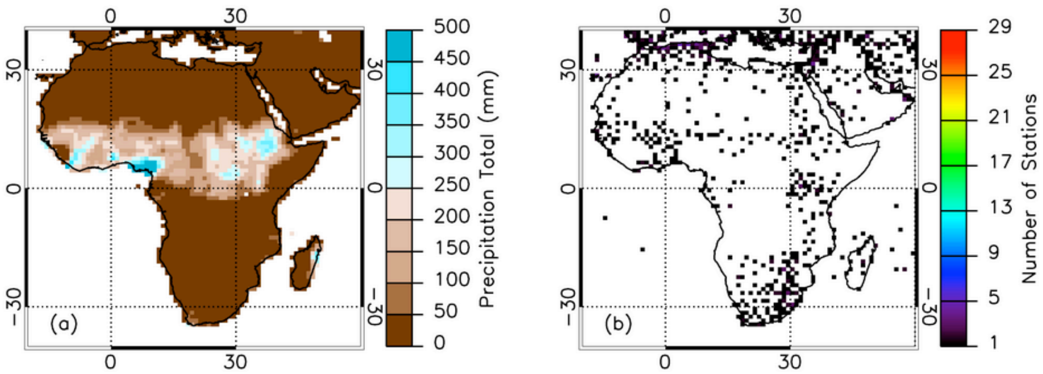

Figure A21: GPCC reanalysis V4 monthly precipitation for July 2007 and the gauges used.

A2.3.5. GPCC – 50-year VASClimO (V1.1)

GPCC also produce VASClimO, a 50-year dataset that is optimised for temporal analysis of rainfall patterns (e.g. trends). This dataset is based on about 9400 stations' records that have undergone rigorous quality control and homogenisation. Unlike the other GPCC products, gauge density is not explicitly provided within the data files. However, a grid density file and monthly 'Jack-knife-error' estimates are provided for the 0.5-degree grid are provided in separate files. The Jack-knife error is the difference of the interpolated value of the location of the nearest station (taking only other stations into account) and the observation at that station. This therefore provides some estimate of the grid cell error.

Coverage of study region:

- Complete, although gauge network is sparse in some regions so product may be unreliable in places (Figure A22).

Dataset update information:

- Planned to be updated in early-mid 2011 to cover 1951-2005 and renamed Homogenized Precipitation Analysis (HOMPRA).

Further information:

- See Section A2.3.2.

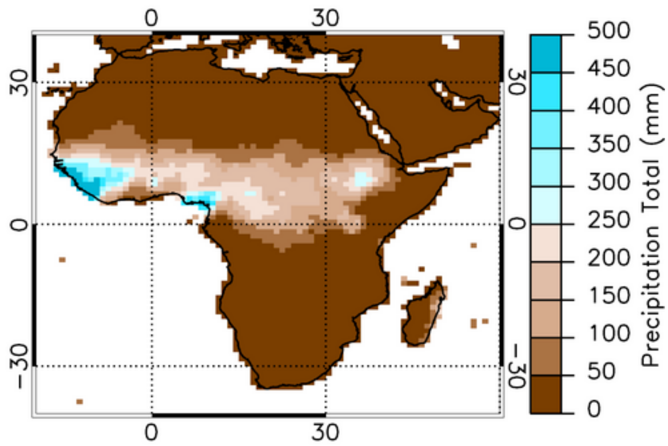

Figure A22: GPCC VASClimO V1.1 monthly precipitation for July 2000.

A2.3.6. GPCC – Gridded Climatological Normals

The final product available from the GPCC is the Gridded Climatological Normals product, which is based on monthly station means between1951-

2000 from about 64,000 stations. Where the station records do not cover the entire 50-year period, 30-year periods are used (e.g. 1961-1990 or 1971-2000) with at least 20 years of available data. Where data covering a 30-year period are unavailable, 10 years of data from any other period is used.

Coverage of study region:

- Complete, although gauge network is sparse in some regions so product may be unreliable in places.

Dataset update information:

- Static.

Further information:

- See Section A2.3.2.

A2.3.7. CRU TS v2.1 and 3.0

Also see Section A1.2.7. The correlation decay distance used for the interpolation of the precipitation data is 450 km.

Coverage of study region:

- Complete (Figure A23), but data may be unreliable where station density is low. Users are advised to consult the station density information supplied with the dataset.

Dataset update information:

- Essentially static, with a new version released every few years.

Further information:

- See Section A1.2.7.

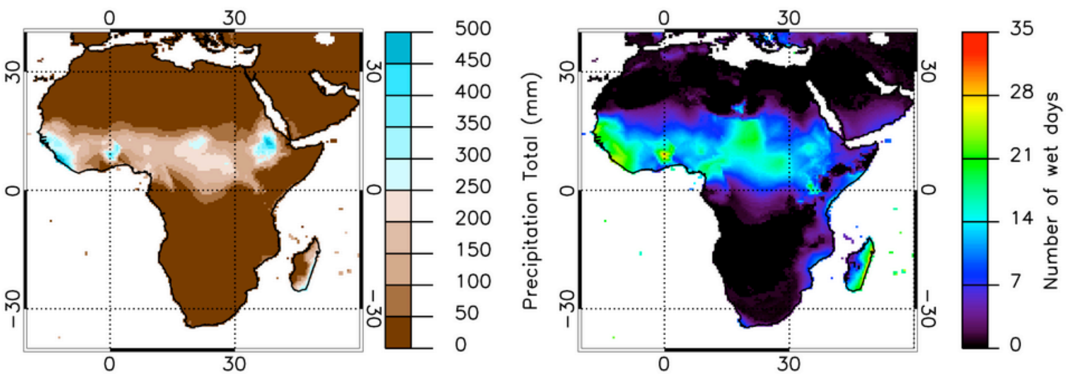

Figure A23: CRU TS3.0 total precipitation and number of wet days for July 2005

A2.3.8. University of Delaware (v2.01)

Also see Section A1.2.6. The University of Delaware gridded precipitation data is based on station observations from a number of different sources, including GHCN (Section A2.3.1). Over Africa, data were also sourced from ‘Sharon Nicholson’s archive of African precipitation data’ (second internet reference below), which contains 1,338 station records between 1950 and 1996. The total number of global stations used in the analysis varies between about 4,100 and 22,000. The method employed for generating the final gridded product is essentially the same as that used for air temperature (Section A1.2.6). The gridded dataset is produced through interpolation of station anomalies (in mm, not %), which are then combined with an equivalent gridded climatology to produce estimates of the absolute mean monthly precipitation. The dataset includes interpolation error estimates derived from the results of cross validation experiments.

Coverage of study region:

- 100% coverage of global land (Figure A24). However, note, where station density is low the data are heavily interpolated and may be unreliable.

Dataset update information:

- Infrequently updated – essentially static.

Further information:

- http://climate.geog.udel.edu/~climate/html_pages/Global2_Ts_2009/README.global_p_ts_2009.html
- http://climate.geog.udel.edu/~climate/html_pages/Tropics_files/README.tropic_precip_ts.html

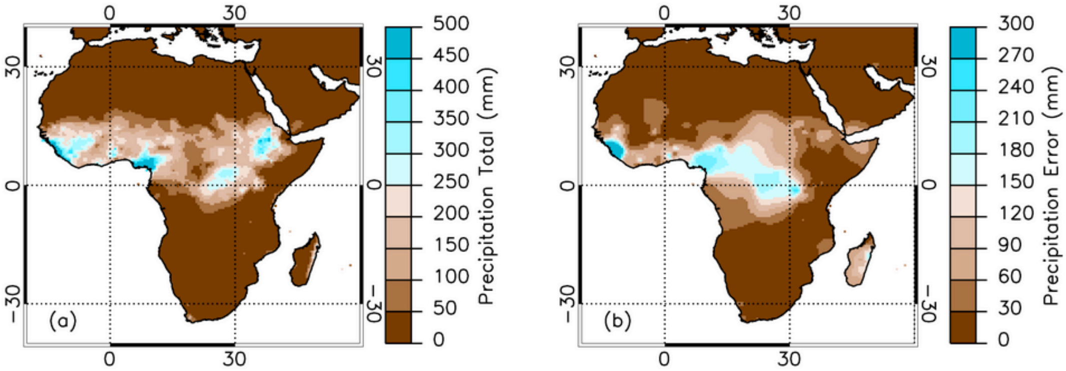

Figure A24: University of Delaware total precipitation and estimated error for July 2007.

A2.3.9. NCDC

Like the NCDC gridded air temperatures, the gridded precipitation product is derived from GHCN station data (Section A1.2.1). Data from approximately 22000 stations are used in the analysis; homogenised records (from the U.S., Canada, and Former Soviet Union) are used in preference to the 'raw' station records. The data set is an anomaly data set, reference to the baseline period 1961-1990. Station anomalies are gridded by averaging values recorded within each grid box.

Coverage of study region:

- Poor (Figure A25).

Data set update information:

- Updated monthly to include most recent month.

Further information:

- http://www.ncdc.noaa.gov/temp-and-precip/ghcn-gridded-products.php
- Also see Section A1.2.1.

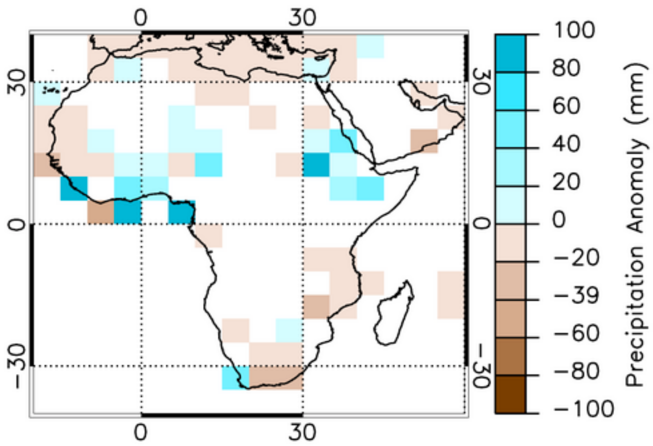

Figure A25: NCDC gridded precipitation analysis for July 2007.

A2.3.10. GPCP (v2.1)

The GPCP monthly precipitation dataset is derived from a combination of satellite and gauge data. The gauge component is sourced from the GPCC reanalysis (1979-2007; Section A2.3.4) and monitoring (2008 onwards; Section A2.3.3) products. The GPCC reanalysis data replace the GHCN+CAMS (Climate Anomaly Monitoring system) analysis implemented in previous versions of the GPCP product for the period 1979-1985. The satellite data originate from a number of sources, including microwave observations from the SSM/I and infrared observations from several geostationary and polar-orbiting sensors. The full dataset includes 27 products, including input and intermediate datasets. The principal final product is the combined gauge-satellite and associated error estimates. There has been considerable effort put into making the dataset consistent through time to account for the different satellite sensors used in the product. However, it is noted that there are still discrepancies between the different satellite instrument phases and the user is referred to the dataset documentation for further information.

Coverage of study region:

- Complete (Figure A26).

Dataset update information:

- Updated with a few months delay (e.g. dataset covered up to September 2009 inclusive in mid March 2010).

Further information:

- http://precip.gsfc.nasa.gov/gpcp_v2.1_comb_new.html
- http://www1.ncdc.noaa.gov/pub/data/gpcp/gpcp-v2.1/doc/V2.1_doc.pdf
- Adler et al. (2003).

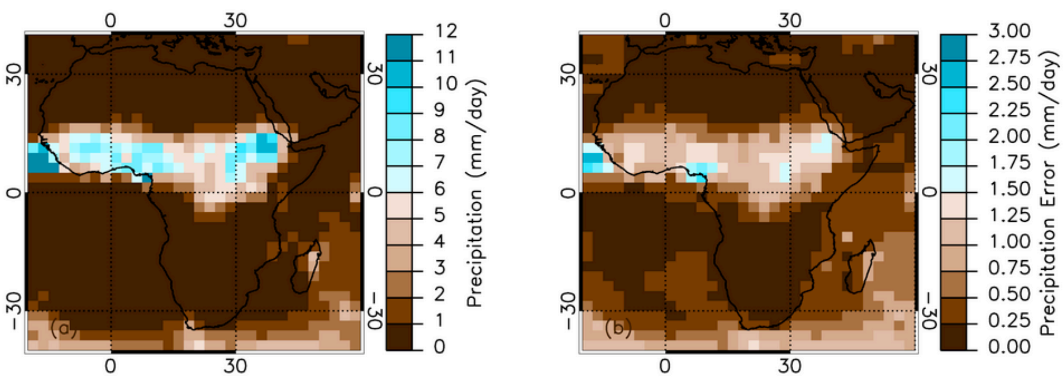

Figure A26: GPCP monthly combined satellite-gauge analysis for July 2007 showing precipitation rate (left) and its estimated absolute error (right).

A2.3.11. GPCP-Int (GPCP Intermediate Products)

In addition to the main GPCP satellite-gauge estimate, the GPCP make several of its intermediate products available to users. Most users will only require the final satellite-gauge combined product described in Section A2.3.10. However, the intermediate product may be useful in some cases where a more in-depth investigation is required. The GPCP intermediate products consist of eleven different analyses that include individual satellite precipitation estimates, a combined multi-satellite estimate and a GPCC gauge-only estimate. For several of the analyses, error estimates, source and sampling information are also available. All data files are available in the same format; users are advised to consult the dataset documentation for further details.

Further information:

- See Section A2.3.10 and Figure A27.

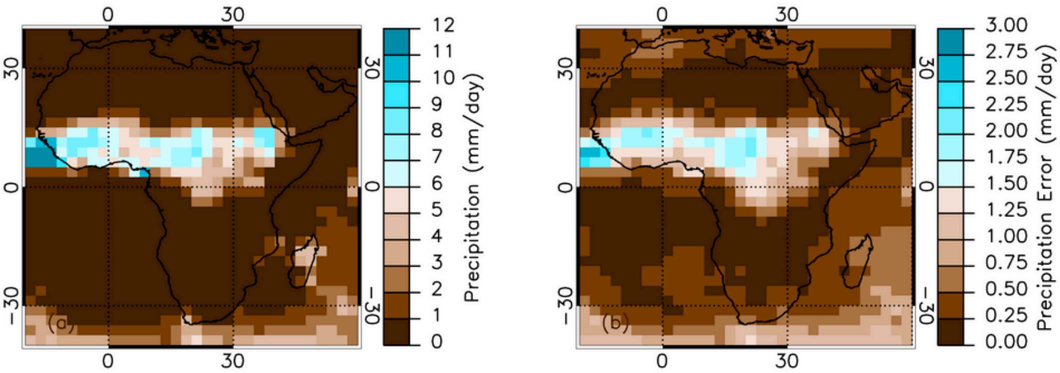

Figure A27: GPCP monthly multi-satellite analysis for July 2007 showing precipitation rate (left) and its estimated absolute error (right).

A2.3.12. CMAP (v1001)

The CPC Merged Analysis of Precipitation (CMAP) is derived from a combination of satellite and rain gauge data. Although similar, it is not the same as the GPCP combined analysis. Several data sources are used in the analysis and these, particularly the satellite datasets, change with time, introducing the possibility of temporal inhomogeneities. For the gauge data, the GPCC monthly monitoring product is used from 1986. Prior to this, an analysis constructed by interpolating station data from Climate Anomaly Monitoring system (CAMS) and GHCN, using the same algorithm as the GPCC analysis, is used. Over the oceans, the gauge data are sourced from atolls and small islands. The satellite precipitation estimates include both infrared and microwave estimates from secondary sources such as the GPI

(Section A2.1.5). Precipitation distributions from the NCEP-NCAR Reanalysis are also used to improve global coverage, particularly over oceans at mid to high latitudes. The final merged product is generated by combining the satellite and reanalysis data into an analysis field, which is then 'pinned' to the gauge observations in an attempt to reduce the overall bias of the final analysis. The data files available include the final merged product, and a secondary version which does not include the reanalysis data. Both datasets include error estimates. The dataset authors warn that the CMAP dataset may contain an artificial downward trend after 1996 (see dataset announcement from the CPC, available on ftp://ftp.cpc.ncep.noaa.gov/precip/cmap/monthly/.

Coverage of study region:

- Complete (Figure A28).

Dataset update information:

- Irregular – not real time.

Further information:

- http://www.cpc.ncep.noaa.gov/products/global_precip/html/wpage.cmap.shtml
- http://www.esrl.noaa.gov/psd/data/gridded/data.cmap.html
- Xie and Arkin (1997).

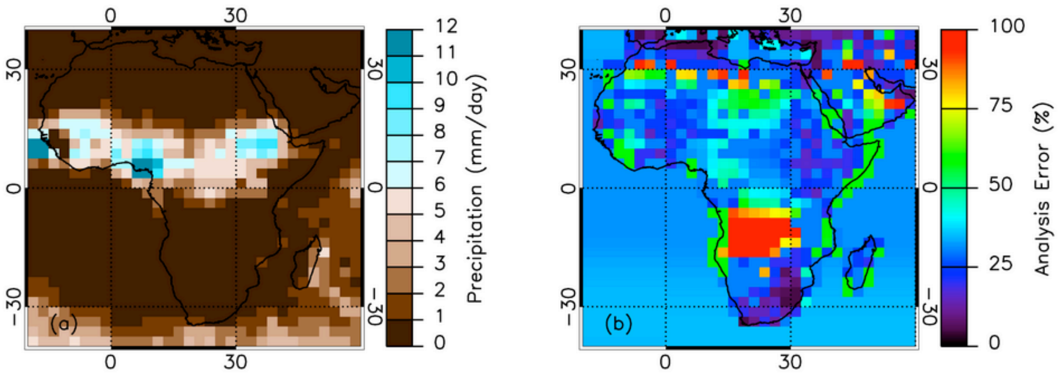

Figure A28: CMAP precipitation rate (left) and its estimated error (right) for July 2007.

A2.3.13. TRMM Combination (3B 43 Version 6)

The TRMM combination monthly product is a combination of the 3-hourly TRMM 3B 42 precipitation estimates (Section A2.1.3) and gauge analyses from the GPCC and Climate Assessment and Monitoring System (CAMS). The gauge data are used first to bias-correct the satellite data fields and then, during merging with the satellite data, to provide the final product using inverse error variance weighting. The product includes RMS error estimates.

Coverage of study region:

- Complete (Figure A29)

Dataset update information:

- Updated about two weeks after the end of each month.

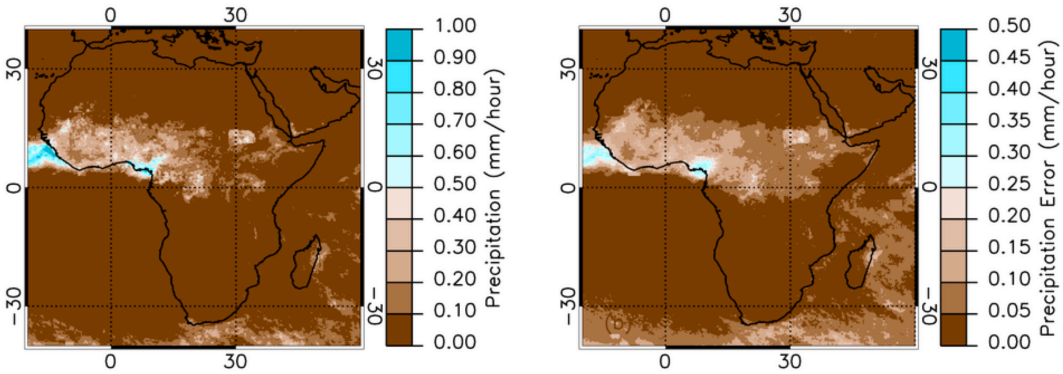

Figure A29: TRMM monthly satellite-gauge analysis for July 2007 showing average rain rate per hour (left) and its estimated RMS error (right).

A2.3.14. CAMS-OPI

The Climate Anomaly Monitoring System (CAMS) and OLR Precipitation Index (OPI) (CAMS-OPI) product is another precipitation dataset originating from the CPC. The product is comparable with the GPCP (Section A2.3.10) and CMAP (Section A2.3.12) monthly products, in that it combines both satellite and gauge precipitation estimates. However, unlike these products, which are available after a few months' delay, the CAMS-OPI is updated in near-real time. As the name suggests, the analysis involves gauge data from the CAMS and satellite precipitation estimates from the outgoing longwave radiation (OLR) Precipitation Index (OPI) developed by Xie and Arkin (1998). The merging technique is similar to that used for CMAP (Section A2.3.12). The CAMS-OPI analysis is primarily designed for real-time monitoring and that users wishing to source data for research purposes should focus on the GPCP and/or CMAP. These datasets are better quality-controlled and include both satellite microwave and infrared estimates of precipitation. The data files made publicly available include the combined analysis, CAMS-only and OPI-only analyses, and the CAMS gauge density.

Coverage of study region:

- Complete for combined and satellite-only analyses, incomplete for CAMS-only analysis (Figure A30).

Dataset update information:

- Updated in near-real time; previous month available shortly after the end of the month.

Further information:

- http://www.cpc.ncep.noaa.gov/products/global_precip/html/wpage.cams_opi.html
- Janowiak and Xie (1999); Xie and Arkin (1998).

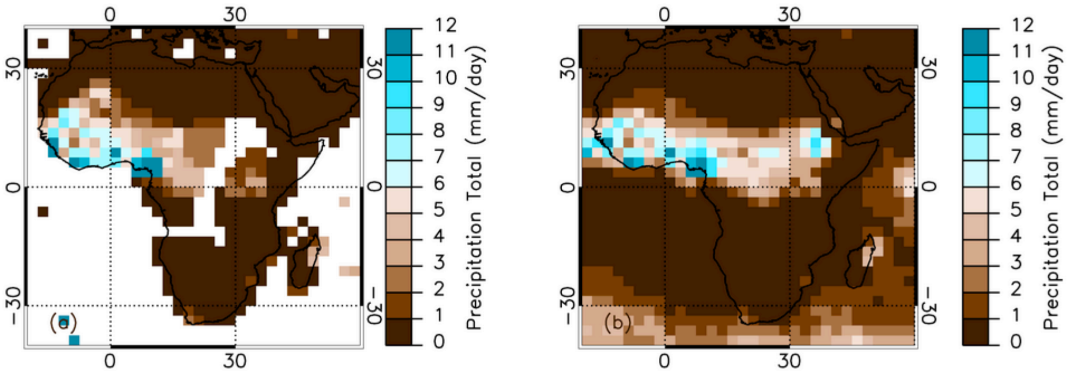

Figure A30: Rain rate for July 2007 from the CAMS-only analysis (left) and the combined satellite-gauge (CAMS-OPI) analysis (right).

A2.3.15. CPC Africa RFE (FEWS-NET)

This dataset is produced from a composite of the daily RFE dataset described in Section A2.1.7.

A2.3.16. TAMSAT

This dataset is produced by aggregating the 10-day rainfall estimates described in Section A2.2.1.

A2.3.17. Intercomparison studies

Nicholson *et al.* (2003a, b) compared monthly GPCC, GPCP, GPI and TRMM 3B-42 and 3B 43 rainfall with a dense gauge dataset over West Africa. They found that in general products that included gauge information outperformed satellite-only products. TRMM 3B 43 was the best of the combined gauge-satellite methods, and TRMM 3B 42 the best of the satellite-only products. All products were more reliable at seasonal than at monthly scales; where even the gauge and gauge-satellite products showed large departures from the reference data. The combined satellite-gauge products did not appear to add much value to the gauge-only GPCC product, a result also found by Ali *et al.* (2005) when comparing GPCP and GPCC over the Sahel. However, this result may not be relevant to the rest of Africa as the GPCC contains a greater density of gauges over West Africa than it does over much of the continent.

Dinku *et al.* (2007) compared monthly 2.5 degree estimates from GPCP, CMAP and TRMM 3B 43 with gridded gauge estimates over Ethiopia. They found that the overall quality of the estimates was good, with CMAP and TRMM 3B 43 performing the best. Again combined satellite-gauge products were less biased than satellite-only products.

Lamptey (2008) compared gridded GPCC gauge estimates and the GPCP-Int satellite-only product over West Africa. The two products agreed on the spatial distribution of rainfall over the region, but not on the magnitude of inter-annual variability. McCollum *et al.* (2000) found that GPCP blended estimates have approximately double the magnitude of gauge estimates over central Africa, and concluded that the GPCP calibration is inappropriate for the region.

Paeth *et al.* (2010) compared station *in situ*, GPCC *in situ*, GPCP blended satellite-*in situ* and TRMM 3B 42 V6 satellite monthly precipitation anomalies over northern Africa during June to August 2007, and found substantial differences, especially dry biases in TRMM over Sudan and Ethiopia.

A2.3.18. Discussion of monthly rainfall datasets

Over the southern Sahel, a rain-gauge density of at least 10 gauges per 2.5° grid box is required to give a monthly precipitation error of less than 10%. The requirement exceeds 20 gauges per 2.5° grid box in the northern Sahel where rainfall is less coherent (Ali *et al.* (2005) cited in Section A2.3.17). The gauge density needed for a given % monthly precipitation error elsewhere in Africa depends on the spatial coherence of monthly rainfall. Mountainous areas (e.g. most of East Africa) and all zones on the edge of monsoon-penetration (e.g. parts of the Sudan, Namibia etc.) are likely to have highly intermittent, patchy rainfall so it is likely that >20 gauges will be needed per 2.5° grid box, and even more for 10-day rainfall which suffers greater, less coherent variations. These requirements are very rarely met in available data for Africa, so in general monthly gauge errors are likely to be much larger than 10%. So improvement of *in situ* data availability is an important objective. This applies for monthly and longer timescales because attribution and seasonal hindcast studies need *in situ* rainfall data predating the satellite era. But satellite-era gauge data are also needed, because the validation studies cited in Section A2.2.3 show the value of local calibration or constraint of satellite rainfall products. Release of monthly data will benefit stakeholders in that long-term trends and their causes can be better assessed; release of daily data will benefit them in that nationally-applicable satellite-based algorithms can be developed and applied to monitoring, studies of extremes, and seasonal forecast verification. This may be best achieved by bilateral agreements in relation to the proposed CSRP study fellowships.

Regarding currently available gauge data, the University of Delaware dataset benefits from the inclusion of Sharon Nicholson's African gauge data for 1950-1996. She has extensive links with African Met services and it is likely that her gauge dataset contains a large number of gauges that have not been included in other datasets. For future development, the University of Cape Town data base includes a dense network for South Africa. At present, the GPCC first guess is recommended for near real time, and the University of Delaware, GPCC reanalysis or GPCC-VASClimO for historical analysis.

Owing to sparsity of available gauge data, we recommend that, at present, datasets with a satellite component should usually be used in preference to those based on gauge-only observations. However we note that Yin and Gruber (2010) demonstrate that even blended satellite-*in situ* datasets may suffer biases owing to changes of *in situ* gauge coverage (Section A2.1.7). Of the satellite-gauge datasets, CMAP and GPCP are long enough (Table A5) and may be sufficiently stable for historical analysis. CMAP is reported to agree well with the GPCP merged analysis over land, and tropical and subtropical oceanic areas, but with some differences over extratropical oceans (Xie and Arkin (1997) cited in Section A2.3.12). TAMSAT appears to be the best option for a historical satellite-only dataset (Section A2.2.3).

For near-real time applications, the CAMS-OPI analysis, operational RFE, TAMSAT or TRMM 3B 43 are recommended.

A3. References

Adler, R.F., 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, and E. Nelkin, 2003: The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present). *J. Hydrometeorol.*, **4**, 1147-1167.

Ali A., A. Amani, A. Diedhiou, and T. Lebel, 2005: Rainfall estimation in the Sahel. Part 2: Evaluation of Raingauge Networks in the CILSS Countries and Objective Intercomparison of Rainfall Products. *J. Appl. Meteorol.*, **44**, 1707-1722.

Arkin, P. A., and B. N. Meisner. 1987: The relationship between large-scale convective rainfall and cold cloud over the western hemisphere during 1982-84. *Mon. Wea. Rev*., **115**, 51-74.

Bergès, J.C., I. Jobard, F. Chopin, and R. Roca, 2010: EPSAT-SG: a satellite method for precipitation estimation; its concepts and implementation for the AMMA experiment, *Ann. Geophys*., **28**, 289-308

Brohan, P., J.J. Kennedy, I. Harris, S.F.B. Tett, and P.D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *J. Geophys. Res.*, **111**, D12106, doi:10.1029/2005JD006548.

Caesar, J., L. Alexander, and R. Vose, 2006: Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded dataset, *J. Geophys. Res*., **111,** D05101, doi:10.1029/2005JD006280.

Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. Wayne Higgins, and J. E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation, *J. Geophys. Res*., **113**, D04110, doi:10.1029/2007JD009132.

Dinku, T., P. Ceccato, E. Grover-Kopec, M. Lemma, S. Connor, and C. Ropelewski, 2007: Validation of satellite rainfall products over East Africa's complex topography. *Int. J. Remote Sensing*, **28**(7), 1503–1526.

Dinku, T., S. Chidzambwa, P. Ceccato, S. Connor, and C. Ropelewski, 2008: Validation of high-resolution satellite rainfall products over complex terrain. *Int. J. Remote Sensing*, **29**(14), 4097–4110.

Dugdale, G., V. McDougall, and J. Milford, 1991: Rainfall estimates in the Sahel from cold cloud statistics: accuracy and limitations of operational systems. In Soil Water Balance in the Sudano-Sahelian zone (Proceedings of the Niamey workshop, Feb 1991). IAHS Publ. no. 199, 1991.

Durre, I., M. J. Menne, and R. S. Vose, 2008: Strategies for evaluating quality-control procedures. *J. Appl. Meteorol. Climatol.,* **47**, 1785-1791.

Fan, Y., and H. van den Dool, 2008: A global monthly land surface air temperature analysis for 1948–present, *J. Geophys. Res*., **113**, D01103, doi:10.1029/2007JD008470.

GCOS, 2006: Systematic Observation Requirements for Satellite-based Products for Climate. Supplemental details to the satellite-based component of the Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC. GCOS -107, WMO/TD No. 1338. Page 21: "Calibration and validation of satellite measurements and related products by high-quality surface-based instruments is a major issue, not only with regard to the capability of the *in situ* instruments, but especially as regards regional variations in precipitation characteristics." and "Such validation is required on a continuous, long-term basis, and needs improved data exchange by all countries."

Grimes, D., E. Pardo-Iguzquiza, and R. Bonifacio, 1999: Optimal areal rainfall estimation using raingauges and satellite data, *J. Hydrol.*, **222**, 93-108

Hansen, J., R. Ruedy, J. Glascoe, and M. Sato, 1999: GISS analysis of surface temperature change. *J. Geophys. Res.,* **104**, 30997-31022.

Hansen, J., R. Ruedy, M. Sato, M. Imhoff, W. Lawrence, D. Easterling, T. Peterson, and T. Karl, 2001: A closer look at United States and global surface temperature change. *J. Geophys. Res*., **106**, 23947-23963.

Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change. Submitted for publication.

Herman, A., V. B. Kumar, P .A. Arkin, and J.V. Kousky, 1997: Objectively Determined 10-Day African Rainfall Estimates Created for Famine Early Warning Systems. *Int. J. Remote Sensing*, **18,** 2147-2159.

Hirpa, F.A. and M. Gebremichael, 2010: Evaluation of high-resolution satellite precipitation products over very complex terrain in Ethiopia. *J. Appl. Meteorol. Climatol*., **49**, 1044-1051.

Huffman, G.J., R.F. Adler, D.T. Bolvin, G. Gu, E.J. Nelkin, K.P. Bowman, E.F. Stocker, and D.B. Wolff, 2007: The TRMM Multi-satellite Precipitation Analysis: Quasi-Global, Multi-Year, Combined-Sensor Precipitation Estimates at Fine Scale. *J. Hydrometeorol*., **8**, 38-55.

Hewitson, B.C. and R.G. Crane, 2005: Gridded area-averaged daily precipitation via conditional interpolation. *J. Climate,* **18,** 41-57

Huffman, G.J., R.F. Adler, M. Morrissey, D.T. Bolvin, S. Curtis, R. Joyce, B McGavock, and J. Susskind, 2001: Global Precipitation at One-Degree Daily Resolution from Multi-Satellite Observations. *J. Hydrometeorol*., **2**, 36-50.

Janowiak, J. E. and P. Xie, 1999: CAMS_OPI: a global satellite-raingauge merged product for real-time precipitation monitoring applications. *J. Climate*, **12**, 3335-3342

Jobard, I., F. Chopin, J. Bergès, A. Ali, T. Lebel, and M. Desbois, 2007: Presentation of the EPSAT-SG method and comparison with other satellite precipitation estimations in the frame of Precip-AMMA. In EGU conference proceedings, Vienna, 15-20 April 2007

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol*., **5**, 487-503

Lamptey, B., 2008: Comparison of gridded multisatellite rainfall estimates with gridded gauge rainfall over West Africa. *J. Appl. Meteorol. Climatol*., **47**, 185–205.

Laws, K., J. Janowiak, and G. Huffman, 2004: Verification of rainfall estimates over Africa using RFE, NASA MPA-RT and CMORPH. In American Meteorological Society Annual Conference, Seattle, Washington, January 11 - 15, 2004.

Lott, J.N. and R. Baldwin, 2002: The FCC integrated surface hourly database, a new resource of global climate data. 13th Symposium on Global Change and Climate Variations, 13-17 January, 2002, Orlando, Florida, American Meteorological Society, Boston, MA, 70-72.

Lott, N., 2004: The quality control of the integrated surface hourly database. 84th American Meteorological Society Annual Meeting, 2004, Seattle, WA, American Meteorological Society, Boston, MA, 7.8 (7pp.)

McCollum, J., A. Gruber, and M. Ba, 2000: Discrepancy between gauges and satellite estimates of rainfall in equatorial Africa. *J. Appl. Meteorol.*, **39**, 666–679.

Mitchell, T.D. and P.D. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.*, **25**, 693-712, doi: 10.1002/joc.1181.

New M, M. Hulme and P.D. Jones, 2000: Representing twentieth century space–time climate variability. Part 2: development of 1901–96 monthly grids of terrestrial surface climate. *J. Climate*, **13**, 2217–2238.

Nicholson, S., and Co-authors. (2003a). Validation of TRMM and other rainfall estimates with a high-density gauge dataset for West Africa. Part 1: Validation of GPCC rainfall product and pre-TRMM satellite and blended products. *J. Appl. Meteorol.,* **42**, 1337–1354.

Nicholson, S., and Co-authors. (2003b). Validation of TRMM and other rainfall estimates with a high-density gauge dataset for West Africa. Part II: Validation of TRMM Rainfall Products. *J. Appl. Meteorol.,* 42, 1355–1368.

Paeth, H., A.H. Fink, S. Pohle, F. Keis, H. Mächel and C. Samimi, 2010: Meteorological characteristics and potential causes of the 2007 flood in sub-Saharan Africa. *Int. J., Climatol.*, **30**, doi: 10.1002/joc2199, in press.

Peterson, T.C., and R.S. Vose, 1997: An overview of the Global Historical Climatology Network temperature database. *Bull. Amer. Meteorol. Soc.*, **78**, 2837-2849.

Peterson, T.C., T.R. Karl, P.F. Jamason, R. Knight, and D.R. Easterling, 1998a: The first difference method: maximizing station density for the calculation of long-term global temperature change. *J. Geophys. Res.*, **103** (D20), 25967-25974.

Peterson, T.C., R. Vose, R. Schmoyer, and V. Razuvaev, 1998b: Global Historical Climatology Network (GHCN) quality control of monthly temperature data. *Int. J. Climatol.*, **18**, 1169-1179.

Roca, R., P. Chambon, I. Jobard, P.-E. Kirstetter, M. Gosset and J.C. Bergès, 2010: Comparing Satellite and Surface Rainfall Products over West Africa at Meteorologically Relevant Scales during the AMMA Campaign Using Error Estimates. *J. Appl. Meteorol. Climatol.*, **49**, 715-731.

Schneider, U., A. Meyer-Christoffer, M. Ziese, A. Becker, B. Rudolf, 2010: Status report and outlook of the Global Precipitation Climatology Centre

(GPCC). Presentation to 2nd Pan-GEWEX-Meeting, 23-27 August 2010, Seattle, WA, USA. Available as www.gewex.org/2010pangewex/GPCC.pdf

Xie, P., and P.A. Arkin, 1996: Analysis of Global Monthly Precipitation Using Gauge Observations, Satellite Estimates, and Numerical Model Prediction, *J. Climate* , **9**, 840-858.

Xie P., and P.A. Arkin, 1997: Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteorol. Soc*., **78**, 2539-2558

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

Xie, P., A. Yatagai, M. Chen, T. Hayasaka, Y. Fukushima, C. Liu, and S. Yang, 2007: A Gauge-Based Analysis of Daily Precipitation over East Asia, *J. Hydrometeorol*., **8**, 607-626.

Yin, X. and A. Gruber, 2010, Validation of the abrupt change in GPCP precipitation in the Congo River Basin. *Int. J. Climatol*., **30**, 110–119. doi: 10.1002/joc.1875

A4. Acronyms and abbreviations

AMMA	African Monsoon Multidisciplinary Analysis
AMSR-E	Advanced Microwave Scanning Radiometer - Earth Observing System
AMSU	*Advanced Microwave Sounding Unit*
CAMS	Climate Anomaly Monitoring System
CMAP	CPC Merged Analysis of Precipitation
CMORPH	CPC MORPHing technique
CORDEX	A COordinated Regional climate Downscaling EXperiment
CPC	Climate Prediction Center
CRU	Climate Research Unit
DWD	Deutscher Wetterdienst
ECMWF	European Centre for Medium-Range Weather Forecasts.
EPSAT-SG	Estimation of Precipitation by Satellites – Second Generation
ERA	ECMWF Reanalysis
GCOS	Global Climate Observing System
GHCN	Global Historical Climate Network
GOES	Geostationary Operational Environmental Satellite
GPCC	Global Precipitation Climatology Centre
GPCP	Global Precipitation Climatology Project
GPI	GOES Precipitation Index
GSFC	Goddard Space Flight Centre
GTS	Global Telecommunication System
HOMPRA	Homogenized Precipitation Analysis
ISD	Integrated Surface Daily
JAXA	Japan Aerospace Exploration Agency
NASA	National Aeronautics and Space Administration
NCDC	National Climatic Data Center
NOAA	National Oceanic and Atmospheric Administration
PERSIANN	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks
PR	Precipitation Radar
RFE	CPC African Rainfall Estimate
SSM/I	Special Sensor Microwave/Imager
TAMSAT	Tropical Applications of Meteorology using SATellite data
TMI	TRMM Microwave Imager
TRMM	Tropical Rainfall Measuring Mission satellite
UNFCCC	United Nations framework Convention on Climate Change
VASClimO	Variability Analysis of Surface Climate Observations (a GPCC monthly *in situ* rainfall dataset designed for studies of long-term changes)

Annex B. Catalogue of Moisture and Aerosol Datasets for Africa.

B1. Surface Moisture

Table B1: Surface Moisture observation data sets

Data Set	Spatial Res.	Spatial Coverage	Temporal Coverage	Temporal Frequency	Variables
Jones-AMSR-E	25 km	Global	2002-2008	12-hourly	Soil Moisture ≤2cm
Njoku - AMSR-E	25 km 1º	Global	2002-now	12-hourly & Daily Monthly	Soil moisture ~ top 1 cm (g cm$^{-3)}$
University of Delaware	0.5º	Global	1900-2008	Monthly	Evapotranspiration, snow melt, surplus, soil moisture, snow cover
CPC	0.5º	Global	1948-now	Monthly	Soil moisture, evaporation and runoff
FEWS-NET	0.1º	Africa	recent	dekadal	Moisture index
AQUASTAT	0.5 º 5 arc min	Global Global	1961-1990 N/A	Mon. Clim.	Evapotranspiration Soil Moisture
SMOS	50 km	Global	2009-now	12-hourly	Soil Moisture

B1.1.1. AMSR-E (NSIDC: Jones algorithm)

NSIDC produce AMSR-E soil moisture estimates using two different algorithms: Jones et al (2009), described in this section, and Njoku et al (2003), described in Section B1.1.2. The Jones et al (2009) data set is a static data set and consists of orbit-based (approximately 12-hourly) estimates of near-surface soil moisture from the AMSR-E satellite instrument, a passive microwave radiometer in polar orbit. Ascending and descending orbits are processed separately so it is possible to obtain separate night and day soil moisture information. Retrievals are performed over land for non-precipitating, non-snow, and non-ice covered conditions using the 6.9 GHz frequency, unless radio interference is present, in which case the 10.7 GHz channel is used. The retrieval includes a dynamic correction for open water. The native AMSR-E pixels are interpolated to a 25-km grid through inverse distance squared weighting. Validation of the data has been performed through comparisons with in situ soil moisture estimates and satellite observations of related variables (e.g. precipitation). These results suggest the accuracy is quite good under conditions of low vegetation optical depth. Performance is poor where vegetation is dense, with the exception of tundra regions.

Coverage of study region:

- Fairly sparse, depending on orbital tracks and data availability (Figure B1).

Data set update information:

- Static.

Further information:

- http://nsidc.org/data/nsidc-0451.html
- Jones, L. A., J. S. Kimball, E. Podest, K. C. McDonald, S. K. Chan, E. G. Njoku. 2009. A Method for Deriving Land Surface Moisture, Vegetation, and Open Water Fraction from AMSRE. IEEE Int. Geosci. Rem. Sens. Symp. IGARSS '09, July 13-17, Cape Town, South Africa.

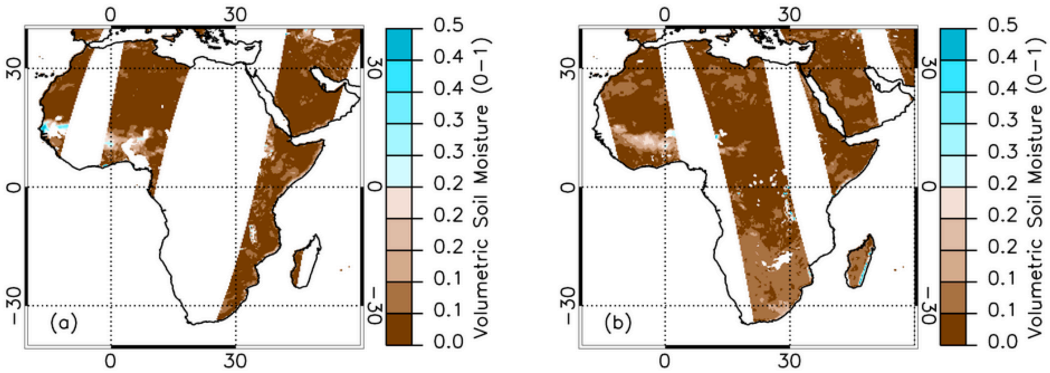

Figure B1: Example of Jones AMSR-E 12-hourly soil moisture product for 01 July 2007 for (a) descending (approx 01:30 local time) and (b) ascending (approx 13:30 local time) orbits.

B1.1.2. AMSR-E (NSIDC: Njoku algorithm)

NSIDC produce AMSR-E soil moisture estimates using two different algorithms: Jones et al (2009), described in Section B1.1.1 and Njoku et al (2003), described in this section. The Njoku soil moisture estimates are available as half-orbit granules (most locations viewed twice per day, every 12 hours), and as gridded daily and monthly products. The data represent the top ~1 cm only and are derived from polarisation ratios at both 6.9 and 10.7 GHz. The retrieval process involves simultaneously estimating soil moisture, vegetation water content and surface temperature by using a microwave radiative transfer model to simulate the observed brightness temperatures (BT). An iterative procedure is used to minimise the weighted-sum of squared differences (chi squared) between observed and modelled BTs, with the Levenberg-Marquardt algorithm used in for minimisation. Once the retrieval has converged, the corresponding values of soil moisture are assigned to the distributed data product. Errors are likely to increase with increasing vegetation cover and for very dense vegetation, soil moisture cannot be retrieved. High chi-squared values indicate poor minimisation and increased errors in the retrieval. A large number of iterations is also indicative of less reliable data.

The daily products include the chi-squared values and number of iterations used in the estimates and should be consulted for quality purposes. Both these products are available from the NSIDC. The monthly product does not contain this information, providing only the standard deviation of all the daily data in a 1 degree grid cell. The monthly product is spatially averaged to 1° resolution and is available from the GSFC.

Coverage of study region:

- Reasonably complete for daily product. Complete for monthly product (Figure B2).

Data set update information:

- Updated daily/monthly in near real time.

Further information:

- http://nsidc.org/data/ae_land.html (half-orbit)
- http://nsidc.org/cgi-bin/get_metadata.pl?id=ae_land3 (gridded daily)
- http://gcmd.nasa.gov/KeywordSearch/Metadata.do?Portal=GCMD&KeywordPath=&NumericId=24107&MetadataView=Full&MetadataType=0&lbnode=mdlb1 (gridded monthly)
- Njoku, E. G., T. L. Jackson, V. Lakshmi, T. Chan, and S.V. Nghiem. 2003. Soil moisture retrieval from AMSR-E. IEEE Transactions on Geoscience and Remote Sensing 41 (2): 215-229.

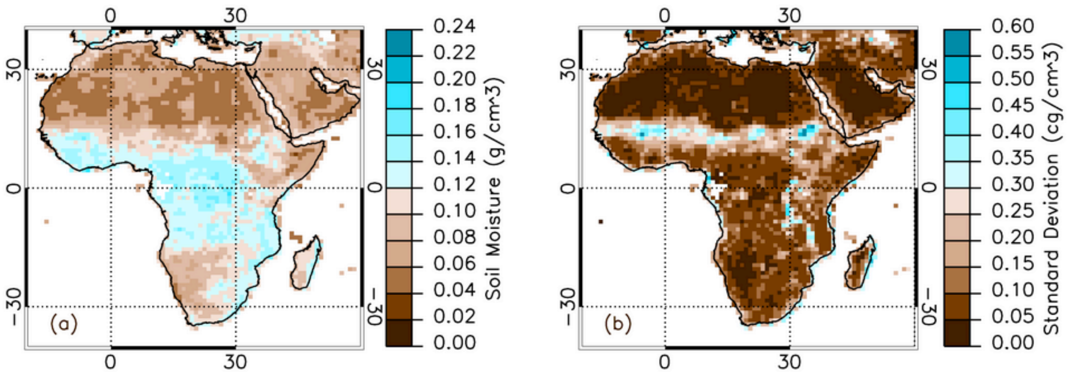

Figure B2: Example of Njoku AMSR-E monthly soil moisture product for July 2007 showing (a) average soil moisture and (b) the standard deviation in each grid cell.

B1.1.3. University of Delaware Datasets

Various hydrological data sets are also available from the University of Delaware. Of note is the Terrestrial Water Budget data archive (V1.03 is the most recent: 1900-2008), which includes monthly potential evapotranspiration (E^o: in mm), monthly actual evapotranspiration (E: in mm), average-monthly deficit (E^o- *E:* in mm), mid-monthly soil-moisture depth (*w:* in mm), mid-monthly water equivalent of the snow pack (w^s: in mm), monthly snow melt (*M*: in mm), and monthly surplus (*S*: in mm). The parameters are estimated from the University of Delaware gridded monthly temperature and precipitation fields described in Annex A.. The estimated parameters are based on observed average monthly precipitation and an estimate of potential evapotranspiration derived from observed average monthly surface temperature*.* Both monthly climatologies and time series are available from the website.

Coverage of study region:

- Complete (Figure B3).

Data set update information:

- Static. Updates periodically.

Further information:

- http://climate.geog.udel.edu/~climate/html_pages/download.html#im2
- Willmott, C. J., C. M. Rowe, and Y. Mintz, 1985b. Climatology of the Terrestrial Seasonal Water Cycle. Journal of Climatology, 5, 589-606.

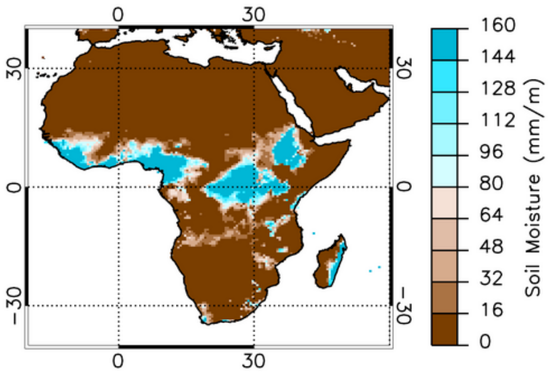

Figure B3: Example of University of Delaware monthly soil moisture product for July 2007.

B1.1.4. CPC Soil Moisture

The CPC soil moisture product is estimated via a one-layer hydrological model, which uses observed monthly temperature and precipitation fields as input. The model parameters are tuned to reproduce runoff in several small river basins in Oklahoma, USA; constant values of these parameters are used for the globe. The analysis uses CPC-observed temperatures, from the CPC GHCN-CAMS dataset (Annex A), and observed precipitation, from the data set of Chen et al (2002).

http://www.cpc.noaa.gov/products/Soilmst_Monitoring/Papers/2003JD004345.pdf

Coverage of study region:

- Complete (Figure B4).

Data set update information:

- Updated in real time, about eight days after the end of the month.

Further information:

- http://www.cpc.ncep.noaa.gov/soilmst/descrip.htm
- http://www.cpc.noaa.gov/products/Soilmst_Monitoring/introduction.shtml
- Yun Fan and Huug van den Dool, 2004, Climate Prediction Center global monthly soil moisture data set at 0.5° resolution for 1948 to present, JGR, Vol. 109, D10102, doi:10.1029/2003JD004345.
- Chen, M., P. Xie, J. E. Janowiak and P. A. Arkin, 2002: Global land precipitation: A 50-yr monthly analysis based on gauge observations. J. Hydrometeor., 3, 249-266

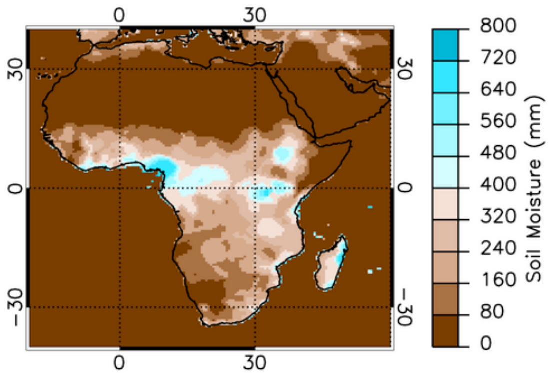

Figure B4: Example of CPC monthly soil moisture product for July 2007.

B1.1.5. FEWS-NET Moisture Index

A 10-day (dekadal) moisture index is produced within the framework of the Famine Early Warning Systems Network (FEWS-NET) project (http://earlywarning.usgs.gov/adds/index.php). The moisture index is produced according to the following formulation:

[(Precipitation + Available Soil Water) / Potential Evapotranspiration] * 100

The precipitation data is sourced from the CPC RFE product described in Annex A. The potential evapotranspiration is calculated from daily Global Data Assimilation System (GDAS) analysis fields. The available soil water is calculated on a dekad-by-dekad basis using the value for the previous month, and the precipitation and potential evapotranspiration values for the current month.

Coverage of study region:

- Near complete (Figure B5).

Data set update information:

- Updated in real time, about eight days after the end of the month.

Further information:

- http://earlywarning.usgs.gov/adds/readme.php?symbol=mi
- Shuttleworth, J., 1992. Evaporation. Chapter 4 in Handbook of Hydrology. (D.Maidment, ed.). McGraw-Hill, Inc., New York.

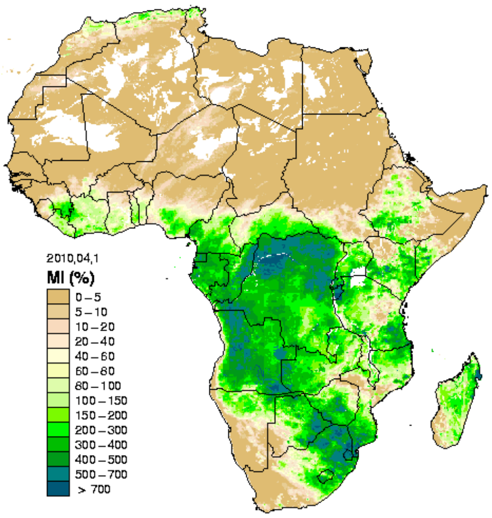

Figure B5: Example of FEWS-NET moisture index for first ten days of April 2010.
Source: http://earlywarning.usgs.gov/adds/index.php?img1=mi&extent=af

B1.1.6. AQUASTAT

AQUASTAT provides global information on water and agriculture for the Food and Agricultural Organisation of the United Nations (FAO). Several static data sets are available from their website, including evapotranspiration and maximum available soil moisture capacity. The evapotranspiration data are provided as mean monthly values for the globe, excluding Antarctica, for the 1961-1990 baseline period. The calculation of evapotranspiration utilises data from the CRU. The soil moisture data (mm/m) were calculated from the Derived Soil Properties of the Digital Soil Map of the World (FAO, 1998).

Coverage of study region:

- Complete (Figure B6).

Data set update information:

- Static.

Further information:

- **http://www.fao.org/nr/water/aquastat/watresafrica/index3.stm**.
- FAO. 1998. Digital soil of the world and derived soil properties. FAO Land and Water Digital Media Series No 1. Rome

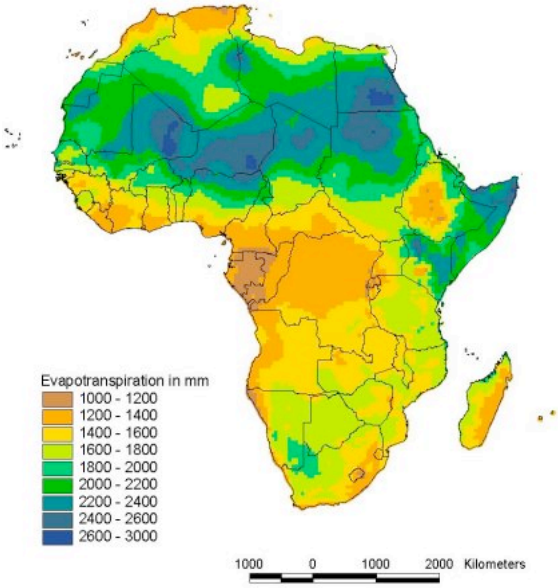

Figure B6: Average annual reference evapotranspiration (1961-1990) from Aquastat.
Source: http://www.fao.org/nr/water/aquastat/watresafrica/index3.stm**.**

B1.1.7. SMOS

Launched in November 2009, the ESA SMOS mission is designed to provide global observations of soil moisture and ocean salinity (SMOS). These parameters are estimated from the L-band (1.4 GHz) microwave emission measurements made by the instrument. The instrument is in polar orbit and provides global coverage in 3 days. The mission was to be in the commissioning phase until the end of April 2010, after which it would become operational and the data will be released for public use. The soil moisture data (orbit - level-2, global – level 3 products) will available through ESA's online catalogue, EOLI (http://envisat.esa.int/earth/www/object/index.cfm?fobjectid=5225). The data have a spatial resolution of 50 km, and are expected to have an accuracy of better than 4 % m^3/m^3 for vegetation water content less than 4 kg/m^2.

Coverage of study region:

- Expected to be near complete every three days.

Data set update information:

- Available in near real time.

Further information:

- http://www.esa.int/esaLP/ESAMBA2VMOC_LPsmos_0.html
- Kerr, Y.H., P. Waldteufel, J.P. Wigneron, J.M. Martinuzzi, J.M. Font and M. Berger (2001), Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. IEEE Transactions on Geoscience and Remote Sensing, 39(8): 1729−1735.

B2. Aerosol

Table B2: Aerosol data sets

Data Set	Spatial Res.	Spatial Coverage	Temporal Coverage	Temporal Frequency	Variables
AERONET	Stations	Global	1993-now	Sub-daily	Visible AOD, size distributions.
MODIS	10 km 1°	Global	2000-now	Orbital Daily, 8-day, monthly	Visible AOD, aerosol type, Mass Concentration
MISR	1.1-17.6 km (orbital) 0.5° (global)	Global	2000-now	Daily, monthly, seasonally, annually	Visible AOD, aerosol composition & size.
GLOB AEROSOL	10 km	Global	1995-2007	Orbital, daily.	AOD at 555 and 870nm, Angstrom coefficient and speciation
TOMS	1x1.25°	Global	1996-2005	Orbital, daily.	Aerosol Index

B2.1.1. AERONET (v2.0)

AERONET is a network of more than 800 ground-based radiometers that are capable of observing aerosol optical thickness (AOD) and size distributions. Aerosol parameters are estimated at several visible and near-infrared wavelengths (exact wavelengths can vary from station-to-station). Direct solar radiometric measurements are made between 7 am and 7 pm at these wavelengths, and the AOD is calculated from spectral extinction at each wavelength based on the Beer-Bouguer Law. Various corrections are required, including accounting for the effects of Rayleigh scattering, Ozone, and other trace gases. Three levels of AOD data are available from NASA: Level 1.0, which are unscreened for cloud, Level 1.5, which are cloud screened and available in near real time, and Level 2, which are cloud screened and quality assured, but are not available until about 12 months are acquisition. Level 2.0 data include both fine and coarse mode AOD, in addition to fine mode fraction. The most recent version of the data is version 2.0, which was released in July 2006.

Coverage of study region:

- Sparse (Figure B7).

Data set update information:

- Level 1.5 data updated in near real time with a couple of days lag.

Further information:

- http://aeronet.gsfc.nasa.gov/
- Holben B.N., Eck, T.F., Slutsker, I., Tanre, D., Buis, J.P., Setzer, A., Vermote, E., Reagan, J.A., Kaufman, Y., Nakajima, T., Lavenu, F., Jankowiak, I., & Smirnov, A. (1998). AERONET - A federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment*, *66*, 1-16.

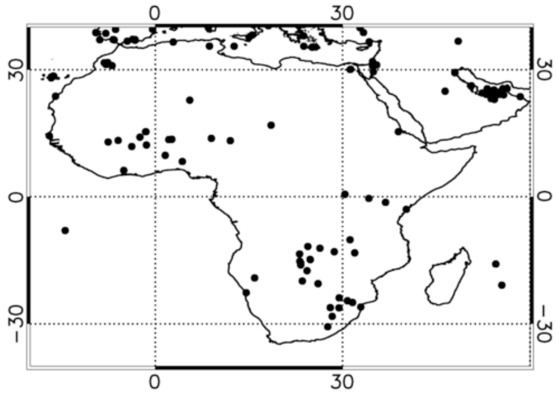

Figure B7: Location of AERONET sites in study region.

B2.1.2. MODIS (collection 005)

MODIS is an imaging radiometer that measures radiances at the top of the atmosphere over a range of visible to infrared wavelengths. AOD is retrieved in several spectral bands in the visible to near-infrared over both land and ocean. Retrievals are performed through fitting the observed radiances to those simulated assuming various aerosol models (e.g. small particles, marine aerosol, dust aerosol). Contributions from both a coarse and fine aerosol mode are assumed. The exact contents of the MODIS aerosol data files depend on the product used, and whether the retrieval is land- or ocean-based. As the retrievals utilise visible wavelengths, the data are day-time only and are not available in cloudy conditions. Orbital products at 10 km resolution are available for both the MODIS/Terra (overpass: 10:30 am/pm) and MODIS/Aqua (overpass: 1:30 am/pm) instruments. The global daily, 8-day and monthly products are generated from the orbital products, and are gridded at 1-degree resolution.

Coverage of study region:

- Patchy – depends on cloud and orbital coverage (Figure B8)

Data set update information:

- Available in near real time.

Further information:

- http://experts.nasa.gov/MOD04_L2/index.html
- https://wist.echo.nasa.gov/
- Remer, L. A., Tanré, D., & Kaufman, Y. J. (2009). Algorithm for Remote Sensing of Tropospheric Aerosol from MODIS: Collection 005, Revision 2, Product ID: MOD04/MYD04
- Tanré, D., Kaufman, Y. J., Herman, M., Mattoo, S. (1997). Remote sensing of aerosol properties over oceans using the MODIS/EOS spectral radiances. Journal of Geophysical Research, 102(D14), 16971-16988.

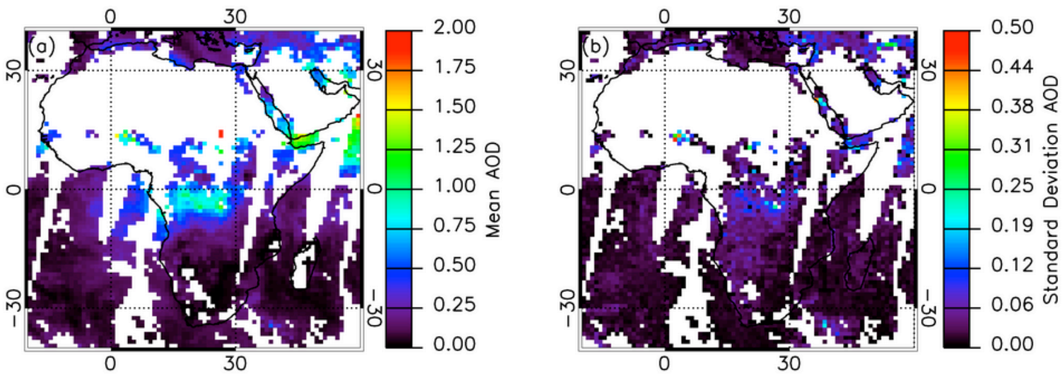

Figure B8: AOD from MODIS (MOD08 Product for Terra) showing (a) mean AOD and (b) the standard deviation of the AOD per 1º grid cell for 01 July 2007.

B2.1.3. MISR

MISR is a multi-angle radiometer on board NASA's Terra satellite platform. It observes the earth at nine different angles in four wavebands (blue, green, red and near-infrared). This multi-angle approach allows comparable measurements to be made through multiple atmospheric path lengths, thus providing detailed information on the state of the atmosphere. In essence, the retrieval process consists of fitting simulated radiances, using various aerosol models and auxiliary data sets, to those observed by MISR in order to estimate AOD. Both orbital (Level 2) and global (Level 3) products are available to users via the NASA WIST system (https://wist.echo.nasa.gov/api/). The spatial resolution of the orbital products ranges from 1.1 km to 17.6 km. These orbital products are spatially and temporally averaged to produce the global products at 0.5° resolution. Global coverage is achieved in 9 days.

Coverage of study region:

- Patchy – depends on cloud and orbital coverage (Figure B9)

Data set update information:

- Near real time.

Further information:

- http://gcmd.nasa.gov/KeywordSearch/Metadata.do?Portal=GCMD&MetadataType=0&MetadataView=Full&KeywordPath=&EntryId=MIL2ASAE2
- Diner, D.J., J.C. Beckert, T.H. Reilly, C.J. Bruegge, J.E. Conel, R. Kahn, J.V. Martonchik, T.P. Ackerman, R. Davies, S.A.W. Gerstl, H.R. Gordon, J-P. Muller, R.B. Myneni, R.J. Sellers, B. Pinty, and M.M. Verstraete (1998). Multiangle Imaging SpectroRadiometer (MISR) description and experiment overview. IEEE Trans. Geosci. Rem. Sens., Vol. 36, pp 1072-1087.
- Martonchik, J.V., D.J. Diner, R. Kahn, T.P. Ackerman, M.M. Verstraete, B. Pinty, and H.R. Gordon (1998). Techniques for the retrieval of aerosol properties over land and ocean using multi-angle imaging, IEEE Trans. Geosci. Rem. Sens., VOl. 36, pp 1212-1227.

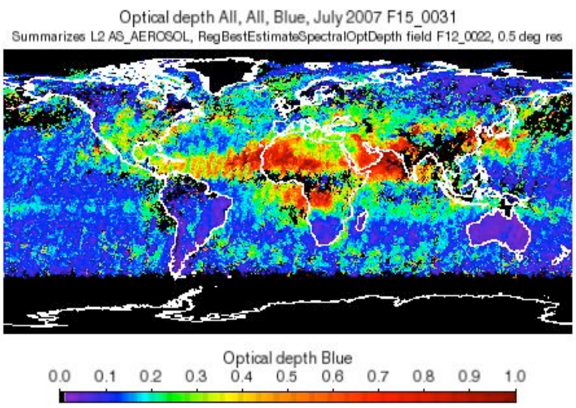

Figure B9: MISR AOD for July 2007. Source: http://eosweb.larc.nasa.gov/PRODOCS/misr/level3/level3_CGAS_small.html**.**

B2.1.4. GLOBAEROSOL

GLOBAERSOL is an ESA Data User Element (DUE) project designed to provide users with global aerosol observations from satellite sensors ATSR-2, AATSR, MERIS and SEVIRI. AOD estimates are provided for the individual sensors on an orbit-by-orbit basis at the resolution of the radiance products for each sensor, and as a global gridded product. GLOBAEROSOL also produces a merged product, which combines the AOD from all sensors at 10 km spatial resolution. Statistical products, summarising aerosol over previous weeks or years, are also available to users. Retrievals are performed using two different algorithms. The Oxford-RAL Aerosol and Cloud (ORAC) retrieval is used for the ATSR-2/AATSR and SEVIRI sensors. This algorithm utilises optimal estimation whereby simulated satellite radiances are fitted to the equivalent observations in order to provide a best estimate of the aerosol parameters. For the MERIS data, the operational ESA algorithm is used. Validation of products has been performed through comparison with in situ observations from AERONET stations (see Section B2.1.1).

Coverage of study region:

- Patchy (Figure B10) – influenced by sensor coverage and presence of cloud.

Data set update information:

- Static.

Further information:

- http://www.globaerosol.info/
- G.E. Thomas et al., 2009, “Oxford-RAL Aerosol and Cloud (ORAC): Aerosol retrievals from satellite radiometers” In “Satellite aerosol remote sensing over land”, Kokhanovsky and De Leuuw (eds).
- R. Santer et al., 1999, “Atmospheric correction over land for MERIS” Int. J. Remote Sensing, 20:9, 1819–1840
- D. Antoine and A. Morel, 1999, “A multiple scattering algorithm for atmospheric correction of remotely sensed ocean colour (MERIS instrument): principle and implementation for atmospheres carrying various aerosols including absorbing ones”, Int. J. Remote Sensing, 20:9, 1875–1916

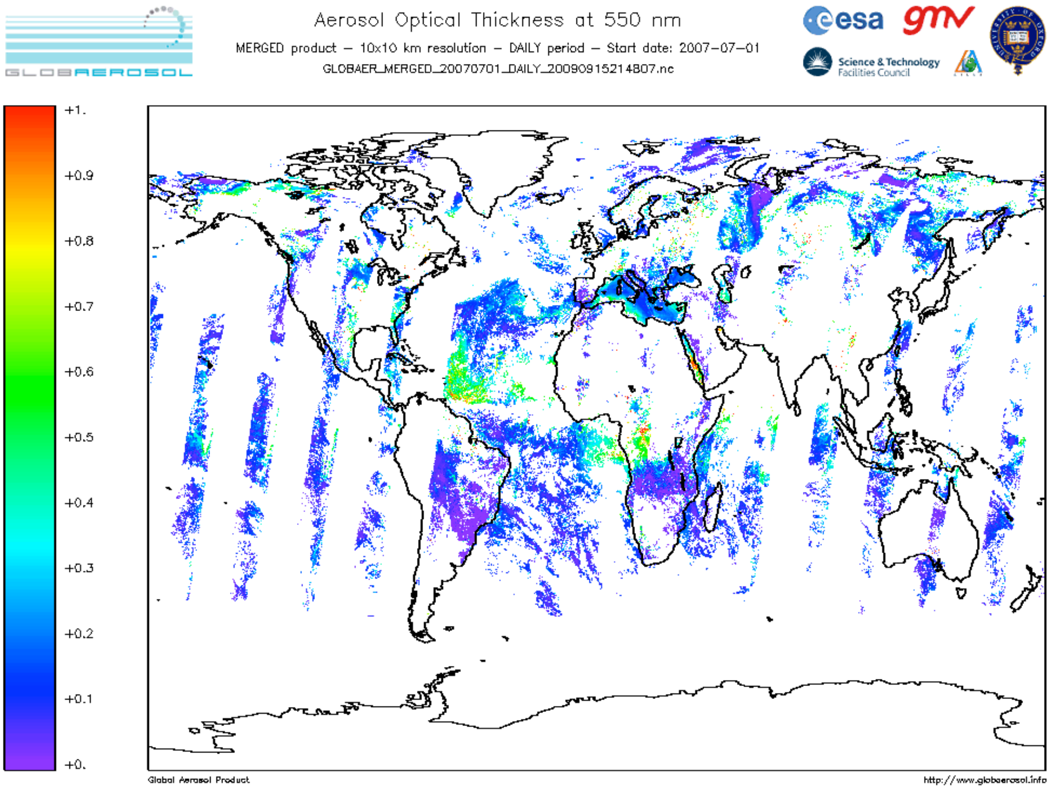

Figure B10: GLOBAEROSOL merged aerosol product for July 2007. Source: http://www.globaerosol.info/project_description/data.htm.

B2.1.5. TOMS

Coverage of study region:

- Nearly complete daily (Figure B11).

Data set update information:

- The http://toms.gsfc.nasa.gov/ site cited below has data in near-real time (about 2 days after the event).

Further information:

- http://toms.gsfc.nasa.gov/aerosols/aerosols_v8.html
- http://gcmd.nasa.gov/KeywordSearch/Metadata.do?Portal=GCMD&MetadataType=0&MetadataView=Full&KeywordPath=&EntryId=GES_DISC_TOMSEPL3_V008

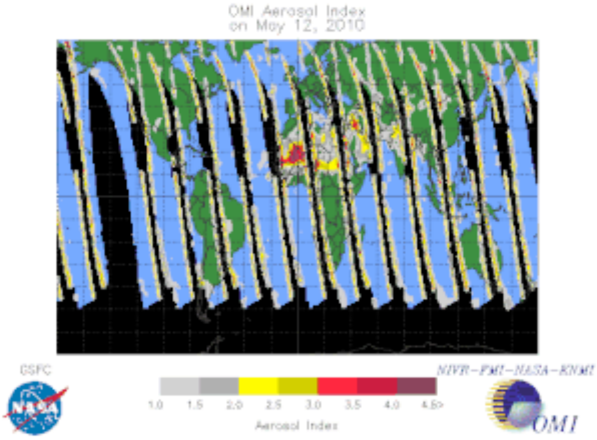

Figure B11. TOMS aerosol index, 12th May 2010. Note the high values in the western Sahara.
Source: http://toms.gsfc.nasa.gov/aerosols/aerosols_v8.html

Acronyms and abbreviations

AERONET	AErosol RObotic NETwork
AMSR-E	Advanced Microwave Scanning Radiometer - Earth Observing System
ATSR-2	Along-Track Scanning Radiometer - 2
AATSR	Advanced Along-Track Scanning Radiometer
AOD	Aerosol Optical Depth
CPC	Climate Prediction Center
CRU	Climate Research Unit
ESA	European Space Agency
GSFC	Goddard Space Flight Centre
MERIS	MEdium Resolution Imaging Spectrometer
MISR	Multi-angle Imaging SpetroRadiometer
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NSIDC	National Snow and Ice Data Center
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SMOS	Soil Moisture and Ocean Salinity