Earth Syst. Sci. Data, 7, 275–287, 2015
www.earth-syst-sci-data.net/7/275/2015/
doi:10.5194/essd-7-275-2015


Open Access Earth System Science Data

# A global satellite-assisted precipitation climatology

**C. Funk[1,3], A. Verdin[2], J. Michaelsen[3], P. Peterson[3], D. Pedreros[1,3], and G. Husak[3]**

[1]U.S. Geological Survey, Earth Resources Observation and Science Center, Sioux Falls, SD, USA
[2]University of Colorado, Boulder, CO, USA
[3]University of California, Santa Barbara Climate Hazards Group, Santa Barbara, CA, USA

*Correspondence to:* C. Funk (cfunk@usgs.gov)

Received: 6 February 2015 – Published in Earth Syst. Sci. Data Discuss.: 12 May 2015
Revised: 22 September 2015 – Accepted: 23 September 2015 – Published: 13 October 2015

**Abstract.** Accurate representations of mean climate conditions, especially in areas of complex terrain, are an important part of environmental monitoring systems. As high-resolution satellite monitoring information accumulates with the passage of time, it can be increasingly useful in efforts to better characterize the earth's mean climatology. Current state-of-the-science products rely on complex and sometimes unreliable relationships between elevation and station-based precipitation records, which can result in poor performance in food and water insecure regions with sparse observation networks. These vulnerable areas (like Ethiopia, Afghanistan, or Haiti) are often the critical regions for humanitarian drought monitoring. Here, we show that long period of record geo-synchronous and polar-orbiting satellite observations provide a unique new resource for producing high-resolution (0.05°) global precipitation climatologies that perform reasonably well in data-sparse regions.

Traditionally, global climatologies have been produced by combining station observations and physiographic predictors like latitude, longitude, elevation, and slope. While such approaches can work well, especially in areas with reasonably dense observation networks, the fundamental relationship between physiographic variables and the target climate variables can often be indirect and spatially complex. Infrared and microwave satellite observations, on the other hand, directly monitor the earth's energy emissions. These emissions often correspond physically with the location and intensity of precipitation. We show that these relationships provide a good basis for building global climatologies. We also introduce a new geospatial modeling approach based on moving window regressions and inverse distance weighting interpolation. This approach combines satellite fields, gridded physiographic indicators, and in situ climate normals. The resulting global 0.05° monthly precipitation climatology, the Climate Hazards Group's Precipitation Climatology version 1 (CHPclim v.1.0, doi:10.15780/G2159X), is shown to compare favorably with similar global climatology products, especially in areas with complex terrain and low station densities.

## 1 Introduction

Systematic spatial variations in climate have been studied since at least the first century AD, when Ptolemy's *Geographia* identified the earth's polar, temperate, and equatorial temperature zones. Analysis of these climatological surfaces continues to be an important aspect of environmental monitoring and modeling. In the 1960s, computers enabled the automatic interpolation of point data, and several important algorithms such as Shepard's modified inverse distance weighting function (Shepard, 1968) and optimal surface fitting via kriging (Krige, 1951; Matheron, 1963) were developed. The value of spatially continuous ancillary data, such as elevation, was soon recognized (Willmott and Robeson, 1995) and the current state-of-the-science climatologies all use background physiographic indicators combined with in situ observations. The most widely-used current global climatologies, such as those produced by the University of East Anglia's Climatological Research Unit (CRU) (New et al., 1999, 2002), and the Worldclim (Hijmans et al., 2005) global climate layers, typically base their estimates on elevation, latitude, and longitude. Daly et al. (1994) used locally varying regressions fit to the topographic facets, while the CRU

and Worldclim climatologies use thin-plate splines (Hutchinson, 1995) to minimize the roughness of the interpolated field, with the degree of smoothing determined by generalized cross validation. The Global Precipitation Climatology Centre (GPCC) generates their climatology products based on the interpolation of a very large database of precipitation normals (Becker et al., 2013; Schneider et al., 2014).

In Africa, Climate Hazards Group (CHG) scientists have demonstrated the utility of satellite fields as a source of ancillary data for climatological precipitation and air temperatures surfaces (Funk et al., 2012; Knapp et al., 2011). This new approach combines satellite fields, gridded physiographic indicators, and in situ climate normals using local moving window regressions and inverse distance weighting interpolation. Expanding from our work in Africa, we have produced a global 0.05° monthly precipitation climatology, the Climate Hazards Group Precipitation Climatology version 1 (CHPclim v.1.0, http://dx.doi.org/10.15780/G2159X). This paper summarizes our statistical approach and modeling results, and presents a validation of the resulting data set. The CHPclim version 1, Worldclim version 1.4 release 3 (Hijmans et al., 2005), CRU CL 2.0 (New et al., 2002, 1999), and the GPCC CLIM M V2015 (doi:10.5676/DWD_GPCC/CLIM_M_V2015_025, Becker et al., 2013; Schneider et al., 2014) climatologies are compared with independent sets of station normals for Colombia, Afghanistan, Ethiopia, The Sahel, and Mexico. The climatologies are also compared with each other, and with a gridded validation data set in Ethiopia.

## 2 Data

### 2.1 Precipitation normals

Two sets of monthly precipitation normals (long-term averages) were used to create the CHPclim. The first set was a collection of 27 453 monthly station averages obtained from the Agromet Group of the Food and Agriculture Organization of the United Nations (FAO). This extensive collection has a fairly detailed level of representation in many typically data-sparse regions, but suffers from a limitation. The FAO database does not provide the period of record used to calculate the long-term averages, although most observations roughly correspond to averages over the 1950s through the 1980s. This data set, therefore, was augmented with 20 591 station climate normals taken from version two of the Global Historical Climate Network (GHCN) (Peterson and Vose, 1997). We compensated for the FAO database's varied coverage in time by supplementing it with averages from a less dense but more temporally consistent information source – the GHCN. The more extensive FAO normals were used to build the preliminary climate surfaces (as described below in Sect. 3). The differences between this surface and GHCN 1980–2009 averages were then estimated and interpolated, and then used to adjust the final monthly surfaces to a 1980–2009 time period.

Monthly means of four satellite products were evaluated as potential background climate surfaces: Tropical Rainfall Measuring Mission (TRMM) 2B31 microwave precipitation estimates (Huffman et al., 2007), the Climate Prediction Center morphing method (CMORPH) microwave-plus-infrared based precipitation estimates (Joyce et al., 2004), monthly mean geostationary infrared (IR) brightness temperatures (Janowiak et al., 2001), and Land Surface Temperature (LST) estimates (Wan, 2008). The TRMM and CMORPH precipitation estimates are based primarily on passive microwave observations from meteorological satellites in asynchronous orbits. The monthly mean infrared brightness temperatures, on the other hand, are derived from a combination of multiple geostationary weather satellites. The LST estimates are derived from multispectral observations from Moderate Resolution Imaging Spectrometers (MODIS) aboard the Terra and Aqua satellites. The LST fields are global, while the CMORPH, TRMM, and IR brightness temperatures span 60° N/S. For each month, for all available years (typically ∼2001–2010), the satellite data were averaged. All four products were convolved to a common 0.05° grid. A fifth predictor was created based on the average of the CMORPH and TRMM precipitation fields.

### 2.2 Topographic and physiographic surfaces

Mean 0.05° elevation, compound topographic index, flow accumulation, aspect, and slope were calculated from global 30 arcseconds GTOPO30 elevation grids following the methodology developed for the HYDRO1K (Verdin and Greenlee, 1996). While the utility of all the topographic fields was explored, only elevation and slope were used in the final analysis because they proved to be the most robust predictors. Latitude and longitude were also included as potential predictor variables.

## 3 Methods – the CHG climatology modeling process

The modeling methodology involved three main steps that were repeated for each month for a set of 56 modeling regions. The extent of the regions was based on (a) station density, (b) homogeneity of predictor response, and (c) availability of the predictor fields. The first step used a series of moving window regressions (MWR) to create an initial prediction of a 0.05° precipitation grid. The second step calculated the at-station residuals from step 1 (station observations minus regression estimates), and then interpolated these values using a modified inverse-distance weighting (IDW) interpolation scheme to create grids of MWR model residuals. The gridded MWR estimates and gridded residuals were combined to create an initial set of climatological surfaces based on the FAO normals. In the third step, these surfaces were then adjusted using the 1980–2009 GHCN station averages.

The differences (ratios) from 1980–2009 GHCN climate normals were computed and used to produce final surfaces corresponding to a 1980–2009 baseline period.

### 3.1 Localized correlation estimates

Our process relies heavily on local regressions between our target variable and background field. We begin by explaining the bivariate standardized case of this process, which corresponds to a localized correlation. At a certain location we can sample a number of points and background variables that fall within a certain distance ($d_{\max}$) and calculate their distance weighted (localized) correlation. The localized correlation process finds a set of $n$ neighboring points (within $d_{\max}$), and estimates their weighted correlation. This study uses a cubic function of the distance ($d$) and a user-defined, regionally variable, maximum distance ($d_{\max}$).

$$w(d) = 0,\ d > d_{\max}$$
$$w(d) = \left[(1 - d/d_{\max})^3\right]^3 d \le d_{\max}. \quad (1)$$

These weights are then used to estimate a localized correlation.

$$r_{x,y} = (n-1)^{-1}\left(\sum_{i=1}^{n} w(d_i)\right)^{-1} \sum_{i=1}^{n} w(d_i) \left[(x_i - \bar{x})\sigma_x^{-1}\right]\left[(y_i - \bar{y})\sigma_y^{-1}\right]. \quad (2)$$

The localized correlation ($rx$, $y$) at some location ($x$, $y$) corresponds with the standardized cross-product of the neighboring points, weighted by their distance. This process can be used to generate correlation maps (Fig. 1). Typically, the direct physical relationship between the station normals and a satellite field, such TRMM/CMORPH precipitation, results in a stronger correlation pattern than that which is produced by an indirect physiographic indicator such as elevation. Figure 1 provides an example of this by contrasting the local correlations between station precipitation, elevation and TRMM/CMORPH precipitation.

### 3.2 Localized moving window regressions

The core of the CHG climatology modeling process is based on a series of local regressions between in situ observations and spatially continuous predictor fields. For each location, a set of neighboring observations is obtained, and a regression model constructed using weighted least squares, with the weight of each observation determined by its distance from the regression centroid (Eq. 1). For each region and month, a grid of center points is defined on a regular 1° grid over land-only locations. Figure 2 shows the modeling regions. At each center-point, station values within the radius ($d_{\max}$) are collected, and a regression model is fit based on weights determined by Eq. (1). The $d_{\max}$ values are defined individually for each model region, varying from 650 km for the larger or data-sparse regions (e.g. Australia, northwest Asia) to 300 km for Central America and the Galapagos.

### 3.3 Model fitting

For each modeling region and month, regression models were determined through a combination of automated regression subset selection and visual inspection of the output. In some cases, visual inspection indicated that a combination of statistically powerful predictors produced obvious artifacts. In these cases, the selection pool was reduced by hand. Based on the boundaries of the interpolation window, certain predictors were omitted (TRMM, CMORPH, IR) because the satellite range did not extend northward or southward enough for these areas.

### 3.4 Interpolation of model residuals

Following the MWR modeling procedure, at-station anomalies (the arithmetic difference between the FAO station normals and the nearest 0.05° regression estimate) are calculated and interpolated using a modified IDW interpolation procedure. For each 0.05° grid cell, the cube of inverse distances is used to produce a weighted average of the surrounding station residuals, $r$. This value is then modified based on a local interpolation radius, $d_{\mathrm{IDW}}$ and the distance to the closest neighboring station ($d_{\min}$).

$$r^* = \left(1 - d_{\min}/d_{\mathrm{IDW}}\right) r. \quad (3)$$

This simple thresholding procedure forces the interpolated residual field to relax towards zero, based on the distance to the closest station. The $d_{\min}$ values were defined by modeling region, and ranged from 350 to 100 km, based on station density. All tiles were allowed to overlap with their neighbors, and locations within these areas of overlap were blended based on weights that were linear functions of the distances from tile edges. This helped to produce smooth transitions from tile to tile.

### 3.5 Rescaling by GHCN ratios

In the final stage, for each month, the regional tiles are composited on a global 0.05° grid and compared with 1980–2009 GHCN climate normals. The ratio of the GHCN and gridded climatology is calculated at each station location. These ratios are capped between 0.3 and 3.0, and interpolated to a 0.05° grid for each month. The values were capped to limit the potential influence of poor station data. A modified IDW procedure, similar to Eq. (3), is used, but instead of relaxing to zero, the interpolation is forced to a ratio of 1 (no change) as the distance to the minimum neighbor reaches $d_{\mathrm{IDW}}$. This ratio grid is multiplied against the sum of the MWR and interpolated residuals, producing the final CHG Climatology field.

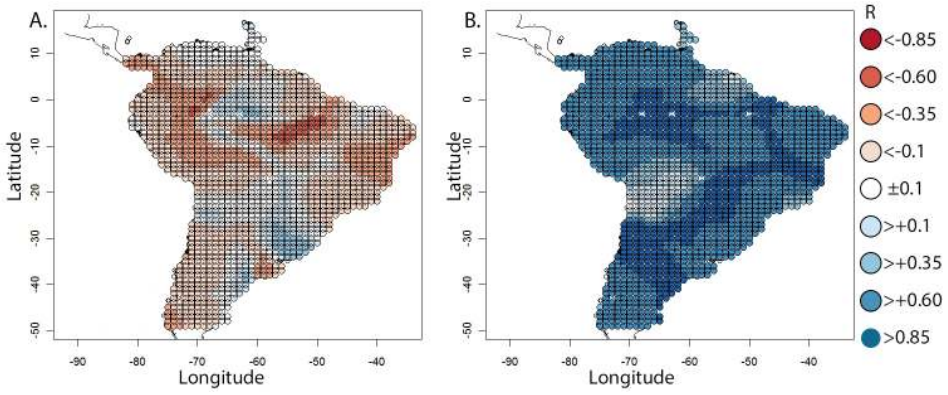


**Figure 1.** Local correlations with July station means. **(a)** Elevation. **(b)** Combined TRMM/CMORPH precipitation.


**Figure 2.** Best predictor, by model region, with station locations.

### 3.6 Cross-validation

Selection bias can inflate the estimated accuracy of statistical estimation procedures, producing artificial skill (Michaelsen, 1987). To limit such inflation, this study uses cross-validation. This technique removes 10 % of the station data, fits the model using the remaining 90 % of the values, and evaluates the accuracy for the withheld locations. This process is repeated ten times, eventually withholding all of the data, to produce a robust estimate of the model accuracy.

### 3.7 Independent validation studies

As additional validation, high quality climatology data sets were obtained for five focus regions: Afghanistan, Colombia, Ethiopia, Mexico, and the Sahel region of western Africa (Senegal, Burkina Faso, Mali, Niger and Chad). Means, spatial $R^2$ values, mean bias errors (MBE [mm]), mean absolute errors (MAE [mm]), percent MBE, and percent MAE statistics were evaluated. These regions (as opposed to the continental United States or Europe) were chosen to represent challenging estimation domains.

## 4 Results

### 4.1 Model fitting results

Figure 2 shows the best predictor for each individual modeling region and the FAO station locations. For regions between 60° N and 60° S, the combined CMORPH and TRMM field tended to be the most useful predictor. The TRMM-only precipitation was selected, however, for southern Africa. Regions beyond 60° N and 60° S could not be modeled with the TRMM or CMORPH means. These regions were generally best fit with LST, slope, or elevations from a digital elevation model (DEM). Figures 3 and 4 show the proportion of modeled cross-validated variance for the MWR and interpolated residuals components for each of the modeling regions. These results are averaged across the 12 months. For most regions, the MWR accounted for over 80 % of the total variance. The interpolated residuals typically accounted for another 10–25 %. Most regions of the globe had average monthly percent errors of between 15 and 25 % (Fig. 5). Figure 6 shows monthly mean CHPclim precipitation fields. As discussed later, these seem generally quite similar, in most places, to the GPCC M V2015, the CRU CL v2.0, and the Worldclim version 1.4 release 3 products. The blending of the overlapping tiles creates generally smooth transitions from tile to tile. These products will be compared more closely later in this paper.

### 4.2 Validation studies

We next present results from our validation studies for Afghanistan, Colombia, Ethiopia, Mexico, and the Sahel (Senegal, Burkina Faso, Mali, Niger, and Chad). In each case, additional high-quality gauge data were obtained from national meteorological agencies (Table 1). These data were screened, and only values not in the FAO or GHCN archive were retained. Table 1 summarizes the number of independent stations and presents the monthly validation statistics, averaged across all 12 months. For each validation sta-

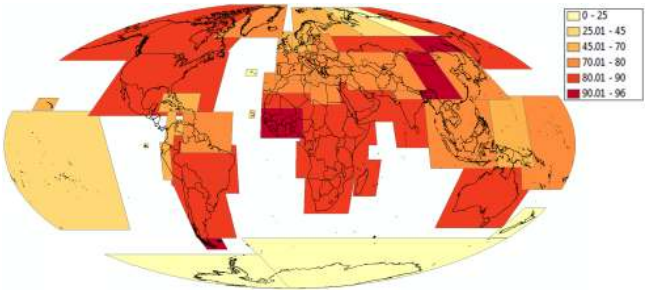


**Figure 3.** Percent of variance explained by cross-validated moving window regression.

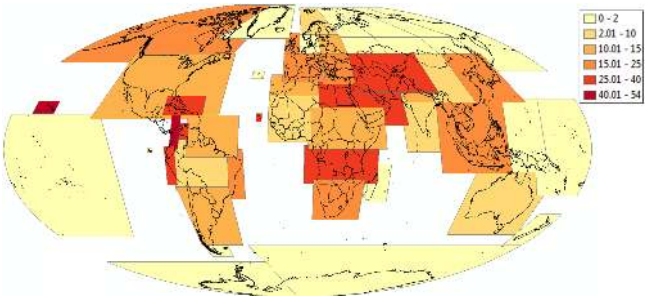


**Figure 4.** Percent of variance explained by cross-validated inverse distance weighting.

tion, the closest CHPclim, CRU, or Worldclim grid cell was extracted. The CHPclim percent biases were substantially smaller in magnitude than the CRU or Worldclim biases, ranging between −2 to +5 %, as compared to −28 to +16 % (CRU) or −16 to 0 % (Worldclim) or −1 to −17 % (GPCC). Note that the GPCC gauge observations were corrected for systematic under-catch errors (Becker et al., 2013; Schneider et al., 2014). While all the climatologies did well in regions with a large number of stations (e.g. Mexico and Colombia), CHPclim's performance was substantially better in data-sparse areas like the Sahel, Ethiopia, and Afghanistan. Averaged across these study regions, the CHPclim/CRU/Worldclim/GPCC data sets had overall mean absolute error (MAE) values of 16, 26, 20 and 20 mm month$^{-1}$, respectively. The average spatial $R^2$ values for the four climatologies were 0.77 (CHPclim), 0.58 (CRU), 0.67 (Worldclim), and 0.51 (GPCC). Overall, the CHPclim compared favorably to the CRU, Worldclim and GPCC data sets.

Plotting the monthly validation statistics provides more temporal information. Figure 7 shows monthly time series of the MAE values for each region and for each set of climatological estimates. In Afghanistan, data were only obtained for the rainy season. The low spatial correlations with the CRU and Worldclim estimates (Table 1) translate into high MAE scores (Fig. 7). In Colombia, the spatial $R^2$ (Table 1) and MAE time series of the CHPclim and Worldclim are similar – both perform well. In Ethiopia, the Worldclim and CRU MAE peak in concert with the seasonal rainfall maxima, while the CHPclim values remain substantially lower. This pattern is recreated for the Sahel and, to a lesser extent, for Mexico. We postulate that the CHPclim performance benefits from the fact that satellite precipitation estimates do a good job of representing heavy convection in these countries during the heart of the precipitation season. Conversely, the thin plate spline fitting procedure, combined with low gauge density in Ethiopia and the Sahel, may make it difficult to statistically represent precipitation gradients in these countries, degrading the performance of the CRU and Worldclim climatologies. Thin plate splines fit polynomial surfaces through point data, creating a generalized surface fit to latitude, longitude, and elevation. The suitability of this fitting process may be problematic when the density of the gauge data is very low. Later in our paper we compare different climate products over Ethiopia.

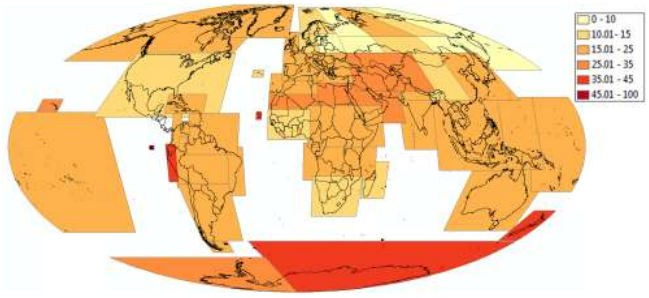


**Figure 5.** Percent standard error explained by cross-validation.

Figure 8 shows similar time series for the spatial $R^2$ statistics. In Afghanistan, Ethiopia, and the Sahel, the CHPclim appears substantially better at representing spatial gradient information. In Colombia and Mexico, CHPclim and Worldclim performance is similar. This may relate to the number of climate normals available in each region (cf. Fig. 2). In Colombia and Mexico, relatively dense gauge networks result in similar Worldclim and CHPclim performance. In regions with fewer stations, the correlation structure of the satellite precipitation data (Fig. 1) probably helps boost the relative performance of CHPclim.

### 4.3 Product comparisons

Here we briefly examine differences between quasi-global total annual precipitation from the CHPclim, GPCC M V2015, CRU CL 2.0 and Worldclim version 1.4 release 3 (Fig. 9) and their global and continental averages (Table 2). The left hand panels show differences between the CHPclim and the three other products. The largest differences appear over the north half of South America, where annual precipitation is very high (Fig. 6). These differences may arise from the local influence of the satellite rainfall fields, which are well correlated with station observations in this region (Fig. 1). Note that the GPCC, CRU, and Worldclim also vary substantially amongst themselves in this area. In Eu-

**Table 1.** CHPclim validation results.

| | Region | N-stns | Station Mean | Climatology Mean | MBE | MAE | Pct MBE | Pct MAE | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|
| CHPclim | Colombia | 194 | 168 | 159 | 8 | 30 | 5 | 18 | 0.84 |
| | Afghanistan | 22 | 35 | 34 | 1 | 9 | 3 | 25 | 0.53 |
| | Ethiopia | 76 | 97 | 94 | 3 | 10 | 4 | 10 | 0.91 |
| | Sahel | 28 | 55 | 53 | 0 | 6 | 0 | 10 | 0.93 |
| | Mexico | 1814 | 77 | 78 | −1 | 23 | −2 | 30 | 0.65 |
| CRU | Colombia | 194 | 168 | 174 | −6 | 47 | −4 | 28 | 0.59 |
| | Afghanistan | 22 | 35 | 45 | −10 | 20 | −28 | 57 | 0.18 |
| | Ethiopia | 76 | 97 | 101 | −4 | 23 | −4 | 24 | 0.68 |
| | Sahel | 91 | 55 | 65 | −11 | 14 | 16 | 21 | 0.87 |
| | Mexico | 1814 | 77 | 75 | 2 | 24 | 2 | 31 | 0.60 |
| Worldclim | Colombia | 194 | 168 | 178 | −11 | 31 | −6 | 19 | 0.82 |
| | Afghanistan | 22 | 35 | 41 | −6 | 18 | −17 | 52 | 0.18 |
| | Ethiopia | 76 | 97 | 97 | 0 | 20 | 0 | 21 | 0.72 |
| | Sahel | 28 | 55 | 65 | −10 | 14 | −16 | 22 | 0.86 |
| | Mexico | 1814 | 77 | 79 | −2 | 18 | −2 | 23 | 0.78 |
| GPCC | Colombia | 194 | 168 | 185 | −17 | 51 | −10 | 31 | 0.75 |
| | Afghanistan | 22 | 35 | 43 | −8 | 15 | −23 | 43 | 0.29 |
| | Ethiopia | 76 | 97 | 99 | −3 | 22 | −2 | 22 | 0.84 |
| | Sahel | 28 | 55 | 70 | −15 | 8 | −27 | 15 | 0.78 |
| | Mexico | 1814 | 77 | 78 | −1 | 21 | −1 | 28 | 0.84 |

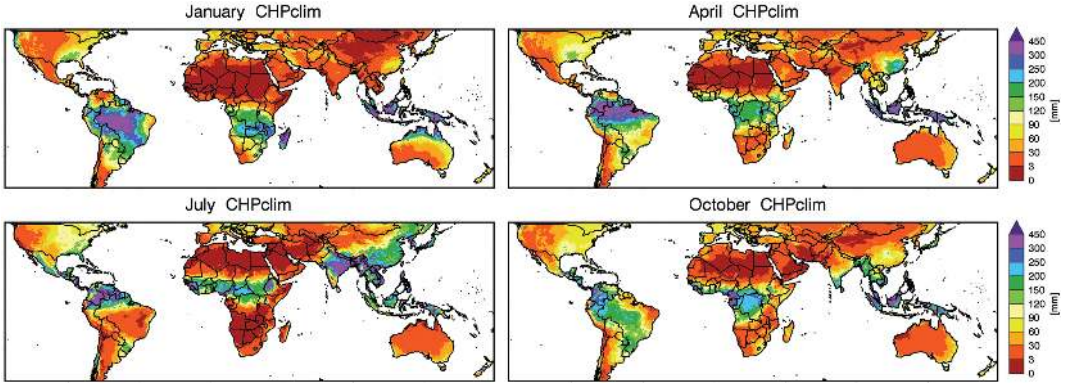


**Figure 6.** CHPclim monthly means for January, April, July and October. While CHPclim is global, we show 50° S–50° N images to facilitate visualization.

rope, northern Asia, North America, and Australia, the differences are fairly limited, most likely due to the high station density in these regions. There are large differences near the Himalayas. The CHPclim appears to be producing more precipitation across the Himalayan plateau and less precipitation on the south-facing mountain slopes. More research will be required to evaluate if this is appropriate or not. CHPclim also appears to be substantially drier over some parts of Africa. A recent study in Mozambique (Toté et al., 2015) of the Climate Hazards group Infrared Precipitation with Stations (CHIRPS, Funk et al., 2014b), which is based on the CHPclim, found low bias over that country. Stations in Africa tend to be biased towards wet locations, and the use of satellite fields as guides to interpolation may help limit this bias. We explore this idea in more detail in the next section, which focuses on an Ethiopia test case.

Before proceeding to that analysis, we note that the global (excluding Antarctica) and continental averages from our four products are in quite close agreement (Table 2), even in Africa. The two outlier's appear to be the GPCC M V2015 averages for Asia (688 mm) and for the globe (880 mm). The global GPCC M V2015 value of 880 mm is close to the 850 mm figure reported in Schneider et al. (2014). The discrepancy between the GPCC results and the other prod-

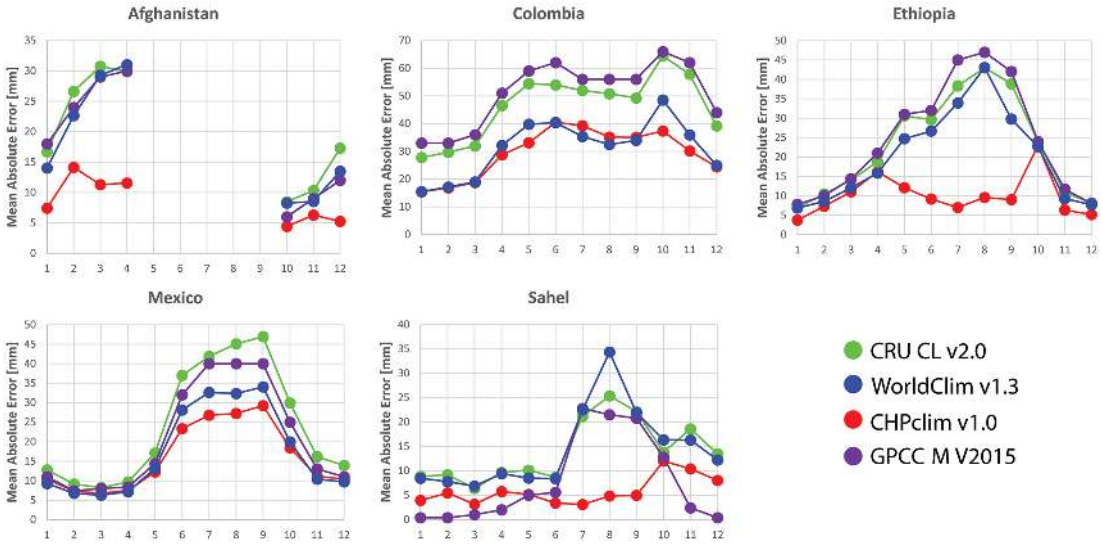


**Figure 7.** Mean absolute error time series [mm month$^{-1}$].

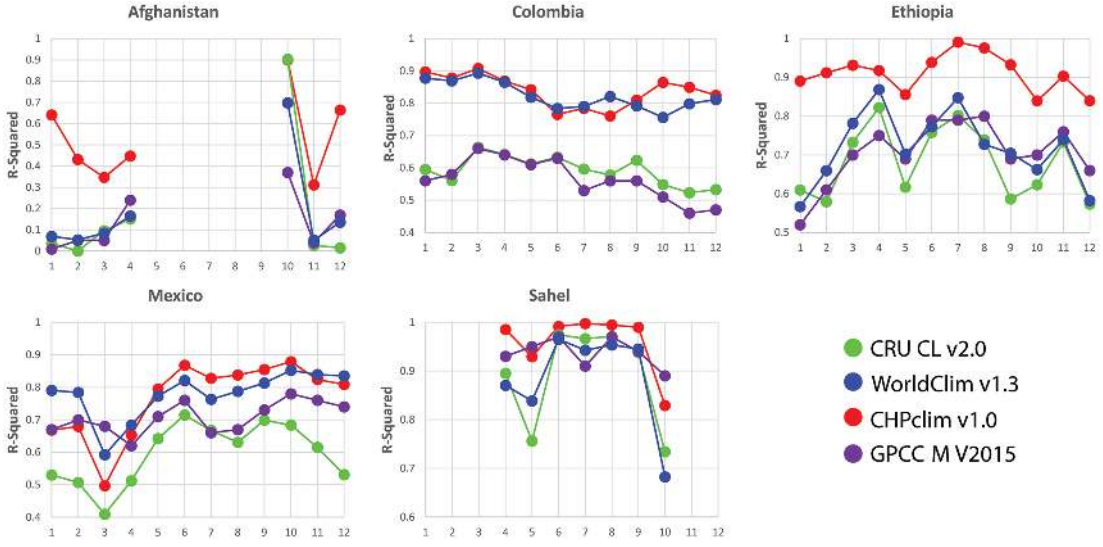


**Figure 8.** Spatial $R^2$ time series.

ucts is likely due to the way they corrected for systematic under-catch by the rainfall gauges. The CHPclim does not incorporate this correction which increases precipitation observations based on estimated under catch values. The global (excluding Antarctica) total annual rainfall values, expressed in "units" of global precipitation of 103 km$^3$ yr$^{-1}$ as in Trenberth et al. (2007) agree quite well with that study's reported values (110 units CRU; 112 units GPCP). We found the CHPclim precipitation resulted 120 units. This difference may relate to CHPclim's interpolation procedure in northern South America and the Maritime Continent where the CHPclim is wetter (Fig. 9, Table 2), perhaps because of guidance provided by satellite observations (Fig. 1).

### 4.4 An Ethiopian validation study

In February of 2015 one of the co-authors led a rainfall gridding workshop in Addis Ababa, in collaboration with lead scientists from the Ethiopian National Meteorological Agency (NMA). This workshop used the GeoCLIM tool to blend CHIRPS satellite rainfall estimates with 208 quality-controlled gauge observations (Figs. 10 and 11, top left) to generate monthly 1981–2014 grids of precipitation. In this section we compare the 1981–2014 average of these blended CHIRPS/NMA station data to the CHPclim, GPCC, CRU and Worldclim data sets. We acknowledge that since the CHPclim is used in the CHIRPS as a background climatology the NMA and CHPclim data sets are not completely indepen-

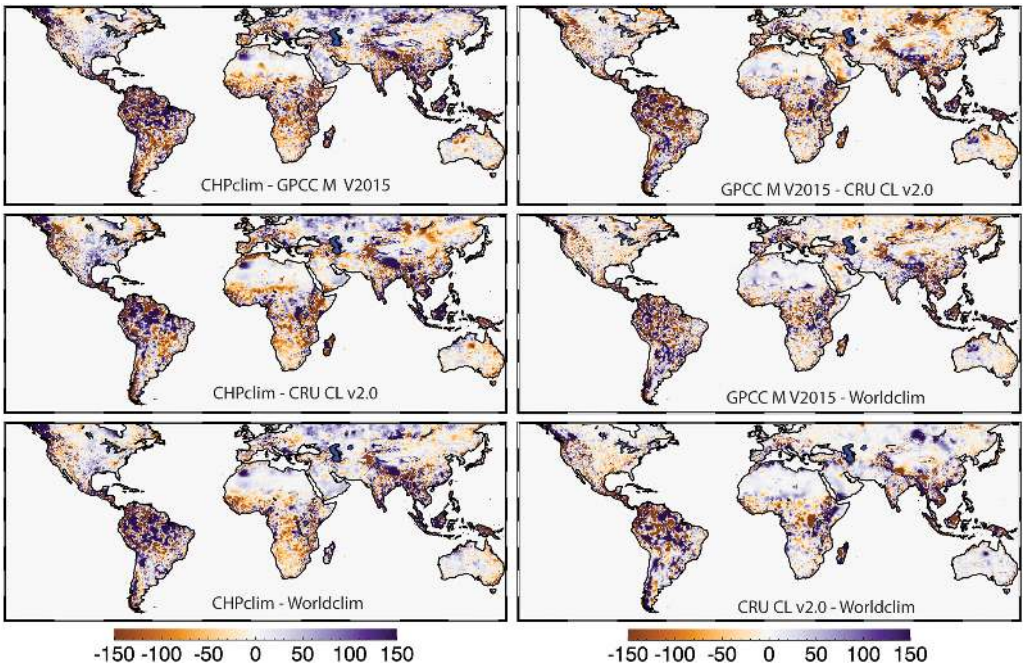


**Figure 9.** Differences in annual total precipitation for CHPclim, the 0.25° GPCC M V2015 climatology, the 0.17° CRU CL v2.0, and the 0.042° version 1.4 release 3 Worldclim climatology.

**Table 2.** Comparison on annual total precipitation [mm] for different regions.

| | Globe, excl Antartica | Europe | Asia | Australia | Maritime Continent | North America | South America | Africa |
|---|---|---|---|---|---|---|---|---|
| CHPclim | 810 | 707 | 625 | 485 | 2829 | 702 | 1594 | 613 |
| GPCC CL2.0 | 880 | 710 | 688 | 576 | 2702 | 732 | 1563 | 631 |
| CRU | 804 | 707 | 607 | 496 | 2756 | 695 | 1580 | 624 |
| World Clim | 796 | 693 | 596 | 487 | 2750 | 682 | 1556 | 611 |

dent. Nonetheless, the 35 years of 208 NMA rain gauge observations have not been included in the CHPclim, and hence provide a valuable validation data set, especially within the areas with good gauge density.

Figure 10 shows the mean 1981–2014 annual rainfall totals based on the gridded NMA data, and similar maps from the CHPclim, GPCC M V2015, CRU CL 2.0, and Worldclim version 1.4 release 3. Also shown are elevation, annual totals of CMORPH/TRMM precipitation and annual average MODIS LST. These fields were used in the CHPclim modeling process. Annual mean MODIS Normalized Difference Vegetation Index (NDVI) values are also shown as an independent proxy for moisture availability. All the precipitation products and the NDVI agree on the broad patterns of spatial rainfall variability, which are extreme. The wettest regions receive more than 2 m of rainfall each year while the driest receive less than 200 mm. The CMORPH/TRMM satellite observations seem to capture these dry areas well – with no ground data at all, i.e. the brown areas in the CMORPH/TRMM agree quite closely with the NMA validation data. The CMORPH/TRMM fields delineate dry area effectively. Within wet areas, the discriminatory power of the satellite observations seems to diminish, indicating (incorrectly) that northwest Ethiopia is as wet as southwest Ethiopia. The similarity between the completely independent NDVI and NMA/CHPclim fields is quite compelling. Many subtle features, such as the humid highlands in north-central, east-central, and southeastern Ethiopia appear well demarcated by these precipitation fields. These seem fairly well captured by the Worldclim and CRU as well.

Note that there are important differences between, on one hand, the elevation and similar LST field and, on the other, the NMA/CHPclim precipitation and NDVI mean fields. While there are certainly some important correspondences, there are also critical differences, such as in north-central Ethiopia which is high and cool, but dry. Conversely, northwest Ethiopia is relatively wet, but relatively low. There are times and locations when elevation is a poor indicator of mean precipitation.

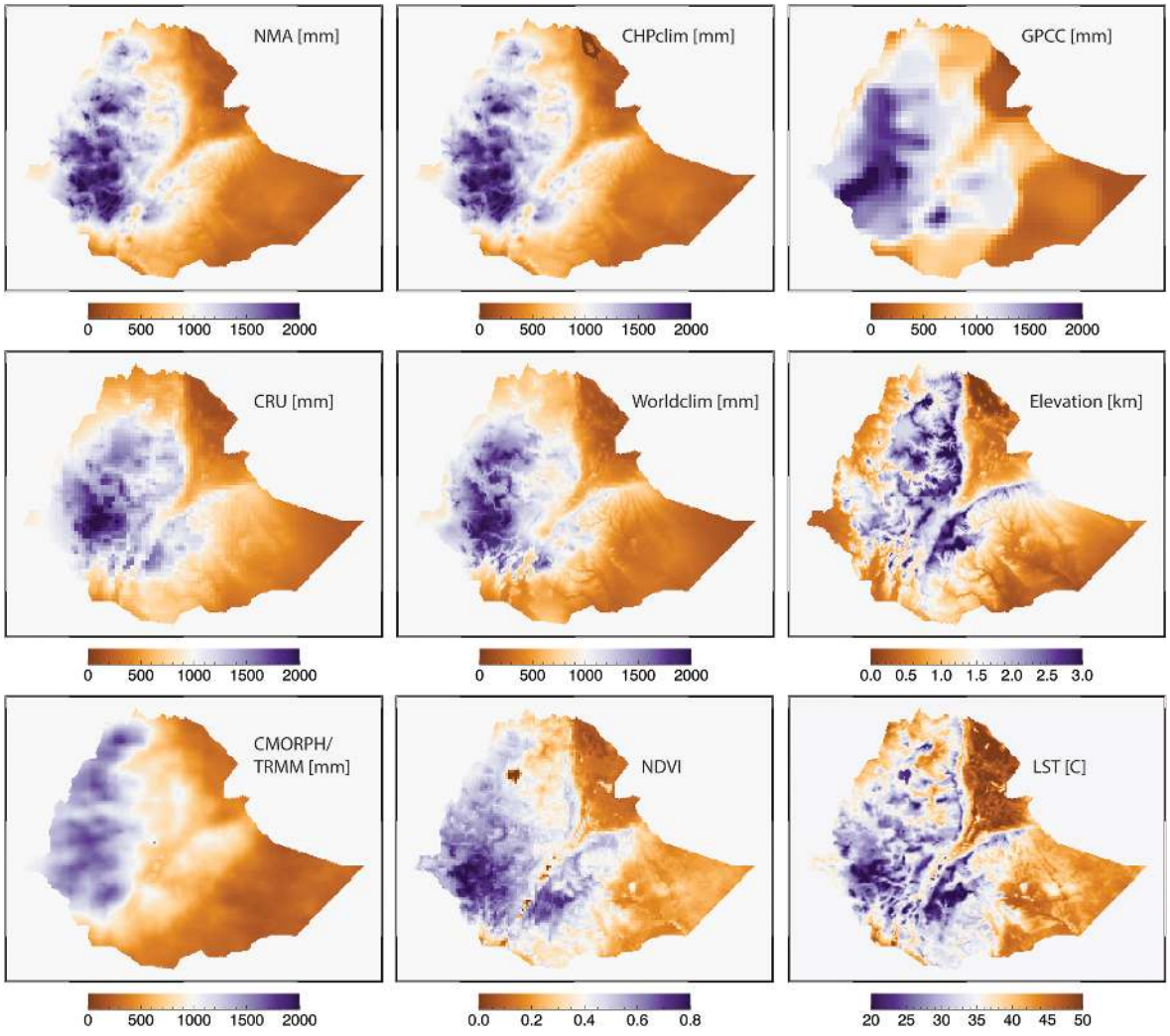


**Figure 10.** Total annual rainfall, elevation, NDVI and LST for Ethiopia. Rainfall totals are from the Ethiopian National Meteorological Agency (NMA), CHPclim, the GPCC M V2015 climatology, the CRU CL v2.0, the version 1.4 release 3 Worldclim climatology, and the blended CMORPH/TRMM data used in the CHPclim modeling process.

Figure 11 shows the differences from NMA validation data. Also shown, to support analysis, are the NMA mean precipitation and elevation data. Purple lines have been drawn showing transects plotted in Fig. 12. The CHPclim follows the NMA climatology closely. The GPCC, CRU, and Worldclim all exhibit substantial ($> |300\,\text{mm}|$) deviations, with the Worldclim performing substantially better than the GPCC and CRU. This helps to confirm the visual impression from Fig. 10 that the Worldclim data follows the NMA data quite closely. The GPCC, CRU and Worldclim all underestimate precipitation in the blue regions in the northwest and southwest of these maps, which are relatively low areas. The CMORPH/TRMM finds rainfall in these areas (Fig. 10), and the CHPclim MBE in these areas is quite modest (Fig. 11). Conversely, dark brown areas in the bottom panels of Fig. 11 denote areas where rainfall is substantially overestimated in the GPCC, CRU, and Worldclim. This appears to be of gravest concern in the center and center-east of the country, which has high elevations and extremely steep rainfall gradients. While not perfect, the CMORPH/TRMM (Fig. 10) seems to capture these gradients with reasonable fidelity, and building on these gradients produces a CHPclim with low bias in these areas.

We explore this topic more fully in Fig. 12, which shows transects of our data sets at 10 and 7° N. We have multiplied the NDVI data by 1500 and divided the elevation data by 5 to facilitate visualization. Begin by noting in the top two panels the similarities between the mean NMA data, the CMORPH/TRMM, and the NDVI. This reinforces the utility of the TRMM/CMORPH, and that the NMA fields are an effective representation of the "true" climatology. The CRU and Worldclim seem to follow the NMA transect quite well, with some substantial deviations shown in the bottom panels. Some of these errors appear to coincide with areas

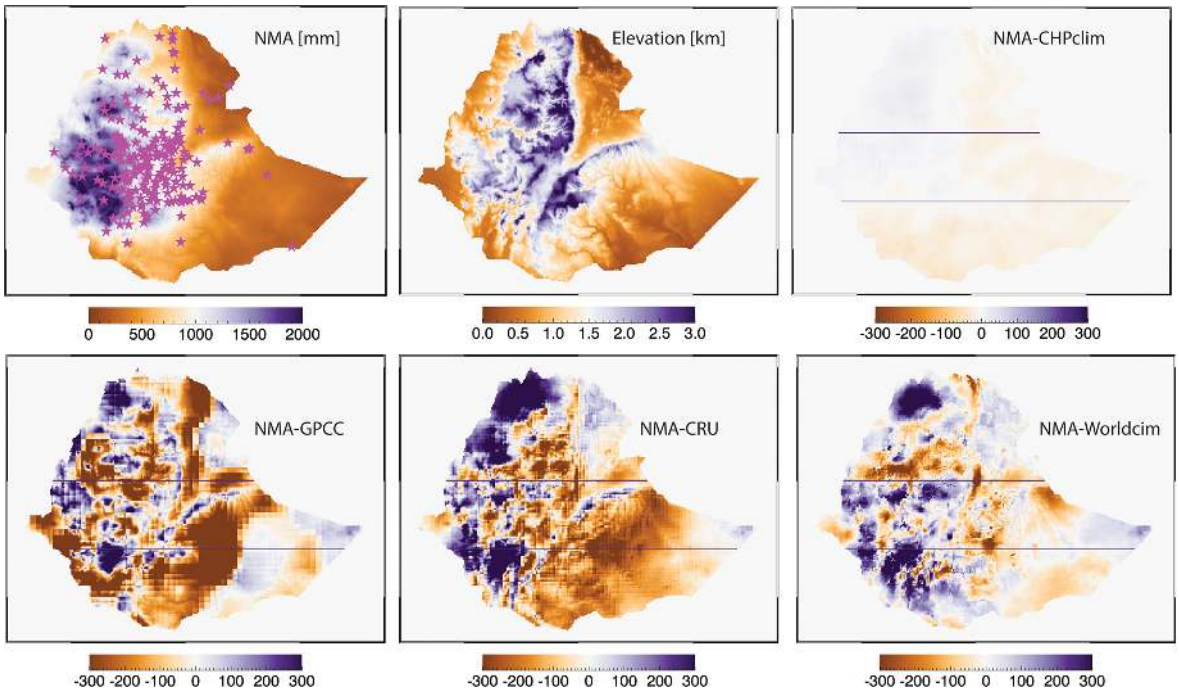


**Figure 11.** Total annual NMA rainfall, elevation and MBE maps based on the NMA minus CHPclim, the NMA minus GPCC, the NMA minus CRU and the NMA minus Worldclim.

having extreme elevation changes, such as 36.5° E, 37.5° E and 40° E at 10° N. At 37° E, 7° N, the CRU, GPCC and Worldclim substantially underestimate rainfall. The CHPclim, assisted by the CMORPH/TRMM, which is quite wet in this region, captures the rainfall well. In the eastern part of the country, where we find the largest percent discrepancies, we find overestimates at 41° E, 10° N and 41.5° E, 7° N. Estimates of rainfall gradients in these poorly instrumented regions are very difficult based on just station data. The CMORPH/TRMM, however, seems to capture these gradients well, and the CHPclim builds on this local gradient information.

## 5 Discussion

This paper has introduced a new climatology modeling process developed by the CHG to support international drought early warning and hydrologic modeling. While this process has been applied to African rainfall and temperatures (Funk et al., 2012; Knapp et al., 2011), we report here for the first time global results, and evaluate the relative accuracy of the CHPclim v1.0 (http://dx.doi.org/10.15780/G2159X). The CHPclim is one part of the CHG's overall strategy to provide improved drought early warning information (Fig. 13). Working closely with early warning scientists from the US Geological Survey's Center for Earth Resources Observation and Science (EROS), the CHG develops improved earth science tools to support food security and disaster relief for the US Agency for International Development's Famine Early Warning System Network (FEWS NET).

These activities fall into two main categories: analytic studies focused on understanding the relationship between local climate variations and large-scale climate drivers (Funk et al., 2008, 2014a; Hoell and Funk, 2013a, b; Liebmann et al., 2014), and the development of integrated data sets and tools supporting agro-climatic monitoring in the developing world. While early precipitation efforts focused on the use of a model (Funk and Michaelsen, 2004) to represent orographic precipitation (Funk et al., 2003), the potential issues produced by spurious model-based trends led us to focus on the use of high-resolution climatologies as proxies for orographic precipitation enhancement (Funk et al., 2007). The global 0.05° CHPclim presented here is the global expansion of that work.

CHPclim provides the first component of our global precipitation monitoring system, which is built on the Climate Hazard Group Infrared Precipitation with Stations (CHIRPS, Fig. 13). The monthly CHPclim fields, described and evaluated here, have been temporally disaggregated to pentadal (5-day) means. These pentadal mean fields are then combined with 1981–near present 0.05° 60° S–60° N IR brightness (Janowiak et al., 2001; Knapp et al., 2011) precipitation estimates to produce the Climate Hazards Group Infrared Precipitation fields (CHIRP). A modified inverse distance weighting procedure is then used to blend these fields with global precipitation gauge station data to produce the CHIRPS (Funk et al., 2014b). These data, which benefit

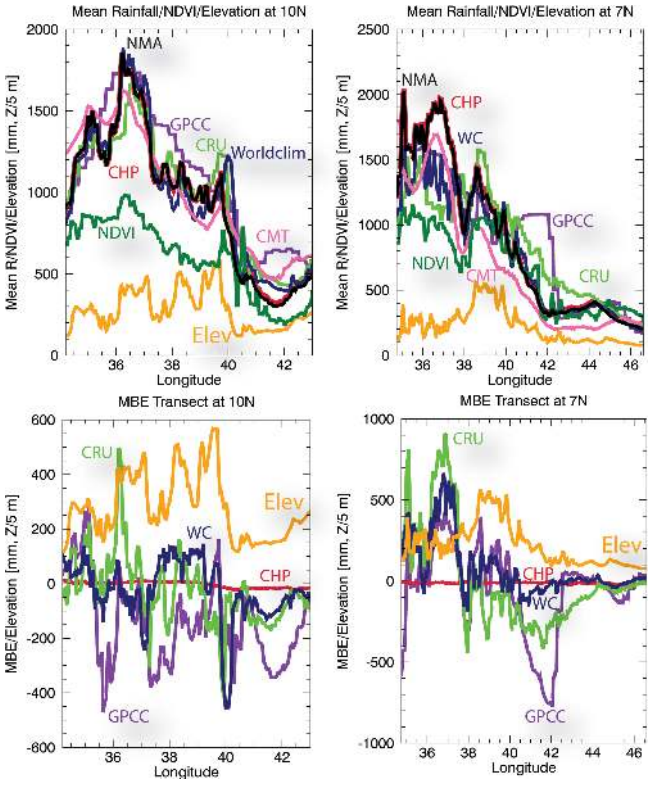


**Figure 12.** The top panels show transects of total annual rainfall at 7 and 10° N. Also shown are transects of elevation in meters divided by 5 and annual mean NDVI, multiplied by 1500. The bottom panels show MBE transects based on CHPclim, GPCC, CRU and Worldclim minus the NMA data. These bottom panels also show elevation in meters divided by 5.

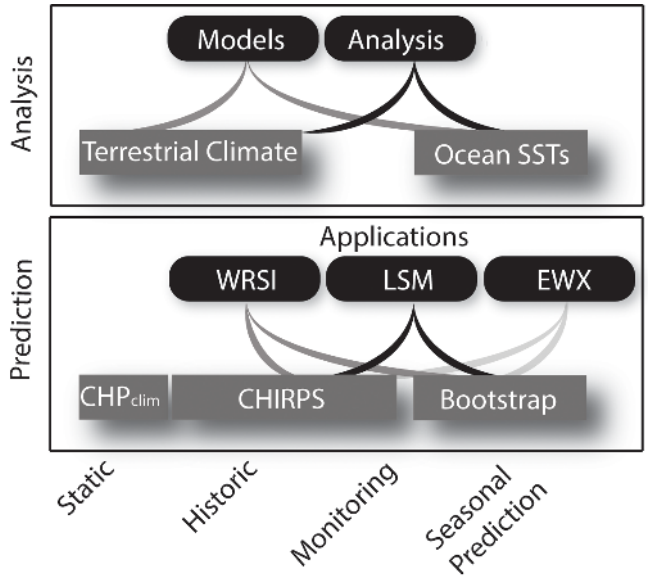


**Figure 13.** Schema of CHG analysis and prediction activities.

from the high-resolution CHPclim climatology, can be used to drive a gridded crop Water Requirement Satisfaction Index model (WRSI) (Verdin and Klaver, 2002), force a special Land Data Assimilation System developed for the US Agency for International Development's FEWS NET (the FLDAS), or populate interactive early warning displays like the Early Warning eXplorer (EWX, http://earlywarning.usgs.gov:8080/EWX/index.html). Improved background climatologies can enhance the efficacy of crop models, increasing their drought monitoring capacity.

Ongoing efforts are being directed towards linking seasonal forecast information with historical CHIRPS archives (Shukla et al., 2014a, b). In East Africa, for example, daily 0.05° rainfall values are used to force a hydrologic model. These results can then be combined with precipitation forecasts that translate large-scale climate conditions into region-specific predictions of CHIRPS rainfall. These rainfall forecasts can be used to drive crop and hydrologic models. In this way, for some high-priority regions like East Africa, CHG scientists hope to combine the climatological constraints described by high-resolution climatologies like the CHPclim, historic precipitation distributions (Husak et al., 2013), the latent information contained in the land surface state as represented by land surface models (Shukla et al., 2014b, 2013), and the foreshadowing of future weather provided by climate forecasts (Funk et al., 2014a; Shukla et al., 2014a, b). The CHPclim, described here, has been designed to provide a good foundation for this, and similar, hydrologic modeling and monitoring systems. The CHPclim data and CHIRPS data sets are available at http://dx.doi.org/10.15780/G2159X and http://dx.doi.org/10.15780/G2RP4Q and http://chg.geog.ucsb.edu.

**Acknowledgements.** This research was supported by US Geological Survey (USGS) cooperative agreement #G09AC000001 "Monitoring and Forecasting Climate, Water and Land Use for Food Production in the Developing World" with funding from the US Agency for International Development Office of Food for Peace award #AID-FFP-P-10-00002 for "Famine Early Warning Systems Network Support", the NASA SERVIR Applied Sciences Team and NOAA Award NA11OAR4310151, "A Global Standardized Precipitation Index supporting the US Drought Portal and the Famine Early Warning System Network".

Edited by: G. König-Langlo

## References


Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., and Ziese, M.: A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present, Earth Syst. Sci. Data, 5, 71–99, doi:10.5194/essd-5-71-2013, 2013.

Daly, C., Neilson, R. P., and Phillips, D. L.: A statistical-topographic model for mapping climate logical precipitation over mountainous terrain, J. Appl. Meteorol., 33, 140–158, 1994.

Funk, C. and Michaelsen, J.: A simplified diagnostic model of orographic rainfall for enhancing satellite-based rainfall estimates in data-poor regions, J. Appl. Meteorol., 43, 1366–1378, 2004.

Funk, C., Michaelsen, J., Verdin, J., Artan, G., Husak, G., Senay, G., Gadain, H., and Magadazire, T.: The collaborative historical African rainfall model: description and evaluation, Int. J. Climatol., 23, 47–66, 2003.

Funk, C., Husak, G., Michaelsen, J., Love, T., and Pedreros, D.: Third Generation Rainfall Climatologies: Satellite Rainfall and Topography Provide a Basis for Smart Interpolation, in: Proceedings of the JRC – FAO Workshop, 27–29 March, Nairobi, Kenya, 283–296, 2007.

Funk, C., Dettinger, M. D., Michaelsen, J. C., Verdin, J. P., Brown, M. E., Barlow, M., and Hoell, A.: Warming of the Indian Ocean threatens eastern and southern African food security but could be mitigated by agricultural development, P. Natl. Acad. Sci. USA, 105, 11081–11086, 2008.

Funk, C., Michaelsen, J., and Marshall, M.: Mapping recent decadal climate variations in precipitation and temperature across eastern Africa and the Sahel, in: Remote Sensing of Drought: Innovative Monitoring Approaches, edited by: Wardlow, B., Anderson, M., and Verdin, J., CRC Press, Boca Raton, Florida, USA, 25 p., 332–356, 2012.

Funk, C., Hoell, A., Shukla, S., Bladé, I., Liebmann, B., Roberts, J. B., Robertson, F. R., and Husak, G.: Predicting East African spring droughts using Pacific and Indian Ocean sea surface temperature indices, Hydrol. Earth Syst. Sci., 18, 4965–4978, doi:10.5194/hess-18-4965-2014, 2014a.

Funk, C., Peterson P., Landsfeld, M., Pedreros, D., Verdin, J., Rowland, J., Romero, B., Husak, G., Michaelsen, J., and Verdin A.: A Quasi-global Precipitation Time Series for Drought Monitoring, US Geological Survey, Reston VA USA, 2014b.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A.: Very high resolution interpolated climate surfaces for global land areas, Int. J. Climatol., 25, 1965–1978, 2005.

Hoell, A. and Funk, C.: The ENSO-relatedWest Pacific sea surface temperature gradient, J. Climate, 26, 9545–9562, 2013a.

Hoell, A. and Funk, C.: Indo–Pacific sea surface temperature influences on failed consecutive rainy seasons over eastern Africa, Clim. Dynam., 43, 1–16, doi:10.1007/s00382-013-1991-6, 2013b.

Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G., Nelkin, E. J., Bowman, K. P., Hong, Y., Stocker, E. F., and Wolff, D. B.: The TRMM Multisatellite Precipitation Analysis (TMPA): quasiglobal, multiyear, combined-sensor precipitation estimates at fine scales, J. Hydrometeorol., 8, 38–55, 2007.

Husak, G. J., Funk, C. C., Michaelsen, J., Magadzire, T., and Goldsberry, K. P.: Developing seasonal rainfall scenarios for food security early warning, Theor. Appl. Climatol., 114, 291–302, 2013.

Hutchinson, M. F.: Interpolating mean rainfall using thin plate smoothing splines, Int. J. Geogr. Inf. Syst., 9, 385–403, doi:10.1080/02693799508902045, 1995.

Janowiak, J. E., Joyce, R. J., and Yarosh, Y.: A real-time global half-hourly pixel-resolution infrared dataset and its applications, B. Am. Meteorol. Soc., 82, 205–217, 2001.

Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P.: 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, 2004.

Knapp, K. R., Ansari, S., Bain, C., L., Bourassa, M. A., Dickinson, M. J., Funk, C., Helms, C. N., Hennon, C. C., Holmes, C., Huffman, G. J., Kossin, J. P., Lee, H.-T., A. Loew, A., and Magnusdottir, G.: Globally gridded satellite (GriSat) observations for climate studies, B. Am. Meteorol. Soc., 92, 893–907, 2011.

Krige, D. G.: A Statistical Approach to Some Mine Valuations and Allied Problems at the Witwatersrand, University of Witwatersrand, Johannesburg, South Africa, 1951.

Liebmann, B., Hoerling, M. P., Funk, C., Bladé, I., Dole, R. M., Allured, D., Quan, X., Pegion, P., and Eischeid, J. K.: Understanding recent eastern Horn of Africa rainfall variability and change, J. Climate, 27, 8630–8645, 2014.

Matheron, G.: Principles of geostatistics, Economic Geology, 58, 1246–1266, 1963.

Michaelsen, J.: Cross-validation in statistical climate forecast models, J. Clim. Appl. Meteorol., 26, 1589–1600, 1987.

New, M., Hulme, M., and Jones, P.: Representing twentieth-century space–time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology, J. Climate, 12, 829–856, 1999.

New, M., Lister, D., Hulme, M., and Makin, I.: A high-resolution data set of surface climate over global land areas, Climate Res., 21, 1–15, doi:10.3354/cr021001, 2002.

Peterson, T. C. and Vose, R. S.: An overview of the Global Historical Climatology Network temperature database, B. Am. Meteorol. Soc., 78, 2837–2849, 1997.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., and Rudolf, B.: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. Theor. and Appl. Clim., 115, 15–40, 2014.

Shepard, D.: A Two-Dimensional Interpolation Function for Irregularly-Spaced Data, in: Proc. 23rd National Conference Association for Computing Machinery, 517–524, 1968.

Shukla, S., Sheffield, J., Wood, E. F., and Lettenmaier, D. P.: On the sources of global land surface hydrologic predictability, Hydrol. Earth Syst. Sci., 17, 2781–2796, doi:10.5194/hess-17-2781-2013, 2013.

Shukla, S., Funk, C., and Hoell, A.: Using constructed analogs to improve the skill of March-April-May precipitation forecasts in equatorial East Africa, Environ. Res. Lett., 9, 094009, doi:10.1088/1748-9326/9/9/094009, 2014a.

Shukla, S., McNally, A., Husak, G., and Funk, C.: A seasonal agricultural drought forecast system for food-insecure regions of East Africa, Hydrol. Earth Syst. Sci., 18, 3907–3921, doi:10.5194/hess-18-3907-2014, 2014b.

Toté, C., Patricio, D., Boogaard, H., van der Wijngaart, R., Tarnavsky, E. and Funk, C.: Evaluation of Satellite Rainfall Estimates for Drought and Flood Monitoring in Mozambique, Remote Sensing, 7, 1758–1776, 2015.

Trenberth, K. E., Smith, L., Qian, T., Dai, A., and Fasullo, J.: Estimates of the global water budget and its annual cycle using observational and model data, J. Hydrometeorol., 8, 758–769, 2007.

Verdin, J. and Klaver, R.: Grid-cell-based crop water accounting for the famine early warning system, Hydrol. Process., 16, 1617–1630, 2002.

Verdin, K. L. and Greenlee, S. K.: Development of Continental Scale Digital Elevation Models and Extraction of Hydrographic Features, in: Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, New Mexico, 1996.

Wan, Z.: New refinements and validation of the MODIS Land-Surface Temperature/Emissivity products, Remote Sens. Environ., 112, 59–74, 2008.

Willmott, C. J. and Robseon, S. M.: Climatologically Aided Interpolation (CAI) of terrestrial air temperature, Int. J. Climatol., 15, 221–229, 1995.