Documentation for the Water Security Indicator Model - Global Land Data Assimilation System (WSIM-GLDAS) Monthly Grids, Version 1

January 2022

Cynthia Crowley, Dan Baston, Joshua Brinks[1], and Center for International Earth Science Information Network (CIESIN), Columbia University

[1] ISciences, L.L.C., 126 College Street, Suite 330, Burlington, VT 05401-8383

Abstract

This document outlines the basic methodology and data sets used to construct the Water Security Indicator Model using Global Land Data Assimilation System (WSIM-GLDAS) data product, along with use cases, limitations, and use constraints. The WSIM-GLDAS data set characterizes surpluses and deficits of fresh water, and parameters determining them, across the global terrestrial surface. Driven by outputs of NASA's Global Land Data Assimilation V2.0 monthly Noah Land Surface Model (Beaudoing, Rodell, & NASA/GSFC/HSL, 2015), WSIM generates anomaly values of temperature, precipitation, soil moisture, potential minus actual evapotranspiration (PETME), runoff, total blue water (flow-accumulated runoff), composite index of water surpluses, and composite index of water deficits in terms of their return periods with respect to a fitted generalized extreme value (GEV) probability distribution function over a historical baseline period of January 1950 to December 2009 at a global spatial resolution of 0.25 degrees over 1-month, 3-month, 6-month, and 12-month periods of integration. The data cover January 1948 through December 2014, inclusive, with a spatial extent of -180.0, -60.0, 180.0, 90.0 (minimum longitude, minimum latitude, maximum longitude, maximum latitude, respectively).

Data set citation: ISciences, and Center for International Earth Science Information Network (CIESIN), Columbia University. 2022. Water Security Indicator Model - Global Land Data Assimilation System (WSIM-GLDAS) Monthly Grids, Version 1. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/z1fn-kf73. Accessed DAY MONTH YEAR.

Suggested citation for this document: ISciences, and Center for International Earth Science Information Network (CIESIN), Columbia University. 2022. Documentation for the Water Security Indicator Model - Global Land Data Assimilation System (WSIM-GLDAS) Monthly Grids, Version 1. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/x7fj-jj41. Accessed DAY MONTH YEAR.

We appreciate feedback regarding this data set, such as suggestions, discovery of errors, difficulties in using the data, and format preferences. Please contact:

NASA Socioeconomic Data and Applications Center (SEDAC) Center for International Earth Science Information Network (CIESIN) Columbia University Phone: 1 (845) 365-8920
Email: ciesin.info@ciesin.columbia.edu

Contents

I. Introduction

Water Security Indicator Model - Global Land Data Assimilation System (WSIM-GLDAS) Monthly Grids, Version 1 is a global gridded data set produced using the land surface model outputs of NASA's Global Land Data Assimilation System (GLDAS) Noah V2.0 monthly model (Beaudoing, Rodell, & NASA/GSFC/HSL, 2015) (Rodell, et al., 2004) as inputs to the Water Security Indicator Model (WSIM) version 2.0.1322[1]. WSIM generates time series of anomalies of fresh water parameters at a 0.25-degree resolution at monthly time steps extending across the global terrestrial surface, excluding Antarctica.

[1] WSIM version 2.0.1322 is associated with Git commit 131e3f6a8864a443645a15288f1aa34019159d6b.

Return periods

The central premise of WSIM is that societies structure their activities based in part on expected climatic provisions of fresh water, and can maintain these activities within a certain degree of variation in the amount, frequency, and timing of these provisions (Isciences, L.L.C., 2019). Variation exceeding that experienced in the recent past, however, may force societies to react in unusual ways. While the relationship between climatic stresses and societal functioning is complex, such reactions could induce internal or transnational water disputes, agricultural shortfalls, electricity shortages, population displacements, infectious disease outbreaks, or political instability.

This premise motivates WSIM to quantify the rarity (anomaly) of an observed value of a fresh water parameter, e.g., monthly precipitation, by expressing it in terms of its return period, an anomaly measure that expresses the expected time interval (in years) between occurrences of events of equal or greater severity.

Use of return periods enables expressing the severity of a water deficit or surplus with respect to the historical hydrological record particular to the locale. For instance, consider two data points: one showing 1100 mm of precipitation over the 12-month period of July to June in Burlington, Vermont, and one showing 1100 mm of precipitation over the same period in Jakarta, Indonesia. Absent additional information about Burlington, Vermont and Jakarta, Indonesia, these figures fail to convey the differential significances among these two locations experiencing physically the same amount of precipitation.

Expressed as return periods, however, the differences are immediately conveyed; a 1-in-2.5-year precipitation surplus in Burlington, and a 1-in-52-year precipitation deficit in Jakarta.

In contrast with physical units, return periods add interpretability not only across space but also across time, allowing immediate comparison across seasons, years, or other timescales.

Beyond interpretability, using return periods has the additional advantage of implicitly correcting for systematic bias in the original data because the raw values on which a return period is calculated need be accurate only relative to one another to generate an accurate return period (Isciences, L.L.C., 2019).

Multiple integration periods

As alluded to above, a second key feature of WSIM is to calculate and represent hydrological anomalies at several timescales. WSIM-GLDAS provides parameter measures in terms of return periods, scientific units, and standardized (standard normal) anomalies, computed over each of multiple temporal periods of accumulation, referred to herein as “integration periods”. As an example, Table 1 illustrates the data input represented by August 2019 precipitation for each of the four integration periods used by WSIM-GLDAS.

Table 1. Integration period coverages for example parameter precipitation ending August 2019

Integration Period	Data
1-month	Total precipitation in the month of August 2019
3-month	Total precipitation summed from June 2019 to August 2019
6-month	Total precipitation summed from February 2019 to August 2019
12-month	Total precipitation summed from September 2018 to August 2019

Return periods for each variable are fit separately for each month, ensuring seasonal independence, and leading to a consistent anomaly classification both across spatial climate regimes and across seasons.

Composite water indices
A third key feature of WSIM is to aggregate the return periods of multiple water parameters into composite indices of overall water surpluses and deficits. Composite surplus and deficit indices provide a succinct summary of the information provided by multiple hydrologic drivers, synthesizing information into a digestible metric.

Derivation from GLDAS-2.0
WSIM-GLDAS is a value-added representation of data outputs from the peer-reviewed, established GLDAS Noah land surface model. Inputs to WSIM-GLDAS comprise ten of the 36 parameters provided by the GLDAS-2.0 Noah data product. WSIM ingests these input data, transforms them or derives new variables from them as applicable, fits a theoretical probability function to the distribution of the time series of parameters within a 60-year historical baseline spanning 1 January 1950 to 31 December 2009, and computes a return period of each observation in the data set, for each of one-month, three-month, six-month, and 12-month time integration periods, for each month spanning January 1948 to December of 2014.

II. Data and Methodology

Input data

The Global Land Data Assimilation System combines ground- and space-based observational measurements and theoretical models to provide high-resolution estimates of terrestrial water and energy storage (Rodell, et al., 2004). GLDAS-2.0 Noah data products provide 0.25-degree resolution at monthly time steps for 36 parameters. Outputs of the Global Land Data

Assimilation System 2.0 Noah monthly data product are described by the GLDAS data set documentation (Beaudoing, Rodell, & NASA/GSFC/HSL, 2015) (Rodell, et al., 2004).

WSIM-GLDAS v1.0 is produced using WSIM 2.0.1322 operating on ten of the GLDAS 2.0 Noah outputs: nominally, `Tair_f_inst`, `Rainf_f_tavg`, `PotEvap_tavg`, `Evap_tavg`, `Qs_acc`, `Qsb_acc`, `Qsm_acc`, `SoilMoi0_10cm_inst`, `SoilMoi10_40cm_inst`, and `SoilMoi40_100cm_inst` (Beaudoing, Rodell, & NASA/GSFC/HSL, 2015).

Methods

Processing of GLDAS data product variables
The following lists each variable in WSIM-GLDAS and the processing steps used to generate them.

Temperature (`temp`): GLDAS temperature data is represented in Kelvin units at the monthly time step. Temperature `T` was converted from degrees Kelvin to degrees Celsius using equation (1).

$$T(°C) = T(K) - 273.15 \tag{1}$$

Precipitation (`precip`): `Rainf_f_tavg` represents the monthly average 3-hourly rate of total precipitation in units of kg m^{-2} s^{-1}. To convert this to total monthly precipitation (`precip`) in units of millimeters, WSIM employs equation (2) for each month, where N represents the number of days in the specified month.

$$\frac{\frac{kg}{m^2} \times \frac{1000\,g}{kg} \times \frac{m^2}{10000\,cm^2} \times \frac{cm^3}{g} \times \frac{10\,mm}{cm}}{sec} \times \frac{86400\,sec}{day} \times \frac{N\,days}{month} = \frac{mm}{month} \tag{2}$$

Potential minus Actual Evapotranspiration (`petme`): GLDAS Potential EvapoTranspiration (PET) is given by `PotEvap_tavg` in Watts per square meter and is converted to units of millimeters per month using equation (3). Actual evapotranspiration `Evap_tavg`, (ET) is provided in units of kilograms per square meter per second, converted to units of millimeters per month using equation (4).

$$\frac{W}{m^2} \times \frac{J}{W\,sec} \times \frac{kg}{2.5 \times 10^6\,J} \times \frac{m^3}{1000\,kg} \times \frac{1000\,mm}{m} \times \frac{86400\,sec}{day} \times \frac{N\,day}{month} = \frac{mm}{month} \tag{3}$$

After converting units, `petme` is generated by subtracting (eq. 4).

$$petme = PET - ET \tag{4}$$

Runoff (`runoff`): GLDAS provides storm surface runoff, baseflow-groundwater runoff, and snow melt (nominally `Qs_acc`, `Qsb_acc`, and `Qsm_acc`, respectively), in monthly average 3-hour accumulations in units of kilograms per square meter per 3 hours. To produce a measure of total monthly runoff in mm, first these are each converted to units of millimeters per month by equation (5). The three converted variables are then summed to produce total runoff.

$$\frac{\frac{kg}{m^2} \times \frac{1000\ g}{kg} \times \frac{m^2}{10000\ cm^2} \times \frac{cm^3}{g} \times \frac{10\ mm}{cm}}{3\ hr} \times \frac{8 * 3\ hr}{day} \times \frac{N\ day}{month} = \frac{mm}{month} \tag{5}$$

Flow-accumulated Runoff (`runoff_accum`): WSIM uses a flow accumulation algorithm to determine the amount of cumulative upstream runoff into each grid cell, known as flow-accumulated runoff or "total blue water". First, per-cell runoff is converted to a volume (in cubic meters) that can be accumulated by dividing ro_mm by 1000 and multiplying by the cell area. WSIM then uses the Deterministic eight-node, "D8" (O'Callaghan and Mark, 1984), a single flow direction model which uses the direction of steepest descent to determine into which of the eight neighboring grid cells the grid cell of interest drains. This procedure models instantaneous flow-accumulated runoff, a simplifying assumption that runoff reaches the sea within one month. Making this assumption absolves the need for accounting for stream-channel properties and river flow modeling but may be unreasonable in certain places.

Soil Moisture (`soil_moisture`): Three of the four available GLDAS soil moisture layers are ingested and summed to produce total soil moisture for the top meter of soil. These soil moisture variables are available in units of kilograms per square meter (equivalently, millimeters; see eq. 6) and in GLDAS are named `SoilMoi0_10cm_inst`, `SoilMoi10_40cm_inst`, `SoilMoi40_100cm_inst`.

$$\frac{kg}{m^2} \times \frac{1000\ g}{kg} \times \frac{cm^3}{g} \times \frac{m^2}{10000\ cm^2} \times \frac{10\ mm}{cm} = mm \tag{6}$$

Return periods and standardized anomalies: Frequency analysis to compute a return period is performed on a pixel-by-pixel basis. For each parameter and integration period, WSIM fits a generalized extreme value (GEV) distribution to observations spanning a 60-year historical baseline of January 1950 to December 2009 using the method of L-moments (Hosking, 2017; Hosking & Wallis, 1997).

WSIM applies the GEV distribution, a flexible, three-parameter distribution which is the limiting distribution of maxima and minima, and is recommended by Stagge et al. (2015) for a climatological index similar to the WSIM composite surplus and composite deficit indices.

The fitted cumulative distribution function is used to find the non-exceedance probability $F(x)$ of each observation of the data set. The return period R is defined as $R = \frac{1}{p}$, where $p = F(x)\ if\ F(x) < 0.5$, and $p = 1 - F(x) if\ F(x) > 0.5$.

The cumulative probability $F(x)$ is further used to produce a standardized anomaly, another measure of event rarity, computed by calculating the inverse probability of a standard normal distribution. This measure can be conceptualized as a z-score, or the number of standard deviations away from the historical expectation (Wilks, 2006, pp. 47-49).

Figure 1 illustrates an example in which a GEV cumulative distribution function is fit to 60 historical observations of precipitation for one grid cell location in the month of April (Isciences, L.L.C., 2019). In this example, the cumulative probability of observing 93 mm of precipitation in April is $F(x) = 0.85$, implying that the exceedance probability is p = $1 - F(x) = 0.15$. The return period in years, then, is $\frac{1}{0.15} = 6.7$. In this case, it was determined that at least 93 mm of precipitation in April is expected to occur, on average, once every 6.7 years.

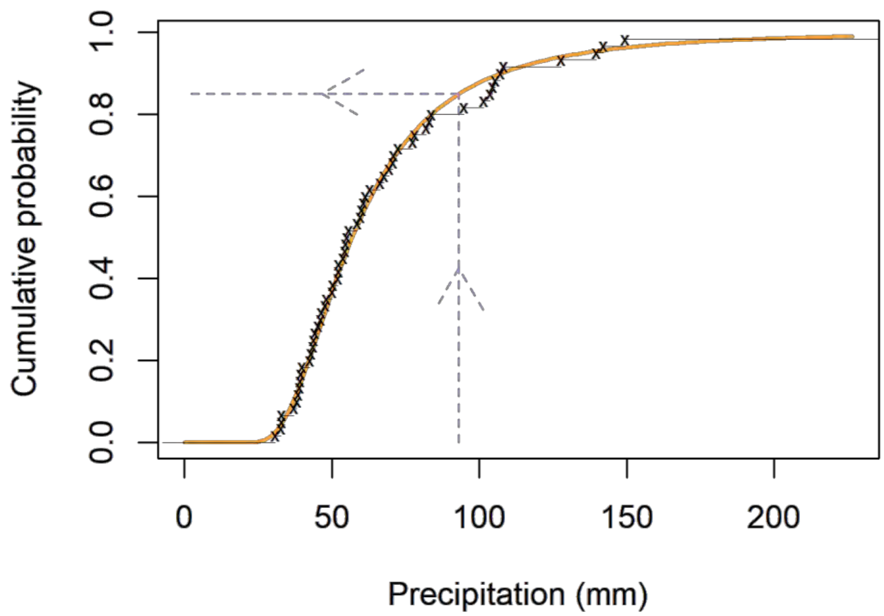

Figure 1. Cumulative probability of 93mm of April precipitation using a fitted GEV distribution function

Composite surpluses and deficits: WSIM computes composite indices of fresh water surpluses and fresh water deficits to report results as digestible indicators of overall hydrological conditions. The composite index is derived from composite anomaly values, which are defined as Composite surplus anomaly value min (`runoff_sa`, `runoff_accum_sa`), where `runoff_sa` is the standardized anomaly of The Composite Surplus Index, and is calculated using the following algorithm:

1. Calculate *F(x)* as described above for each of the following variables: runoff (`runoff`), flow-accumulated runoff (`runoff_accum`), potential minus actual evapotranspiration (`petme`), and soil moisture (`soil_moisture`).
2. Compute the standard normal quantile associated with each cumulative probability *F(x)*. This value represents the "standardized anomaly" for each parameter, denoted e.g., `runoff_sa`.
3. Define the composite surplus anomaly value as the maximum of the standardized anomaly values associated with runoff and flow accumulated runoff, i.e., max (`runoff_sa`, `runoff_accum_sa`). Similarly, compute the composite deficit anomaly value as the minimum of the standardized anomaly values of soil moisture, negative

potential minus actual evapotranspiration, and flow-accumulated runoff: min (`soil_moisture_sa`, `-petme_sa`, `runoff_accum_sa`).
4. Fit a distribution to the historical baseline of composite surplus anomaly values, and likewise for the composite deficit anomaly values. Unlike the distributions of individual hydrological parameters, which are specific to both the integration period and calendar month of the period's end, the historical distributions of each composite index value are specific only to the integration period.
5. Derive a cumulative probability *F(x)* and calculate an associated return period as described above for both the composite surplus anomaly and the composite deficit anomaly. These return periods comprise the composite surplus index and composite deficit index, respectively.

Composite index return periods are capped at 60 years to avoid extrapolation beyond the historical baseline period.

The composite data files also indicate, using a parameter named "both", where there are non-negligible (> 3 year) return periods of both surpluses and deficits. A conceptual example of a scenario in which both water surpluses and water indexes are occurring is in a period of low soil moisture resulting from a low precipitation and high temperature, which simultaneously spurs a large snow-melt event, causing a surplus of runoff into local streams.

For more on this composite indexing methodology, see the WSIM technical documentation (Isciences, L.L.C., 2019).

Processing of Data Deliverables
Further processing steps were taken to structure the approximately 26,000 WSIM-GLDAS netCDF files and produce the data deliverables available on the SEDAC website. These additional processing steps rely upon python's netCDF4 and xarray modules. For each of Iscience's WSIM-GLDAS variables (excluding those in the fits category), netCDF files are available that correspond to each calendar month from January 1948 until December 2014 and each applicable integration period. These individual netCDF files were given new variables termed 'integration_period_end_month' by extracting the year and month from their filename. The files were grouped by their WSIM-GLDAS variable (composite, composite_anom, petme, precip, runoff, runoff_accum, soil_moisture, temp), aggregate function (avg, sum, min, max), and integration_window_months (1 mo, 3 mo, 6 mo, 12 mo). These files were concatenated into single files for each variable at each integration period. During concatenation, a new time dimension and coordinate was created for each of these files.

III. Data Set Description(s)

The Water Security Indicator Model - Global Land Data Assimilation System (WSIM-GLDAS) Monthly Grids, Version 1 data set contains grids of fresh water parameters for each month spanning January 1948 to December 2014, and are integrated over 1-month, 3-month, 6-month, and 12-month integration periods at 0.25-degree resolution.

For more information on the parameters, variables, and data types, refer to these appendices:

- Appendix 3: WSIM-GLDAS Parameter Metadata
- Appendix 4: WSIM-GLDAS Data Set Metadata
- Appendix 5: WSIM-GLDAS Grid Files

Data set web page:
SEDAC URL: https://sedac.ciesin.columbia.edu/data/set/water-wsim-gldas-v1
Permanent URL: https://doi.org/10.7927/z1fn-kf73

Data set format:
The data are available in netCDF format as a downloadable zip file, one netCDF file per parameter/integration period that comprises the complete time series for that parameter and integration period. The downloadable is a compressed zip file, containing: 1) global raster, 2) Readme, and 3) PDF documentation.

Data set downloads:
The 38 downloadable data zip files follow this file-naming pattern:

- water-wsim-gldas-v1-parameter-integrationperiod-netcdf.zip

For example, "water-wsim-gldas-v1-composite-three-month-netcdf.zip" contains composite water surplus and deficit indices for each three-month period spanning January 1948 to December 2012.

Missing data values are described by the `_FillValue` attribute in the netCDF file. Coordinate variables `lat` (float64), `lon` (float64) and `crs` (int32) are stored for all files, as are units and other information conforming to the CF Conventions.

IV. How to Use the Data

The multidimensional data in netCDF format can be used in major software interfaces or through an available netCDF application for scientific data analysis and charting. This data set includes geographic coordinate variables that can be used in any standard Geographical Information System (GIS) for mapping and geospatial analysis.

V. Potential Use Cases

The WSIM-GLDAS data set can be used to create series of "hotspot" maps that present a succinct visualization of the evolution of short- or long-term water deficits and surpluses in any region of land on the globe. Figure 2 presents a 60-year history of composite surpluses and deficits integrated over the 12-month period covering January through December for each of the

60 years spanning 1951 to 2010 in the southwestern United States. Water surpluses are in blue, water deficits are in red.

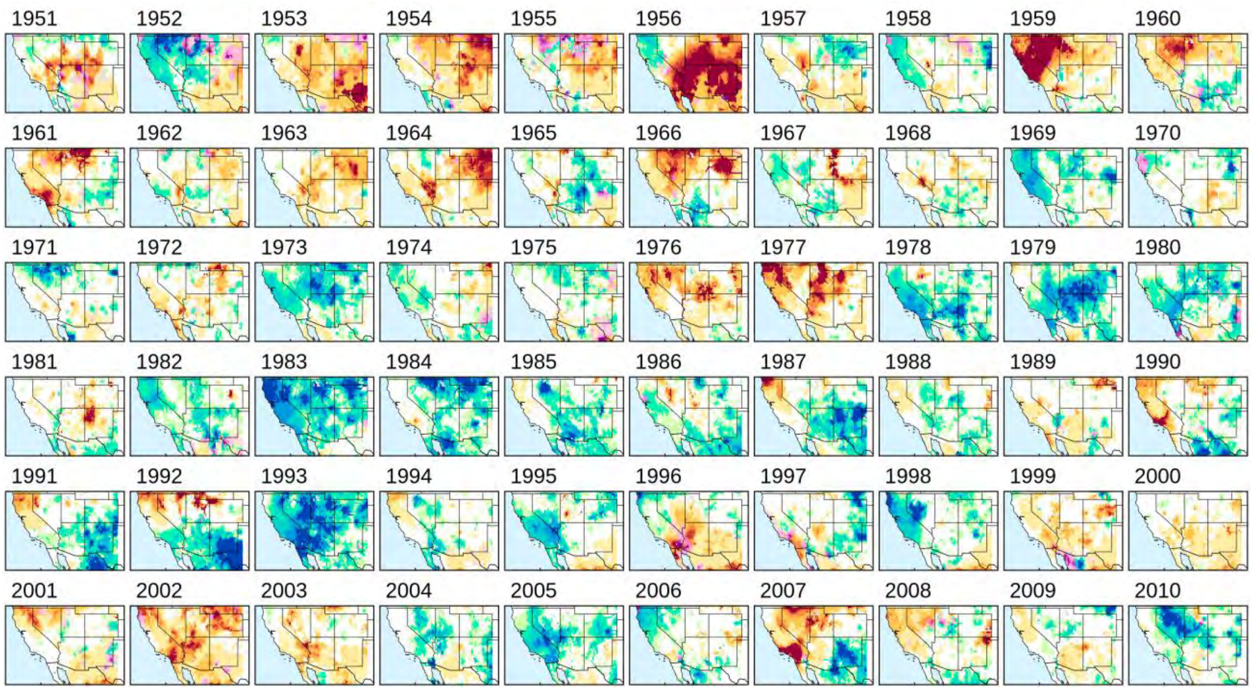

Figure 2. 60-year history of annualized composite water surpluses and deficits

WSIM-GLDAS data set is well-poised to be paired with other raster or polygonal data sets for socio-environmental analysis. For example, zonal statistics summarizing rainfed agricultural areas experiencing water stress over time can be computed using basin polygons, agricultural grids, and an efficient zonal statistics calculator such as *exactextract* (Baston, 2018).

VI. Limitations

Frequency analysis in which the probability of the magnitude of an event is calculated by fitting a probability function assumes independence and stationarity of the observations. In other words, events are viewed as randomly distributed over time following a particular probability distribution, which assumes that environmental factors behave the same way in the past, present, and future to control the underlying hydrological processes (Du, et al., 2015).

The use of return periods, or any anomaly measure, relies on choosing a historical baseline period of observations to use as a reference for past, present, and future events. The choice of baseline is naturally constrained by data availability, and beyond that the choices of the number of years to include and the set of that number of years are subject to judgement. WSIM uses a 60-year baseline period to cover the second half of the 20th century and a decade into the 2000's; this covers nearly all data available (excluding calendar years 1948 and 1949) in the GLDAS Noah V2.0 product used to build WSIM-GLDAS.

While the premise motivating the use of return periods is that societal impacts of an event will be proportional to the rarity of the event, the rarity of a weather event relative to some historical baseline may not be directly proportional to its societal impact. It is plausible that, for example, growing awareness of climate change spurs societal developments which, on average, increase resilience over time to an event that exceeds any given return period as measured with respect to a baseline that exists statically in the past.

Representing data in terms of return periods also requires an estimation of the unknown (and unknowable) cumulative probabilities of events, which requires choosing an estimator of probability. There are several empirical estimations of probability in the *quantile* function in the statistical software R alone, and among other statistical software packages (Hyman & Fan, 1996). A parametric approach, taken herein, involves choosing a theoretical probability distribution form, of which parameters are fit to the observed data. Parametric fits smooth the finite observed data to the most likely distribution of a specified form but assume that the distributional form is an appropriate one.

VII. Acknowledgments

Funding for development and dissemination of this data set was provided under the U.S. National Aeronautics and Space Administration (NASA) contract 80GSFC18C0111 for the continued operation of the Socioeconomic Data and Applications Center (SEDAC), which is operated by the Center for International Earth Science Information Network (CIESIN) of Columbia University. Funding for development of WSIM version 2.0 was provided by the United States Army Corps of Engineers.

VIII. Disclaimer

CIESIN follows procedures designed to ensure that data disseminated by CIESIN are of reasonable quality. If, despite these procedures, users encounter apparent errors or misstatements in the data, they should contact SEDAC User Services at ciesin.info@ciesin.columbia.edu. Neither CIESIN nor NASA verifies or guarantees the accuracy, reliability, or completeness of any data provided. CIESIN provides this data without warranty of any kind whatsoever, either expressed or implied. CIESIN shall not be liable for incidental, consequential, or special damages arising out of the use of any data provided by CIESIN.

IX. Use Constraints

Users are free to use, copy, distribute, transmit, and adapt the work for commercial and non-commercial purposes, without restriction, if clear attribution of the source is provided.

X. Recommended Citation(s)

Data set(s):

ISciences, and Center for International Earth Science Information Network (CIESIN), Columbia University. 2022. Water Security Indicator Model - Global Land Data Assimilation System (WSIM-GLDAS) Monthly Grids, Version 1. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/z1fn-kf73. Accessed DAY MONTH YEAR.

Scientific publication:

ISciences, L.L.C. 20 August 2019. *Water Security Indicator Model*. https://wsim.isciences.com/.

XI. Source Code

Source code used to develop the WSIM-GLDAS data set is available at https://gitlab.com/isciences/wsim/wsim, using the WSIM version. It was produced using R version 3.6.1 (R Core Team, 2019).

XII. References

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control,* 19(6):716-723. https://doi.org/10.1109/TAC.1974.1100705.

Baston, D. 2018. GitHub Repository: https://github.com/isciences/exactextract.

Beaudoing, H., M. Rodell, and NASA/GSFC/HSL. 2015. *GLDAS Noah Land Surface Model L4 monthly 0.25 x 0.25 degree V2.0*. Greenbelt, MD: Goddard Earth Sciences Data and Information Services Center (GES DISC). https://doi.org/10.5067/9SQ1B3ZXP2C5.

Du, T., L. Xiong, C.-Y Xu, C. J. Gippel, S. Guo, S., and P Liu. 2015. Return period and risk analysis of nonstationary low-flow series under climate change. *Journal of Hydrology*, 527:234-250. https://doi.org/10.1016/j.jhydrol.2015.04.041.

GDAL/OGR contributors. 2019. *GDAL/OGR Geospatial Data Abstraction software Library*. Open Source Geospatial Foundation. https://gdal.org.

Hosking, J. R. 2017. *L-Moments*. CRAN. https://cran.r-project.org/web/packages/lmom/.

Hosking, J. R., and J. R. Wallis. 1997. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge, UK: Cambridge University Press.

Hyman, R. J., and Y. Fan. November 1996. Sample quantiles in statistical packages. *The American Statistician*, 50(4):361-365. https://www.jstor.org/stable/2684934.

O'Callaghan J. F., and D. M. Mark. 1984. The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing*. 28(3):323-344. https://doi.org/10.1016/S0734-189X(84)80011-0.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rodell, M., P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C. Meng, and D. Toll. 2004. The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, 85(3):381-394. https://doi.org/10.1175/BAMS-85-3-381.

Stagge, J. H., L. M. Tallaksen, L. Gudmundsson, A. F. Van Loon, and K. Stahl. 2015. Candidate Distributions for Climatological Drought Indices (SPI and SPEI). *International Journal of Climatology*, 35(13):4027-4040. https://doi.org/10.1002/joc.4267.

Wilks, D. S. 2006. *Statistical Methods in the Atmospheric Sciences,* 2nd edition. Burlington, MA: Academic Press.

XIII. Documentation Copyright and License

Appendix 1. Data Revision History

No revisions have been made to this data set.

Appendix 2. Contributing Authors & Documentation Revision History

Revision Date	ORCID	Contributors	Revisions
January 20, 2022	0000-0002-8004-5385 0000-0002-2195-1716	Kytt MacManus, John Squires, Cynthia Crowley, Dan Baston, Joshua Brinks	This document is the 1st instance of documentation.

Appendix 3. WSIM-GLDAS Parameter Metadata

Name	**Description**	**Units**
`both`	Parameter indicating that both the composite water deficit and composite water surplus indices exceed a value of 3 years	
`deficit (composite_anom)`	The standardized anomaly from which the composite deficit index is derived	dimensionless
`deficit (composite)`	Composite water deficit index as described above, measured in return periods	year
`deficit_cause`	Predominant driver of composite water deficits	Index: 1=`-petme_sa`, 2=`soil_moisture_sa`, 3=`runoff_accum_sa`
`surplus (composite)`	Composite surpluses index as described above, measured in return period	years
`surplus (composite_anom)`	The standardized anomaly from which the composite surplus index is derived	dimensionless
`surplus_cause`	Predominant driver of composite water surpluses	Index: 1= `runoff_sa` 2= `runoff_accum_sa`
`integration_period_end_month`	Represents the month in which the temporal period of accumulation (integration period) ends	YYYYMM
`precip`	Monthly total precipitation	mm

`precip_sum`	Monthly total precipitation summed over the integration period	mm
`petme`	Total monthly potential minus actual evapotranspiration	mm
`petme_sum`	Total monthly potential actual evapotranspiration, summed over the integration period	mm
`runoff`	Total monthly runoff	mm
`runoff_accum`	Total blue water, or flow accumulated runoff	cubic meters (m^3)
`runoff_accum_sum`	Total blue water summed over the integration period	cubic meters (m^3)
`runoff_accum_min`	Minimum monthly total blue water over all months in the integration period	cubic meters (m^3)
`runoff_accum_max`	Maximum monthly total blue water over all months in the integration period	cubic meters (m^3)
`runoff_sum`	Total runoff summed over the integration period	mm
`soil_moisture`	Total water content in the top meter of soil	mm
`soil_moisture_avg`	Soil moisture averaged over the integration period	mm
`time`	Represents the number of days past the first date of the first month of the data set time-series (for instance: 1948-12-01 00:00:00 for data sets measuring 12 month integration periods); netCDF tools such as Python's xarray module can convert these integer values into datetime format (see: http://xarray.pydata.org/en/stable/time-series.html)	Integer number of days
`temp`	Monthly average air temperature	degrees Celsius (°C)
`temp_avg`	Monthly air temperature averaged over the integration period	degrees Celsius (°C)

Appendix 4. WSIM-GLDAS Data Set Metadata

Name	Description	Integration Period	Variable (Units)

`composite`	These files provide composite surplus and composite deficit index data as the return periods; includes values for the following variables: composite combined surplus & deficit index (both), composite deficit index, cause of deficit, composite surplus index (composite), cause of surplus.	1 month
		3 month
		6 month
		12 month
			`both`
			`deficit_cause(index)`
			`deficit(years)`
			`surplus (index)`
			`surplus_cause (years)`

`composite_anom`	These files provide values for composite surplus anomalies and composite deficit anomalies; includes data on the following variables: composite combined surplus & deficit index (both), composite deficit index, cause of deficit, composite surplus index, cause of surplus.	1 month
		3 month
		6 month
		12 month
			`both`
			`deficit_cause(index)`
			`deficit`
			`surplus`
			`surplus_cause(index)`

`Petme_sum`	These files provide data for the monthly potential minus actual evapotranspiration parameter summed over the integration period.	1 month
		3 month
		6 month
		12 month
			`scientific (mm)`
			`return_period(years)`
			`anomaly`

`Precip_sum`	These files provide data for monthly total precipitation summed over the integration period.	1 month
		3 month
		6 month
		12 month
			`scientific(mm)`
			`return_period(years)`
			`anomaly`

`Runoff_accum_max`	These files provide data for the maximum monthly total blue water, or flow accumulated runoff, over all months in the integration period. Due to simplifying assumptions in the Deterministic eight-node "D8" (O'Callaghan and Mark, 1984) flow accumulation algorithm, data at the 1 month integration period resolution is not provided.	3 month
		6 month
		12 month
			scientific(m^3)
			`return_period(years)`
			`anomaly`

`Runoff_accum_min`	These files provide data for the minimum monthly total blue water, or flow accumulated runoff, over all months in the integration period. Due to simplifying assumptions in the Deterministic eight-node "D8"	3 month
		6 month
		12 month
			scientific(m^3)
			`return_period(years)`
			`anomaly`

	(O'Callaghan and Mark, 1984) flow accumulation algorithm, data at the 1 month integration period resolution is not provided.		

`Runoff_accum_sum`	These files provide data for the total blue water, or flow accumulated runoff, summed over the integration period. Due to simplifying assumptions in the Deterministic eight-node "D8" (O'Callaghan and Mark, 1984) flow accumulation algorithm, data at the 1 month integration period resolution is not provided.	3 month
		6 month
		12 month
			`scientific(m³)`
			`return_period(years)`
			`anomaly`

`Runoff_sum`	These files provide data for the total runoff summed over the integration period.	1 month
		3 month
		6 month
		12 month
			`scientific(mm)`
			`return_period(years)`
			`anomaly`

`Soil_moisture_avg`	These files provide data for soil moisture averaged over the integration period.	1 month
		3 month
		6 month
		12 month
			`scientific(m³)`
			`return_period(years)`
			`anomaly`

`Temp_avg`	These files provide data for the monthly air temperature averaged over the integration period.	1 month
		3 month
		6 month
		12 month
			`scientific (°C)`
			`return_period (years)`
			`anomaly`

`fits`	Corresponding fits data for the parameter values (location, scale, shape) of the fitted GEV distribution are provided as one file per parameter per integration period per calendar month. The fit files are named using the syntax:<parameter>_<integration period>_month_<mm>.nc	All integration periods available, 1 per calendar month	

Appendix 5. WSIM-GLDAS Grid Files

Parameter	Variables	Data Type
composite	both(composite combined surplus & deficit index) composite deficit index cause of deficit composite surplus index cause of surplus integration_period_end_month time	float32 float32 float32 int8 int8 string int64
composite_anom	Both(composite combined surplus & deficit index) composite deficit index cause of deficit composite surplus index cause of surplus integration_period_end_month time	float32 float32 float32 int8 int8 string int64
petme_sum	Scientific return_period anomaly integration_period_end_month time	float32 float32 float32 string int64
precip_sum	scientific return_period anomaly integration_period_end_month time	float32 float32 float32 string int64
runoff_accum_max	scientific return_period anomaly integration_period_end_month time	float32 float32 float32 string int64
runoff_accum_min	scientific return_period anomaly integration_period_end_month time	float32 float32 float32 string int64
runoff_accum_sum	scientific return_period anomaly integration_period_end_month time	float32 float32 float32 string int64

runoff_sum	scientific return_period anomaly integration_period_end_month time	float32 float32 float32 string int64
soil_moisture_avg	scientific return_period anomaly integration_period_end_month time	float32 float32 float32 string int64
temp_avg	scientific return_period anomaly integration_period_end_month time	float32 float32 float32 string int64
fits	location scale shape	float64 float64 float64