Africa Rainfall Data Documentation

Release 1.0.0

Ashley Cryan

Apr 29, 2022

CONTENTS:

The African Rainfall Project derives accurate rainfall estimates over Sub-Saharan Africa, with the help of a high-resolution (1km) application of the Weather Research and Forecasting Model (WRF) running on the IBM World Community Grid.

Simulation data are made available for reuse under a CC BY-NC 4.0 International license. For more information about the project and how to access the data, please see below.

ABOUT THE AFRICA RAINFALL PROJECT

1.1 Project goals

The goal of the African Rainfall Project (ARP) is to derive accurate rainfall estimates over Sub-Saharan Africa, with the help of a high-resolution (1km) application of the Weather Research and Forecasting Model (WRF). Such a resolution will allow the model to better represent rainfall, and in particular, convective precipitation.

This is a unique experiment that has never been performed at such a scale. The model runs on the IBM World Community Grid (WCG), an SCR activity of IBM. These data can be used to:

- Help scientists better understand these storms and improve forecasting models

- Produce more accurate rainfall forecasts for sub-Saharan Africa

- Give farmers more timely information about when to plant, help them obtain insurance

- Support resilience in the face of climate change

    "This is the first time we'll be able to map huge parts of Africa for a whole rainy season, and has never been done before at this level of resolution. This is only possible because of the amount of computing power we'll have through World Community Grid."* - Nick van de Giesen, Principal Investigator

1.2 The Weather Research and Forecasting Model (WRF)

These simulation data were produced using the Weather Research and Forecasting Model (WRF) from the National Center for Atmospheric Research (NCAR) Mesoscale & Microscale Meteorology Laboratory.

From the NCAR website: "The Weather Research and Forecasting (WRF) Model is a next-generation mesoscale numerical weather prediction system designed for both atmospheric research and operational forecasting applications.

For researchers, WRF can produce simulations based on actual atmospheric conditions (i.e., from observations and analyses) or idealized conditions. WRF offers operational forecasting a flexible and computationally efficient platform, while reflecting recent advances in physics, numerics, and data assimilation contributed by developers from the expansive research community.

WRF is currently in operational use at NCEP and other national meteorological centres as well as in real-time forecasting configurations at laboratories, universities, and companies."

- Read more about The Weather Research and Forecasting Model

1.3 The World Community Grid

The WRF model runs on the World Community Grid (WCG), a SCR initiative of IBM.

The WCG relies on volunteers who download a secure software program to their computer. When the computer is idle or not using its full computing power, it will run a simulated experiment in the background. Then, the computer contacts the World Community Grid server to let it know that it has completed the simulation, which is then uploaded to an IBM server. All of this happens unobtrusively.

World Community Grid receives the results volunteers send back (often called work units or research tasks), combines them with hundreds of thousands of results from other volunteers all over the world, and sends them to the Delft research team. The researchers then begin the difficult work of analyzing the data. While this process can take years, it accelerates that would otherwise take decades or might even be impossible.

- Africa Rainfall Project Overview on The World Community Grid website
- Join The World Community Grid so you and your computer can help accelerate this important research

1.4 ARP project news and updates

- Lastest news and updates on the ARP project from WCG
- Project statistics (data production and model runs)

1.5 More resources

- Article (*newscientist.com*, 2015): "Sensors to give early storm warnings to people near deadly lake"
- Article (*TU Delft*, 2018): "Super computing power for rainfall modelling in Africa"
- Article (*bi-platform.nl*, 2019): "IBM en TU Delft starten Africa Rainfall Project"
- Article (*de Volkskrant*, 2019): "Boeren in Afrika hebben jouw computer nodig voor hun weersvoorspellingen"
- Trans-African Hydro-Meteorological Observatory (TAHMO) Data Portal
- TAHMO School2School program

ACCESSING ARP DATA

2.1 THREDDS server

Data are stored as netCDF files (netCDF 3 classic 64-bit) on a THREDDS server hosted at TU Delft. These data can be accessed for anlysis purpose, using the OPeNDAP protocol. The main data catalog is available at: https://africarain.ceg.tudelft.nl:9010/thredds/catalog.html

To work with these files on your local system via OPeNDAP, you first need to obtain an OPeNDAP-enabled client program. Some common client programs include NCO, MATLAB, R, ArcGIS, Python and others. Please see this information page from the NOAA Physical Sciences Library for more information.

2.2 Access ARP data using Python

This is an example of how to access data from the Africa Rain THREDDS server using Python. This relies on the netCDF4 library, a powerful library for working with netCDF data in general. This example uses a dummy dataset, which endpoint is `https://africarain.ceg.tudelft.nl:9010/thredds/dodsC/demos/demo.nc`

```python
import netCDF4

# endpoint for specific file

url = 'https://africarain.ceg.tudelft.nl:9010/thredds/dodsC/demos/demo.nc'

# read dataset
nc = netCDF4.Dataset(url)

# read variables
ncv = nc.variables
print(ncv.keys())

## a subset of the data set can be retrieved using coordinate.
## Use relevant coordinates for your dataset, the values below are just an example
# lon = ncv['longitude'][10:-10:2,20:-10:2]
# lat = ncv['latitude'][10:-10:2,20:-10:2]

# read the nth time step
itime = 10
tair = ncv['air_temperature'][itime]

# print data
```

(continues on next page)

```
print(tair)

## Other examples: https://publicwiki.deltares.nl/display/OET/
↪Reading+data+from+OpenDAP+using+python
```

2.3 Contact

Processed data will be available directly from the THREDDS URL listed above. Data must be cited appropriately and used in accordance with licensing requirements. See *Licensing and citation*

Raw data files (see *Production of raw data files*) are available upon request from Nick van de Giesen, n.c.vandegiesen@tudelft.nl.

VISUALIZING DATA USING WMS

This is an example on how connect to the Web Map Service of the AfricaRain THREDDS server to vizualize data in interactive maps. We assume that the *demo.nc* contains a layer called **air_temperature**

Requirements: The *ipyleaflet* module, which you can install with `$ pip install ipyleaflet`

```
[25]: from ipyleaflet import Map, WMSLayer

wms = WMSLayer(
    url='https://africarain.ceg.tudelft.nl:9010/thredds/wms/demos/demo.nc?
→COLORSCALERANGE=273,317', #Use COLORSCALERANGE=273,317 only within this example.␣
→Omit for any other case.
    layers='air_temperature',
    format='image/png',
    transparent=True,
    attribution='Africa Rain Project, TU Delft'
)

center=(40.1, -104.5)

m = Map(center=center, zoom=7.5)

m.add_layer(wms)

display(m)
```

```
Map(center=[40.1, -104.5], controls=(ZoomControl(options=['position', 'zoom_in_text',
→'zoom_in_title', 'zoom_o...
```

PRODUCTION OF RAW DATA FILES

Data available for download have been processed from their original raw form. This page describes the generation of the raw simulation data, including which model and parameter values were used to generate it. Steps taken to produce the processed data are described in the *Production of processed data files* section.

4.1 Raw simulation data production

Data are high-resolution computer simulations of localized rainstorms in sub-Saharan Africa produced using massive, crowd-sourced computing power from World Community Grid (see documentation on *The World Community Grid*).

The amount of raw simulation data produced is about 0.5 PB or, in more nostalgic terms, a pile of floppy disks of over 1000 km. That pile would weigh over 6700 tons and would be over 1200 km high. About twenty variables of direct interest are stored and uploaded to the central WCG facility. These data are stored in netCDF files.

4.2 Forcing data

Forcing data used as input for the simulation results come from:

> National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce. 2015, updated daily. NCEP GDAS/FNL 0.25 Degree Global Tropospheric Analyses and Forecast Grids. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. DOI: 10.5065/D65Q4T4Z

The data are available from UCAR here. Data are free, but registration is required.

4.3 Model architecture

The model used to produce these simulations is the Weather Research and Forecasting Model (WRF) V3.9.1.1 from the National Center for Atmospheric Research (NCAR).

ARP divides the African continent into over 35609 WRF modelling units. For each unit of 50x50 cells, WRF is run on a personal computer of a volunteer, who shares spare computing resources via the WCG. It calculates episodes of two days' worth of weather with output for every 15 minutes (193 total time steps each).

Each WRF unit is triply nested. Hence, it first calculates at a coarse resolution of 9km x 9km, covering a 468km x 468km region with historical boundary conditions from NOAA's Global Forecast System (GFS-ANL). Within the

centre of this domain, it calculates the next domain at the intermediate resolution of 3 km (156 km x 156 km) with the boundary conditions set by the coarser domain calculation.

Finally, a unit of 52 km x 52km is calculated at the centre of the intermediate domain. Vertically, the atmosphere is divided into 51 layers so that the output is produced on a 52x52x51 grid.
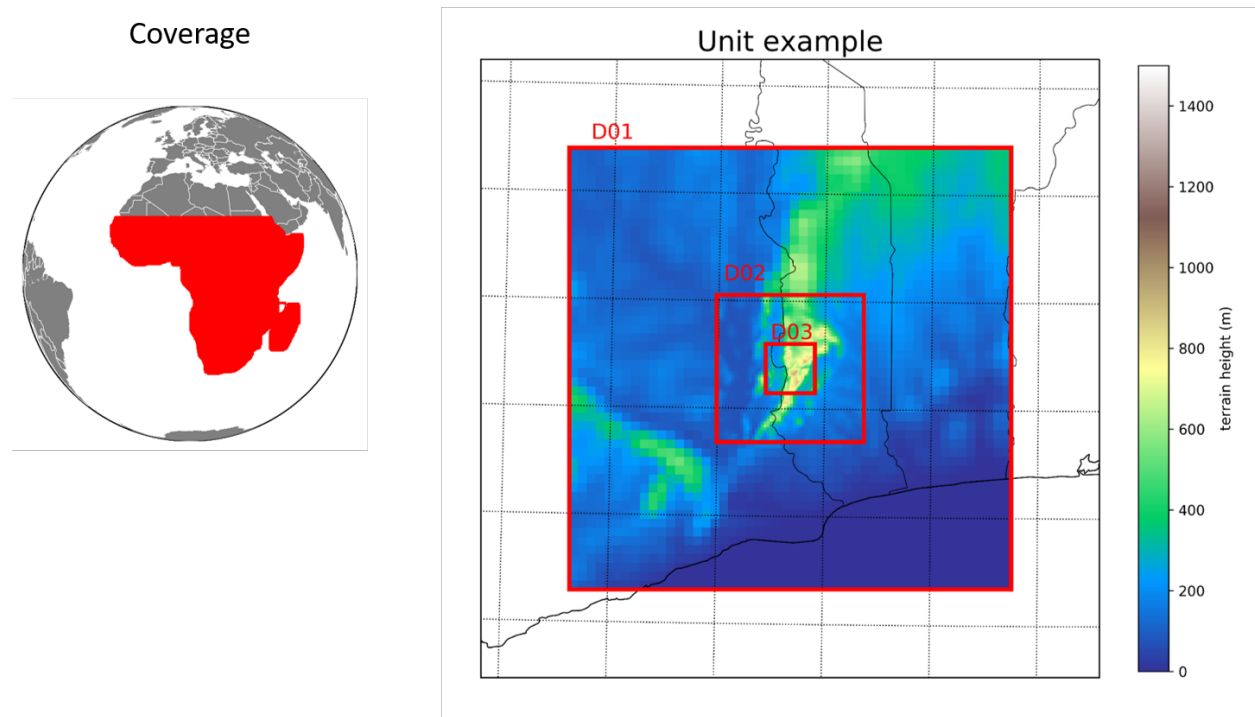
4.3.1 Model input parameter values

For a complete list of parameter values that were used as inputs for the WRF model in order to produce the resulting simulation data, please see *Model input parameter values*.

4.4 Time period

The period of simulation data covered will ultimately run from 1 June 2018 until 31 May 2019. Raw simulation data are generated on a rolling basis subject to volunteer participation. If the current pace continues, the dataset is expected to be complete in mid-2022. Simulation data are generated at a 15-minute time interval.

4.5 Units

Units define geographic areas for which simulation results are available or will be available. A total of 35.609 square units cover Sub-Saharan Africa. For each unit, simulation results are produced at three spatial granularities called domains. Thus, a domain can also be described as a subset of a unit with a particular spatial resolution.

The domains used in the simulation have the following resolutions:

- Domain 1: 9 km

- Domain 2: 3 km

- Domain 3: 1 km

The centroid of each unit is separated by 15.3 minutes of arc in both latitude and longitude. Each unit partially overlaps with adjacent units; all domains contain 51 X 51 grid points. The model results are non-deterministic, so units were designed to overlap and create redundancy for a given geographic location, i.e., more than one value for a specific geographic location at a given time. These values will treated in the processing steps to remove the effect of the overlapping values. More information about the status of processed data in *Production of processed data :docs:`data_processing*

4.6 Georeferencing information

Raw datasets were produces using the **WRF Lambert Conformal** projection. For details, consult the WRF Model Manuals

---

4.7 Variables in raw simulation dataset

VARI-ABLE	DESCRIPTION	DATA TYPE	UNITS	GEO-GRAPHIC DATA
Times	–	char	–	No
HFX_FORCE	SCM ideal surface sensible heat flux	float	W/m2	No
NEST_POS	–	float	–	Yes
Q2	Water vapor mixing ratio (QV) at 2m	float	Kg/Kg	Yes
T2	Air temperature at 2m	float	K	Yes
TH2	Potential temperature at 2m	float	K	Yes
PSFC	Surface air pressure	float	Pa	Yes
U10	U component of the wind speed at 10m (X surface wind)	float	m/s	Yes
V10	V component of the wind speed at 10 m (Y surface wind)	float	m/s	Yes
ITIMESTEP	–	int	–	No
XTIME	Minutes since 2018-07-01 00:00:00	float	min-utes	No
SMOIS	Soil moisture	float	m3/m3	Yes
P_TOP	Pressure top of the model	float	Pa	No
RAINC	Accumulated total cumulus precipitation (convective precipitation)	float	mm	Yes
RAINSH	Accumulated shallow cumulus precipitation (large-scale precipitation)	float	mm	Yes
RAINNC	Accumulated total grid scale precipitation (non-convective precipitation)	float	mm	Yes
SWDOWN	Downward short wave flux at ground surface (surface downwelling shortwave radiation)	float	W/m2	Yes
GLW	Downward long wave flux at ground surface (surface downwelling longwave radiation)	float	W/m2	Yes
OLR	Top of atmosphere outgoing longwave radiation	float W/m2	Yes	
SR	Fraction of frozen precipitation	float	–	Yes
SST	Sea surface temperature	float	K	Yes

FIVE

PRODUCTION OF PROCESSED DATA FILES

5.1 Data processing steps

Data that will be available for download will be processed and aggregated from their original form. The following steps were taken to produce processed data:

> **Attention:** This is a work in progress. Processed data will be available in the near future.

Raw data files are maintained separately and, due to their large volume, are not made available via the THREDDS server along with the processed data. If your research project requires access to raw simulation data, please contact Nick van de Giesen (N.C.vandeGiesen@tudelft.nl) to request access.

5.2 Georeferencing information

Processed data will be produced, most likely, using the **Lambert Conformal** projection and the WGS 1984 datum.

5.3 Variables in the processed dataset

VARIABLE	DESCRIPTION	DATA TYPE	UNITS	GEOGRAPHIC DATA

5.4 Time period

Processed data will be created from raw simulation outputs as the WCG volunteer program generates them on a rolling basis. Processed data will ultimately be available from 1 June 2018 until 31 May 2019. Simulations are expected to be complete in mid-2022. Processed data are available at a 1-hour time interval.

5.5 Spatial resolution

Processed data will be available, initially, at 1-km resolution.

FILE METADATA

Title:

Africa Rainfall Project Data

Data format:

netCDF3 64-bit offset

Summary:

Data are high-resolution computer simulations of localized rainstorms in sub-Saharan Africa produced using massive, crowd-sourced computing power from the World Community Grid.

Keywords:

Africa, rainfall, precipitation, modelling

Institution:

TU Delft

Conventions:

CF 1.8

Projection:

Lambert Conformal

Source:

Weather Research and Forecasting (WRF) Model V3.9.1.1 run on the IBM World Community Grid (WCG)

License:

This data is provided under a CC BY-NC 4.0 International license. Specific terms for sharing and use can be found at creativecommons.org.

6.1 CF-1.8 Convention

These data conform to CF convention 1.8.

Note: For more information about the CF conventions and specifications related to CF-1.8, please see cfconventions.org.

6.2 Viewing file metadata with ncdump

The metadata for files that have been downloaded from this server can be found using multiple tools specific to working with netCDF files.

One recommended tool for viewing this metadata and working with netCDF files, in general, is a set of command-line programs called the NetCDF Operators (NCO). These can be downloaded from http://nco.sourceforge.net/ and installed following instructions on that page.

Once you have the NCO programs installed, you can use the `ncdump` command to view the metadata for any netCDF file.

```
$ ncdump -h filename.nc
```

Note: For more options when using `ncdump`, see unidata.ucar.edu.

MODEL INPUT PARAMETER VALUES

Simulation datasets were produced using the following parameters for the WRF model:

PARAMETER	VALUE	DESCRIPTION
SIMULATION START DATE	2018-07-01_00:00:00	The date and time when the simulation comm
WEST-EAST_GRID_DIMENSION	52	Dimension of the grid in the West-East direc
SOUTH-NORTH_GRID_DIMENSION	52	Dimension of the grid in the South-North dir
BOTTOM-TOP_GRID_DIMENSION	51	Dimension of the grid in the vertical directio
DX	9000.f	Grid resolution in the X direction (m).
DY	9000.f	Grid resolution in the Y direction (m).
SKEBS_ON	0	Stochastic kinetic energy backskatter scheme
SPEC_BDY_FINAL_MU	1	Whether to call spec_bdy_final for mu.
USE_Q_DIABATIC	0	Whether to include QV and QC tendencies i
GRIDTYPE	C	Type of grid used by the model.
DIFF_OPT	1	Turbulence and mixing option.
KM_OPT	4	Eddy coefficient option.
DAMP_OPT	0	Upper-level damping flag (0 = no damping).
DAMP_COEFF	0.2f	Damping coefficient.
KHDIF	0.f	Horizontal diffusion constant (m2/s).
KVDIF	0.f	Vertical diffusion constant (m2/s).
MP_PHYSICS	10	Microphysics scheme.
RA_LW_PHYSICS	4	Longwave radiation scheme.
RA_SW_PHYSICS	4	Shortwave radiation scheme.
SF_SFCLAY_PHYSICS	2	Surface layer scheme.
SF_SURFACE_PHYSICS	2	Land surface scheme.
BL_PBL_PHYSICS	2	Planetary boundary for layer scheme.
CU_PHYSICS	3	Cumulus parameterization scheme.
SF_LAKE_PHYSICS	0	Lake physics scheme.
SURFACE_INPUT_SOURCE	3	Landuse and soil category.
SST_UPDATE	1	Option to use time-varying SST, seaice, vege
GRID_FDDA	0	Grid nudging option (0 = none).
GFDDA_INTERVAL_M	0	Time interval (minutes) vetween analyses fo
GFDDA_END_H	0	Time (hours) to stop nudging after the start o
GRID_SFDDA	0	Surface FDDA switch (0 = off).
SGFDDA_INTERVAL_M	0	Time interval (minutes) between surface ana
SGFDDA_END_H	0	Time (hours) to stop surface nudging after th
HYPSOMETRIC_OPT	2	Hypsometric option.
USE_THETA_M	0	Whether to use theta (1+1.61QV).
GWD_OPT	0	Gravity wave drag option (0 = off).

Table 7.1 – continued from previous

PARAMETER	VALUE	DESCRIPTION
SF_URBAN_PHYSICS	1	Urban surface model option.
SF_OCEAN_PHYSICS	0	Ocean model option.
SHCU_PHYSICS	0	Shallow convection option.
MFSHCONV	0	Turns on day-time EDMF for QNSE (0 = off
FEEDBACK	0	For nested domain: 0 = one-way nesting, 1 =
SMOOTH_OPTION	2	Smoothing option for the parent domain in th
SWRAD_SCAT	1.f	Scattering turning parameter for ra_sw_phys
W_DAMPING	1	Vertical velocity damping flag.
DT	36.f	Time step (seconds).
RADT	1.f	Minutes between radiation physics calls.
BLDT	0.f	Minutes between boundary-layer physics cal
CUDT	0.f	Minutes between cumulus physics calls.
AER_OPT	0	Aerosol input option (RRTMG only).
SWINT_OPT	0	Interpolation of shortwave radiation based on
AER_TYPE	1	Aerosol type to be used
AER_A0D550_OPT	1	
AER_ANGEXP_OPT	1	
AER_SSA_OPT	1	
AER_ASY_OPT	1	
AER_AOD550_VAL	0.12f	
AER_ANGEXP_VAL	1.3f	
AER_SSA_VAL	0.85f	
AER_ASY_VAL	0.9f	
MOIST_ADV_OPT	1	Advection options for moisture.
SCALAR_ADV_OPT	1	Advection options for scalars.
TKE_ADV_OPT	1	Advection options for TKE.
DIFF_6TH_OPT	0	6th-order numerical diffusion (0 = none).
DIFF_6TH_FACTOR	0.12f	6th-order numerical diffusion non-dimension
OBS_NUDGE_OPT	0	obs-nudging fdda (0 = off).
BUCKET_MM	-1.f	Bucket reset values for water accumulation (
BUCKET_J	-1.f	Bucket reset value for energy accumulations
PREC_ACC_DT	0.f	Bucket reset time interval between outputs fo
ISFTCFLX	0	Alternative Ck (exchange coefficient for tem
ISHALLOW	0	Turns on shallow convection (default is 0 =
ISFFLX	1	Heat and moisture fluxes from the surface fo
ICLOUD	1	Cloud effect to the optical depth in radiation
ICLOUD_CU	0	
TRACER_PBLMIX	1	Mix tracer fields consistent with PBL option
SCALAR_PBLMIX	0	Mix scalar fields consistent with PBL option
YSU_TOPDOWN_PBLMIX	0	Turns on top-down radiation-driven mixing (
GRAV_SETTLING	0	Gravitational settling of fog/cloud droplets (
DFI_OPT	0	Digital filter initialization (default 0 = none).
SIMULATION_INITIALIZATION_TYPE	REAL DATA CASE	
WEST-EAST_PATCH_START_UNSTAG	1	
WEST-EAST_PATCH_END_UNSTAG	51	
WEST-EAST_PATCH_START_STAG	1	
WEST-EAST_PATCH_END_STAG	52	
SOUTH-NORTH_PATCH_START_UNSTAG	1	
SOUTH-NORTH_PATCH_END_UNSTAG	51	

Table 7.1 – continued from previous

PARAMETER	VALUE	DESCRIPTION
SOUTH-NORTH_PATCH_START_STAG	1	
SOUTH-NORTH_PATCH_END_STAG	52	
BOTTOM-TOP_PATCH_START_UNSTAG	1	
BOTTOM-TOP_PATCH_END_UNSTAG	50	
BOTTOM-TOP_PATCH_START_STAG	1	
BOTTOM-TOP_PATCH_END_STAG	51	
GRID_ID	1	Domain identifier (can be 1, 2 or 3).
PARENT_ID	0	ID of the parent domain.
I_PARENT_START	1	The starting lower-left corner i-indice from t
J_PARENT_START	1	The starting lower-left corner j_indice from t
PARENT_GRID_RATIO	1	Parent-to-nest domain grid size ratio.
CEN_LAT	12.99997f	Latitude of the domain's center.
CEN_LON	-4.950012f	Longitude of the domain's center.
TRUELAT1	20.f	Projection parameter - true latitude 1.
TRUELAT2	0.f	Projection parameter - true latitude 2.
MOAD_CEN_LAT	12.99997f	Mother of all domains center latitude.
STAND_LON	5.f	Projection parameter - standard longitude.
POLE_LAT	90.f	The pole latitude.
POLE_LON	0.f	The pole longitude.
GMT	0.f	
JULYR	2018	
JULDAY	182	
MAP_PROJ	1	Map projection.
MAP_PROJ_CHAR	Lambert Conformal	Map projection.
MMINLU	MODIFIED_IGBP_MODIS_NOAH	Related to land use category.
NUM_LAND_CAT	21	Number of land categories in input data.
ISWATER	17	Related to land use category.
ISLAKE	21	Related to land use category.
ISICE	15	Related to land use category.
ISURBAN	13	Related to land use category.
ISOILWATER	14	Related to land use category.
HYBRID_OPT	-1	Option related to the hybrid vertical coordina
ETAC	0.f	Option related to the hybrid vertical coordina

EIGHT

LICENSING AND CITATION

8.1 License

These data are provided under a CC BY-NC 4.0 International license. Under this license, you are free to:

- **Share** - copy and redistribute the material in any medium or format

- **Adapt** - remix, transform, and build upon the material

The following terms apply:

- **Attribution** - You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

- **NonCommercial** - You may not use the material for commercial purposes.

Read more about the terms of this license at creativecommons.org.

8.2 How to cite these data

van de Giesen, N. (2021). The African Rainfall Project. http://africarain.ceg.tudelft.nl:9010/thredds/catalog.html. [DATE ACCESSED].

NINE

ATTRIBUTIONS

- Nick Van De Giesen, Delft University of Technology, Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft, Netherlands (Principal Investigator)
- Camille Le Coz, Delft University of Technology, Delft, Netherlands
- Lloyd A. Treinish, IBM Research USA, Yorktown Heights, NY, United States
- Qidi Yu
- Rick Hagenaars, Delft University of Technology, Data Analyst and Software Developer
- John S. Selker, Oregon State University, Professor of Biological and Ecological Engineering
- World Community Grid
- IBM Corporation
- The Weather Company