The High Asia Refined analysis (HAR)
User guide

HAR Version 1
User Guide Version V1.4 (31/10/2014)

History:
V 1.4, 31/10/2014, name changed (was "High Asia Reanalysis")
V 1.3, 23/06/2014, small typos, new contact list
V 1.2, 26/02/2014, updated references with final publication in J. Climate
V 1.1, 10/01/2014, fixed a bug where proj central lons and lats were inverted. Added full CF support
V 1.0, 05/12/2013, first published version

Author:
Fabien Maussion
Chair of Climatology
Technische Universität Berlin

Contact:
J. Curio: julia.curio@tu-berlin.de
D. Scherer: dieter.scherer@tu-berlin.de
F. Maussion: fabien.maussion@tu-berlin.de

When using the HAR dataset in your presentations/reports/manuscripts please refer to the HAR as following:

High Asia Refined analysis (HAR, Maussion et al., 2014)

Maussion, F., D. Scherer, T. Mölg, E. Collier, J. Curio, and R. Finkelnburg: Precipitation Seasonality and Variability over the Tibetan Plateau as Resolved by the High Asia Reanalysis. J. Climate, 27, 1910–1927, doi:10.1175/JCLI-D-13-00282.1, 2014.

Terms of use:

The HAR dataset is provided by the Chair of Climatology, TU Berlin. Users of the HAR dataset are expected to carefully read the User Guide and the corresponding publications (Maussion et al., 2011, 2014).

The HAR dataset is generated by dynamical downscaling of global analysis data (Final Analysis data from the Global Forecasting System; dataset ds083.2) using the Weather Research and Forecasting (WRF-ARW) model, run with the daily reinitialization strategy and without further assimilation of observations during the simulation time. It is therefore strongly dependent on (i) the quality of the global input data and (ii) the capacity of the WRF model to simulate the atmospheric processes. We recommend that users evaluate the dataset before using it in own research and applications.

For questions about the dataset or to discuss your results, please use the contact information provided above. Further HAR versions may be produced in the future if considered necessary to improve the quality of the dataset. Your feedback may help us in this process.

The Chair of Climatology of the TU Berlin and the HAR authors cannot be hold responsible for any misinterpretation or misuse of HAR data.

1. Introduction

This document is a reference guide for the users of the HAR dataset. It shortly summarizes the methods used to generate the HAR (Section 2), presents the file format and how to read the data (Section 3), the available variables (Section 4) and the key strengths and weaknesses of the HAR known to date (Section 5).

2. The HAR dataset

The HAR is generated with the advanced research version of the Weather Research and Forecasting model (WRF-ARW version 3.3.1). The dataset consists of consecutive reinitialized model runs of 36 h time integration. Each run starts at 12:00 UTC. The first twelve hours from each run are discarded as spin-up while the remaining 24 h of model output provide one day of the 12-year long time series. The model is forced with the Operational Model Global Tropospheric Analyses (Final Analyses, FNL; dataset ds083.2), which are available every six hours and have a spatial resolution of one degree. For the HAR we have not performed data assimilation on the regional level, as is the case with " regional reanalyses like the North American Reanalysis (NARR). To our knowledge, such a comprehensive and process-based dataset is currently unique for the TP, and therefore the HAR is intended to fill a gap where other regional reanalyses are not available. However it is not free of errors or shortcomings.

The model configuration used to produce HAR is summarized in Table 1 from Maussion et al. (2014). The surface temperature of lakes on the Plateau is obtained using the recently introduced inland water surface temperature initialization module. The land cover data is obtained form the WRF USGS dataset, but we updated the ice mask with the Randolph Glacier Inventory (RGI V1, Arendt et al. 2012) for glaciers.

Please refer to Section 2 from Maussion et al. (2014) for a comprehensive description of the HAR.

3. How to read the data

3.1 File format

The data are delivered as binary NetCDF files (Network Common Data Format):
http://www.unidata.ucar.edu/software/netcdf/

They are delivered in NetCDF Version 4 format (compressed). Yearly and monthly variables are also available in NetCDF Version 3 (uncompressed, therefore significantly larger in size).

Several free softwares can be used to explore the content of a file, for example:
- ncdump, ncview (linux packages)
- Panoply: http://www.giss.nasa.gov/tools/panoply/
- IDV: http://www.unidata.ucar.edu/software/idv/

A comprehensive list of the softwares and programming languages with NetCDF support can be found here:
http://www.unidata.ucar.edu/software/netcdf/software.html

3.2 File naming convention

There is **one** file per **variable** and per **year**, independently of the spatial or temporal resolution and of the variable type. The file names follow the convention:

har_spatial resolution_aggregation level_vertical resolution_variable name_year.nc

Examples:
har_d10km_d_3d_press_geopotential_2002.nc
har_d30km_y_2d_psfc_2010.nc
har_d30km_static_hgt.nc

Spatial resolution tags:
d30km, *d10km*

Aggregation level tags:
h (hourly, model output), *d* (daily means), *m* (monthly means), *y* (yearly means), *static* (time invariant fields)

Vertical resolution tags:
2d (surface variables), *3d_eta* (model eta levels), *3d_press* (pressure levels), *3d_soil* (soil levels)

The files are stored on the ftp server using the corresponding directory structure:
spatial resolution / *aggregation level* / *vertical resolution*

3.3 File Content

As an example, we reproduced the output from *ncdump -h* on a 2d variable file at the end of this document (Appendix A). This should be self-explanatory.

The time field is following the COARDS convention:
http://ferret.wrc.noaa.gov/noaa_coop/coop_cdf_profile.html

The data is available from October 2000 to December 2012. **The first months of year 2000 and the areas below topography in the 3d pressure level fields are replaced with floating Not-A-Number (NaN) values.**

3.4 Map projection

The HAR data is distributed in the original map projection used by the WRF model. The projection parameters are indicated by the global attributes:

PROJ_NAME, PROJ_CENTRAL_LON, PROJ_CENTRAL_LAT, PROJ_STANDARD_PAR1, PROJ_STANDARD_PAR2, PROJ_SEMIMAJOR_AXIS, PROJ_SEMIMINOR_AXIS, PROJ_FALSE_EASTING, PROJ_FALSE_NORTHING

The WRF projection ellipsoid is actually a sphere, therefore the A and B semi-axes are the same. **However, the longitudes and latitudes computed by transforming the grid-points eastings and northings are defined in the WGS-84 datum**. It is very important to remind that **no datum shift is needed when mapping data in WGS-84 to a WRF projection map**, but only a projection transformation.

The HAR GRID is regularly distributed on the WRF projection. The GRID parameters are indicated by the global attributes:

GRID_DX, GRID_DY , GRID_X00, GRID_Y00, GRID_X01, GRID_Y01, GRID_NX, GRID_NY

where [X00, Y00] are the coordinates of the lower left corner and [X01, Y01] the coordinates of the upper left corner.

For convenience, we also provide the longitudes and latitudes of each grid point as a single variable in each NetCDF file. They are not regular and correspond to curvilinear grid coordinates. Tools like CDO (https://code.zmaw.de/projects/cdo) can handle curvilinear grids correctly.

4. Variables

A list of the variables available on the server is provided at the end of this document (Appendix B). Currently, 3d_eta and hourly data are not available online. Contact us if you need additional data.

5. Validation: key results

In comparison to rain gauges on the Tibetan Plateau, standard statistics such as MAD and MD are good and consistent with similar studies using WRF. However, results indicate that TRMM is closer to observations than HAR on monthly sums. However, features such as orographic effects, precipitation statistics (frequency), or snowfall are better represented by HAR10 than by TRMM. HAR10 is closer to observations than HAR30. Precipitation seasonality is well represented (Maussion et al. 2014). In areas with very steep topography (e.g. Himalayas) the precipitation gradients can be unrealistic.

Temperature validation studies show that HAR correlates well with observations on a daily basis (single events are well caught) up to an hourly basis (daily cycle). No spatially consistent bias was found in summer, but a probable cold bias in winter on the TP could be detected (unpublished work).

Mölg et al. (2014) drove a surface energy balance glaciological model with HAR data only and could reach a similar accuracy as when driven with AWS observations. However, a scaling factor of approx. 0.65 had to be applied to correct for possible HAR precipitation errors and wind drift effects on snowfall.

Due to the use of final analysis data as input and the short period of time where HAR data is available, **trend calculations are not robust and should not be conducted with HAR**. The focus of HAR based analyses should be hold on decadal averages, inter-annual to intra-seasonal variability, and physical process understanding (see Mölg et al 2014, Kropacek et al 2012, Dietze et al 2014 for examples of use of HAR).

Dataset publications:

Maussion, F., D. Scherer, T. Mölg, E. Collier, J. Curio, and R. Finkelnburg (2014): Precipitation Seasonality and Variability over the Tibetan Plateau as Resolved by the High Asia Reanalysis. J. Climate, 27, 1910–1927, doi:10.1175/JCLI-D-13-00282.1, 2014

Maussion, F., Scherer, D., Finkelnburg, R., Richters, J., Yang, W., and Yao, T. (2011): WRF simulation of a precipitation event over the Tibetan Plateau, China - an assessment using remote sensing and ground observations, Hydrol. Earth Syst. Sci., 15, 1795-1817, doi:10.5194/hess-15-1795-2011, 2011.

Related publications:

Mölg, T., Maussion, F. and Scherer, D. (2014): Mid-latitude westerlies as a driver of glacier variability in monsoonal High Asia, Nature Climate Change, 4, 68-73, doi:10.1038/nclimate2055, 2014.

Kropacek, J., Maussion, F., Chen, F., Hoerz, S., and Hochschild, V. (2013): Analysis of ice phenology of lakes on the Tibetan Plateau from MODIS data, The Cryosphere, 7, 287-301, doi:10.5194/tc-7-287-2013, 2013.

Mölg, T., Maussion, F., Yang, W., and Scherer, D. (2012): The footprint of Asian monsoon dynamics in the mass and energy balance of a Tibetan glacier, The Cryosphere, 6, 1445-1461, doi:10.5194/tc-6-1445-2012, 2012.

Dietze, E., Maussion, F., Ahlborn, M., Diekmann, B., Hartmann, K., Henkel, K., Kasper, T., Lockot, G., Opitz, S. and Haberzettl, T. (2014): Sediment transport processes across the Tibetan Plateau inferred from robust grain-size end members in lake sediments, Clim. Past, 10(1), 91–106, doi:10.5194/cp-10-91-2014, 2014.

APPENDIX A: NetCDF file dump (2m temperature)

```
netcdf har_d10km_m_2d_t2_2010 {
dimensions:
        time = UNLIMITED ; // (12 currently)
        west_east = 270 ;
        south_north = 180 ;
variables:
        int time(time) ;
                time:long_name = "Time" ;
                time:units = "days since 2010-01-01 00:00:00" ;
        float west_east(west_east) ;
                west_east:long_name = "x-coordinate in Cartesian system" ;
                west_east:units = "m" ;
        float south_north(south_north) ;
                south_north:long_name = "y-coordinate in Cartesian system" ;
                south_north:units = "m" ;
        float lon(south_north, west_east) ;
                lon:long_name = "Longitude" ;
                lon:units = "degrees_east" ;
        float lat(south_north, west_east) ;
                lat:long_name = "Latitude" ;
                lat:units = "degrees_north" ;
        float t2(time, south_north, west_east) ;
                t2:long_name = "temp at 2 m" ;
                t2:units = "k" ;
                t2:agg_method = "mean" ;
                t2:coordinates = "lon lat" ;
// global attributes:
                :TITLE = "High Asia Refined analysis - HAR V1 - d10km" ;
                :DATA_NOTES = "File generated with the output of successive model runs of
36H (first 12 hours discarded for spin-up)" ;
                :WRF_VERSION = " OUTPUT FROM WRF V3.3.1 MODEL" ;
                :CREATED_BY = "Fabien Maussion - fabien.maussion@tu-berlin.de" ;
                :INSTITUTION = "Technische Universität Berlin, Institut für Ökologie,
Fachgebiet Klimatologie" ;
                :CREATION_DATE = "19.12.2012 03:50:18" ;
                :SOFTWARE_NOTES = "IDL V7.1.1, WAVE post V0.1" ;
                :VARNAME = "t2" ;
                :DOMAIN = "2" ;
                :NESTED = "YES" ;
                :TIME_ZONE = "UTC" ;
                :PRODUCT_LEVEL = "M" ;
                :LEVEL_INFO = "H: original simulation output; D: daily; M: monthly; Y:
yearly; S: static" ;
                :PROJ_NAME = "Lambert Conformal Conic" ;
                :PROJ_STANDARD_PAR1 = " 30.0" ;
                :PROJ_STANDARD_PAR2 = " 35.0" ;
                :PROJ_SEMIMAJOR_AXIS = " 6370000.0" ;
                :PROJ_SEMIMINOR_AXIS = " 6370000.0" ;
                :PROJ_FALSE_EASTING = " 0.0" ;
                :PROJ_FALSE_NORTHING = " 0.0" ;
                :PROJ_DATUM = "WGS-84" ;
                :PROJ_ENVI_STRING = "4, 6370000.0, 6370000.0, 30.0, 87.0, 0.0, 0.0, 30.0,
35.0, WGS-84, WRF Lambert Conformal" ;
                :GRID_INFO = "Grid spacing: GRID_DX and GRID_DY (unit: m), Down left
corner: GRID_X00 and GRID_Y00 (unit: m), Upper Left Corner: GRID_X01 and GRID_Y01 (unit:
m)" ;
                :GRID_DX = 10000.f ;
                :GRID_DY = 10000.f ;
                :GRID_X00 = -1434999.f ;
                :GRID_Y00 = -534999.9f ;
                :GRID_X01 = -1434999.f ;
                :GRID_Y01 = 1255000.f ;
                :GRID_NX = 270 ;
                :GRID_NY = 180 ;
                :PROJ_CENTRAL_LON = " 87.0" ;
                :PROJ_CENTRAL_LAT = " 30.0" ;
}
```

APPENDIX B: variables list

NAME	DESCRIPTION	UNIT	TYPE
cosalpha	local cosine of map rotation	-	static
hgt	terrain height	m	static
isltyp	dominant soil category	-	static
ivgtyp	dominant vegetation category	-	static
lu_index	usgs landuse dominant category	-	static
sinalpha	local sine of map rotation	-	static
albedo	albedo	-	2d
canwat	canopy water	kg m-2	2d
emiss	surface emissivity	-	2d
et	evapotranspiration (step-wize)	mm h-1	2d
graupel	grid scale graupel (step-wize)	mm h-1	2d
grdflx	ground heat flux	w m-2	2d
hail	grid scale hail (step-wize)	mm h-1	2d
hfx	upward heat flux at the surface	w m-2	2d
lh	latent heat flux at the surface	w m-2	2d
lwdown	downward long wave flux at ground surface	w m-2	2d
lwup	upward long wave flux at ground surface	w m-2	2d
netrad	net radiation at ground surface (+ = downward)	w m-2	2d
pblh	pbl height	m	2d
potevap	potential evaporation (step-wize)	w m-2	2d
prcp	total precipitation (step-wize)	mm h-1	2d
prcp_c	cumulus precipitation (step-wize)	mm h-1	2d
prcp_nc	grid scale precipitation (step-wize)	mm h-1	2d
psfc	sfc pressure	pa	2d
q2	qv at 2 m	kg kg-1	2d
qfx	upward moisture flux at the surface	kg m-2 s-1	2d
scld	total column clouds	-	2d
scldfra	Surface cloud fraction computed in a 50km radius FOV	-	2d
sfroff	surface runoff	mm	2d
slp	sea level pressure	hpa	2d
snow	snow water equivalent	kg m-2	2d
snowfall	grid scale snow and ice (step-wize)	mm h-1	2d
snowh	physical snow depth	m	2d
sst	sea surface temperature	k	2d
swdown	downward short wave flux at ground surface	w m-2	2d
swup	upward short wave flux at ground surface	w m-2	2d
t2	temp at 2 m	k	2d
tsk	surface skin temperature	k	2d
u10	u at 10 m	m s-1	2d
udroff	underground runoff	mm	2d
ust	u* in similarity theory	m s-1	2d
v10	v at 10 m	m s-1	2d
ws10	10 m wind speed	m s-1	2d
geopotential	full model geopotential on mass points	m2 s-2	3d_press
qliquid	liquid water mixing ratio	kg kg-1	3d_press
qsolid	solid water mixing ratio	kg kg-1	3d_press
qvapor	water vapor mixing ratio	kg kg-1	3d_press
theta	potential temperature (theta)	k	3d_press
u	x-wind component	m s-1	3d_press
v	y-wind component	m s-1	3d_press
w	z-wind component	m s-1	3d_press
ws	horizontal wind speed on mass grid points	m s-1	3d_press
sh2o	soil liquid water	m3 m-3	3d_soil
smcrel	relative soil moisture	-	3d_soil
smois	soil moisture	m3 m-3	3d_soil
tslb	soil temperature	k	3d_soil