GRRATS (Global River Radar Altimeter Time Series) Dataset Handbook

Introduction

The Global River Radar Altimeter Time Series (GRRATS) are river heights from OSTM/Jason-2 and Envisat that are conformed to look like river gauges via virtual stations (VS). The purpose of these heights are to provide satellite altimetric river height data in a form that is more recognizable to the observational community and as a way to get users use to using satellite data for river hydrology. GRRATS provides data from 914 VS on 39 of the world's largest rivers (wider than 900m). River heights were processed with limits established by DEM data from inside the VS. When applicable, times of ice cover are also flagged consistently. To allow for maximum usability, all processing data is included (original L2 data, filtering limits, etc.). When possible, data was validated with in situ gauges. Other locations were assigned a qualitative letter grade, based on the amount of missing data, agreement with nearby VS and identifiable seasonal cycle. Validation information (quantitative or qualitative) is packaged with each VS's data to aid the end user in selection the best time series for their particular task.

Potential virtual Stations and their locations

Virtual stations (locations sampled by the altimeters) were located by overlaying river centerlines from the GRWL dataset (citation) on Landsat imagery. The centerlines were filtered so that only those representing river widths 900m or more were remaining. These were then overlain with either the ENIVSAT or OSTM/ Jason-2 ground track. Once identified these locations were used to construct a mask for data extraction.

The data extraction masks were drawn on Landsat imagery from a list compiled for the RivWidth project (Pavelsky and Smith, 2008). These images were selected because they were temporally representative of median flow and had very little in the way of visual obstruction. Polygons were then drawn using an unlimited vertices format to allow for the best fitting shape within the median flow edge limits of the river. The polygons were drawn to a 2 km extent perpendicular to the ground track to allow for deviation from nominal ground track which satellites experience (Beckley et al., 2013). Each VS location was also assigned a flow distance based on the river center lines from the Hydro1k dataset (EROS, 1996).

Data extraction from the GDR (JASON-2 GDR-D, ENVISAT GDR v2.1)

The mask was then used to extract all of the data from a given altimeter ground track, whose (nominal) footprint center point fell within the polygon for a particular virtual station. DEM data from SRTM, ASTER, and GMTED2010, pertaining to the area inside the mask are also collected at this point. Altimeter data was extracted from the GDR using the following flags that were constructed using guidelines from the Jason-2, and ENVISAT Handbooks (Dumont et al., 2009; Batoula Soussi and ESA). Jason-2 data can be accessed at avisoftp.cnes.fr. ENVISAT data can be accessed at ESA(European Space Agency) Earth Online: http://earth.esa.int.

A. Editing criteria of the JASON-2 GDR-D data

A1. Conditions of 1-Hz data parameters

Flag orbit_state_flag_rest = 3 (adjusted preliminary/precise orbit)

Range correction model_dry_tropo_corr is available

Range correction model_wet_tropo_corr is available

Range correction iono_corr_gim_ku is available

Range correction solid_earth_tide is available

Range correction pole_tide is available

A2. Conditions of 20-Hz data parameters

Altitude alt_20hz is available

Ku band altimeter range (ice retracking) ice_range_20hz_ku is available

Flag ice_qual_flag_20hz_ku = 0 (good, Ku band ice retracking quality)

Ku band backscatter coefficient (ice retracking) ice_sig0_20hz_ku is available and not negative (in dB)

A3. The number of 20-Hz data points to be compressed to 1-Hz in a selected pass segment has to be 2 or more.

A4. The time span of each pass segment should not be longer than 1.5 seconds (this limit was removed in processing some segments exceptionally large virtual Stations).

B. Editing criteria of the ENVISAT GDR, Version 2.1 or later

B1. Conditions of 1-Hz data parameters

Bits 30-31 of meas_conf_data_flags = 00 (adjusted preliminary/precise DORIS orbit)

Range correction mod_dry_tropo_corr is available

Range correction mod_wet_tropo_corr is available

Ku band range correction of GIM ionosphere (not in Ofline GDR) is available

Range correction solid_earth_tide_ht is available

Range correction geocen_pole_tide_ht is available

B2. Conditions of 18-Hz data parameters

Altitude hz18_diff_1Hz_alt is available

Flag bit of map_18hz_ku_trk_flags = 0 (valid, Ku band tracker range validity)

Ku band altimeter range hz18_ku_ice1 is available (ice1 retracking)

Flag bit of ku_ice1_retrk_qua_flags = 0 (good, Ku band ice1 retracking quality)

Ku band backscatter coefficient hz18_ku_ice1_bscat is available and not negative (in dB, ice1 retracking)

B3. The number of 18-Hz data points to be compressed to 1-Hz in a selected pass segment has to be 2 or more.

B4. The time span of each pass segment should not be longer than 1.5 seconds (this limit was removed in processing some segments exceptionally large virtual Stations).

Processing of mask filtered GDR data

Altimeter data, having been sorted by location, associated with relevant DEM data and assigned a flow distance were then processed to produce the time series in the dataset. Each location was processed as a member of its particular river/altimeter combination. The shape file was used as an index for processing each river/ altimeter group.

The DEM data for each location were used to establish a base line height used in generating height limits for filtering the altimeter returns at that location. Though data were collected from SRTM, GMTED and ASTER for each location when available, only one set was used for filter construction. It was selected by availability of SRTM, GMTED and ASTER in that preferential order. The median DEM return for the area inside each was used as a baseline for that virtual station. These baseline heights were then processed with their flow distances using a linear optimization algorithm to force the baseline heights to be lower in the direction of the mouth of the river. The final height filter was constructed from these constrained baseline heights.

The height filter limits used in processing of the dataset are 15 m above or 10 m below the constrained baseline height. These limits were established after careful consideration of over 150 USGS stations on watersheds of 20,000 km^2 in size. They represent reasonable limits in the USGS data for flood and base flow conditions for rivers of this scale (Steve Tuozzolo, personal communication, June 29, 2015). Altimeter returns that remained were filtered again to remove erroneous returns on the low end of that data. Any heights that fell below 2 m below the 5^{th} percentile of height were removed. This removed only .03% of data but created a more realistic decay in height approaching base flow. After height filtering, any virtual station that was missing 50% or more altimeter cycles were removed as they were not viable for constructing a time series. Rivers that experience seasonal freeze up were also processed to indicate times where the river was frozen.

Though frozen river heights are a part of the unprocessed GDR data product, they do not appear in the time series. River freeze up were determined from USGS data when available. A USGS station near the mouth of the river was selected as breakup occurs last at the mouth of a river. When Freeze and thaw data were not directly available (Russia), broad limits were established using dates found in the literature.

Validation

On rivers where in-situ gauges were available, each VS was compared with each gauge through relative height comparison (mean difference removed). The maximum and average Nash-Sutcliffe efficiency, minimum and average standard deviation of error (STDE), and maximum R (correlation coefficient) were calculated and can be found in the root of each VS file. Because not all rivers were able to be validated, a qualitative system was also put in place.

VS on rivers that had no gauges were assigned a qualitative letter grade (A-D) to guide the end user in choosing the most likely VS to be accurate. Letter grades were assigned based on data comprehensiveness (percentage of data post filtering), the presence of seasonal osculation, and relative comparison with nearby VS. The qualitative letter grades can be found in the file root in place of validation statistics.

Data packaging and Variable identification

Sample Altimetry Data (NetCDF format)

Format: netcdf4 Title = 'Altimetry Data for virtual station Yukon_Jason2_0'

Global Variables:

Variable	Dimension	Datatype	Units	Name
lon	X	double	degrees east	longitude
lat	Y	double	degrees north	latitude
ID	root	char	-	Reference VS ID
sat	root	char	-	satellite
Flow_Dist	distance	double	km	Distance from river mouth
rate	root	double	Hz	sampling rate
pass	root	int32	-	pass number
nse	grade	double	-	Max Nash Sutcliffe efficiency
nse AVG	grade	double	-	Average Nash Sutcliffe efficiency
R	grade	double	-	Correlation coefficient
std	grade	double	m	Minimum standard deviation of error
stdAVG	grade	double	m	Average standard deviation of error
grade	grade	char	-	qualitative letter grade

The global variables are: longitude and latitude of the center of the virtual station, the virtual station ID, the satellite name, flow distance, sampling rate , the satellite pass number and a suite of fit statistics, or a qualitative letter grade. Qualitative letter grades were assigned based on amount of data points, seasonal pattern, and similarity to nearby VS. This was done, only when validation data was unavailable. When validation was possible, the VS was evaluated with all gauges on the river through relative height comparison. Maximum Nash-Sutcliffe Efficiency (NSE), Average NSE, maximum R(correlation coefficient), minimum standard deviation of error (STDE), and average STDE are reported.

Groups:

/Unprocessed GDR Data[/				
Variable	Dimension	Datatype	Units	Name
lon	X	double	degrees east	Longitude
lat	Y	double	degrees north	latitude
h	Z	double	meters above EGM2008 geoid	Unprocessed heights
sig0	UGDR	double	dB	Sigma0
pk	UGDR	double	unknown	peakiness
cycle	UGDR	int32	unknown	Altimeter cycle
time	T	double	Days since Jan-1-1900 00:00:00	
heightfilter	UGDR	int32	-flag-	Good heights flag
icefilter	UGDR	int32	-flag-	No ice flag
allfilter	UGDR	int32	-flag-	Ice free heights that passed height filter

This includes the data from each return: lon and lat, the height of the water level in meters, the signal strength, sigma0, in decibels, a 'peakiness' value, the cycle number, the time of the return, and filter flags that signal 1 for data that should

be included and 0 for data that should be excluded. The flags are for a height filter, an ice filter, and the logical intersection of the two (allfilter), with 1 denoting returns that pass through the filter and 0 denoting returns that do not.

/Timeseries/				
Variable	Dimension	Datatype	Units	Name
time	T	double	Days since Jan-1-1900 00:00:00	time
cycle	TS	int32	-	Altimeter cycle
hbar	Z	double	meters above EGM2008 geoid	average height
hwbar	Z	double	meters above EGM2008 geoid	weighted average height
sig0bar	time	double	dB	average sigma0
pkbar	time	double	-	Average peakiness

These are pass-averaged values, having gone through the filter. There are two values that flag data: -9999 for data that is missing from the GDR, and -9998 for data that is missing because of height/ice filters. These flags are only present when none of the values to be averaged can be found. The other values give average height (hbar), in meters, and sigma-0 weighted height using

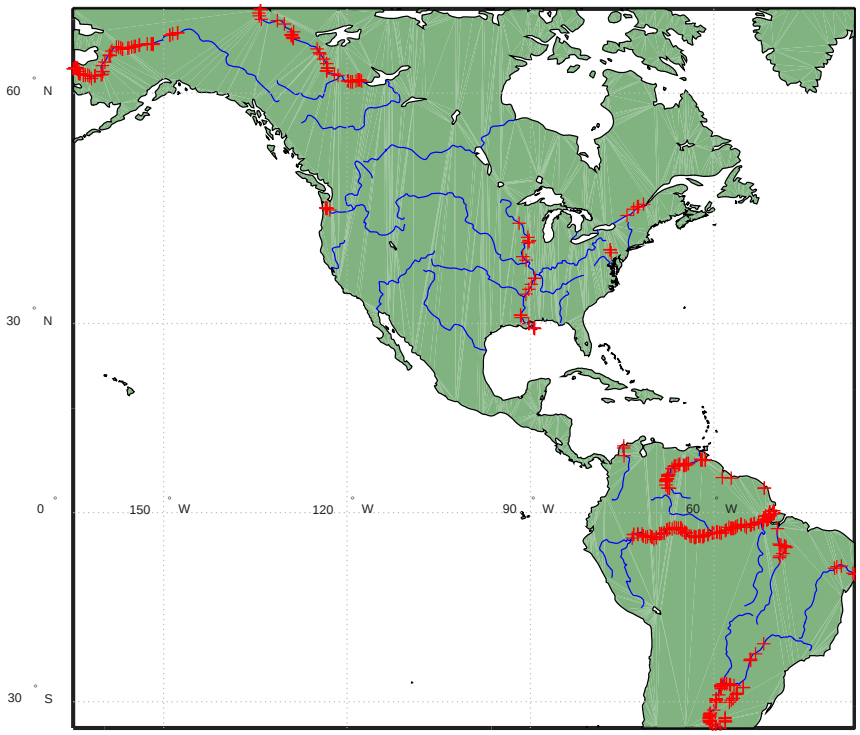
/Sampling/				
Variable	Dimension	Datatype	Units	Name
scene	scene	char	-	Landsat Scene ID
lonbox	X	double	degrees east	Longitude box extent
latbox	Y	double	degrees north	Latitude box extent
island	scene	int32	-flag-	Island flag

This is the data from the polygons, including the Landsat scene ID used to draw the polygons. The island flag is used when islands are visible inside the polygon in the imagery when drawing the mask.

/Filter/				
Variable	Dimension	Datatype	Units	Name
nNODATA	-	int32	count	Number of cycles without data
riverh	Z	double	meters above EGM2008 geoid	River elevation from filter file
maxh	Z	double	meters above EGM2008 geoid	Max elevation allowed by filter
minh	Z	double	meters above EGM2008 geoid	Min elevation allowed by filter
icethaw	T	double	Days since Jan-1-1900 00:00:00	Thaw dates for river
icefreeze	T	double	Days since Jan-1-1900 00:00:00	Freeze dates for river
DEMused	DEM	Char	-	DEM used in height filter

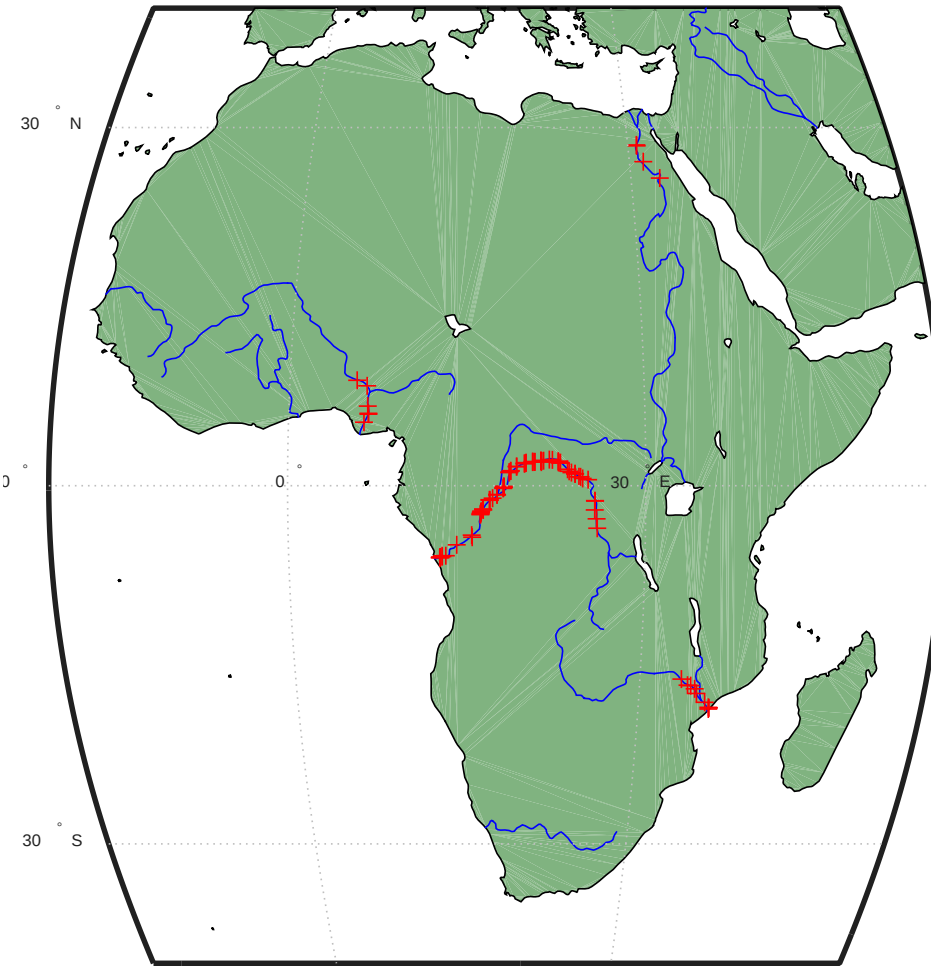
This is the filter data; nNODATA gives the number of cycles that have no data because of a lack of data in the GDR and/or data that is filtered out. riverh gives the river elevation extracted from a 30 arc-second DEM of the region. This is used for the height filter. maxh and minh are the upper and lower bounds of river heights included in the filtered data; we set a +15m, -10m from the DEM river elevation as a first pass, and then removed any data that was 5m below the 5th percentile of river stage heights. icethaw and icefreeze are the thaw and freeze dates, respectively, for the years included in the altimetry dataset. DEM used refers to the DEM that the basline height was taken from.

Coverage
The Americas



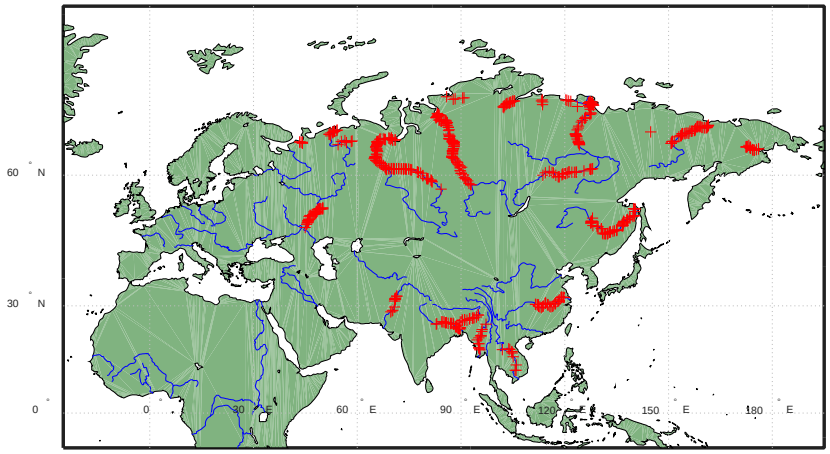
Location	Number of Virtual Stations	
	Envisat	Jason-2
North America		
Columbia	28	10
Mackenzie	13	6
Mississippi	4	2
St Lawrence	2	2
Susquehanna	34	21
Yukon	3	3
South America		
Amazon	92	23
Courantyne	1	0
Essequibo	1	0
Magdalena	2	1
Oiapoque	1	1
Orinoco	30	8
Parana	28	21
Sao Francisco	4	1
Tocantins	7	2
Uruguay	10	4
Totals:	**260**	**105**

Africa



Location	Number of Virtual Stations	
	Envisat	Jason-2
Africa		
Congo	63	12
Niger	5	1
Nile	3	1
Zambezi	5	4
Totals:	**76**	**18**

Eurasia



Location	Number of Virtual Stations	
Eurasia	**Envisat**	**Jason-2**
Amur	47	5
Anabar	5	
Anadyr	10	8
Ayeyarwada	11	3
Brahmaputra	16	3
Ganges	16	5
Indigirka	2	
Indus	12	1
Khatanga	16	
Kolyma	40	3
Kuloy		3
Lena	64	19
Mekong	11	
Menzen	1	4
Ob	71	35
Olenyok	6	
Pechora	15	2
Pyasina	6	
Volga	24	12
yangtze	28	7
Yenisei	70	34
Total:	**471**	**144**

References

Batoula Soussi, C., Urien, S., Soulat, F. Muir, A., Roca, M., Cotton, D., P. Féménias (Editor), 2011, ENVISAT ALTIMETRY Level 2 User Manual, Version 1.4.

Beckley, B., Ray, R., Holmes, S., Zelensky, N., Lemoine, F., Yang, X., Brown, S., Desai, S., Mitchum, G., and Hausman, J., 2013, Integrated multi-mission ocean altimeter data for climate research TOPEX/Poseidon, Jason-1, and OSTM/Jason-2 user's handbook, PO: DAAC, CA, USA. doi, v. 10, p. 5067.

Dumont, J., Rosmorduc, V., Picot, N., Desai, S., Bonekamp, H., Figa, J., Lillibridge, J., and Scharroo, R., 2009, OSTM/Jason-2 products handbook: CNES: SALP-MU-M-OP-15815-CN, EUMETSAT: EUM/OPS-JAS/MAN/08/0041, JPL: OSTM-29-1237, NOAA/NESDIS: Polar Series/OSTM J, v. 400, p. 1.

EROS, U., 1996, HYDRO1k Elevation Derivative Database: http://eros. usgs. gov/#/FindData: Products andDataAvailable/gtopo30/hydro (last accessed: 30 September 2012).

Pavelsky, T. M., and L. C. Smith (2008), RivWidth: A software tool for the calculation of river widths from remotely sensed imagery, *Geoscience and Remote Sensing Letters, IEEE*, *5*(1), 70–73.