R2R Technical Report
Data Processing Report
R2R GeoCSV Navigation

Version: 1.0

Persistent URL: http://get.rvdata.us/product/r2rnav

Contact

R2R Program, info@rvdata.us

Authors

Christopher Olson, Geological Data Center, Scripps Institution of Oceanography, US

About R2R

The Rolling Deck to Repository (R2R) Program works with the US academic research fleet community to routinely document, assess, and preserve the underway sensor data from oceanographic research vessels. For more information, see the R2R website at http://www.rvdata.us/.

Acknowledgements

R2R is funded by the National Science Foundation (OCE 1447797), with additional support from the Office of Naval Research (ONR), National Oceanographic and Atmospheric Administration (NOAA), and Schmidt Ocean Institute, and assistance from the University-National Oceanographic Laboratory System (UNOLS). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF or any other agency/contributor.

Overview of Workflow

A suite of navigation data products is produced by R2R for every cruise that has a suitable navigation dataset. Since many sources of navigation exist, R2R works with technicians for each vessel to identify the instrument that produces the most reliable navigation and treats this as the 'primary' navigation source. On cruises where the primary navigation dataset is significantly incomplete or not parseable, a secondary source of navigation is quality assessed and used to create products.

In general, the navigation data is converted to a standard format, a set of quality assessment checks is performed, bad points are flagged in a quality control step, and products are created from the result. The code used to convert, perform quality tests, and create products is part of a public repository hosted on GitHub.

Reformatting

Metadata is collected about the cruise, including the start and end times and locations. A list of raw files that fit within the temporal bounds of the cruise is created. These files are parsed and converted into a common format.

Quality Assessment

A series of tests are performed on the converted data in an attempt to determine the quality of the navigation data. Specific results of the tests differ in format – some may be simple true/false checks, others may have a specific numeric value – but each test is assigned an associate green, yellow, or red value. Test results are recorded in an xml report that is distributed with the set of products. Detailed information on the specific test performed is listed below in the "Description of Processing" section.

Quality Control

After quality assessment, a quality control operation is performed in an attempt to weed out bad data. Data found to be bad are never discarded, but flagged to indicate that they should not be used.

Sampling and Creating Products

Three navigation products are created from the quality-controlled version of the data. The first product is the full resolution data (i.e. at the recorded sample interval, usually 1Hz). We refer to this version as 'bestres.' The second product is made by sampling the full resolution data at a one minute interval. The last product is a control point version of the data. The Ramer-Douglas-Peucker algorithm is used to create a minimum number of points to represent the navigation as a trackline on a map.

Description of Processing

Quality Control

Data points are flagged for one of four reasons:

- Points with bad GPS quality indicator
- Points contributing to unreasonable velocity
- Points contributing to unreasonable acceleration
- Points that are out of temporal sequence

Thresholds for unreasonable velocities and unreasonable accelerations are ship specific. They are determined by the maximum capabilities of the ship. Accelerations and velocities above these thresholds are flagged. To find the exact values used for each of these values, check the xml report for the associated data product.

XML Report

An XML QA report summarizing the tests performed, the result values of each test (if applicable) and the resulting green/yellow/red grades for each test. The XML QA report also contains a comprehensive set of cruise and fileset level metadata for the source metadata used to produce the product set.The XML QA report is available on the R2R QA dashboard as well as through the cruise catalog and web services.

Description of Product Set

Best Resolution Navigation File

The original converted navigation resolution is preserved and output in a 'bestres' file and flagged according to the quality control criteria laid out in the previous section. This file is saved in ascii GeoCSV format which has important metadata in the header including digital object identifiers (dois) linking to the source dataset, cruise, and repository. Best resolution navigation files follow the naming convention <cruise_id>_bestres.geoCSV.

Example R2R GeoCSV Navigation file:

```
#dataset: GeoCSV 2.0
#title: Processed Trackline Navigation Data: Best Resolution
#field_unit:
ISO_8601,degree_east,degree_north,(unitless),(unitless),(unitless),meter
,meter\second,degree
#field_type:
datetime,float,float,integer,integer,float,float,float,float
#field_standard_name:
iso_time,ship_longitude,ship_latitude,nmea_quality,nsv,hdop,antenna_heig
ht,speed_made_good,course_made_good
#field_long_name: date and time,longitude of vessel,latitude of
vessel,NMEA quality indicator,number of satellite vehicles
observed,horizontal dilution of precision,height of antenna above mean
sea level,course made good,speed made good
#standard_name_cv: http://www.rvdata.us/voc/fieldname
#ellipsoid: WGS-84 (EPSG:4326)
#delimiter: ,
#field_missing: NAN
#attribution: Rolling Deck to Repository (R2R) Program;
```

```
http://www.rvdata.us/
#source_repository: doi:10.17616/R39C8D
#source_event: doi:10.7284/907234
#source_dataset: doi:10.7284/124390
#cruise_id: RR1709
#creation_date: 2017-05-30T17:24:48Z
iso_time,ship_longitude,ship_latitude,nmea_quality,nsv,hdop,antenna_heig
ht,speed_made_good,course_made_good
2017-04-25T23:59:59Z,144.663270,13.428553,2,8,1.4,24,0.00,0.000
2017-04-26T00:00:00Z,144.663270,13.428553,2,8,1.4,24,0.00,270.000
2017-04-26T00:00:01Z,144.663268,13.428553,2,8,1.4,24,0.22,90.000
```

One Minute Sampled Navigation File

The Best Resolution Navigation File is then downsampled to 60 second resolution. The sampling is done simply by iterating through the best resolution file and sampling the first valid record every . Flagged values are ignored during downsampling, therefore there are no flagged points in this product. These files are also in the GeoCSV format and follow the naming convention <cruise_id>_1min.geoCSV.

Abstracted Navigation File

The One Minute Sampled Navigation File is then run through a Ramer-Douglas-Peucker algorithm in order to reduce the navigation to the minimum number of points needed to plot a trackline on a map. This downsampled version is referred to as the ‘control’ file. These files are also in the GeoCSV format and follow the naming convention <cruise_id>_control.geoCSV.

XML Report

An XML DP report summarizing the computational methods used for flagging questionable data values in the final data product is created. This document also includes the source metadata used to produce the product set and a link to the related XML QA Report. The XML DP report is available through the cruise catalog and web services.

References

GitHub code repository: https://github.com/rvdata/NavManager
GeoCSV Format Specification: http://get.rvdata.us/format/100157/
R2R Navigation Quality Assessment:
http://get.rvdata.us/qa_docs/Navigation/R2R_Navigation_QualityAssessment_Description.pdf

Revision History

Version 1.1
Effective processing date span: start of program to present date
Default URL:
http://get.rvdata.us/product/r2rnav/R2R_Navigation_DataProcessing_Description.pdf
Versioned snapshot:
http://get.rvdata.us/product/r2rnav/R2R_Navigation_DataProcessing_Description_20181020.pdf
Older Versions:
http://get.rvdata.us/product/r2rnav/R2R_Navigation_DataProcessing_Description_20180406.pdf