Quality Assurance of Ground Motion Data: Using the IRIS MUSTANG Analytics System

By Tim Ahern¹, Rob Casey¹, Jonathan Callahan², Gillian Sharer¹, Laura Keyson¹, Mary Templeton¹

> ¹ IRIS DMC, Seattle, WA USA ² Mazama Science, Seattle, WA USA

NSF Grant EAR-1261681 - (SAGE)

CTBTO S&T 2017 Conference



The IRIS Archive 444 terabytes growing at 75 terabytes per year 22,587,020 station day files

IRIS DMC Archive as of 1 June 2017 443.6 terabytes



CTBTO SnT2017

CTBTO S&T 2017 Conference

Extremely Effective Data Distribution Mechanisms



1180.07 terabytes projected on May 31, 2017



Characteristics of the IRIS DMC

- Data from 5 decades
 - 1970 present
- IRIS DMC has data from 34,019 stations in its archive
- 3,500 stations in real time as of 15 June 2017
- IRIS DMC has a staff of 23 people
- Assessing quality of this amount of data had to be automated



The Modular Utility for STAtistical Knowledge Gathering (MUSTANG) System

- System that computes metrics related to data quality for "most" channels at the IRIS DMC
- Automated recalculation of metrics when needed for changes in



- Relevant metadata,
- time series version, or
- Changes in algorithms used to compute quality metrics

MUSTANG Metrics - 46 metrics

• State of Health (15)

- amplifier_saturation
- calibration_signal
- clock_locked
- digital_filter_charging
- digitizer_clipping
- event_begin
- event_end
- event_in_progress
- glitches
- missing_padded_data
- spikes
- suspct_time_tg
- telemetry_sync_error
- timing_correction
- Timing_quality

• Data Archiving Metrics (5)

- Real Time Data Latency
- Real Time Feed Latency
- Real Time Total Latency
- Percent Available Per Day
- Station Completeness
 BTO SnT2017

- Simple Data Statistics (6)
 - Max Sample Value
 - Min Sample Value
 - Mean Sample Value
 - Median Sample Value
 - Sample RMS Variance
 - Signal to Noise Ratio

• Data Continuity Metrics (5)

- Channel Up Time
- Max Gap Size
- Max Overlap Size
- Number of Gaps
- Number of Overlaps
- Signal Anomalies (6)
 - Cross Talk
 - DC Offset
 - Dead Channel
 - Max STA/LTA
 - Number of Spikes
 - Pressure Effects

CTBTO S&T 2017 Conference

Noise Analysis (5)

- Power Spectral Density **
- Probability Density Function **
- PDF Mode time series**
- Percent Above HNM
- Percent Below LNM

• Metadata Checks (4)

- Orientation Check
- Polarity Check *
- Timing Drift *
- Transfer Function *

* In test** Special web service

MUSTANG Metrics are Exposed by Web Services

http://service.iris.edu/mustang/measurements/1/query?metric=num_gaps&net =IU&sta=ANMO&loc=00&cha=BH1&format=text&timewindow=2017-06-01T03:15:09,2017-06-16T03:15:09&nodata=404

"Num Gaps Metric"'value", "target", "start", "end", "lddate" "0", "IU.ANMO.00.BH1.M", "2017/06/02 00:00:00", "2017/06/02 23:59:59", "2017/06/04 08:53:53" "0", "IU.ANMO.00.BH1.M", "2017/06/03 00:00:00", "2017/06/03 23:59:59", "2017/06/05 08:23:20" "0", "IU.ANMO.00.BH1.M", "2017/06/04 00:00:00", "2017/06/04 23:59:59", "2017/06/06 08:07:12" "0", "IU.ANMO.00.BH1.M", "2017/06/05 00:00:00", "2017/06/05 23:59:59", "2017/06/07 08:17:35" "0", "IU.ANMO.00.BH1.M", "2017/06/06 00:00:00", "2017/06/06 23:59:59", "2017/06/08 09:08:27" "0", "IU.ANMO.00.BH1.M", "2017/06/07 00:00:00", "2017/06/07 23:59:59", "2017/06/08 09:08:27" "0", "IU.ANMO.00.BH1.M", "2017/06/07 00:00:00", "2017/06/07 23:59:59", "2017/06/09 09:31:10" "0", "IU.ANMO.00.BH1.M", "2017/06/08 00:00:00", "2017/06/07 23:59:59", "2017/06/10 08:55:39" "0", "IU.ANMO.00.BH1.M", "2017/06/09 00:00:00", "2017/06/09 23:59:59", "2017/06/11 09:27:48" "0", "IU.ANMO.00.BH1.M", "2017/06/10 00:00:00", "2017/06/10 23:59:59", "2017/06/12 09:18:55" "0", "IU.ANMO.00.BH1.M", "2017/06/11 00:00:00", "2017/06/11 23:59:59", "2017/06/12 09:18:55" "0", "IU.ANMO.00.BH1.M", "2017/06/11 00:00:00", "2017/06/11 23:59:59", "2017/06/13 09:22:08" "0", "IU.ANMO.00.BH1.M", "2017/06/12 00:00:00", "2017/06/11 23:59:59", "2017/06/13 09:22:08" "0", "IU.ANMO.00.BH1.M", "2017/06/12 00:00:00", "2017/06/12 23:59:59", "2017/06/13 09:22:08" "0", "IU.ANMO.00.BH1.M", "2017/06/13 00:00:00", "2017/06/13 23:59:59", "2017/06/14 09:18:59" "0", "IU.ANMO.00.BH1.M", "2017/06/13 00:00:00", "2017/06/13 23:59:59", "2017/06/14 09:18:59"

MUSTANG Clients

- LASSO a station-monitor style web client to view metrics
- MUSTANG Databrowser a plotting web tool for metrics
- MUSTANGular a mapping tool for station metrics
- PDF Gallery a web tool for exploring noise analyses









CTBTO SnT2017

CTBTO S&T 2017 Conference

Research Ready Data Sets (RRDS)

- RRDS is a filter for timeseries data selection that is based upon quality control criteria provider by a user.
- RRDS uses MUSTANG metrics as the QA value that is compared with a user's criteria
- For example: If a user has an algorithm that is gap intolerant
 - RRDS will return only time series that have no gaps
- Numerous criteria can be specified in one service call
- RRDS enables culling of unwanted data on the data center side

CTBTO S&T 2017 Conference

An RRDS Example

http://service.iris.edu/iriswsbeta/rrds/1/query?net=IM &sta=NV31&start=2016-08-22&end=2016-08-24 &nodata=404&sample_rms_gt=100

• Nodata=404

IM NV31 -- BHN 2016-08-22T00:00 2016-08-24T00:00:00
 Applied Metric(s): sample_rms > 100

Log: The request contained these research-ready parameters + sample_rms_gt : 100

The following records were rejected

- IM NV31 BHE 2016-08-23T00:00:00 2016-08-24T00:00:00 sample_rms=94.378
- IM NV31 BHZ 2016-08-23T00:00:00 2016-08-24T00:00:00 sample_rms=94.082

Summary

- The MUSTANG System is a fully functional and automated system
- MUSTANG identifies all problems that affect any metric
 - New metrics can be added to identify additional types of problems as needed
- When RRDS is completed, users will be able to request data with quality characteristics that meet their scientific needs
- IRIS has released a portable version of the MUSTANG metrics, IRIS System for Portable Assessment of Quality (ISPAQ) that could be incorporated at other data centers