

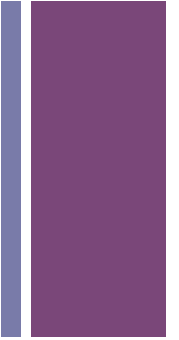


Assessing Data Quality

Dr. Mary Templeton
IRIS Data Management Center

+ What is “good” data quality?

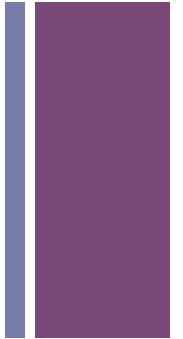
- For time series?
- For metadata?





MUSTANG:

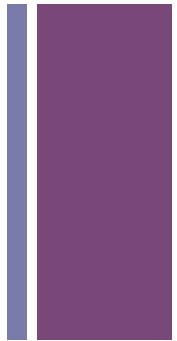
Modular Utility for Statistical Knowledge Gathering



- New IRIS automated data quality assurance system
- Assesses data in the IRIS SEED archive
- Recalculates measurements when data or algorithms change (in development)
- **Web service** design - query MUSTANG using URLs
- Easy to add new metrics
- Can include measurements from other institutions



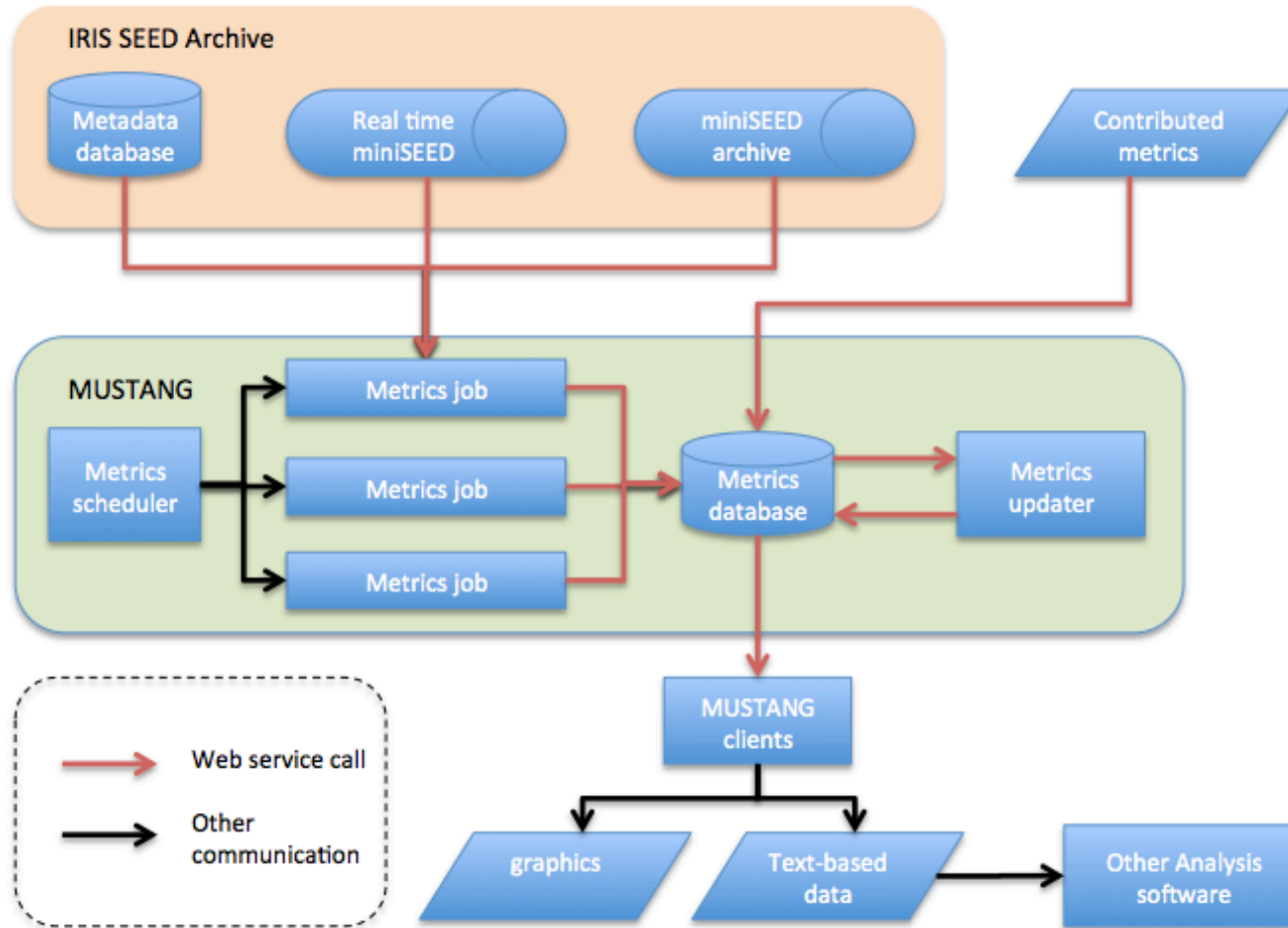
MUSTANG now serves 45 metrics



State of health flags	Data continuity	Time series amplitude statistics	Noise analysis
amplifier_saturation	max_gap	sample_max	noise-psd
calibration_signal	max_overlap	sample_mean	noise-pdf
clock_locked	num_gaps	sample_median	noise-mode-timeseries
digital_filter_charging	num_overlaps	sample_min	pct_above_nhnm
event_begin		sample_rms	pct_below_nlnm
event_end		sample_snr	
event_in_progress			
glitches	Data transmission and archiving	Signal anomalies	Metadata accuracy
missing_padded_data	percent_availability	cross_talk	m2_tides*
spikes	channel_uptime	dc_offset	orientation_check
suspect_time_tag	station_completeness	dead_channel_exp	polarity_check
telemetry_sync_error	data_latency	max_stalta	timing_drift
timing_correction	feed_latency	num_spikes	transfer_function
timing_quality	total_latency	asl_coherence	

* in development

+ The MUSTANG system





Where do I find MUSTANG?

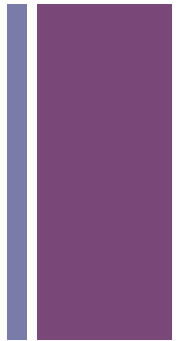
<http://service.iris.edu/mustang/>

IRIS DMC Web Services

Services Implementation: MUSTANG

Request tools

Service interface	Version	Summary	Return options
measurements	v.1	The main MUSTANG web service returning measurements for metrics relating to station data quality.	<ul style="list-style-type: none">• XML (default)• text• CSV• JSON• JSONP
noise-psd	v.1	Returns Power Spectral Density estimates of seismic data and can generate aggregate plots.	<ul style="list-style-type: none">• Text – CSV• XML• Plot (PNG)
noise-pdf	v.1	Returns Probability Density Functions in frequency 'bins' and can generate aggregate plots.	<ul style="list-style-type: none">• Text – CSV• XML• Plot (PNG)
noise-mode-timeseries	v.1	Returns PDF Mode Timelines at select frequencies and can generate plots.	<ul style="list-style-type: none">• Text – CSV• XML• Plot (PNG)
metrics	v.1	The metrics web service returns a description of available metrics in a variety of formats	<ul style="list-style-type: none">• XML• HTML• XSD• JSON• JSONP
targets	v.1	The targets web service returns a list of stations and channels for a given metric.	<ul style="list-style-type: none">• Text



+ Channel Orientations

WebServices Home / MUSTANG / Measurements / Docs / v. 1 / Builder

URL Builder: measurements v.1

Service interface URL builder Help Revisions

Use this form to build a URL to the **measurements** web service. Notice that as you edit the form, the link is automatically updated.

Usage

Metric: Orientation Check Metric

Targets

SNCLQ filter or Target? Filter Target

Network: C

Station: GO08

Location:

Channel:

Format

Format: CSV

Temporal Constraints

Parameter: Time window Start/end

Start time: 2013-01-01T00:00:00

End Time: 2015-01-01T00:00:00

Value Constraints

Parameter:

Condition: None

Value:

Sorting

Both fields must be defined for sorting to be applied.

Parameter: start

Order by: Ascending

Click the link:

http://service.iris.edu/mustang/measurements/1/query?metric=orientation_check&net=C&sta=GO08&format=csv&timewindow=2013-01-01T00:00:00,2015-01-01T00:00:00&orderby=start_asc

■ URL Builder (client) example

- Text boxes provide parameters for the URL at the bottom
- This URL returns a CSV file that Excel can read

More URL parameters described here

+ Channel Orientations



- **orientation_check** finds observed channel orientations for shallow $M \geq 7$ events by
 - Calculating the Hilbert transform of the Z component ($H\{Z\}$) for Rayleigh waves
 - Cross-correlating $H\{Z\}$ with trial radial components calculated at varying azimuths until the correlation coefficient is maximized
 - The **observed channel orientation** is difference between the calculated event back azimuth and observed radial azimuth

Stachnik, J.C., Sheehan, A.F., Zietlow, D.W., Yang, Z, Collins, J. and Ferris, A, 2012, *Determination of New Zealand Ocean Bottom Seismometer Orientation via Rayleigh-Wave Polarization*, Seismological Research Letters, v. 83, no. 4, p 704-712.



Channel Orientation Analysis

orientation_check measurements from 2013 and 2014 for C.GO03 having correlation coefficients > 0.4.

target	start	magnitude	azimuth_Y_obs	azimuth_Y_meta	max_Czr	Yobs-Ymeta	mean	2sd	median95%
C.GO03..BHZ.M	11/25/13 6:40	7.0	284.08	0.00	0.60	-75.92	-1.04	50.53	-1.33
C.GO03..BHZ.M	11/15/14 3:40	7.1	336.06	0.00	0.88	-23.94			
C.GO03..BHZ.M	9/24/13 12:34	7.7	337.38	0.00	0.83	-22.62			
C.GO03..BHZ.M	11/17/13 9:21	7.7	344.20	0.00	0.97	-15.80			
C.GO02..BHZ.M	11/17/13 9:22	7.7	344.66	0.00	0.98	-15.34			
C.GO03..BHZ.M	4/1/14 23:50	8.2	348.36	0.00	0.96	-11.64			
C.GO03..BHZ.M	4/19/14 14:25	7.5	349.88	0.00	0.89	-10.12			
C.GO03..BHZ.M	10/9/14 2:30	7.1	352.64	0.00	0.88	-7.36			
C.GO03..BHZ.M	6/23/14 21:50	7.9	352.65	0.00	0.86	-7.35			
C.GO03..BHZ.M	7/15/13 14:24	7.3	355.56	0.00	0.93	-4.44			
C.GO03..BHZ.M	10/14/14 4:11	7.3	358.29	0.00	0.97	-1.71			
C.GO02..BHZ.M	4/12/14 21:08	7.6	358.67	0.00	0.98	-1.33			
C.GO03..BHZ.M	11/1/14 19:41	7.1	359.51	0.00	0.77	-0.49			
C.GO03..BHZ.M	1/5/13 9:44	7.5	359.58	0.00	0.91	-0.42			
C.GO03..BHZ.M	4/6/13 5:46	7.0	3.35	0.00	0.83	3.35			
C.GO03..BHZ.M	5/23/13 18:01	7.4	3.97	0.00	0.88	3.97			
C.GO03..BHZ.M	4/19/13 4:11	7.2	5.09	0.00	0.70	5.09			
C.GO03..BHZ.M	8/30/13 17:20	7.0	8.72	0.00	0.95	8.72			
C.GO03..BHZ.M	10/25/13 18:19	7.1	12.90	0.00	0.69	12.90			
C.GO03..BHZ.M	9/25/13 16:48	7.1	15.40	0.00	0.98	15.40			
C.GO03..BHZ.M	4/3/14 2:46	7.7	22.01	0.00	0.96	22.01			
C.GO03..BHZ.M	10/15/13 1:25	7.1	47.25	0.00	0.65	47.25			
C.GO03..BHZ.M	4/16/13 11:47	7.7	55.83	0.00	0.58	55.83			

This median observed Y azimuth differed from the metadata by -1.33 degrees

This value was omitted from the median because it exceeded two standard deviations

A discrepancy with the C.GO08 metadata orientation was found and reported using this metric.

GRO-Chile Channel Orientation Analysis Results			
target	median95%		
C.GO01..BHZ.M	7.09		
C.GO02..BHZ.M	-2.28		
C.GO03..BHZ.M	-1.33		
C.GO04..BHZ.M	-1.28		
C.GO05..BHZ.M	-3.06		
C.GO06..BHZ.M	-7.16		
C.GO07..BHZ.M	-8.00		
C.GO08..BHZ.M	23.05		metadata is within 10 degrees
C.GO09..BHZ.M	3.34		metadata differs by 10-20 degrees
C.GO10..BHZ.M	3.21		metadata differs by >20 degrees



Monitoring Mass Position

Latest Assessment of Seismic Station Observations (LASSO)



Basic **Advanced** About

View groups of MUSTANG metrics designed to examine specific aspects of a seismic station's performance. Groupings focus on seismometer component channels and are available by Virtual Network.

Virtual Network: Global Reporting Geophysical Observatories in Chile

View: Metric quality assessment groupings

Metric(s):

Location(s): Specify seismometer location code(s) per station

Ranking: Qualitative Quantitative

Table Type: Snapshot Mean for Period Median for Period

Time: Display metrics closest to this date and time

Show Counts:

Disclaimer: Please note that metrics are still being developed for MUSTANG and back-populated for the entire IRIS archive. Most metrics for many major permanent networks are complete for the last several years, and gaps in metric coverage are being filled. Metric and network coverage reports are periodically updated. Channel labeling conventions are available through IRIS Data Services and Appendix A of the SEED manual.

Mass Positions view of _GRO-CHILE for 2015-05-16T00:00:00 requested at Mon May 18 2015 14:19:45 GMT-0600 (MDT):

Show entries Search:

Target	Rank	m1	m2	m3
C.G002...M	<input type="button" value="Data"/>	00,00,03	1.20	1.58
C.G003...M	<input type="button" value="Data"/>	01,00,02	1.30	0.70
C.G004...M	<input type="button" value="Data"/>	03,00,00	0.10	0.57
C.G007...M	<input type="button" value="Data"/>	02,00,01	-0.40	-0.42
C.G010...M	<input type="button" value="Data"/>	00,01,02	-1.00	-1.00

Showing 1 to 5 of 5 entries

Previous Next

- LASSO (client) example
 - <http://lasso.iris.edu>
 - Extends MUSTANG by
 - converting mass position counts into volts using metadata
 - Displaying “virtual networks”

m1 Rules

- bad >= fix(-Infinity)
- fair >= fix(-0.9)
- good >= fix(-0.75)
- fair >= fix(0.75)
- bad >= fix(0.9)

+ GPS Clock Locking

MUSTANG databrowser

IRIS DMC

Plot Type and Metric

Plot Type

Network Boxplots

Metric

Daily Flag Count: Clock locked

Plot Options for Boxplots

outliers

[Box Plot Explainer](#)

Data Source and Timespan

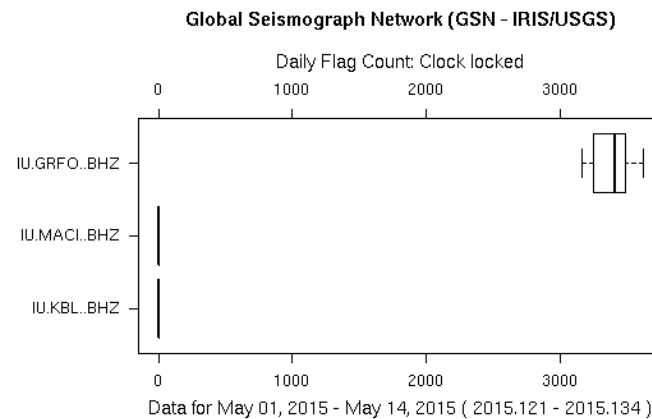
Network - Station - Location - Channel

IU KBL -- BHZ

Time Span

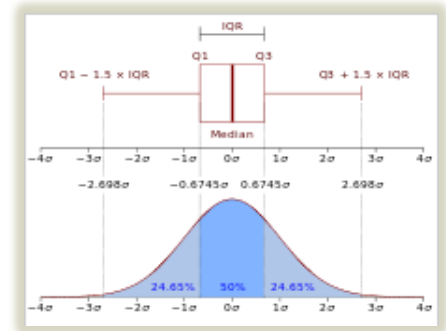
Start: 2015-05-01

End: 2015-05-15



Access Speed (sec): data load=3.27, plot=0.03, R total=3.30

[data URL](#)



■ Databrowser (client) example

■ <http://ds.iris.edu/mustang/databrowser/>

■ Extends MUSTANG by

■ Providing a variety of plot types

■ Plotting related channels (or metrics) as a group



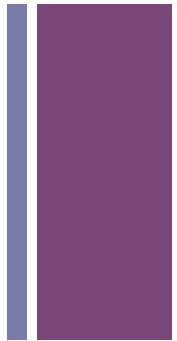
Dead channels

■ Scripting (client) example

- Quickly retrieves one month of measurements for the IU network
- Summarizes potential data problems based on metrics thresholds

```
dead: dead_channel_exp < 0.3 && pct_below_nlnm > 20
dead      IU.TRQA.10.BH1.M      2015-04-22      2015-04-30      9
dead      IU.TRQA.10.BH2.M      2015-04-12      2015-04-19      8
dead      IU.TRQA.10.BH2.M      2015-04-22      2015-04-30      9
dead      IU.TRQA.10.BHZ.M      2015-04-12      2015-04-19      8
dead      IU.TRQA.10.BHZ.M      2015-04-22      2015-04-30      9
dead      IU.WCI.00.BHZ.M      2015-04-03      2015-04-03      1
dead      IU.WCI.00.BHZ.M      2015-04-07      2015-04-30      24
```

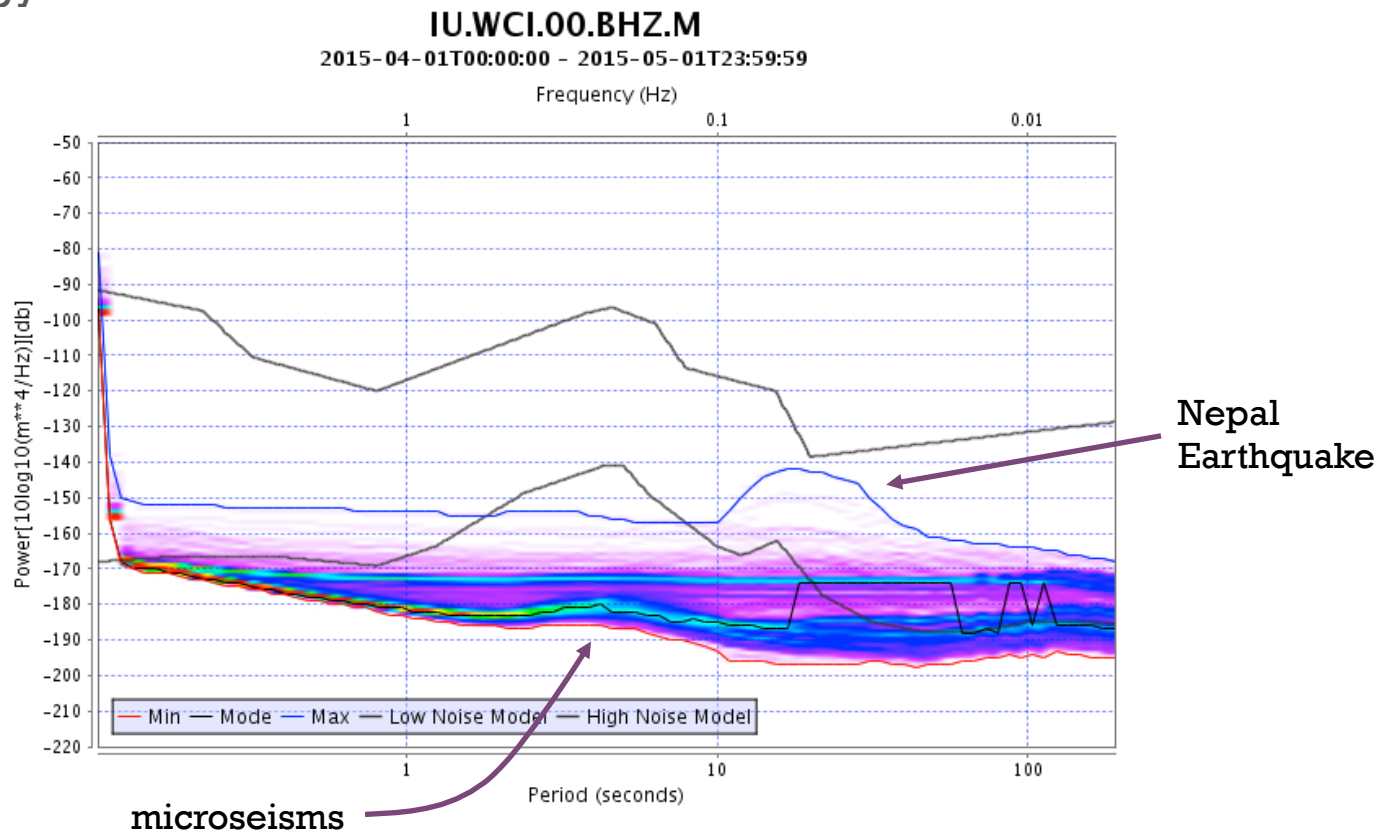
- Analysts can focus on verifying potential problems – what else can we learn about IU.WCI?



+ Dead channels

■ MUSTANG pdf-noise service

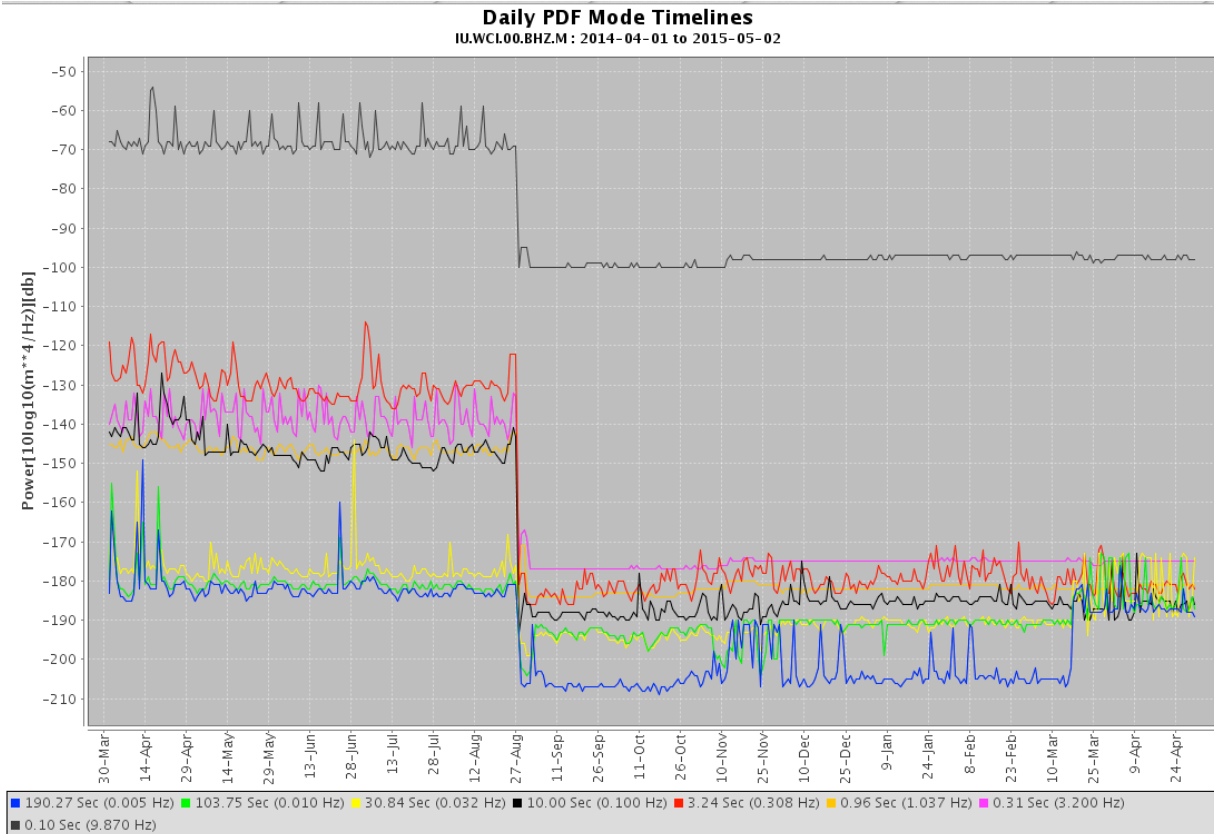
- IU.WCI.00.BHZ isn't completely dead – it still records some energy





Dead channels

- **MUSTANG noise-mode-timeseries service**
 - The problem started on 2014/08/27.





Would you like to learn more about



- Data quality assessment?
- How you can use MUSTANG and its clients?
- Data quality of stations archived at IRIS?

If so, I invite you to talk to me during the week and/or attend the Thursday afternoon breakout session.

+

Thank you –

I look forward to talking with you.

