Exploring MUSTANG Metrics with LASSO

This handout describes one approach for assessing the data quality of the GRO (Global Reporting Observatories) Chile network (right) during May 2015 using MUSTANG metrics and the LASSO client.   The GRO stations belong to the Chilean National Seismic Network (their network code is "C") and their names begin with GO (gee-oh).

In a web browser, enter the URL
    http://lasso.iris.edu/
and select the **Advanced** tab. We will create custom queries to assess data completeness, amplitude health and signal anomalies. To view **completeness** metrics, select the following:

Metrics: channel_uptime, max_gap, num_gaps,
    percent_availability
Specify: By Virtual Network
Virtual Network: _GRO-CHILE
Locations: *
Channels: BH*
Quality: M
Ranking: Qualitative
Table Type: Mean for Period
Start Time: 2015-05-01 00:00:00
End time: 2015-05-26 00:00:00

Select **Get Measurements**.

What is the **channel_up_time** value associated with 100 **percent_availability**?

Sort by **percent_availability** (click the metric name at the top of its column). Which stations have no data?

Which channels have partial data? Are they characterized by many small gaps? A few large gaps?

For channels with the most gaps, is this common or just one bad day? (click a value to plot it)

Next, we'll select a set of metrics that tell us about **amplitude health**. Keeping the rest of the values the same, select

> **Metrics**: dead_channel_exp, pct_below_nlnm, pct_above_nhnm, sample_mean, sample_rms

Select **Get Measurements**.

To the right is a Probability Density Function (PDF) plot of power spectral densities (PSDs) for station GO03 in May 2015 – a graph of many hour-long power spectra with the highest overlap density in blue. The grey lines above and below are the New High Noise Model (NHNM) and New Low Noise Model (NLNM) – the envelope of healthy global digital network signals prior to 1993.

C.GO03.‑‑.BHZ.M
2015-05-01T00:00:00 – 2015-05-18T00:00:00
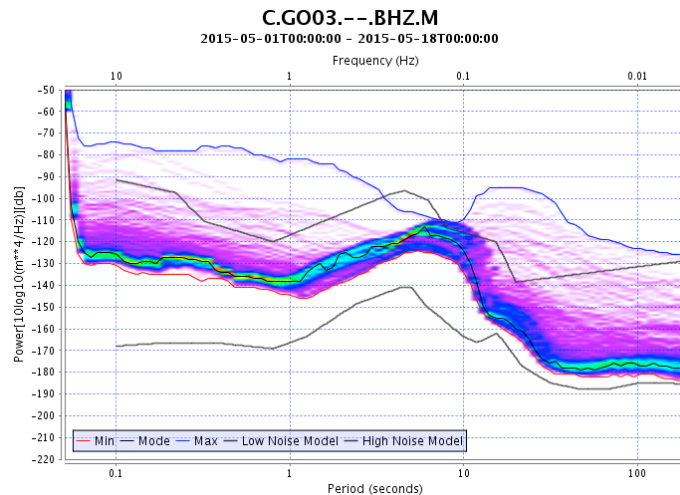
Three metrics (**dead_channel_exp**, **pct_below_nlnm**, **pct_above_nhnm**) are derived from these plots. Their values are not instrument dependent because instrument responses have been removed in PSD data.

Sort by **dead_channel_exp**. This metric fits an exponential curve (the shape of a dead channel on a log(period)-power/dB graph) to the daily PSD mean curve and reports the residual of the errors to this fit. Values <0.3 indicate that the channel may be "dead".

Are there any potentially dead channels among the _GRO-CHILE stations?

pct_below_nlnm also helps identify dead channels – when used in combination with dead_channel_exp, it can indicate the difference between a channel recording no energy and one that is recording strong noise. The PDF graph above is divided into bins. This metric is the percentage of the total bins that lie below the NLNM and contained energy that day. A value >20 percent typically indicates a problem for BH or HH channels.

Sort by **pct_below_nlnm**. What is the largest value for _GRO-CHILE?

pct_above_nhnm gives the percentage of total bins that have energy and lie above the NHNM curve for the day. BH channels with noise issues will have values >30 percent fairly consistently.

Sort by **pct_above_nhnm**. What channels have the largest average values? Plot the values for these channels. Based on the graphs, the station map and the size of the values, do you believe these noise levels to indicate a hardware problem?

The **sample_mean** metric records the daily amplitude mean and is usually a good indicator of off-center masses when mass position data are unavailable. Though it is in units of digitizer counts, a mass against its stops will have values near the maximum absolute count value for the datalogger (near 10E+7 for most Q330s).

Sort by **sample_mean**. Which channels have the largest absolute values for _GRO-CHILE? Do theses masses appear to be near their stops?

sample_rms is a measure of how much amplitude variation a channel has in a day. These values are also in raw digitizer counts. Dead channels will have consistently low values and noisy channels will have consistently high values.

Which channels have the highest average **sample_rms** values? How well do they correlate with high **pct_above_nhnm** values? Click these values to view the graph of daily values. What do you believe to be the cause of the high average values?

Finally, we'll select a set of metrics that tell us about **signal anomalies** that are potentially hardware-related. Keeping the rest of the values the same, select

> **Metrics**: cross_talk, dc_offset, num_spikes

Select **Get Measurements**.

The **cross_talk** metric cross-correlates orthogonal channels from the same instrument to check for signal leakage. Values that are consistently near 1.0 should be investigated. Occasional values near 1.0 may be caused by mass drift, recentering or calibrations.

Sort by **cross_talk**. Are there any _GRO-CHILE channels that should be investigated for cross-talk?

dc_offset looks for long-term shifts in the **sample_mean** metric. It can potentially catch losses of differential output or the sensor. Large values (>50?) should be investigated.

Click the ? button for **dc_offset** – what does the **Uses** section say that this metric also catches? What _GRO-CHILE channel has the largest **dc_offset** average? Approximately, what is this channel's largest daily value?

This approach can be used with any network in the IRIS SEED archive that has MUSTANG metrics available. (We are still in the process of generating initial metrics for many broadband channels.) The Resources on the back of this page list some other tools to help with assessing the quality of data archived at IRIS.

Data Quality Assurance Resources at IRIS

MUSTANG (http://service.iris.edu/mustang/): MUSTANG home page with links to each MUSTANG service, as well as to clients and documentation. The noise-pdf service is especially useful for viewing station noise levels, signal health and instrument response accuracy.

LASSO (http://lasso.iris.edu/): the LASSO client for viewing MUSTANG metrics. There are links to a User Guide and Tutorial at the bottom of the page.

MUSTANG Databrowser (http://ds.iris.edu/mustang/databrowser/): a visualization client for MUSTANG metrics that offers a variety of plotting options. These include network and station boxplots, plotting of related metrics, trace plotting and PDF display. The list of available metrics is partial – Databrowser does not yet include the most recent metrics.

IRIS Quality Assurance (http://ds.iris.edu/ds/nodes/dmc/quality-assurance/): This home page for the IRIS data quality analysis team includes links to tutorials on Getting Started with MUSTANG and interpreting Waveforms and their Power Spectral Density Expressions, as well as other useful products and resources.

IRIS Web Services (http://service.iris.edu/): This home page offers links to web services that return a variety of information. These are particularly useful for data quality assurance:

- station – metadata content
- dataselect – time series in miniSEED format
- timeseries – plots of time series that can optionally be filtered or otherwise processed before display. Timeseries can downloaded or plotted without downloading.
- evalresp – displays instrument response information

Metadata Aggregator (http://ds.iris.edu/mda/): Viewing of metadata for the IRIS SEED archive.

JPlotResp (http://ds.iris.edu/pub/programs/JPlotResp/JPlotResp_v1.77/): A very useful program for viewing and assessing the quality of instrument responses.

MUSTANG QA email list (http://ds.iris.edu/mailman/listinfo/mustang-qa/): Community email list for discussion and questions about MUSTANG and its use for data quality assurance.

IRIS QA Group email (dmc_qa@iris.washington.edu): The best email address for contacting the IRIS Quality Assurance group about the quality of data in the IRIS SEED archive.