

Gillian Sharer, **Mary E. Templeton***, Laura Keyson, and Jerry Carter
Incorporated Research Institutions for Seismology (IRIS): Data Services

S21H-0631

As part of the IRIS Data Service’s commitment to characterizing and optimizing the quality of data within its archive, we provide over 40 automated data quality metrics to our user community through MUSTANG web services (services.iris.edu/mustang/). Metric measurements include amplitude statistics, data latency, completeness, power spectral density-based noise characterization and more. We have recently expanded MUSTANG and related visualization tools to include

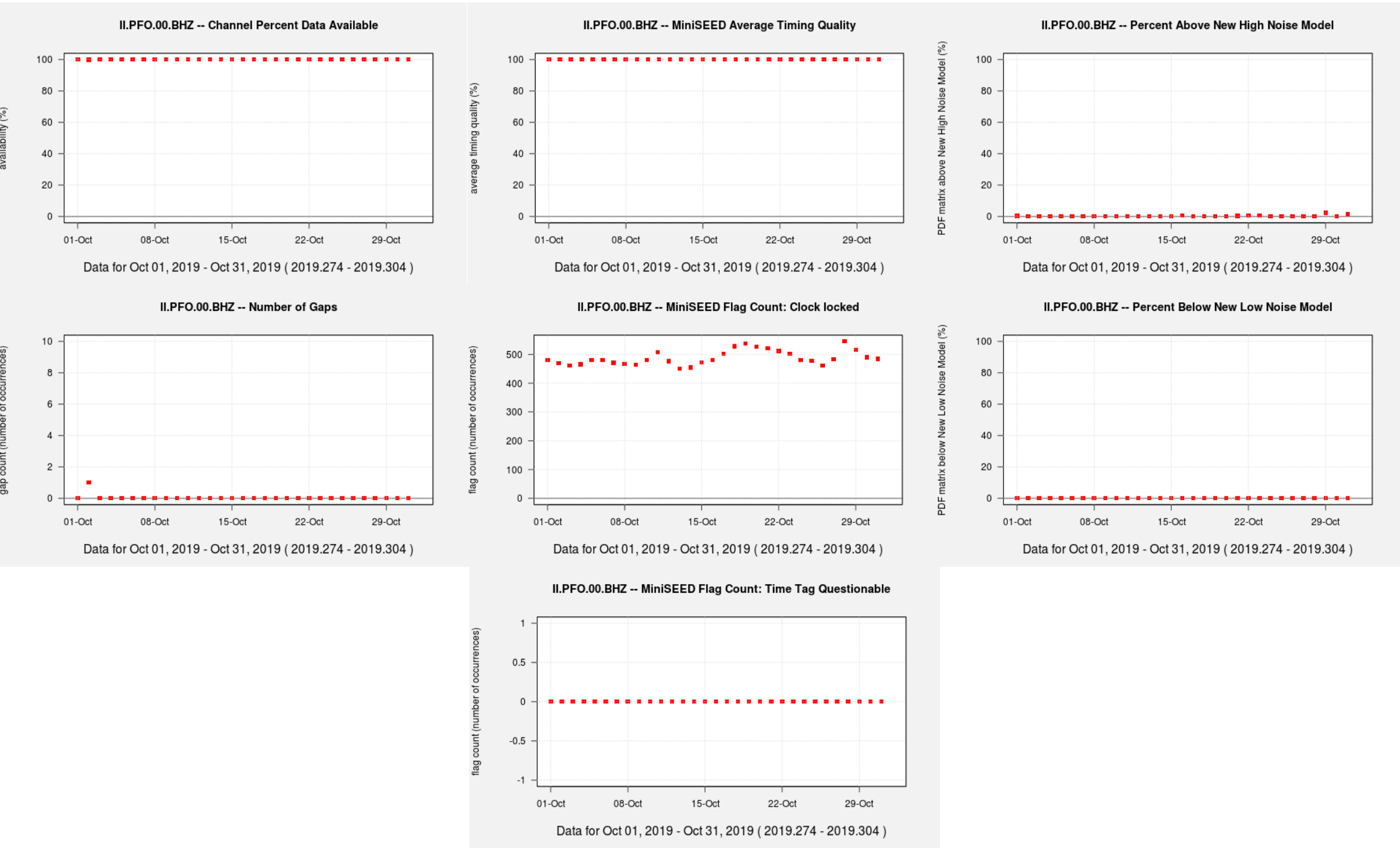
- Gap duration plotting in MUSTANG Databrowser
- QuARG – a utility in development for analyzing, tracking and reporting network data quality issues (available in 2020)
- PDF daily mode spectrograms viewable in an array of color palettes
- Measurements for data stored in PH5 format – in progress
- Improvements to ISPAQ – a Python utility for generating metrics for local data or any data available through FDSN web services.

Here we showcase how to use these enhancements to answer common data quality questions.

Is this channel healthy?

Channels are likely to be healthy if they

- are complete and continuous (*percent_availability*=100; *num_gaps*=0),
- have good timing (*clock_locked*>0; *timing_quality*>>0%; *suspect_time_tag*=0),
- record seismic energy (*pct_below_nlnm*<20), and
- have low noise levels (*pct_above_nhn*<20).



Plots from **MUSTANG Databrowser**, a web-based visualization client (ds.iris.edu/mustang/databrowser/), show that but for one gap, II.PFO.00.BHZ fit these criteria during October 2019. We plan to incorporate these plots into a single view in 2020.

Tip

The **RRDS** (**R**esearch-**R**eady **D**atasets: service.iris.edu/irisws/rrds/1/) web service allows you to retrieve only data complying with metric constraints that you choose. To retrieve gap-free data having amplitudes consistent with the presence of seismic signal:

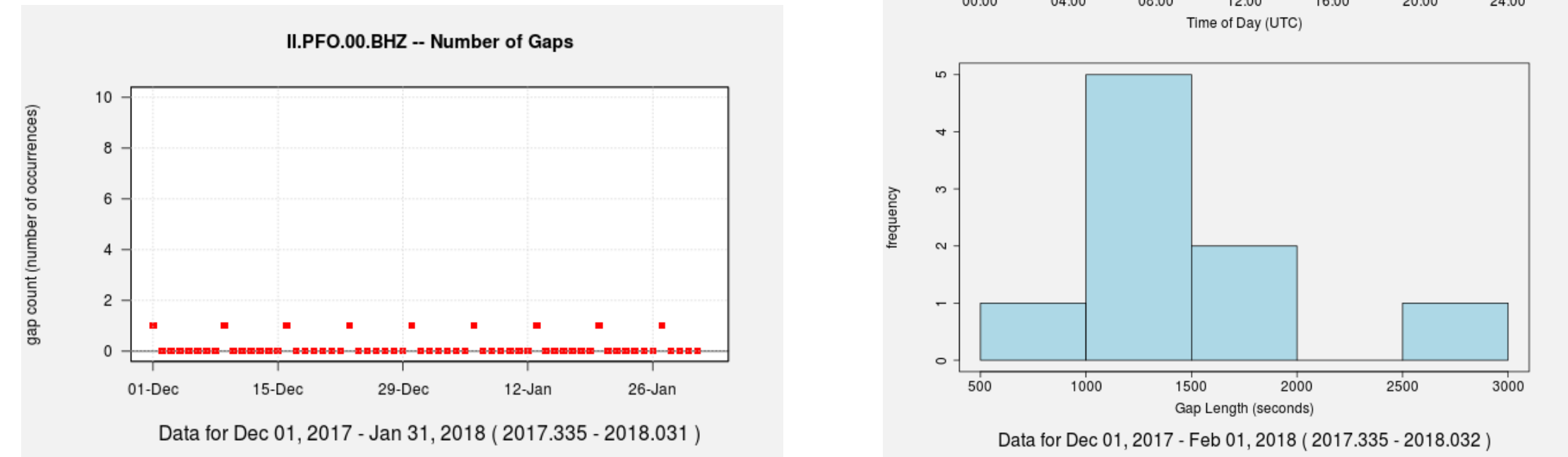
- build an **rrds** query using the URL builder found at the link above
- append metric constraints: `&num_gaps_eq=0&pct_below_nlnm_lt=20`
- send this query by web browser or **curl** to receive a list of suitable channels
- send this list using **FetchData** to receive your data. (**FetchData** is available at seiscode.iris.washington.edu/projects/ws-fetch-scripts/.)

My station has gaps – what can I learn about this?

In late 2017 and early 2018, station II.PFO experienced a weekly data gap during which both dataloggers at the site rebooted.

Plotting gap duration by date, time of day and frequency of occurrence using **MUSTANG Databrowser** shows that gaps with similar durations had occurred each Saturday night.

Discussion revealed that a computer security contractor was probing the PFO dataloggers, causing them to reboot. An additional firewall solved the problem.



For Updates on these Projects

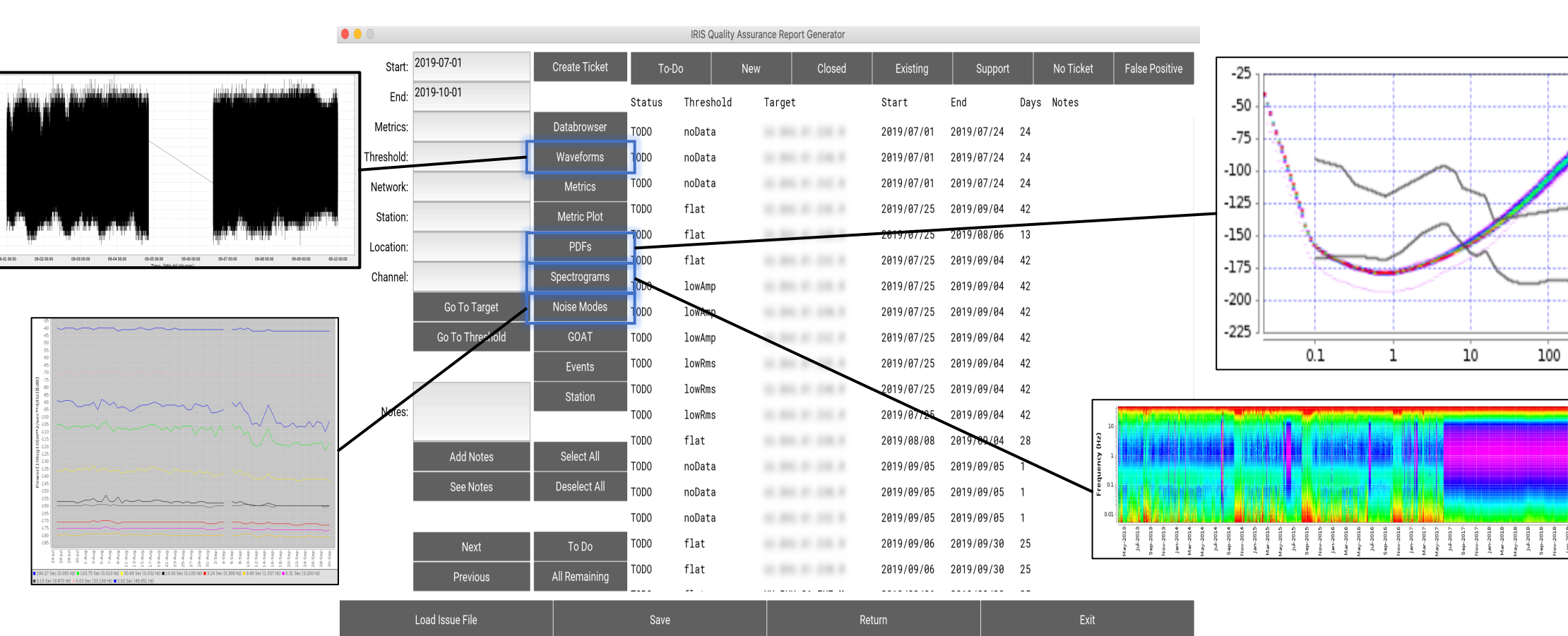
see **News** at ds.iris.edu/ds/nodes/dmc/quality-assurance/.

Which stations in my network have quality issues?

We are developing a new utility called **QuARG** (the **Q**uality **A**ssurance **R**eport **G**enerator) that queries MUSTANG metrics and guides the user through the steps of discovering, examining, tracking and reporting quality issues. It is particularly helpful for examining potential issues by providing an easy way to utilize several quality assurance tools that IRIS provides.

The end result is a **Quality Assurance Report** about problematic data in an easy-to-read format.

Once released, QuARG will be available at github.com/iris-edu.

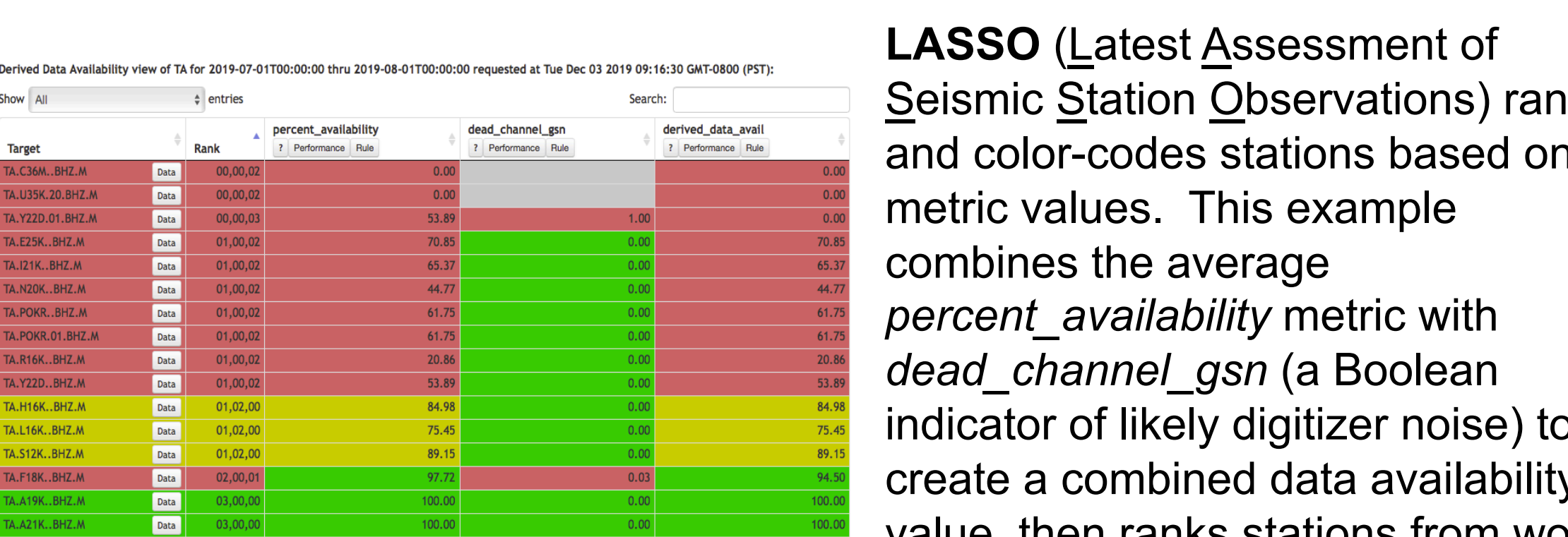


What other tools are available to quickly identify stations in my network that have anomalous metric values?

IRIS has two additional visualization tools that can summarize MUSTANG metrics over a range of time: MUSTANGular and LASSO.

MUSTANGular displays a color-coded summary of station metric values on a map so that you can quickly find problem stations. This example identifies TA and AK network stations that have good to poor average data availability in July 2019.

- Coming in 2020:
- combine all channel-locations from a station into a single value
 - prioritize location code
 - improvements to binning



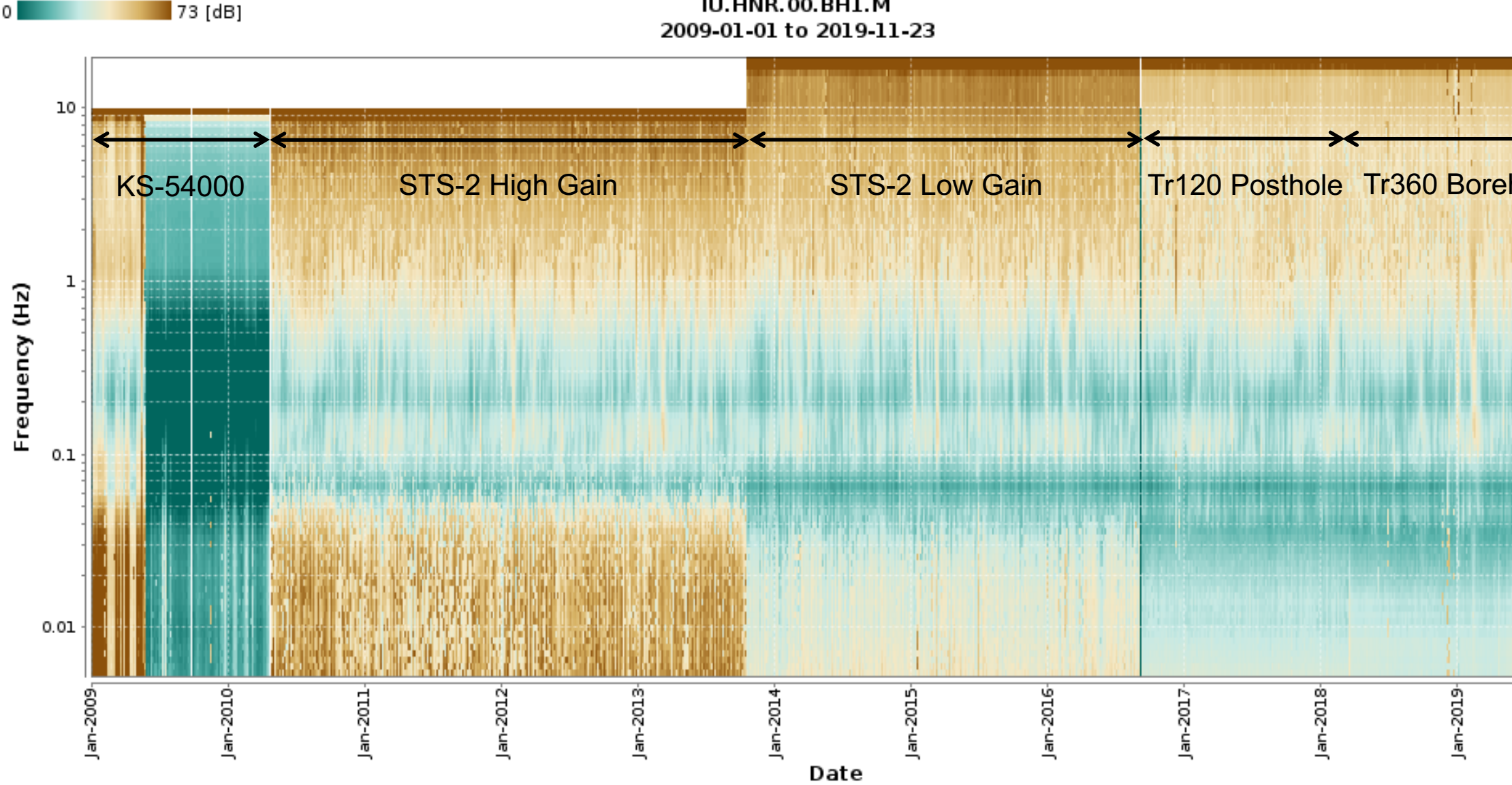
ds.iris.edu/mustang/mustangular

LASSO (**L**atest **A**ssessment of **S**eismic **S**tation **O**bservations) ranks and color-codes stations based on metric values. This example combines the average *percent_availability* metric with *dead_channel_gsn* (a Boolean indicator of likely digitizer noise) to create a combined data availability value, then ranks stations from worst to best for the TA network in July 2019 (not all stations are shown).

How did hardware upgrades affect my station?

Noise spectrograms (service.iris.edu/mustang/noise-spectrogram/1/) of daily Probability Density Function (PDF) modes can elucidate small Power Spectral Density (PSD) changes over time.

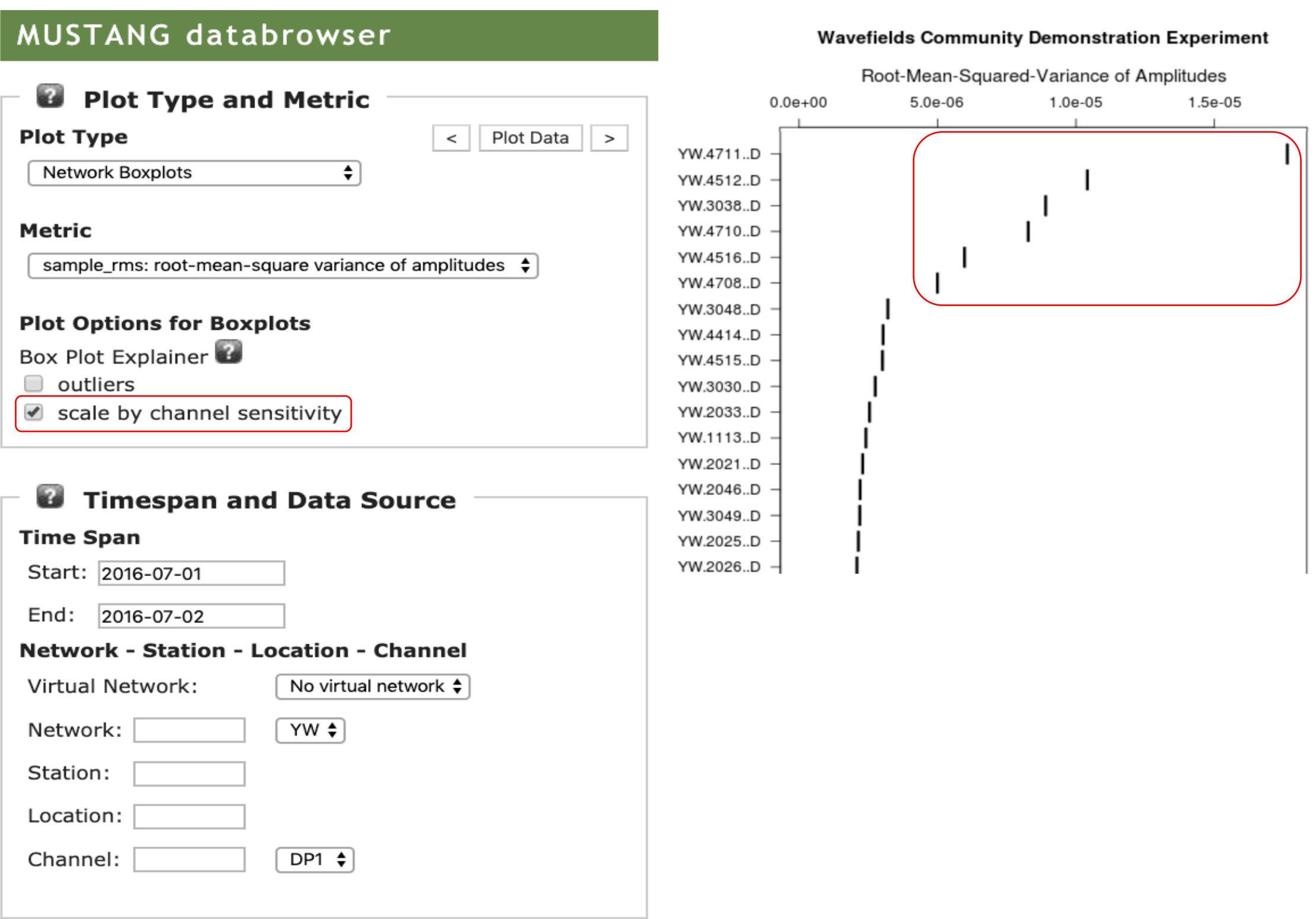
The spectrogram below shows that after the failure of the KS-54000 seismometer in 2010, low-frequency noise levels improved with each of three sensor upgrades. Noise levels for the final Trillium 360 borehole sensor appear comparable to the Trillium 120 posthole that it replaced.



Spectrograms can be displayed in a variety of color palettes that are compatible with greyscale or color printing, and/or are colorblind-friendly.

Which PH5 traces should I omit from my shot gathers?

We now calculate metrics for PH5 data and our next **MUSTANG Databrowser** release will display them. Plotting amplitude metrics as Network Boxplots will make culling geophones with anomalous amplitudes easier. High-amplitude outliers for the metric *sample_rms* appear at the top of the box plots. This metric is in units of Counts, so we recommend that you “scale by channel sensitivity” when comparing heterogeneous instruments.



How can I generate QA metrics for data not housed at IRIS?

ISPAQ (**I**RIS **S**ystem for **P**ortable **A**ssessment of **Q**uality) is a command-line utility that allows users to generate MUSTANG metrics on local seismic data, or data from any of the Federated Data Centers. While ISPAQ is wrapped in Python code, at its core is the exact same R code that the DMC uses in-house to calculate metrics. ISPAQ stores metrics as text csv or PNG files. Improvements over the past year include the ability to create PDFs for arbitrary time periods.

ISPAQ is available from Github: github.com/iris-edu/ispaq.gi.

Acknowledgements

Many thanks to Pete Davis, director of Project IDA for providing the II.PFO station gap example. LASSO is funded through IRIS Instrument Services. IRIS Data Services are funded through the National Science Foundation Cooperative Agreement EAR-1851048 (SAGE).