Data Quality for Gravitational-Wave Detectors

Dr. Andrew Lundgren
Institute of Cosmology and Gravitation
University of Portsmouth

DIRECT OBSERVATION OF GRAVITATIONAL WAVES FROM A BINARY BLACK HOLE MERGER

Image credit: Aurore Simonnet, E/O Sonoma State University
A New Era of Astronomy

- Huge Surveys in optical and radio (SDSS, LSST, Gaia, SKA)
  - Large areas of the sky
  - Millions of galaxies, or millions of stars
  - SKA will be exabyte/day raw, 300 PB/year stored
- Gravitational Waves (LIGO, Virgo)
  - A new window on the Universe
  - Extreme limits of precision measurements
Plan of the Talk

• What is Data Quality?
• The instruments, and data, we use
• Some of our algorithms
Data Quality

Three things to do:

• Mark all times when detector misbehaving

• Hunt down causes so they can be fixed

• Validate detection (not caused by an artefact; not affected by noise)
Data Quality

Three things to do:

• Mark all times when detector misbehaving

• Hunt down causes so they can be fixed

• Validate detection (not caused by an artefact; not affected by noise)
Data Quality

Three things to do:

• Mark all times when detector misbehaving

• Hunt down causes so they can be fixed

• Validate detection (not caused by an artefact; not affected by noise)
Matched Filtering

• Key to most interesting discoveries

• Require:
  • Known waveform / templates (Sascha’s talk)
  • Search over all unknown parameters (Alex’s talk)
  • Stationary Gaussian noise
Stationary Gaussian Noise

- Wold’s Theorem (Wold 1938):
  - Stationary noise = FIR filter convolved with independent samples, plus deterministic term
- Stationary = Power spectrum fixed in time
- Gaussian = Whitened data is Gaussian-distributed
Components of the Noise

- Gaussian shot noise
- Ground motion, suppressed by $10^{10}$
- Lines (electric, vibrational, intentional)
- Nonstationary

Graph showing GW amplitude spectral density vs. frequency.
Improvements

Seismic

Shot noise

Strain noise (Hz$^{-1/2}$)

Frequency (Hz)
LIGO Site (Hanford)
How Does LIGO work?

- Gravitational wave changes differential arm length
- Intensity on readout modulated (Michelson interferometer)
- Direct measurement of strain in 10 to 1000 Hz range
Improvements: Seismic Isolation

Test mass vacuum enclosure

Active seismic isolation

Quadruple suspension

Images by Jeff Kissel
Where stuff is on a BSC-ISI

STAGE 0 (ST0)

L4C

T240

f = ~1 Hz

Inertial Sensors

CPS

Displacement Sensor

STAGE 1 (ST1)

STAGE 2 (ST2)

GS13

f = ~1 Hz

Inertial Sensor

ACT

Electromagnetic Actuators

One corner’s ST0-1 and ST1-2 position sensors and actuators

Jeffrey Kissel, MIT Dec 11th 2009
Auxiliary channels (per site)

- 3,000 fast channels (256 Hz to 16384 Hz)
- 200,000 slow channels (16 Hz)
- 1 terabyte per day
- Archived in 64-second chunks, all channels together, in our own frame format
- Also second and minute trends (min,max,mean,rms)
PEM (Environmental Monitoring)
Cartoon of ML

Target
Features
Constraints
ML: Target

- Detection range
- Stationarity of the noise
- Transient artefacts
ML: Constraints and Regularisation

• We want simple answers
  • and avoid multiple comparisons
• Mechanisms we can test, things we can fix
• Use physics in the model
• How to establish causality, not just correlation?
ML: Features

- All those auxiliary channels
- Whiten (invert spectrum, apply as filter)
- Sine-Gaussian basis: Shourov Chatterji’s thesis, dspace.mit.edu/handle/1721.1/34388
- Band-limited RMS
- Position, velocity
- Multi-channel correlations
Methods and Examples
LASSO

• Regression of the range against many, many auxiliary channels

• L1 norm penalty to encourage sparse weights

• Collaborators: Marissa Walker, Jeff Bidler, Alex Macedo, Joshua Smith (Cal State Fullerton)
LASSO Configuration

Parameters

This analysis used the following parameters:

Start time: 2017-02-18 05:59:42 (1171432800)
End time: 2017-02-18 11:42:42 (1171453380)

Range channel: L1:DMT-SNSH_EFFECTIVE_RANGE_MPC.mean (SenseMonitor_hoft_L1_M)

Channels searched: 415490

Number of flat channels: 344995 (flat channel list)

LASSO coefficient threshold: 0.0001

Model Information

Model: LASSO
Non-zero coefficients: 11
Alpha: 0.22
Zero coefficients: 69572 (zeroed channel list)

<table>
<thead>
<tr>
<th>Channel</th>
<th>LASSO Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1:HPI-ETMY_ISO_X_OUT16.mean</td>
<td>-0.307973</td>
</tr>
<tr>
<td>L1:SUS-ETMY_R0_DAMP_Y_IN1_DQ.rms</td>
<td>-0.070306</td>
</tr>
<tr>
<td>L1:SYS-TIMING_Y_FO_A_PORT_9_SLAVE_GENERIC_PAYL...</td>
<td>0.059067</td>
</tr>
<tr>
<td>L1:HPI-ETMY_BLRMS_Z_30M.mean</td>
<td>0.029452</td>
</tr>
<tr>
<td>L1:OAF-CAL_DARM_DQ.rms</td>
<td>-0.027925</td>
</tr>
<tr>
<td>L1:ISI-BS_ST1_FFB_BLRMS_RY_10_30.mean</td>
<td>-0.021117</td>
</tr>
<tr>
<td>L1:SUS-SR3_M1_VOLTMON_T2_MON.rms</td>
<td>-0.019242</td>
</tr>
<tr>
<td>L1:HPI-ITMX_BLRMS_RX_10_30.rms</td>
<td>-0.005955</td>
</tr>
<tr>
<td>L1:HPI-ETMY_BLRMS_Z_30M.rms</td>
<td>0.000236</td>
</tr>
<tr>
<td>L1:SUS-SR3_M1_VOLTMON_T2_MON.mean</td>
<td>0.000153</td>
</tr>
</tbody>
</table>
LASSO Example: Alignment

25 million light years

Highly Correlated Channels
LASSO Example: Electronics Failure
GravitySpy

• A Zooniverse citizen science project

• gravityspy.org

• Detector experts, sociologists, and machine learning researchers working together

• Zevin et.al. Class. Quantum Grav. 34 (2017) 064003
Eager for more ways to contribute to LIGO science? Check out Einstein@Home, where your computer's idle time is used to help search for gravitational waves in LIGO data. You can even run this program while classifying events in GravitySpy!

Please sign in or sign up to access more glitch types and classification options as well as our mini-course.
GravitySpy Talk

Is this an example of clattering? Found quite a few of these amongst the Hanford O2 images.

In reply to sinjinza1984's comment
April 7th 2018, 2:53 pm

I think it is different from clattering, I call this type #forest. There are really plenty of similar glitches. Just working on a proposal of them.

Hi this is @EcceuElime post re. clattering
April 7th 2018, 6:42 pm

https://astronomy.stackexchange.com/questions/15896/are-you-sure-we-are-listening-for-a-living-planet-10971217
Ground Motion

• Microseismic peak driven by wave motion
  • 7 sec typical period, 1 micron/sec amplitude
  • Local noise 1 Hz and above (human activity)
Scattering Prediction

Time [seconds] from 2016-12-02 23:01:50 UTC (1164754927.0)
Machine Learning - Before

L1:GDS-CALIB STRAIN (Omicron)

Loudest event: Time=1169083974.874511003, Peak Frequency=262, SNR=1.09 \times 10^3

With Fulbright student Hunter Gabbard (now PhD student, Glasgow)
Machine Learning - After

After HIGH SCATTER ML (Omicron)

Loudest event: Time=1169083974.874511003, Peak Frequency= 262, SNR= 1.09 × 10³

Time [hours] from 2017-01-21 12:02:32 UTC (1169035370.0)

with Fulbright student Hunter Gabbard
A Much Harder Case

L1:GDS-CALIB_STRAIN 2015-10-24 08:50:07 - 1,129,711,824 (290s)

Fs=16,384Hz, sec/fft = 1.00, overlap = 0.50, fft length=16,384, #-FFT = 579, bw = 1, in samples = 4,751K, low = 0.20
Auxiliary Channel Correlations
Future Developments

• Try to integrate different ideas together
• Think hard about what to optimise, and constraints on the model
• Use more features together; more complicated models
• Try to keep interpretability
Concluding Thoughts

• Infrastructure is important

• All data (and transforms) available through a single method

• Expert knowledge is very powerful

• Signal processing and control theory still very useful

• Big science projects of the future should design in data mining and diagnostics from Day 1