C02 Calibration Uncertainty Material For Review

Craig Cahillane for the Calibration Team
(edited by Jeff Kissel)

Where's the Original Source Material?

- Craig has done all the work. Thanks Craig!
- For the original source material, see LHO aLOG 26889 and LLO aLOG 25950
- Results have been produced by
  `aligocalibration/trunk/Runs/O1/Common/Scripts/Uncertainty/NumericalUncertaintyBudget.m`
  - Further instructions on locations and how to run the code can be found in comments to above aLOGs
- Kissel was able to reproduce all plots from the aLOGs with a local checkout of repo.
- Kissel's plots (those not found in Craig's aLOGs) are produced from text files attached to the aLOG, and plotted with
  `aligocalibration/trunk/Common/Documents/G1601003_C02Review/plot_GW151225_uncertainty_forG1601003.m`

What’s new in C02 Uncertainty since GW150914

- Now have Uncertainty for GW150914, LVT151012, and GW151226
  - And they’re different on the level of ~5% / 5 deg
- Now have plots where you can easily read off the “pair of numbers” to quote in papers.
  - However, **we** continue to **aggressively encourage search groups to fold all of our knowledge into future analyses**
- Scalar, time-dependent corrections (kappa_c, kappa_tst, and \kappa_pu) have been applied to h(t), but still impact the uncertainty.
  - Informed by noise leftover from de-trending the time series of these parameters
  - Fluctuations are a function of time, because SNR of CAL lines are a function of time
  - Not large, but now included in overall statistical uncertainty
- Time-dependent cavity pole remains uncompensated
  - That means the systematic error (because the value of the pole is known) is time dependent
- Plots showing the uncertainty evolution over the entire run instead of just at event times
  - Spectrograms (systematic error, statistical uncertainty, maximum uncertainty & error)
  - Percentile Plots (maximum uncertainty & error)
  - Time series (of maximum uncertainty & error)
- Eventually (but not yet): including 1-4.5 kHz data

What do we need?

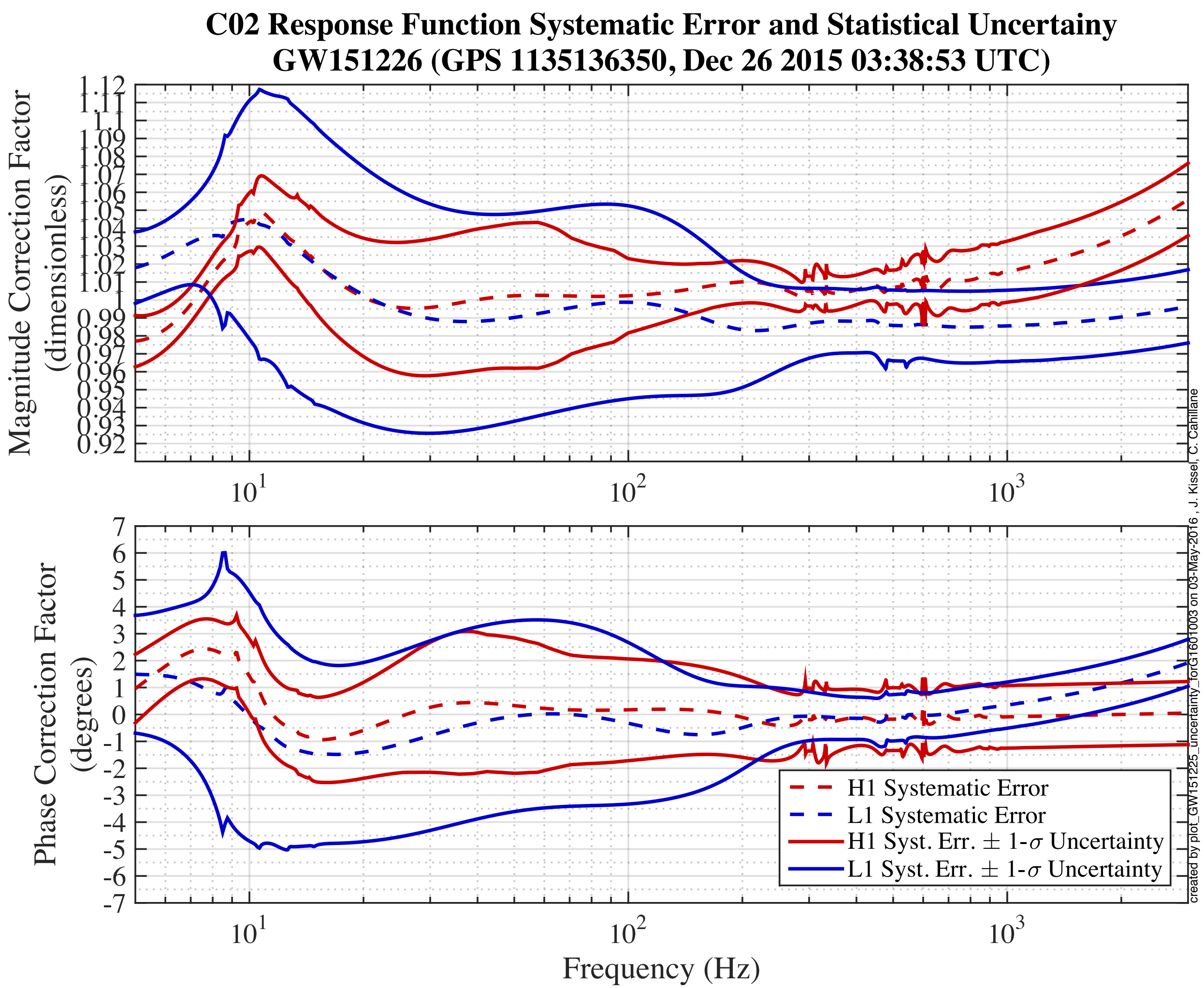
- CBC / PE group needs a one-sentence statement about the uncertainty, in the usual “pair of numbers” format
  - Boxing Day Event Paper (just at time of event; Dec 26, 2015 03:38:53 UTC)
  - BBH Search Paper (For the all three events)
  - Rates Statements (All of O1; focused on how Inspiral Range varies)
- Frequency range needs for all of the above is now confirmed to be **20 – 1000 Hz**
- Current statement: (Boxing Day Paper, Lines 70-73)

“The calibration is continuously monitored and measured to an uncertainty (1 σ) of less than 10% in amplitude and 10 degrees in phase [11].”

GW151226
C02 Total Uncertainty and Error

Components of Systematic Error:
- Time-independent model vs. measurement discrepancy
- Uncompensated change in DARM coupled cavity pole
- Frequency-independent Systematic Error in PCAL

Components of Statistical Uncertainty:
- Weighted Uncertainty (i.e. coherence) from several frequency-domain transfer functions
- Uncertainty in change in time-dependent parameters



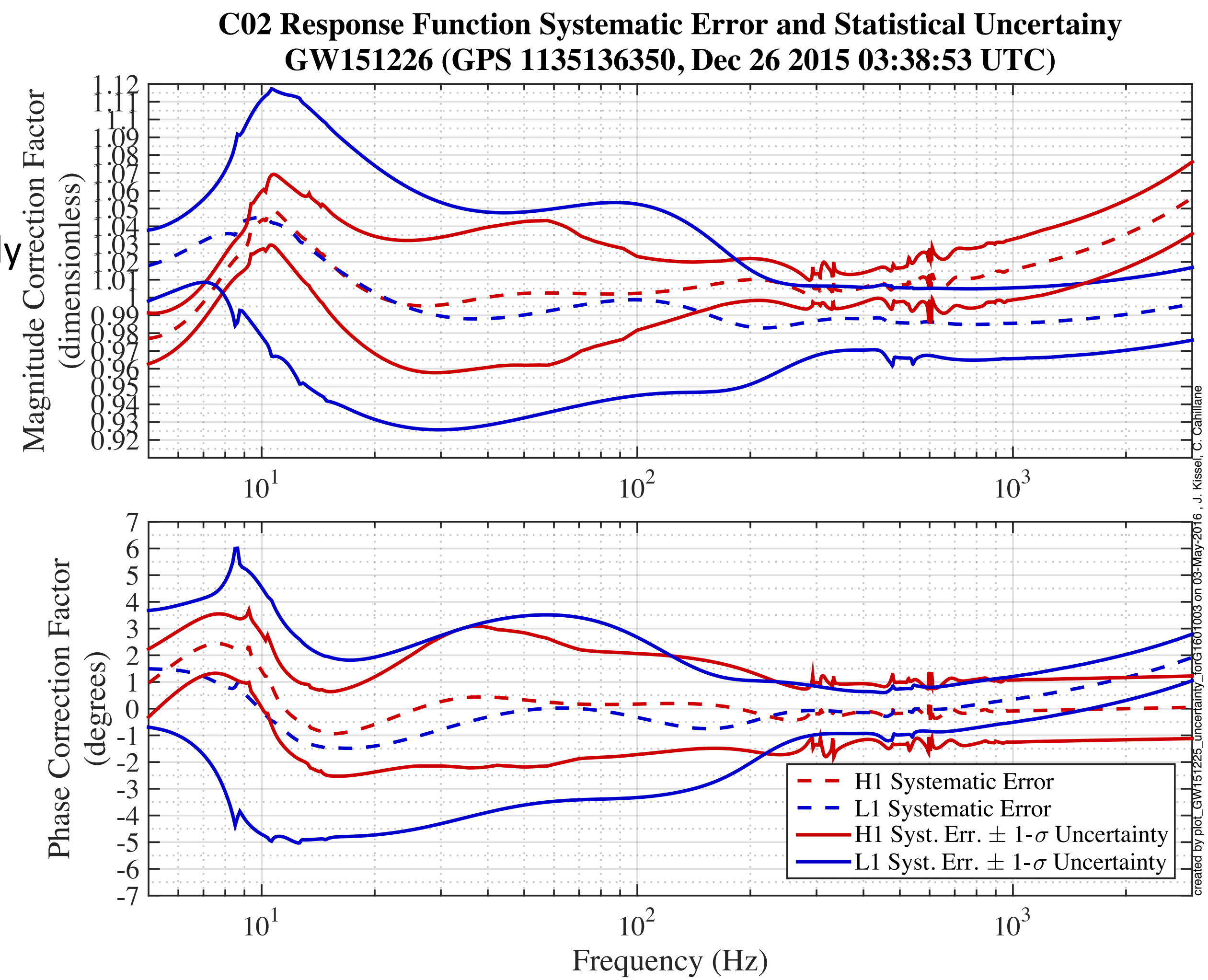
GW151226
C02 Total Uncertainty and Error

L1 Uncertainty is larger than H1 because of the limited number of measurements they were able to obtain

H1 error at high frequency is poorly understood, (More on this later…) but doesn't significantly impact "CBC" or "event-like" frequency band

H1 error below 20 Hz is due to poorly understood deviation from flat sensing function (L1 doesn't have enough low-freq data points to resolve it)

L1's error below 20 Hz is due to poorly resolved frequency dependence of PUM stage (again, due to lack of low-freq measurement)

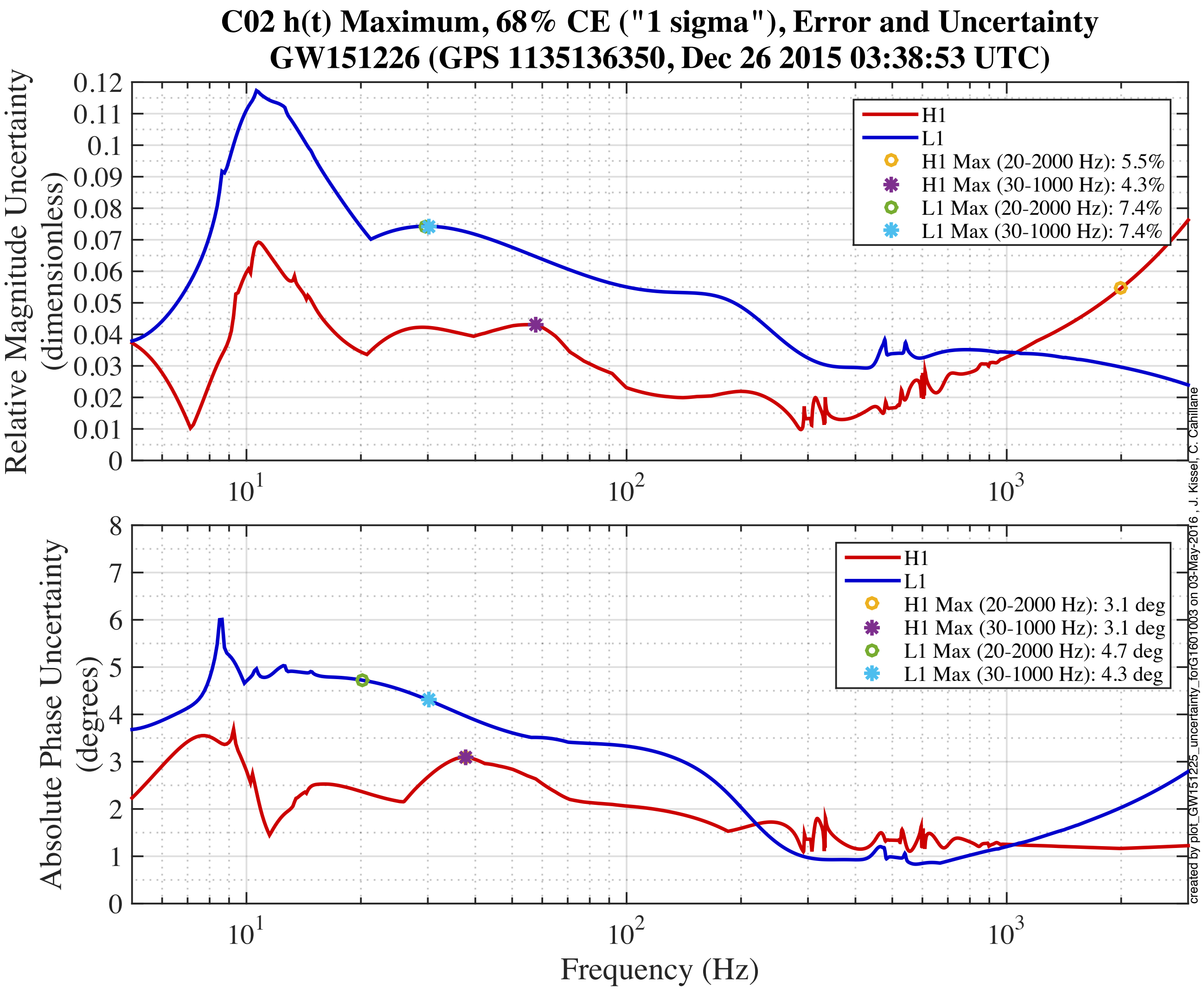


GW151226
Maximum 1-σ Error and Uncertainty

GW Event PRL's don't have the space to say – and/or Searches and PE aren't yet sophisticated enough to include – more than
A pair of frequency independent numbers "Uncertainty is less than XX% and YY deg."

- Take data from previous plot
- Subtract 1.0 (to turn "relative correction factor" error into "additive systematic correction" error)
- Take the absolute value of solid lines; (syst. error + stat. uncertainty) as well as (syst. error - stat. uncertainty)
- Plot the maximum as a function of frequency
- Find the maximum value (for magnitude and phase) over the frequency range of the search and parameter estimation

GW151226
C02 Maximum 1-σ Error and Uncertainty



	20-1000 Hz
H1	**4.3% / 3.1 deg**
L1	**7.4% / 4.7 deg**

This is the final answer for GW151226

Why is L1 larger than H1?

In brief:

- PUM actuator: systematic errors *between* measurements at high freq
- TST actuator: low freq outliers
- Sensing function determined totally differently between sites

SEE APPENDIX B

What about for all of O1?

Components of Systematic Error:

- Time-independent model vs. measurement discrepancy
- Uncompensated change in DARM coupled cavity pole
- Frequency-independent Systematic Error in PCAL

We must at least make plots of how the uncertainty evolves as a function of time, to see if "less than 10% and 10 deg" holds for all data over the entire run.

← This makes the systematic error time-dependent

Components of Statistical Uncertainty:

- Weighted Uncertainty (i.e. coherence) from several frequency-domain transfer functions
- Uncertainty in change in time-dependent parameters

The next pages walk through all of the plots that have helped us quantify this time dependence.

← This makes the statistical uncertainty time-dependent

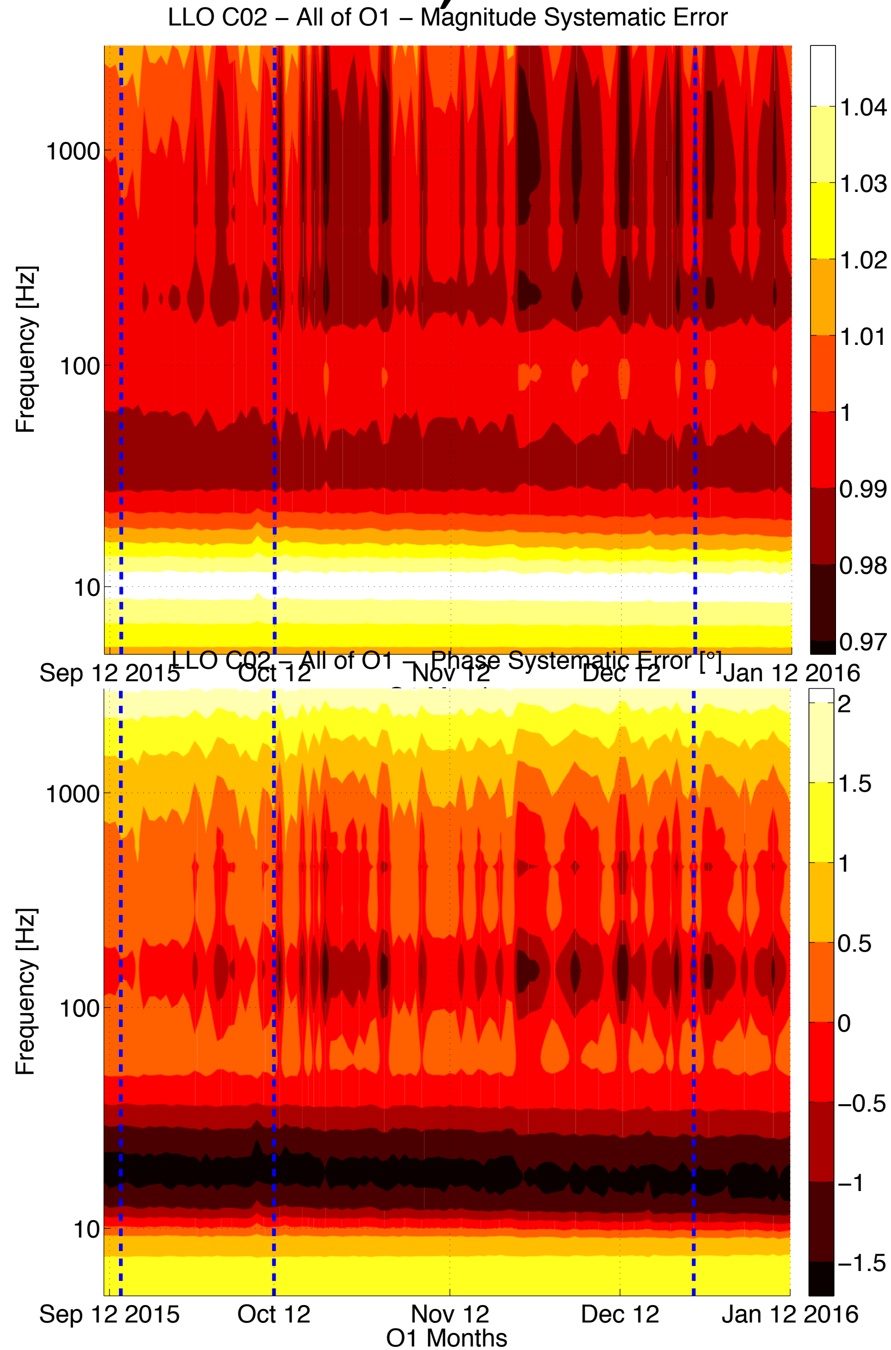
Systematic Error for All of O1, L1

For L1:

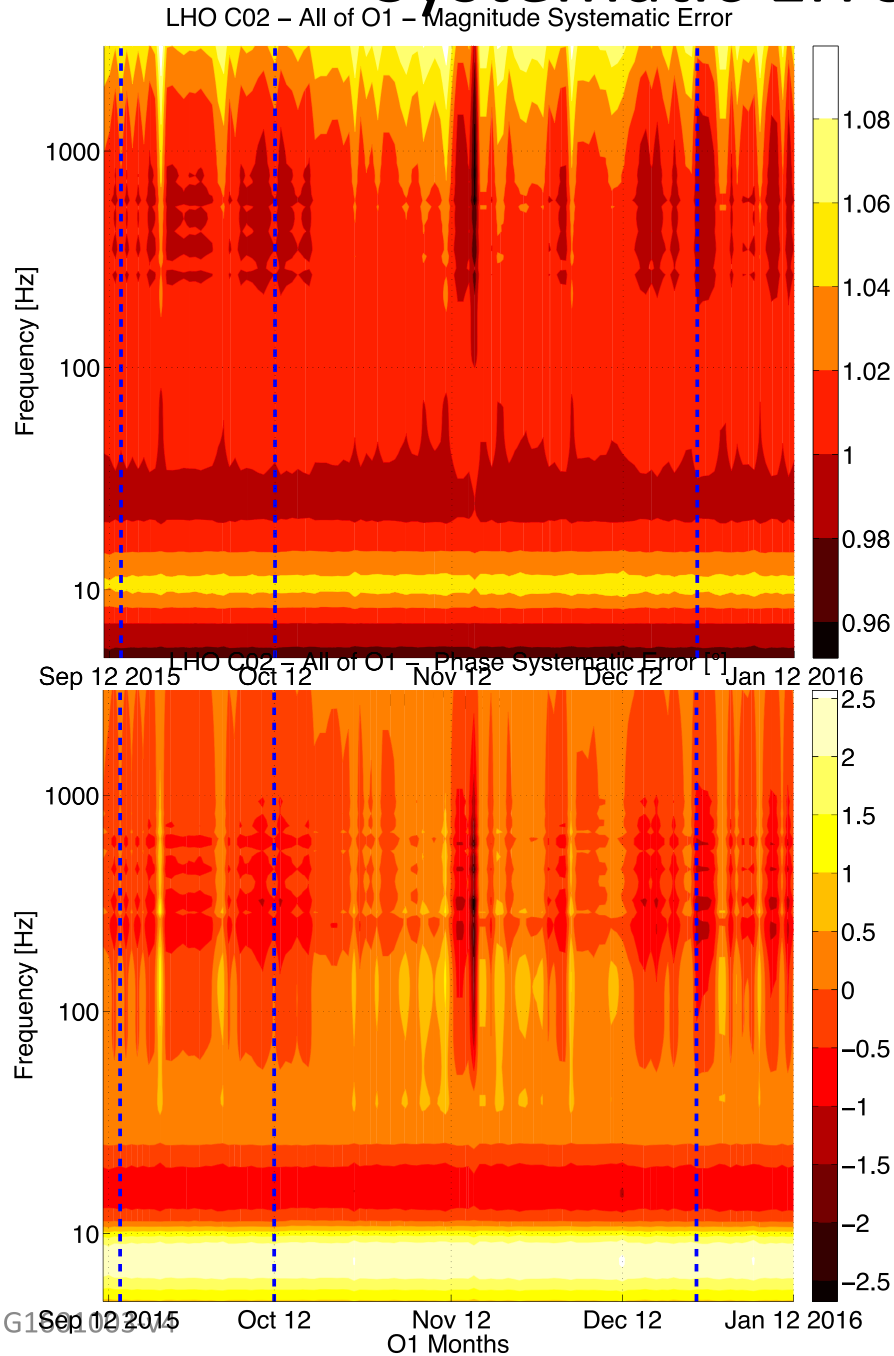
Even though we've compensated for scalar systematic error, the remaining frequency dependent error (from the changing DARM coupled cavity pole) changes as a function of time.

For a given frequency, extreme of change is on the order of 3% / 2 deg at ~300 Hz from beginning to the end of run

Z, color axis range is limited to the max and min of value from 5 to 3000 Hz: **[-3% to +5%] and +/-2 deg**
Change in value is smaller…



Systematic Error for All of O1, H1



Similar but worse story at H1

Even though we've compensated for scalar systematic error, the remaining frequency dependent error (from the changing DARM coupled cavity pole) changes as a function of time.

For a given frequency, extreme of changes are on the order of 3% / 2 deg at ~300 Hz from beginning to the end of run.

Changes from cavity pole are larger (over this frequency range) in H1 because the pole frequency is lower, and it varies more because alignment was more poorly controlled

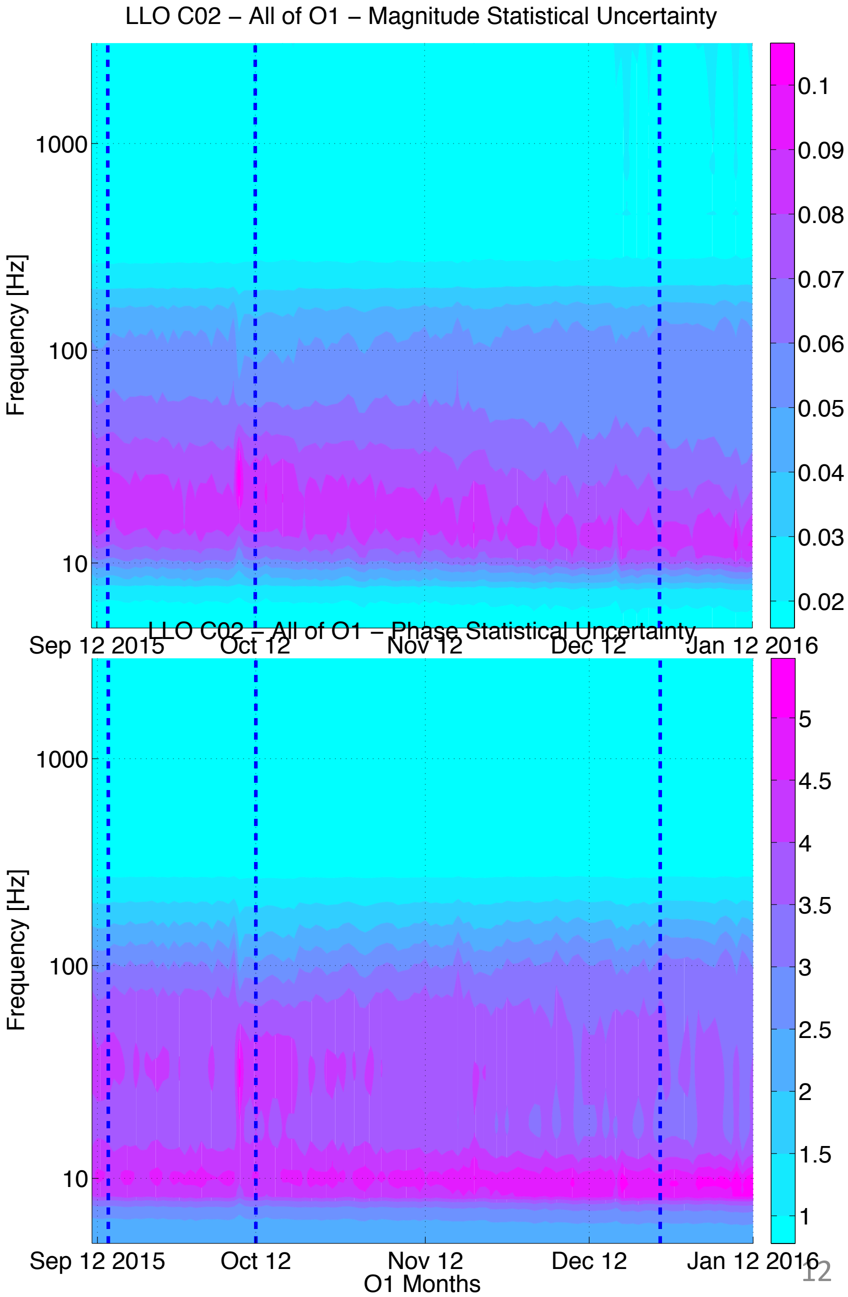
5 to 3000 Hz Range is **[-5% to 9%] and +/- 2.5 deg** but the *change* is smaller

Statistical Uncertainty for All O1, L1

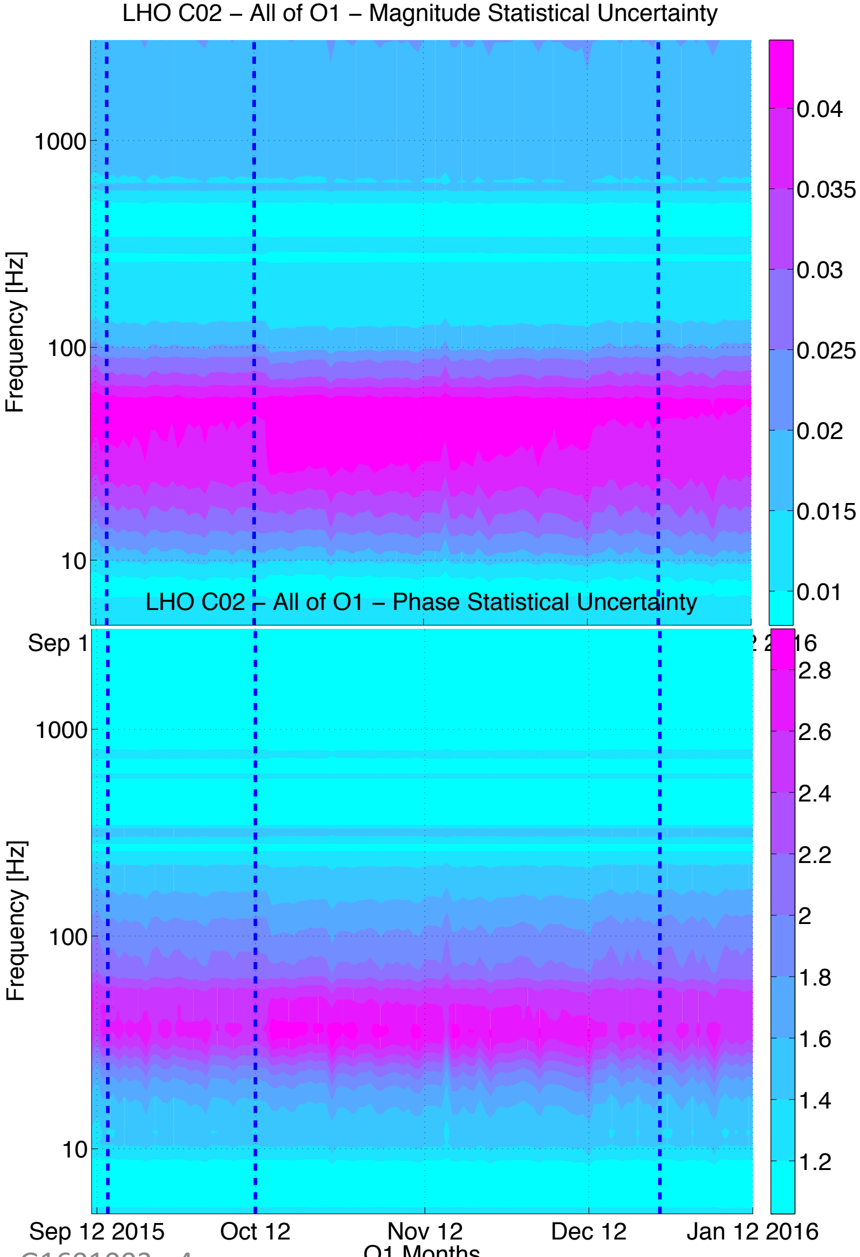
Measurements of the time-dependent parameters are based on SNR of PCAL and ESD Calibration Lines

Sensitivity changes as a function of time → statistical uncertainty is a function of time

Worst at low frequency for L1, where we lost sensitivity throughout the run. At 30 Hz, varies also at the level of 3% / 2 deg.



Statistical Uncertainty for All O1, H1



Measurements of the time-dependent parameters are based on SNR of PCAL and ESD Calibration Lines

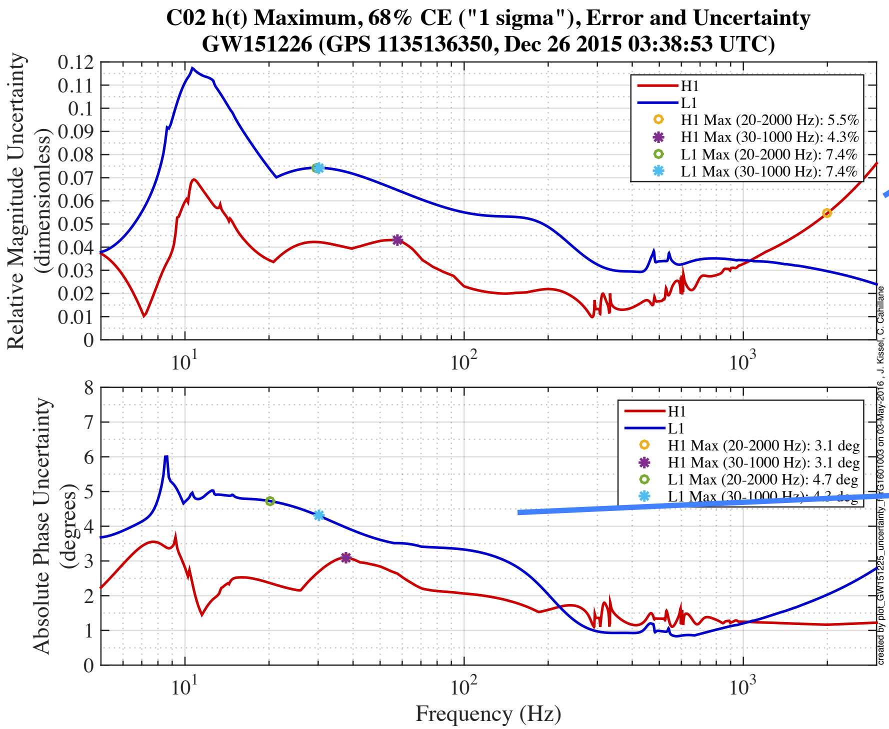
Sensitivity changes as a function of time → statistical uncertainty is a function of time

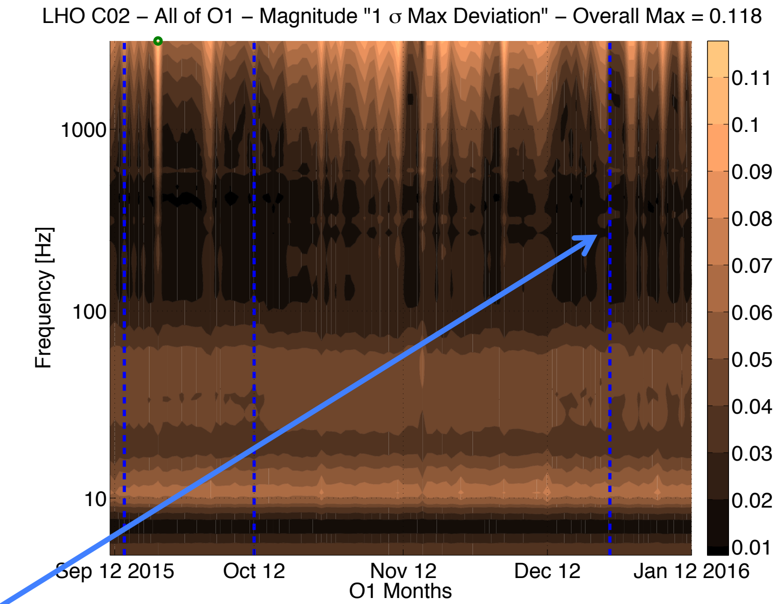
At H1, variation is smaller and most prevalant at the DARM UGF, as the ESD actuation strength changed throughout the run (including BIAS flip in mid-October). At 50 Hz, varies also at the level of 0.5% / 1 deg.

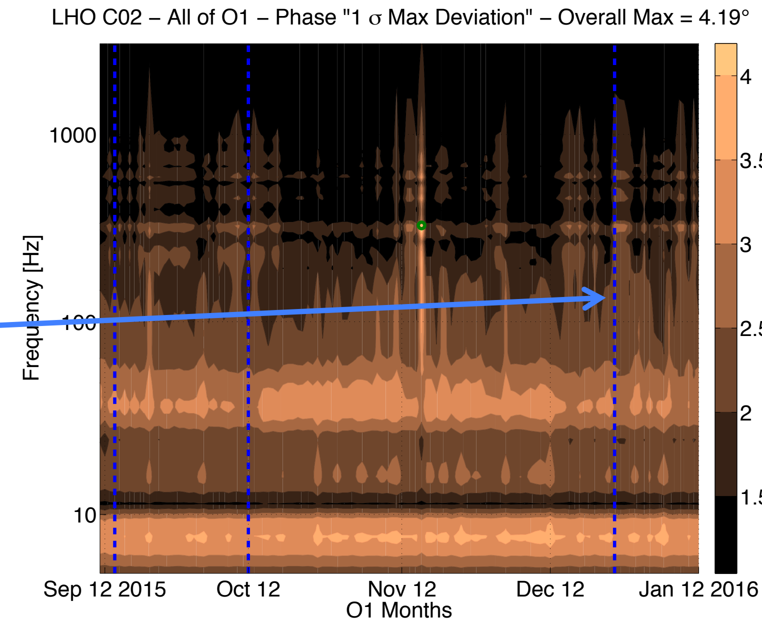
Maximum Statistical Uncertainty and Systematic Error *as a Function of Time*

Because both statistical uncertainty and systematic error are frequency dependent and a function of time – how do we convey this as a pair of numbers for all O1?

Calculate the maximum over time, and show **spectrograms**, time-series of a horizontal slice and frequency-dependent percentiles covering all time

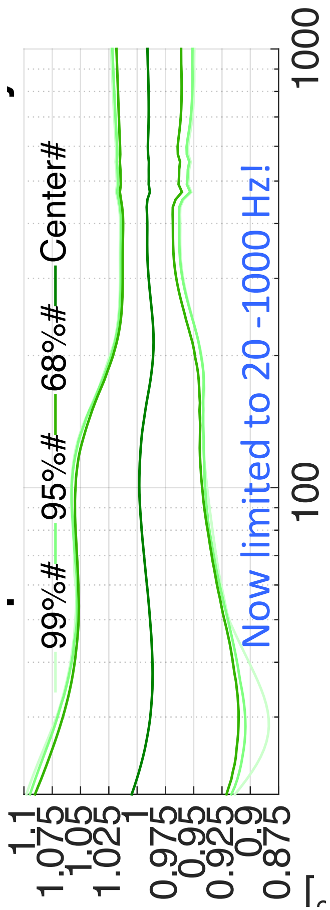


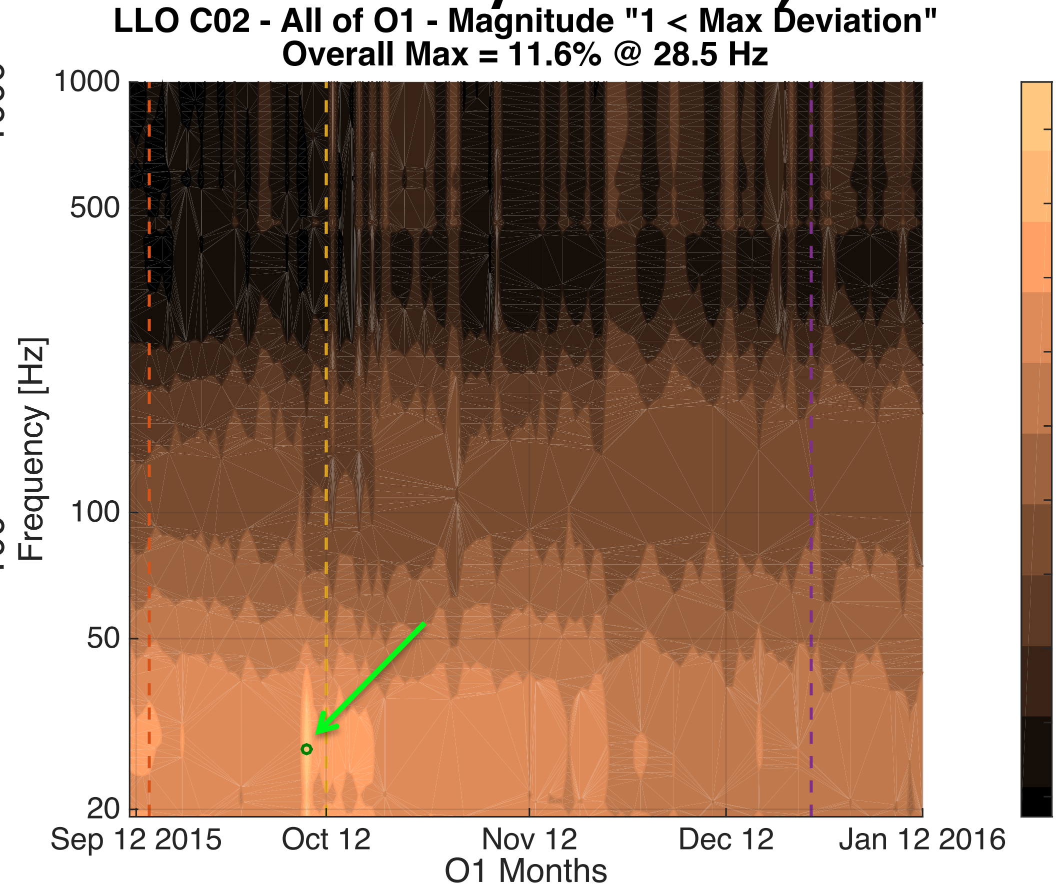




Maximum Stat. Unc. and Syst. Err., L1

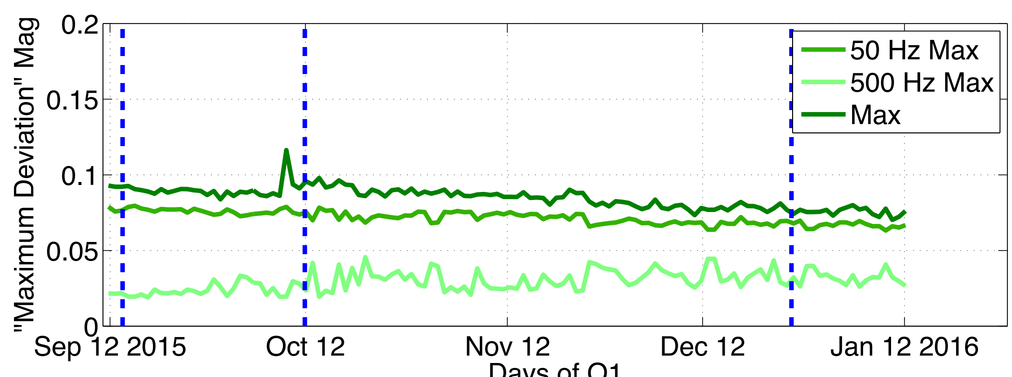
Max Magnitude Uncertainty and Error **Spectrogram**, Time Series, and Percentile plots
(i.e. pg 7 as a function of time)





Because L1's cavity pole doesn't change much (better alignment control)

Uncertainty & Error are large, but quantifiably quite stable.



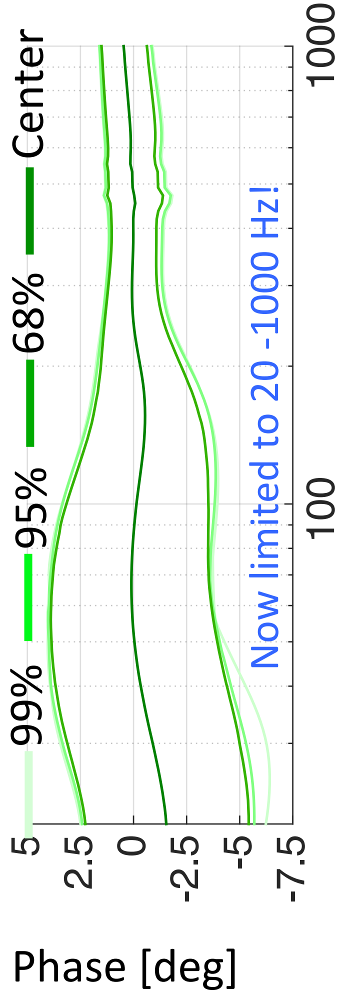
Only significant time-dependence is a trend due to \kappa_TST value.
Even though it's corrected for, it still folds into uncertainty partial derivatives.

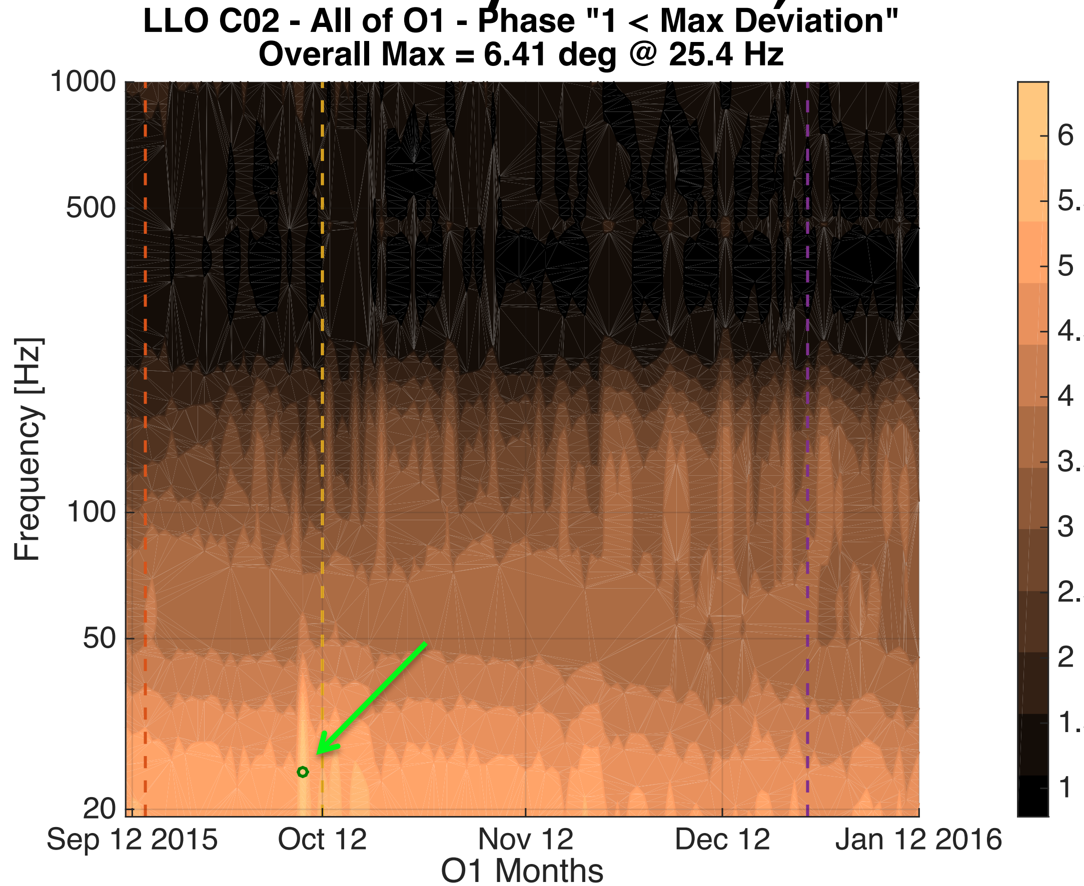
Maximum Stat. Unc. and Syst. Err., L1

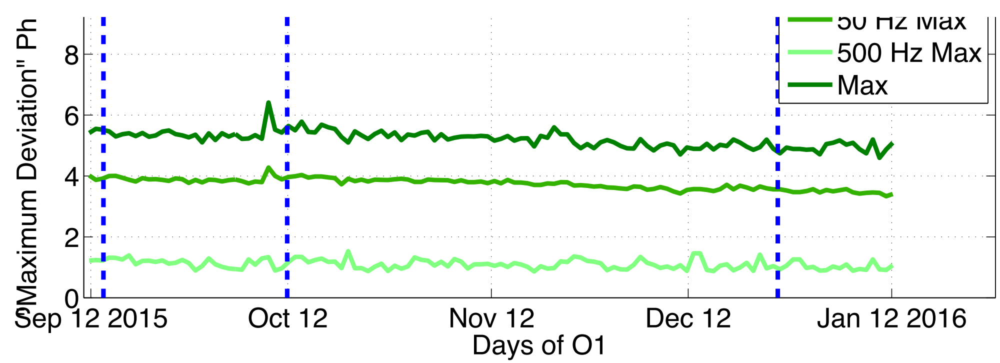
Largest excursion is on Oct 10th (marked w/ green circle and arrow)

This is a result of L1 turning off linearization, and reducing the ESD bias voltage: LLO aLOG 24573

Corrected for in h(t), but

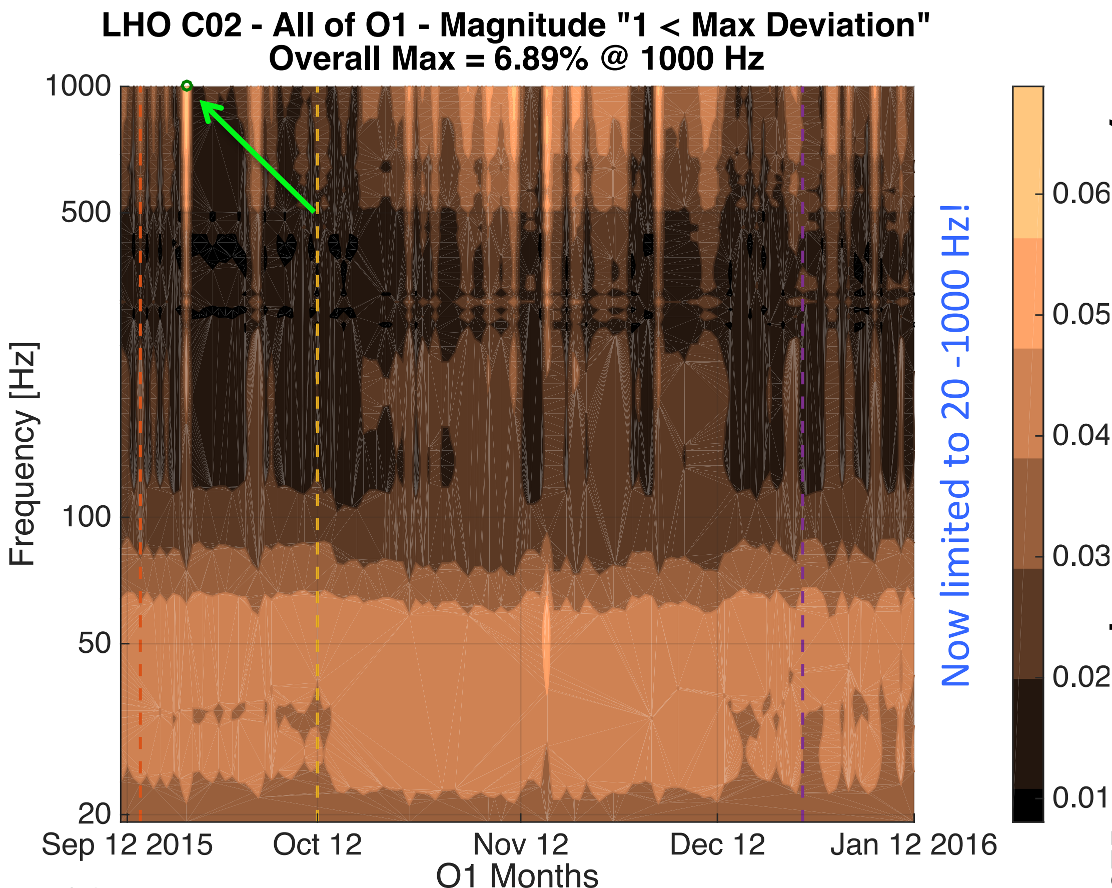


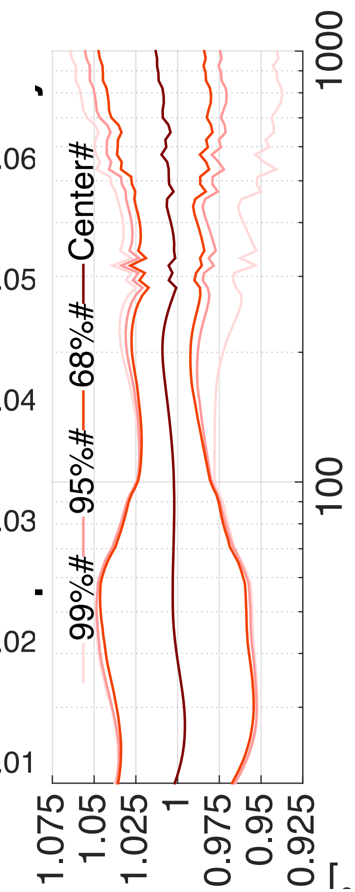




Again, the *value* of \kappa_TST still plays into the uncertainty via partial derivatives

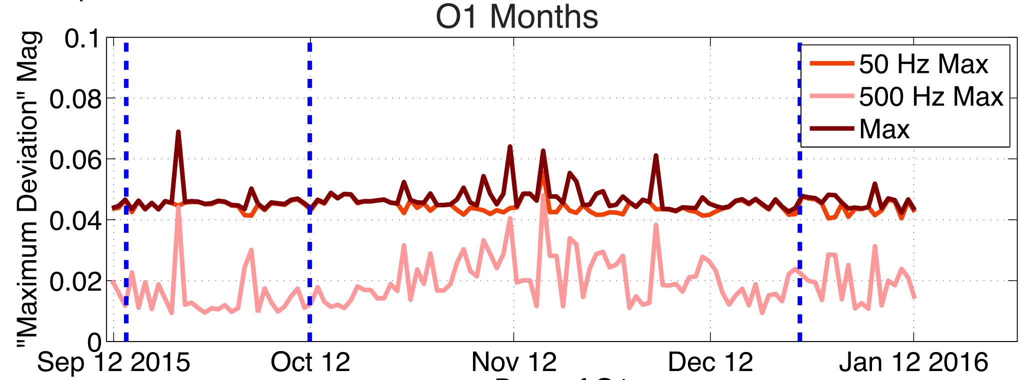
Maximum Stat. Unc. and Syst. Err., H1





H1's has a lower-frequency (~341 Hz) and more dynamic cavity pole

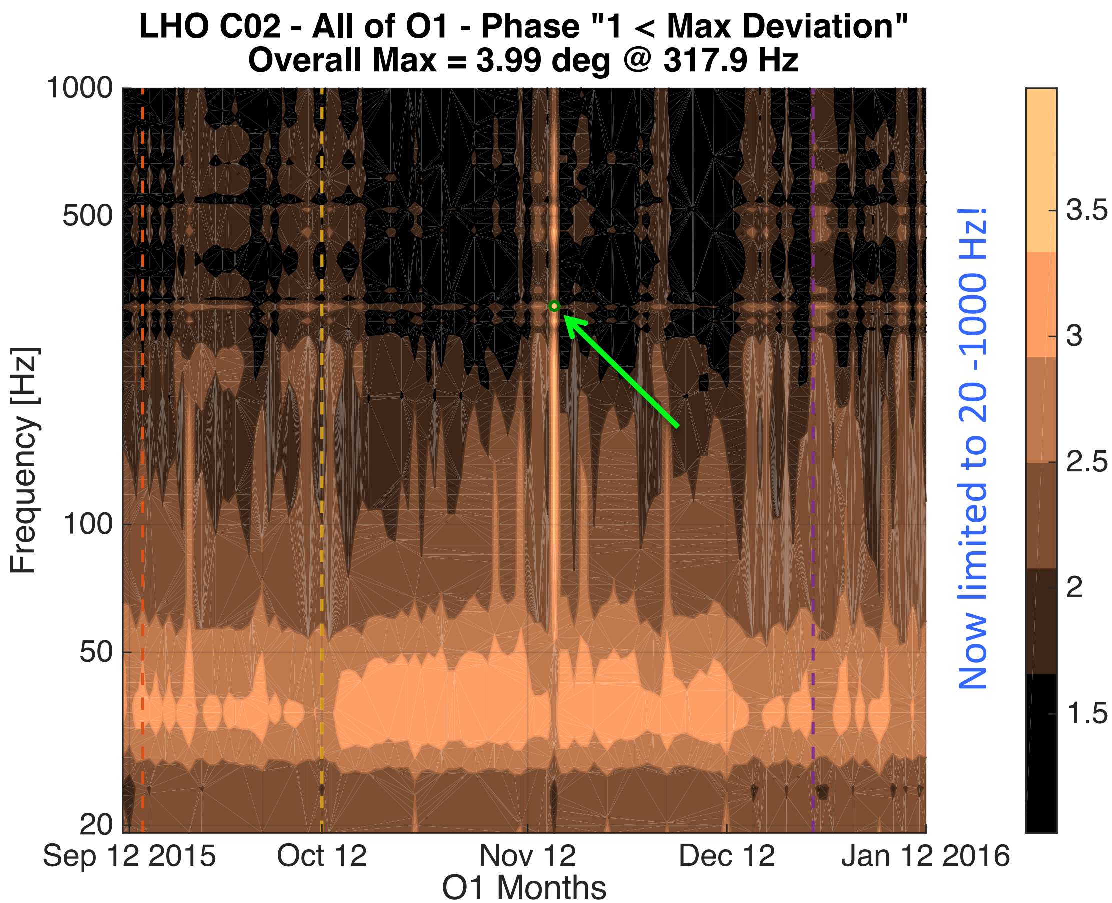
Therefore the percentile bounds for the run are significantly different above cavity pole frequency

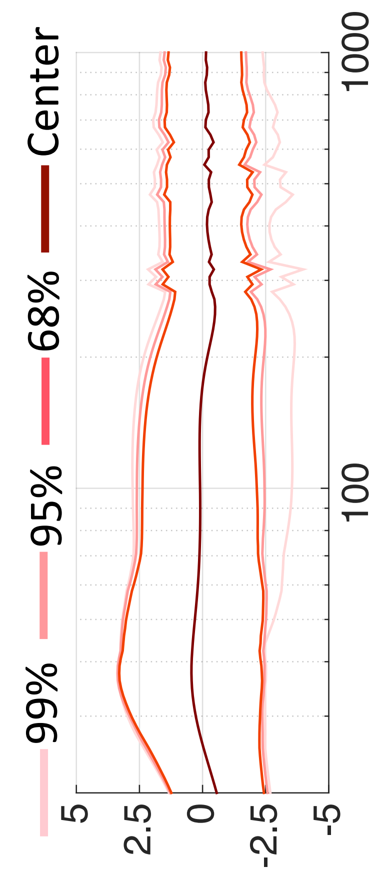


However, the deviations are sporadic, likely because of lock-to-lock differences in alignment

(Note the ylim range on time series is half that of L1)

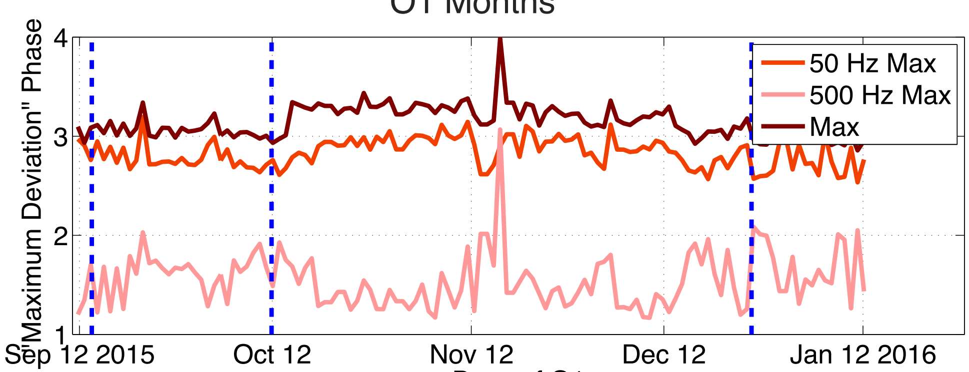
Maximum Stat. Unc. and Syst. Err., H1





The phase shows the same story: the largest deviation is at the cavity pole frequency

Largest deviation is around Nov 14th, which had several problems with wind, TCS laser and RF45 noise, e.g. LHO aLOG 23385



(Again, be weary of ylim range on time series – it's half that of equivalent L1 plot)

Things Left to Do

- Establish a pair of numbers for 20-1000 Hz region using maximum deviation plot for all of O1 **DONE**
  - Actually nail down from search groups what frequency they want! (Still waiting to hear from Burst, CW, and Stochastic Groups)
- Need to include 1-4.5 kHz data **In Progress**
- Request to "think out of the box" for what we might not be accounting **(never really) DONE**

None of these to-do items impact review for GW151226.

C02 Maximum 1-σ Error and Uncertainty

Over a 20 – 1000 Hz frequency range:

	H1 Reviewed, including T.Dep. Statistical Unc.	H1 Used in PE Runs (from EVNT aLOGs)	L1 Reviewed, including T.Dep. Statistical Unc.	L1 Used in PE Runs (from EVNT aLOGs)
GW150914	4.5% / 3.1 deg	4.82% / 3.158 deg	9.3% / 5.4 deg	8.23% / 4.196 deg
LVT151012	4.5% / 3.1 deg	4.19% / 2.667 deg	9.6% / 5.6 deg	8.32% / 4.283 deg
GW151226	4.3% / 3.1 deg	4.22% / 2.71 deg	7.4% / 4.7 deg	6.88% / 3.573 deg
All O1	6.89% / 4.0 deg	(n/a)	11.6% / 6.4 deg	(n/a)

Recall, PE group used numbers for EVNT aLOGs 11578, 11576, and 11580, which was before Craig corrected and implemented the median subtracted time-dependence statistical uncertainty

"Using methods as described in [GW150914 CAL companion paper], the calibration uncertainty (1 sigma) in both detectors is better than 8% in amplitude and 5 degrees in phase at the time of GW151226."

Conclusions

- Here's the "final" (as usual grossly over-simplified) answer for GW151226:

	20-2000 Hz	30-1000 Hz	20-1000 Hz
H1	5.5% / 3.1 deg	4.3% / 3.1 deg	**4.3% / 3.1 deg**
L1	7.4% / 4.7 deg	7.4% / 4.3 deg	**7.4% / 4.7 deg**

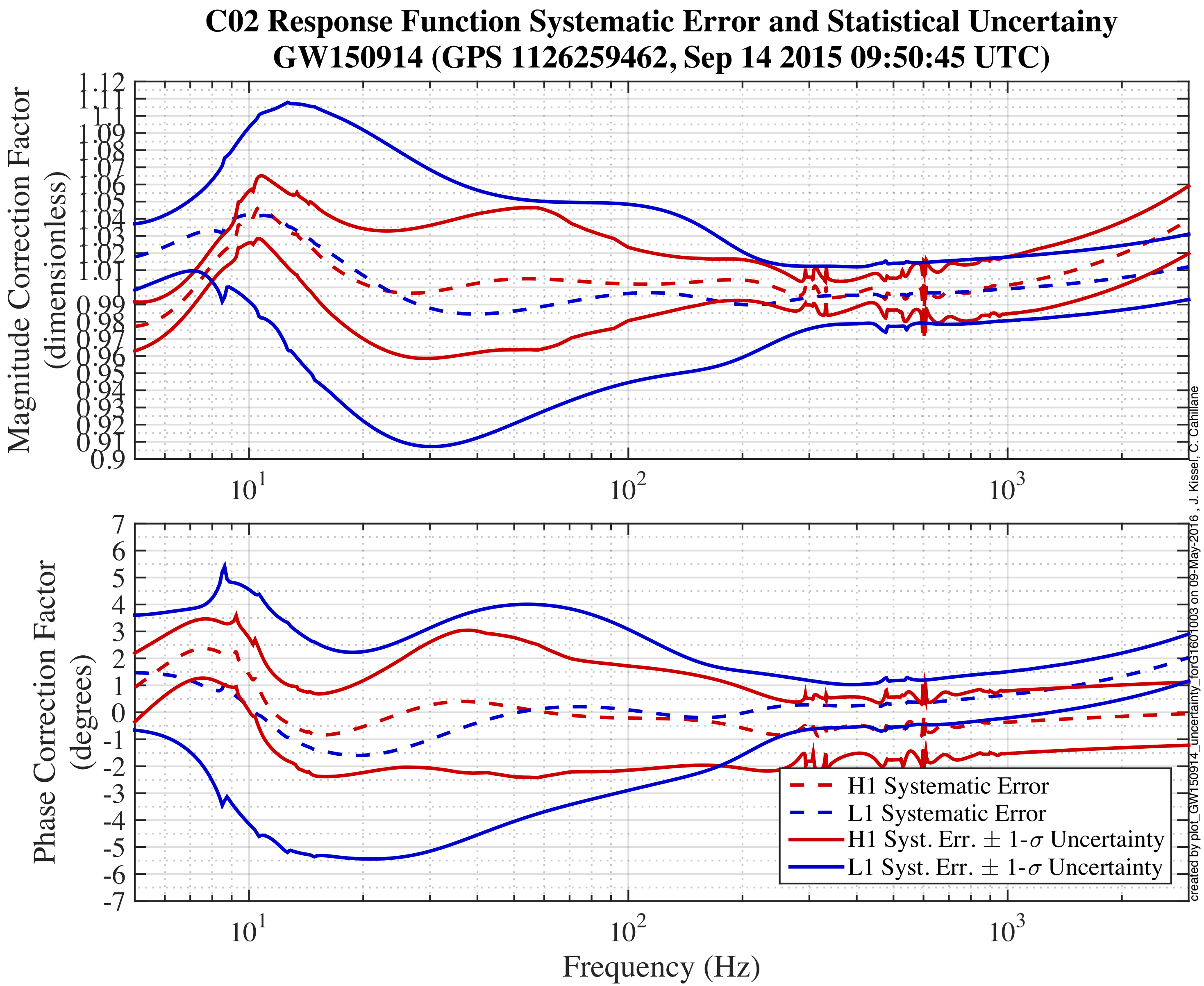
- Given the above frequency range, the calibration uncertainty still satisfies the "less than 10% and 10 deg" statement for all event and candidates (GW150914, LVT151012, and GW151226), *but not for all of O1*.
- **Landry, Kissel, and Vitale have agreed on**

  "Using methods as described in [GW150914 CAL companion paper], the calibration uncertainty (1 sigma) in both detectors is better than 8% in amplitude and 5 degrees in phase at the time of GW151226."

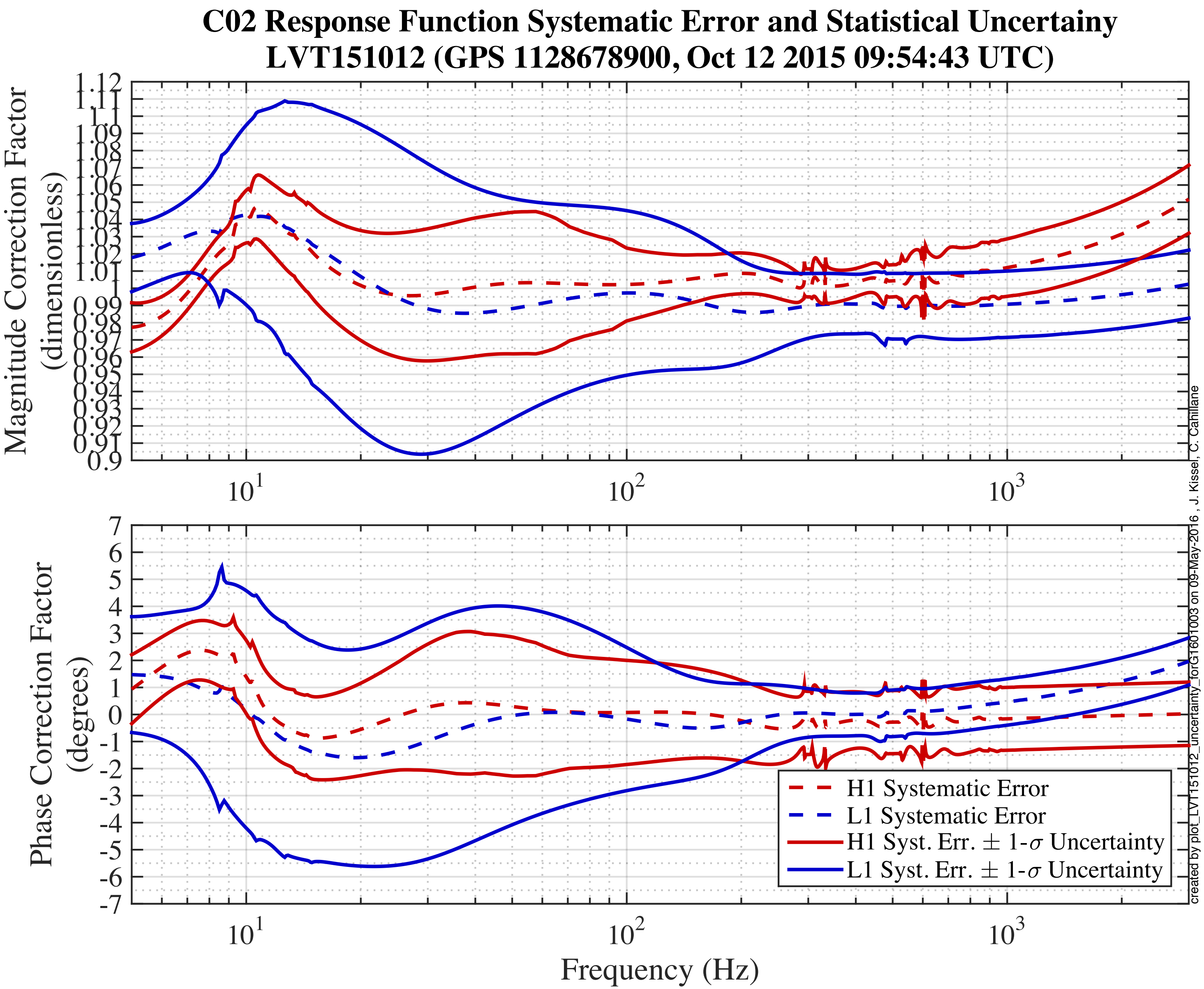
  for Boxing Day Paper.
- There is still more work to do to refine 1-4.5 kHz for other searches
- All of O1 uncertainty is a function of time. We have plots to show how it evolves. We wait for discussion with search groups on how to fold this information into analysis (e.g. Rates group and BBH Paper).

Appendix A:
Overall C02 Uncertainty Plots for GW150914, LVT151012, and GW151226

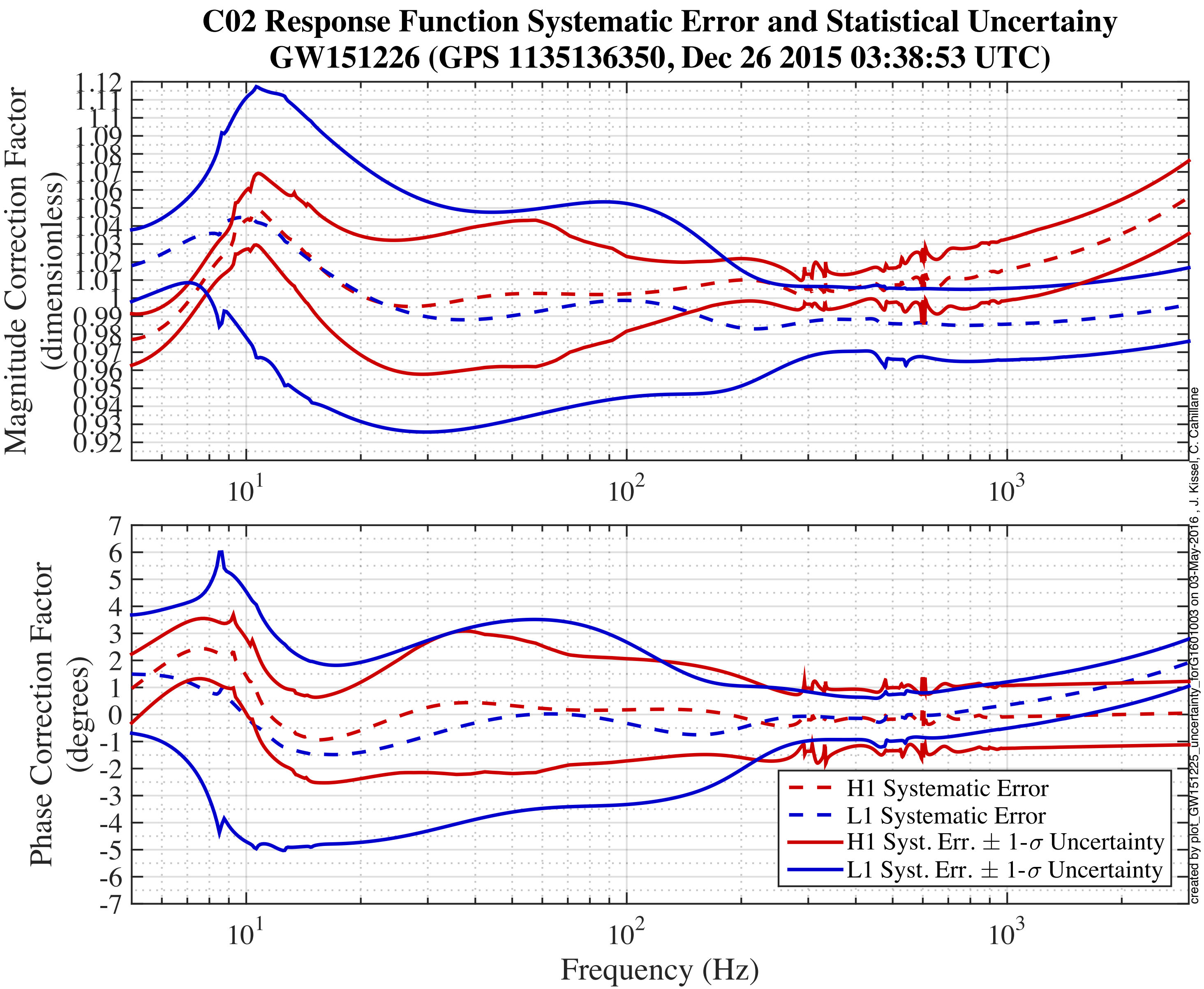
GW150914
C02 Total Uncertainty and Error



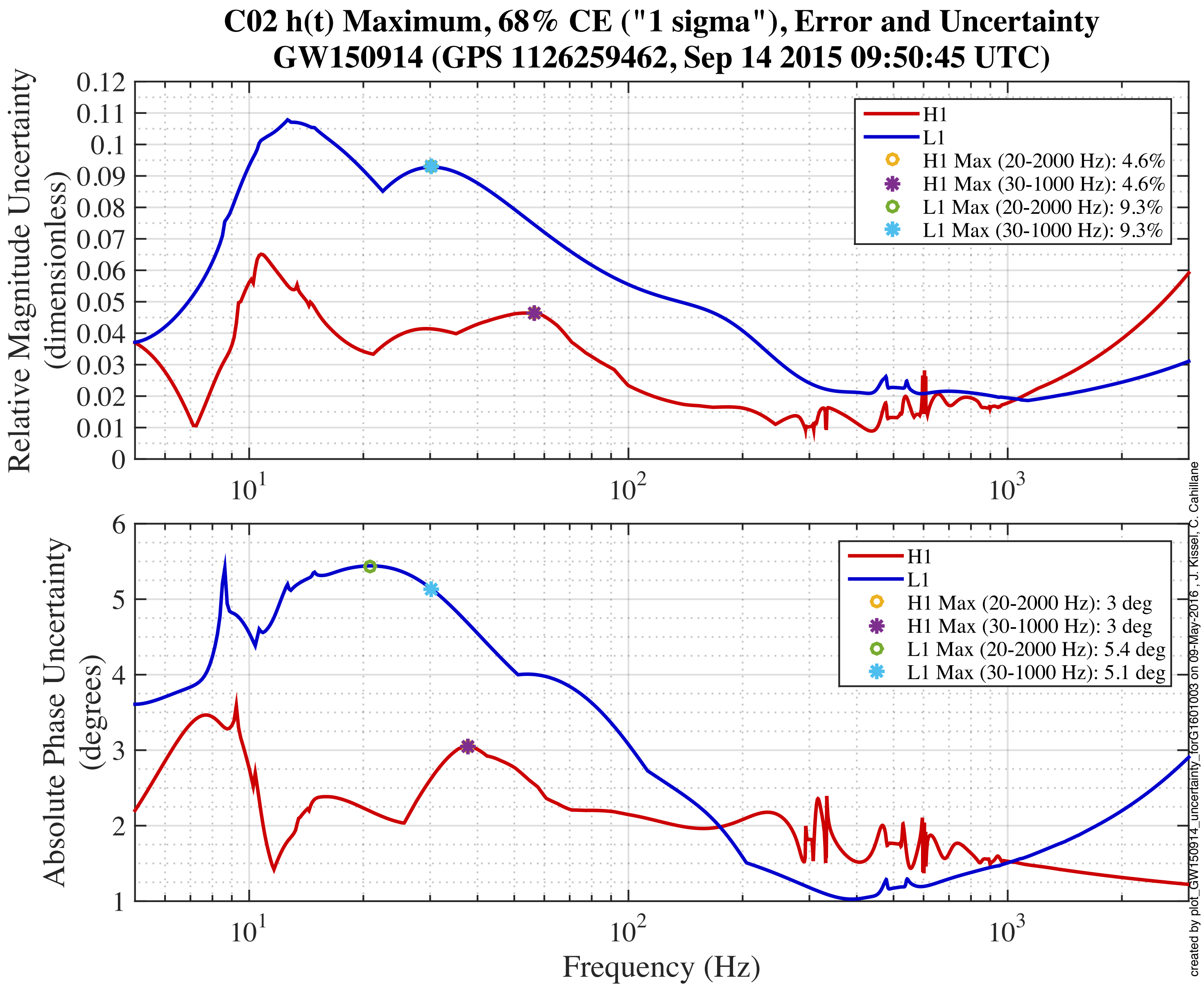
LVT151012
C02 Total Uncertainty and Error



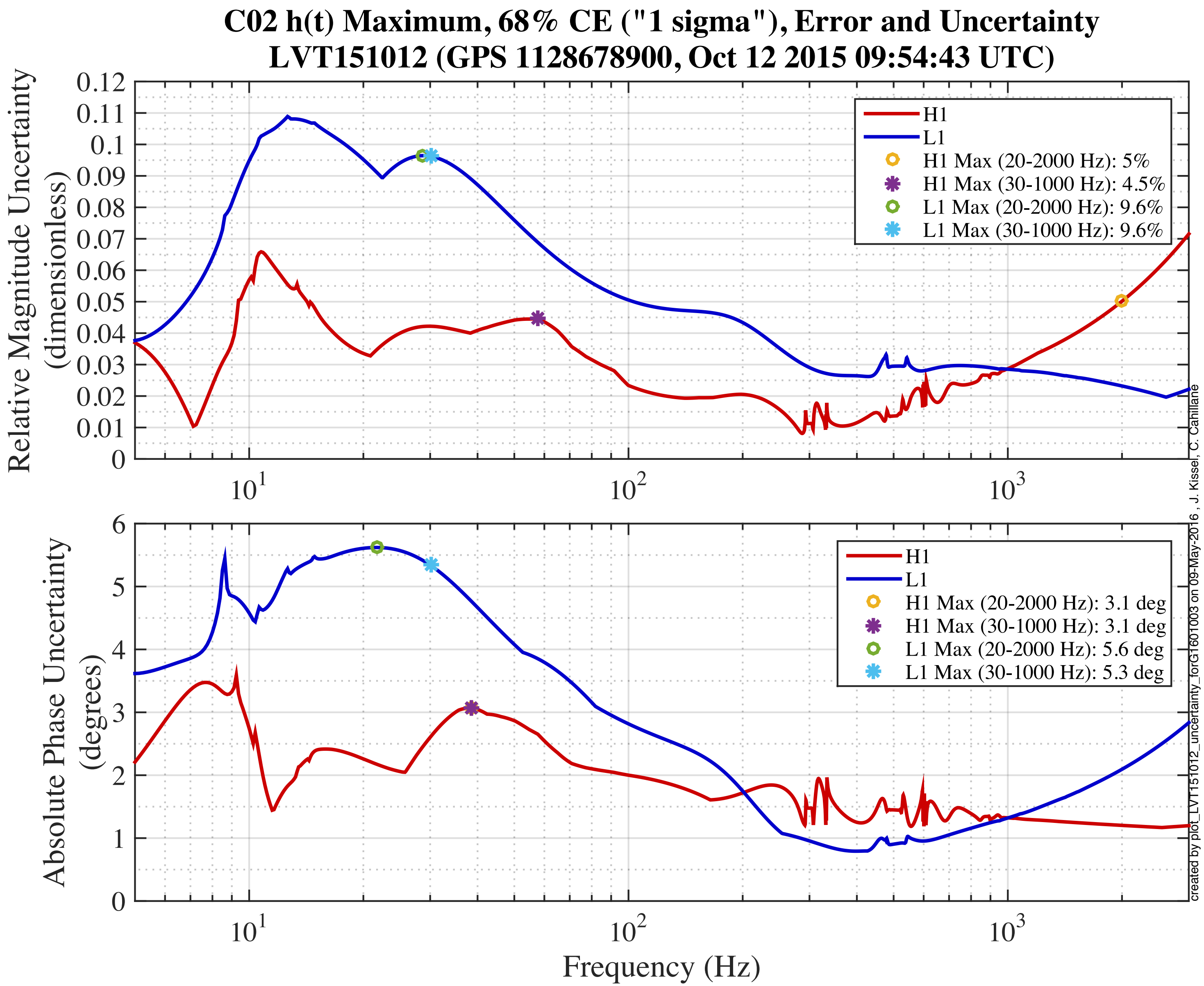
GW151226
C02 Total Uncertainty and Error



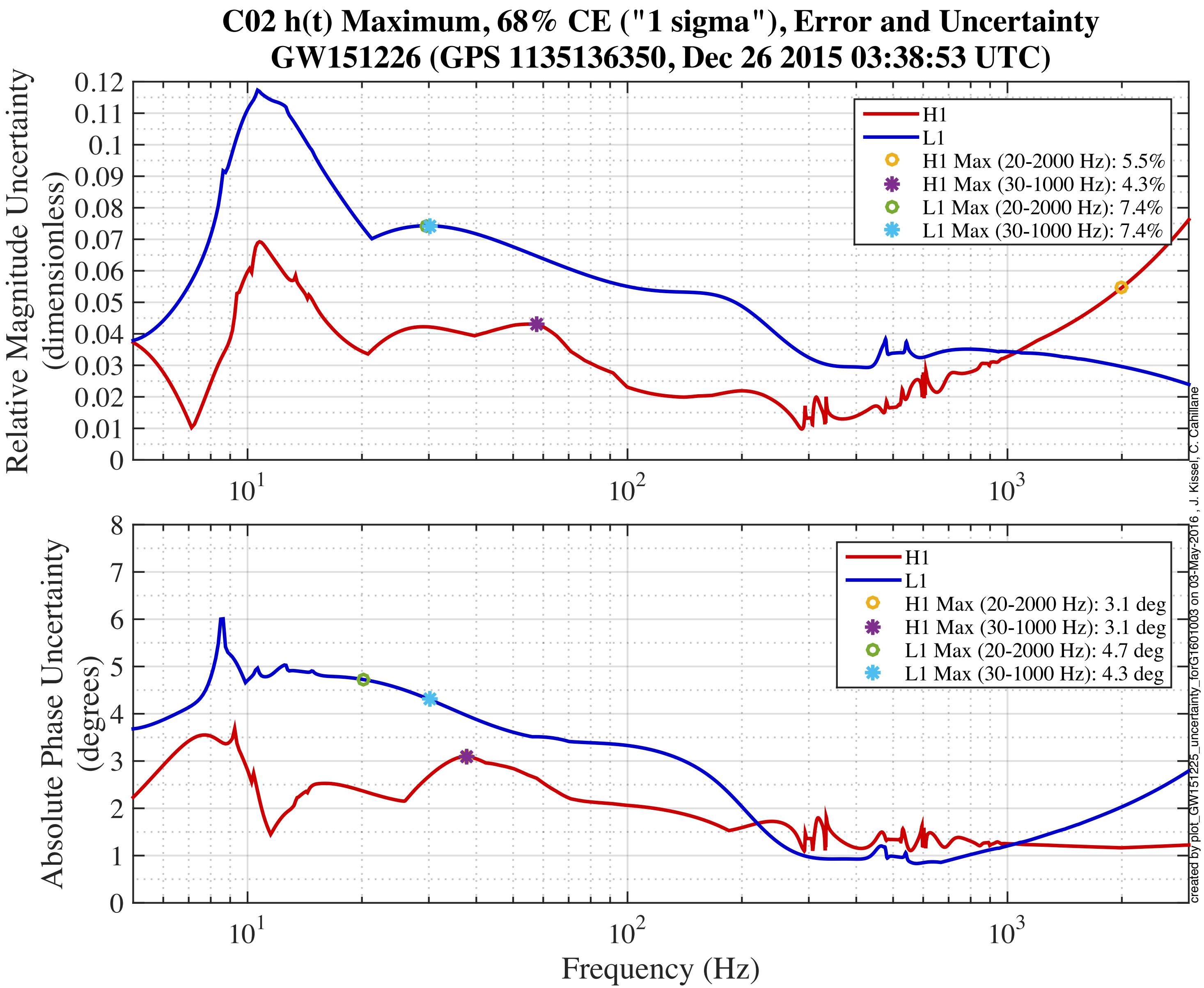
GW150914
C02 Maximum 1-σ Error and Uncertainty



LVT151012
C02 Maximum 1-σ Error and Uncertainty



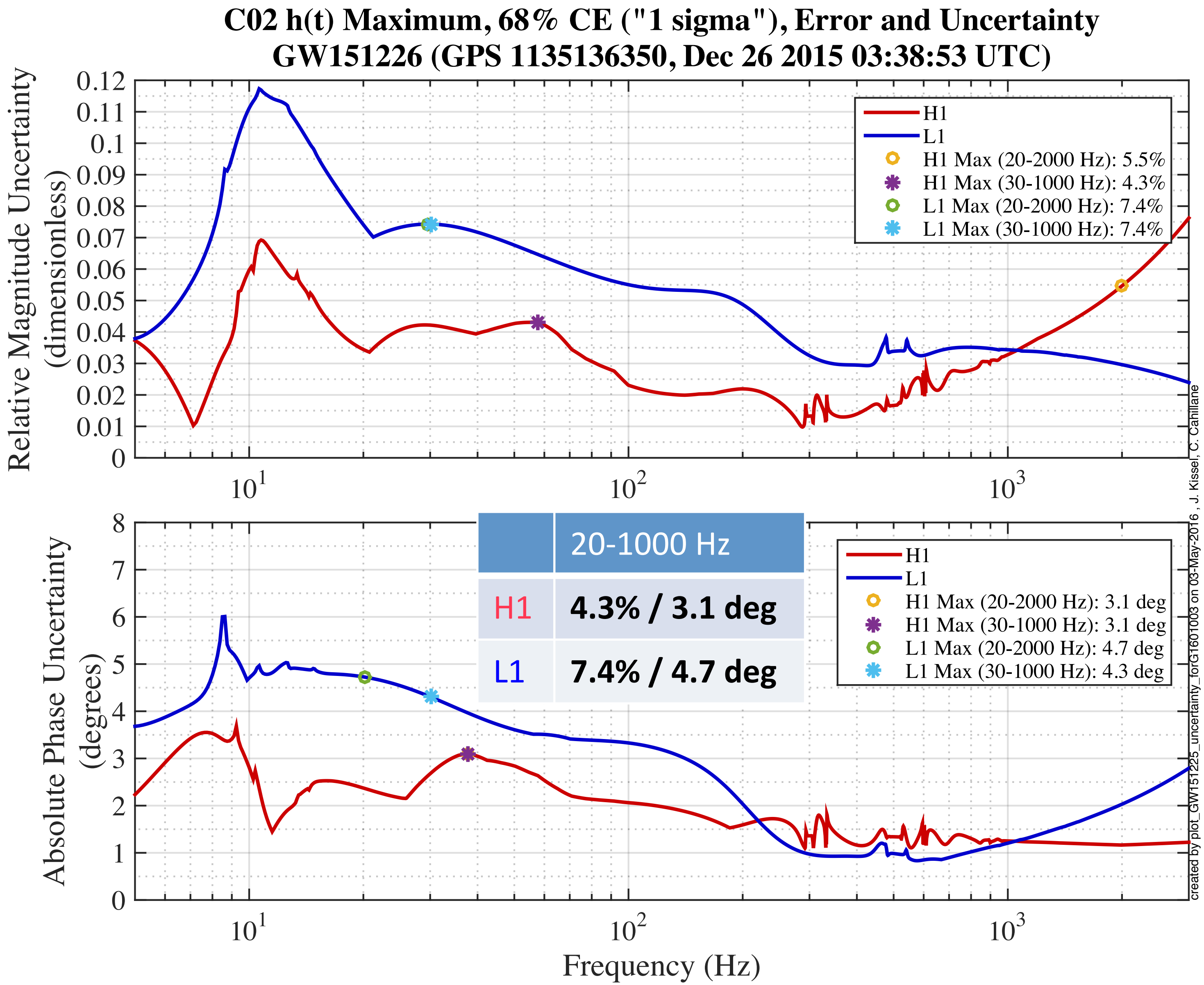
GW151226
C02 Maximum 1-σ Error and Uncertainty



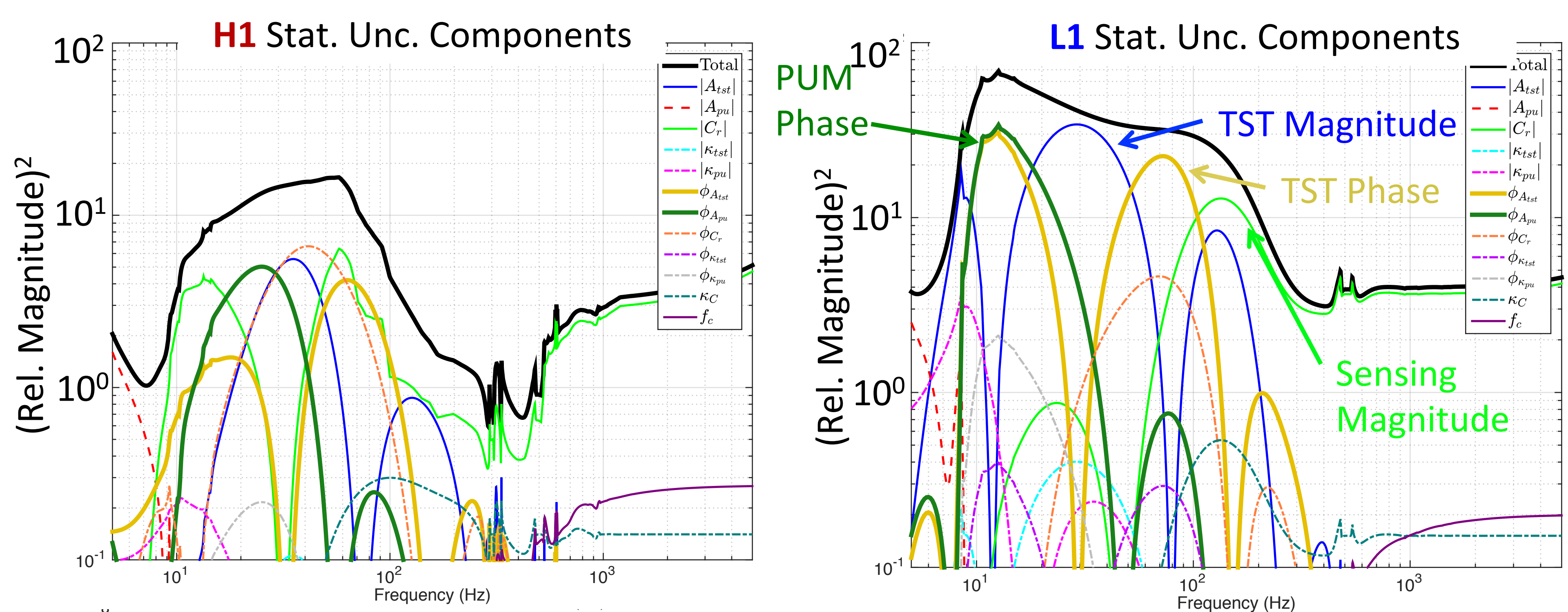
Appendix B:
Why is **L1** Uncertainty Larger than **H1**?

Why is L1's statistical Uncertainty Larger than H1's?

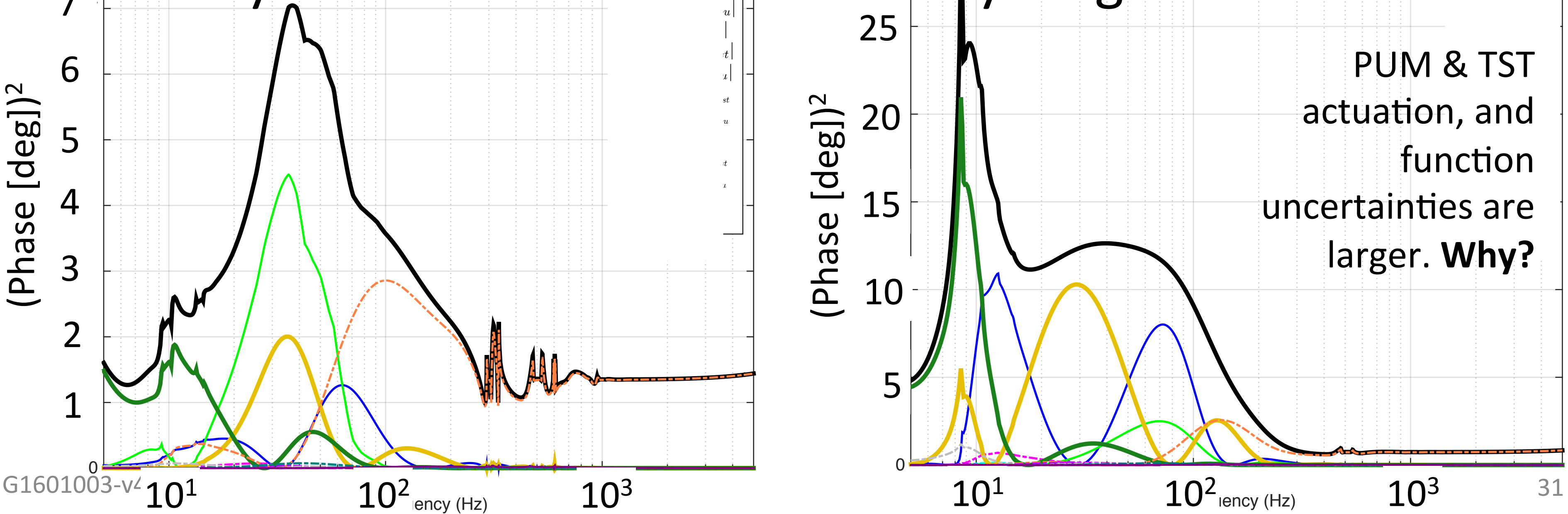
Let's follow the Rabbit Down the Hole, taking GW151226 as the example…



	20-1000 Hz
H1	**4.3% / 3.1 deg**
L1	**7.4% / 4.7 deg**

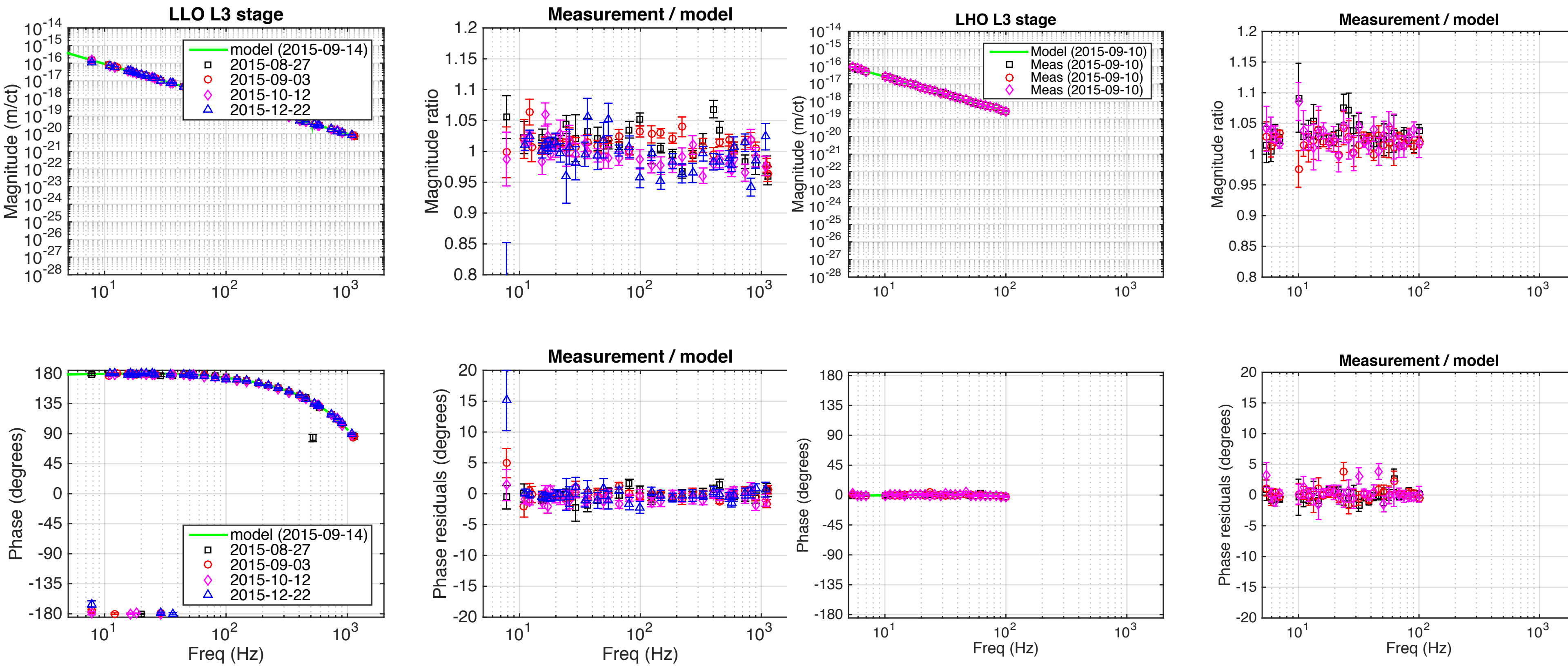


Why is L1's Statistical Uncertainty Larger than H1's?



Why is L1's Statistical Uncertainty Larger than H1's?

NOT because of less measurements!



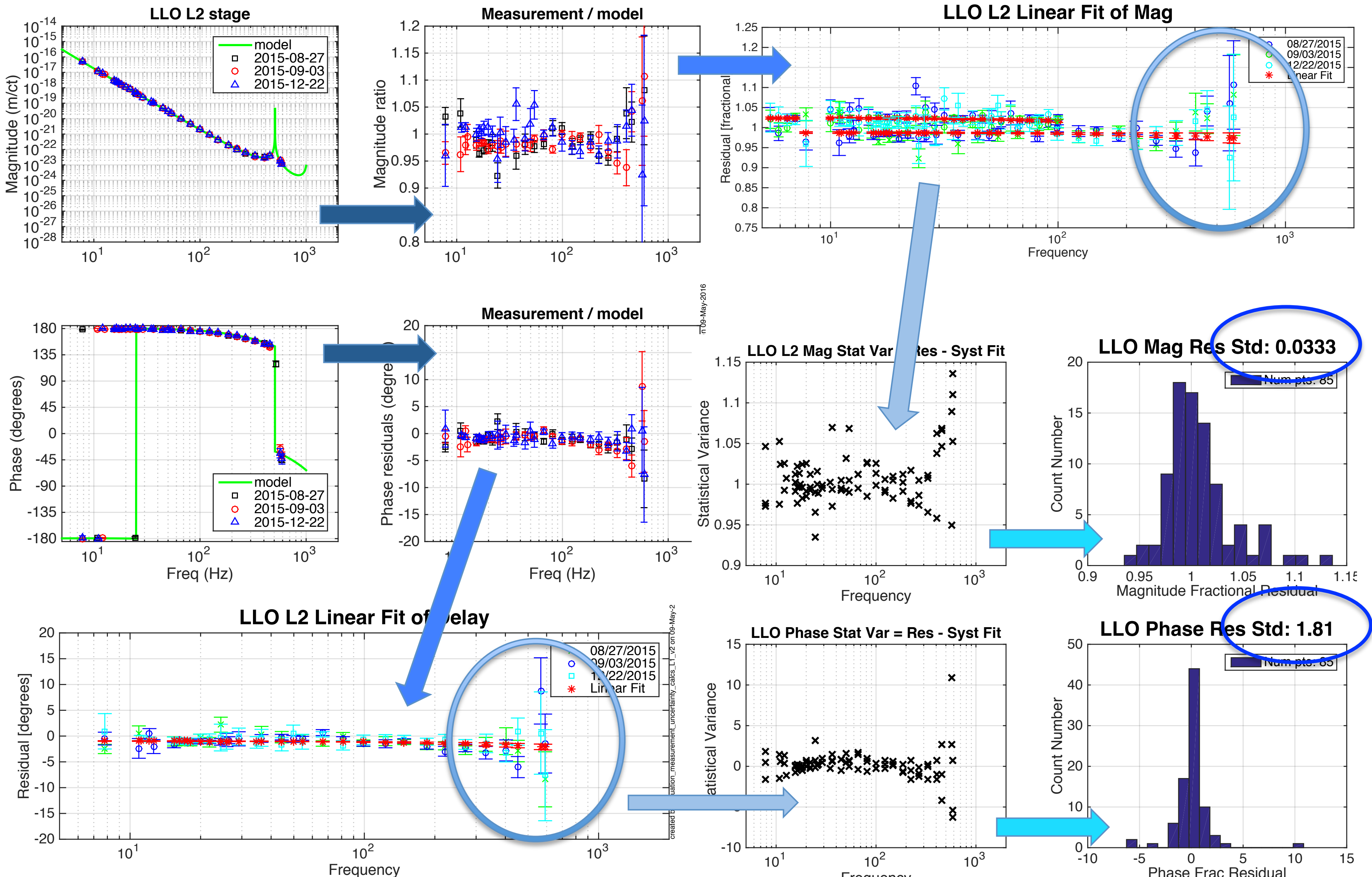
Scatter doesn't too different either!

So why?

- Because L1 took data out to higher frequency, and the residual systematic error between measurements was not consistent
- That means *overall* systematic error removed still is non-Gaussian and just taking "std" of residual-with-overall-systematic-removed is a poor estimate, an over estimate at best, of statistical uncertainty.

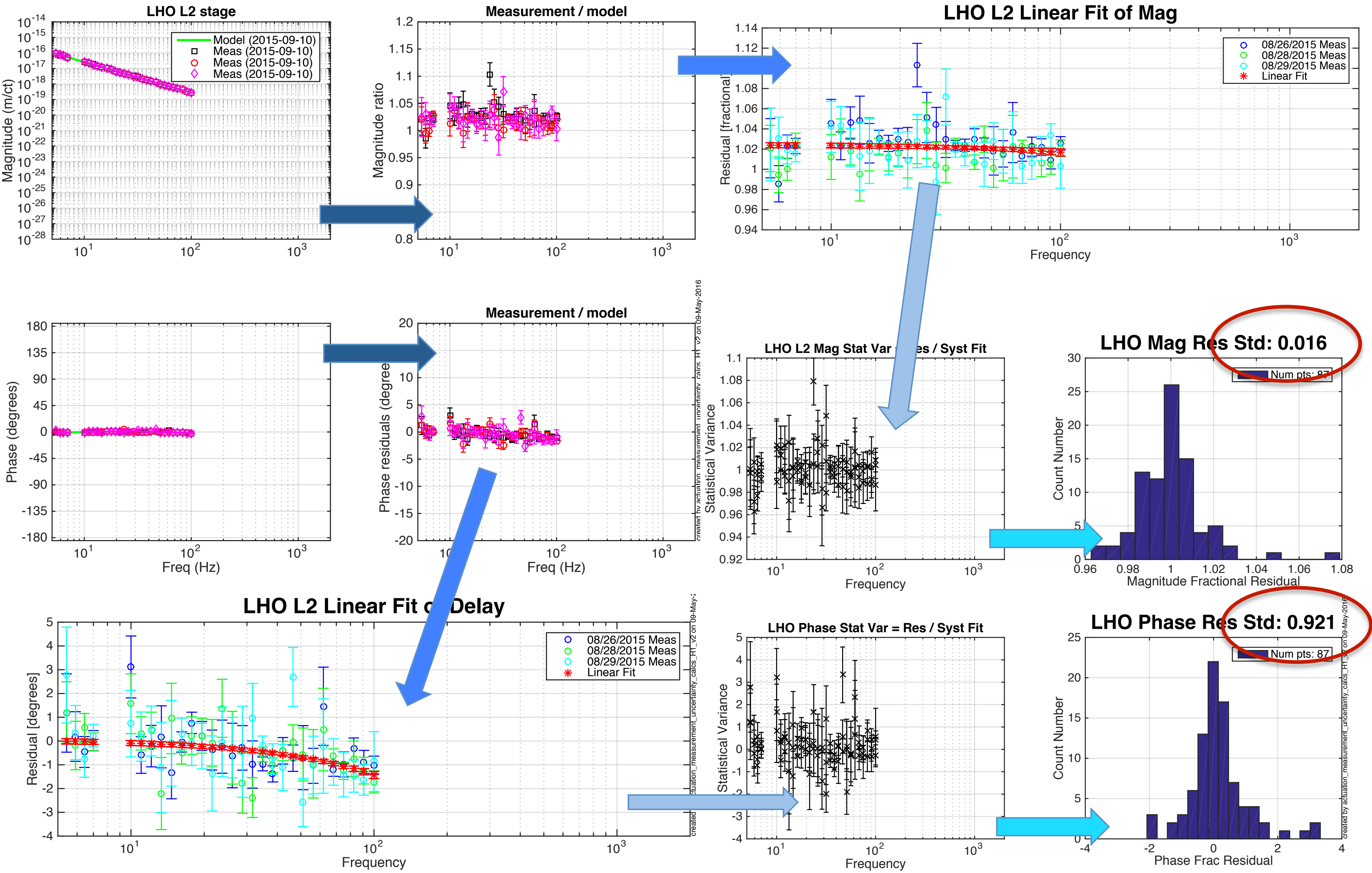
Why is L1's Statistical Uncertainty Larger than H1's?

Because of L1 PUM violin-mode systematics that aren't modeled well.



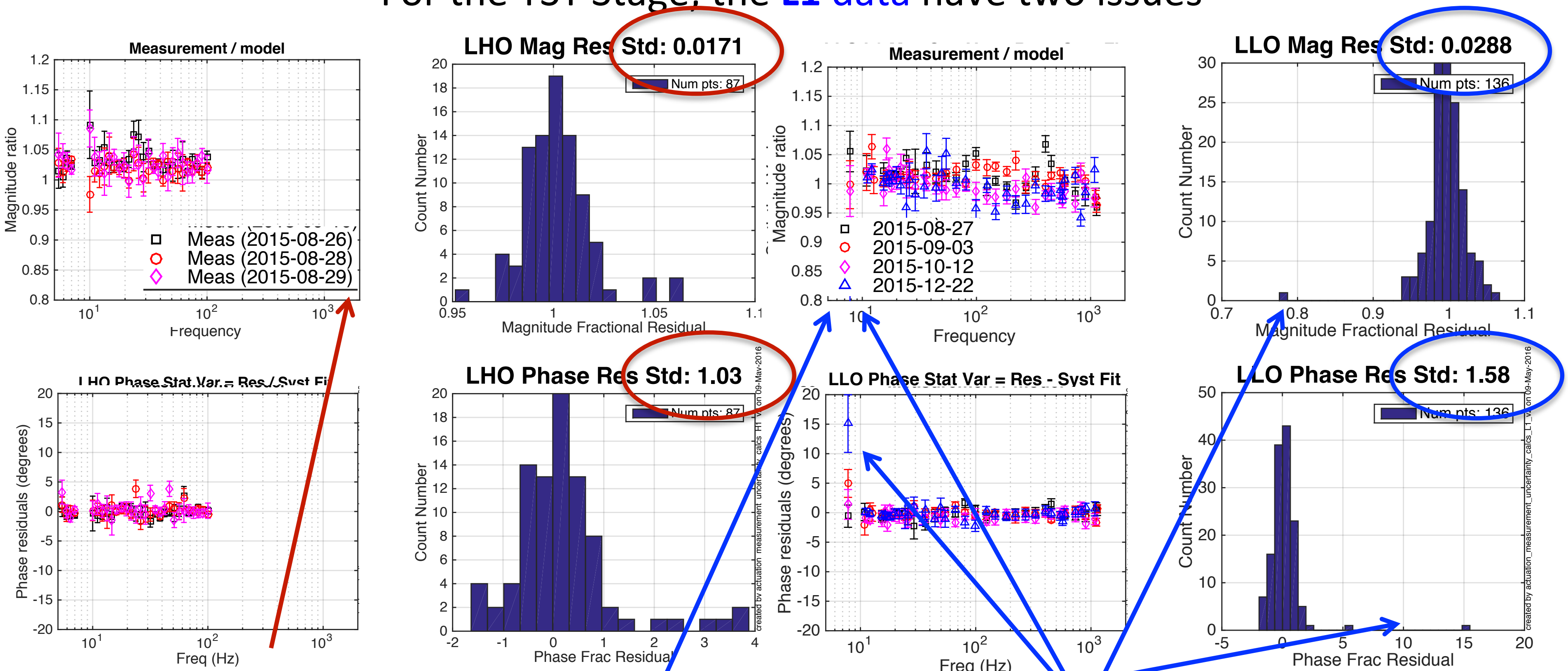
Why is L1's Statistical Uncertainty Larger than H1's?

Compare against H1 PUM results which have much more Gaussian fluctuations.



Why is L1's Statistical Uncertainty Larger than H1's?

For the TST Stage, the L1 data have two issues



H1 measurements tightly grouped in time
L1 measurements spread over months

The low frequency outliers go out to 25% and 15 deg, doubling the standa deviation.

L1 TST Stage

- Standard deviation of *all* (residual – systematic error) frequency points from *all* measurements is used as **frequency-independent** statistical uncertainty.
- **L1** Low-frequency outliers skew the standard deviation
- **L1** data taken over months, and not compensated for known actuation strength change. Makes measurement-to-measurement non-Guassian, which means std is not a good estimate of uncertainty

Why is L1's Statistical Uncertainty Larger than H1's?

Actuation Summary

Actuation Summary:

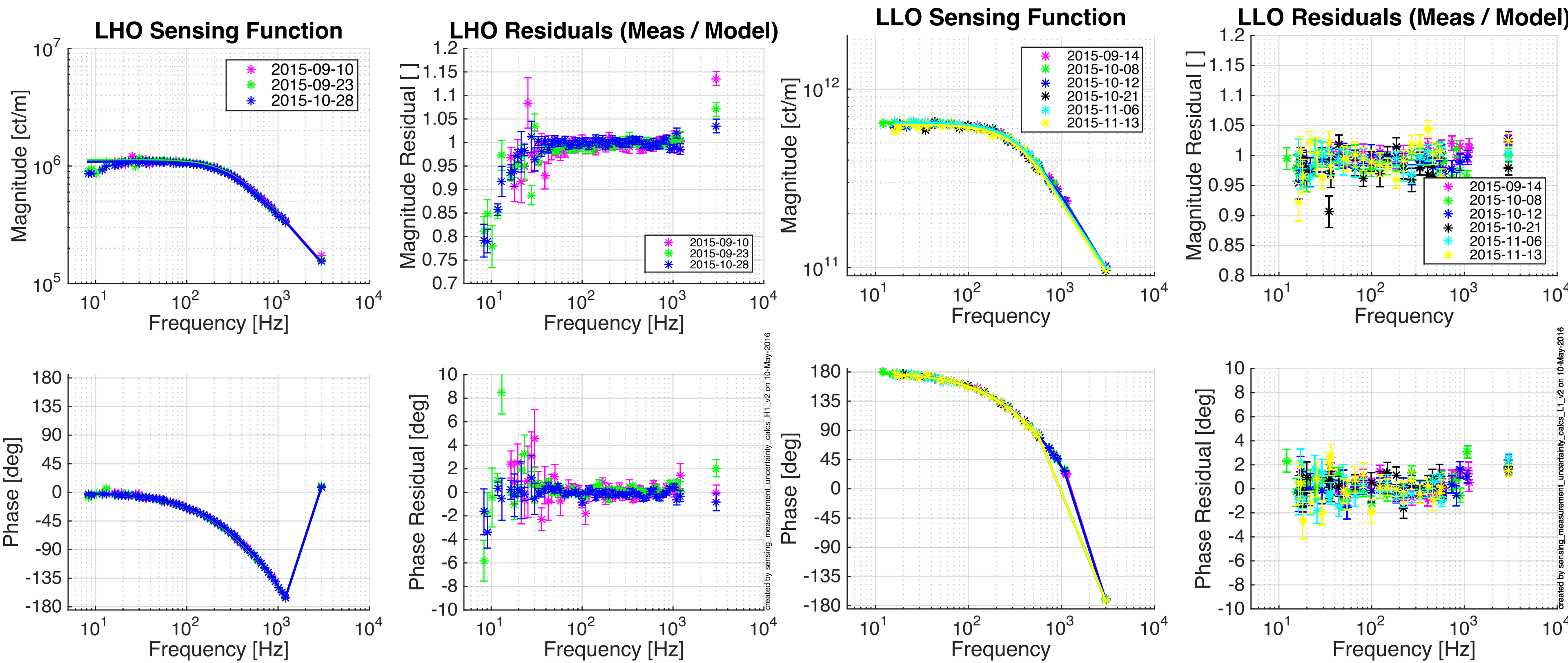
- Standard deviation of all (residual – systematic error) frequency points is used as **frequency-independent** statistical uncertainty.
- **L1 TST** and **PUM** have a factor of 2 larger uncertainty than **H1** **for different reasons**:
    - **PUM** because of high-frequency violin mode systematics between measurements at **L1** that **H1** just didn't measure
    - **TST** because **L1** it has low frequency outliers which skew the standard deviation
    - **L1 TST** also larger, because 4 measurements were taken over 4 months and not compensated for known actuation strength change (**H1** data all taken within 1 week.)

Advice for the future:

- Weight frequency points contributing actuation coefficient uncertainty by
    - coherence
    - distribution filter / contribution to over all actuation strength

Why is L1's Statistical Uncertainty Larger than H1's?

How about the Sensing Function Uncertainty?



Again – **not** at all an issue of lack of measurements.

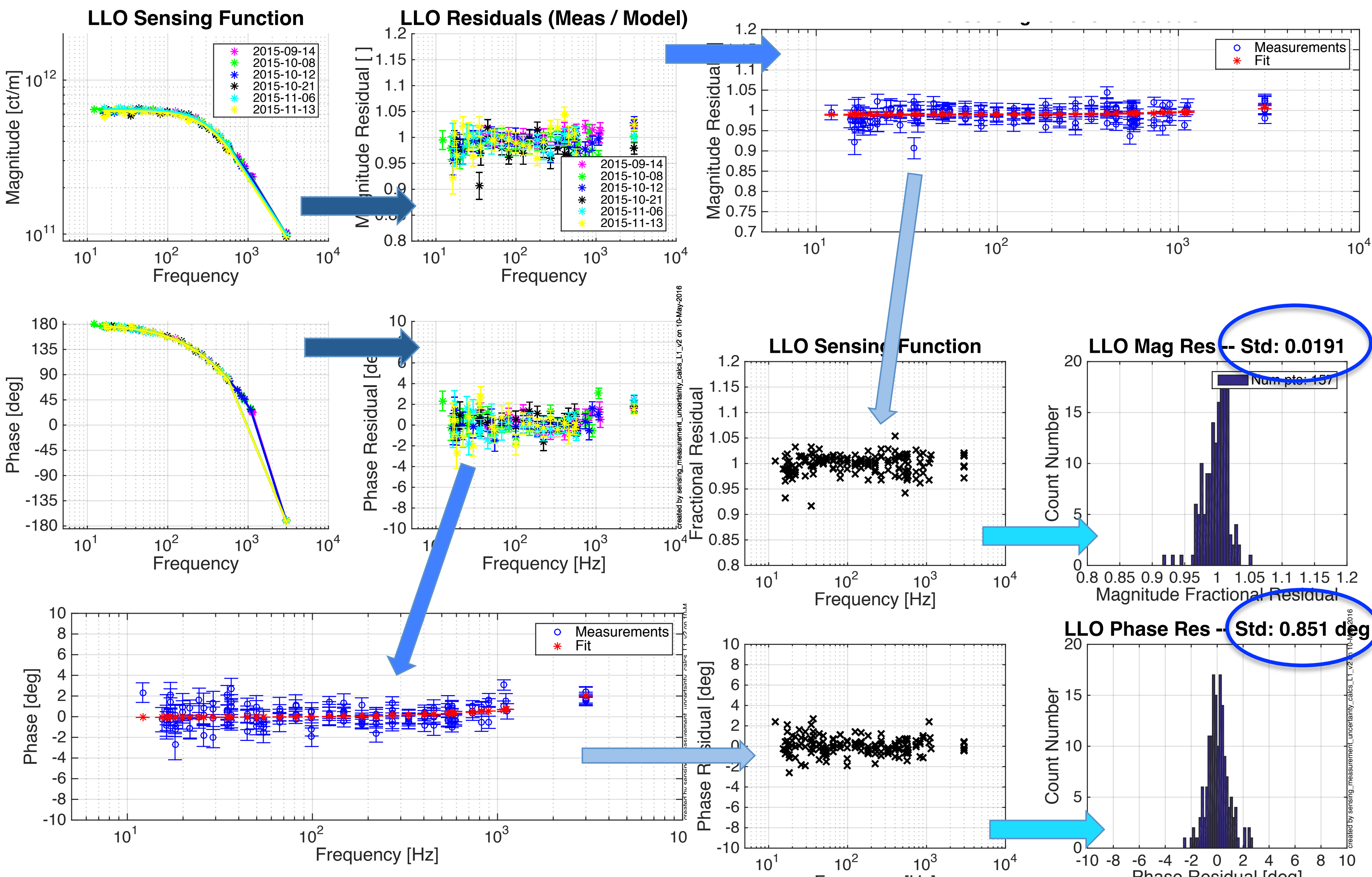
(Actually, H1's uncertainty estimation missed measurements from 2015-10-15, 2015-12-01, and 2015-12-21 ☹)

Why, then?

Because of H1's non-linear magnitude residual, a "rolling standard deviation" was used, so the uncertainty is much smaller in the bucket.

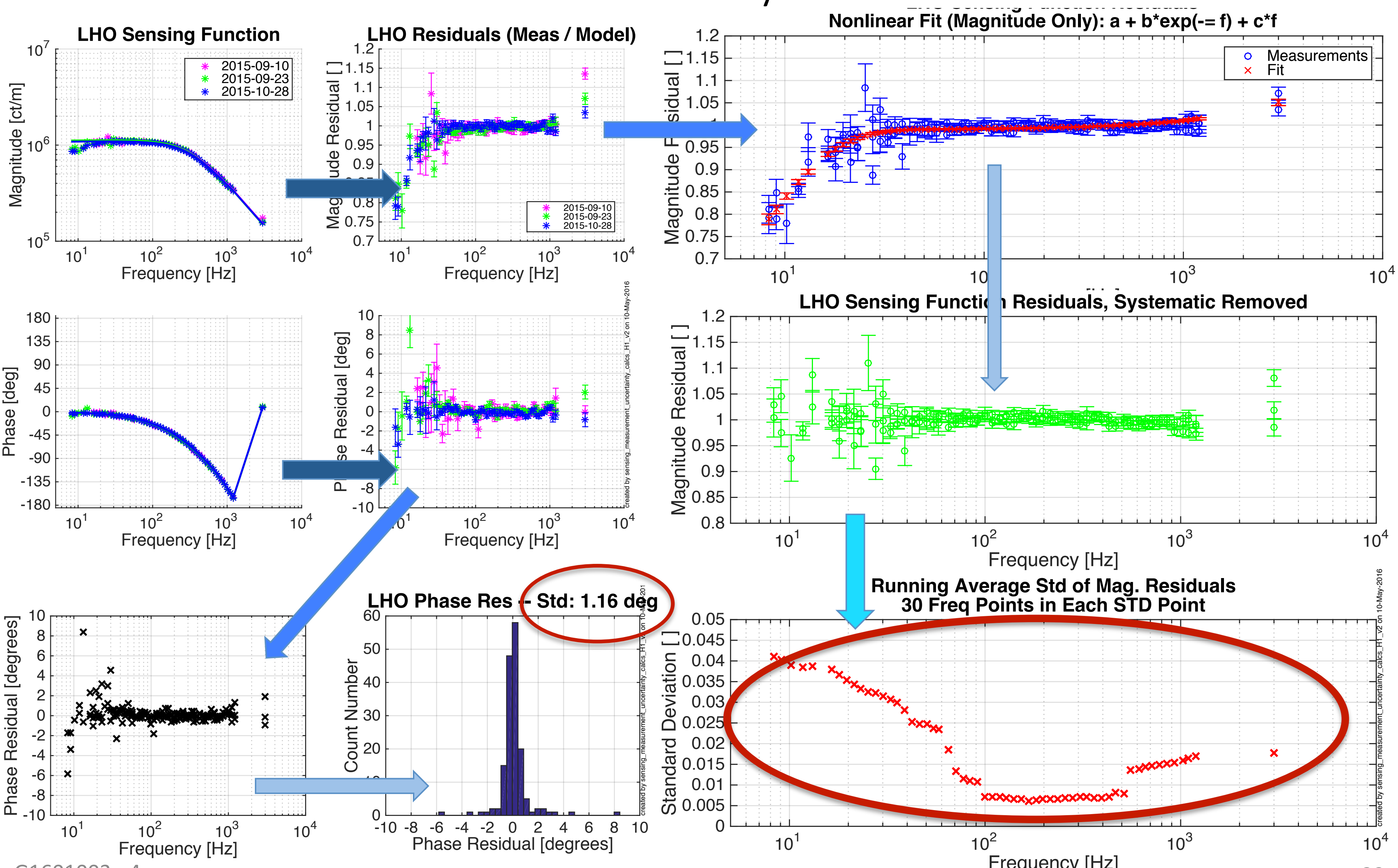
Why is L1's Statistical Uncertainty Larger than H1's?

L1's sensing uncertainty was treated exactly like actuation uncertainty:



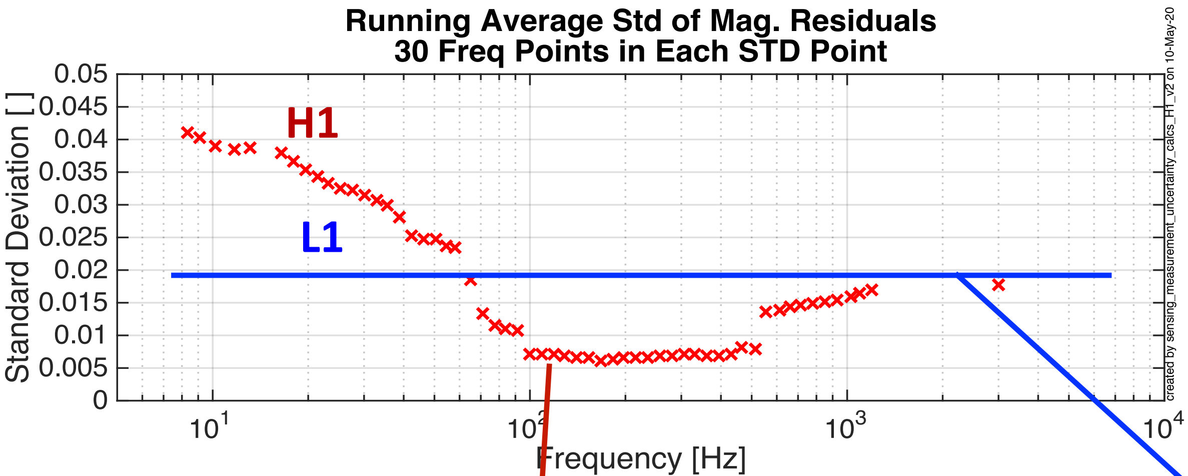
Why is L1's Statistical Uncertainty Larger than H1's?

H1's sensing's magnitude systematic error and uncertainty were treated totally differently:



Why is L1's Statistical Uncertainty Larger than H1's?

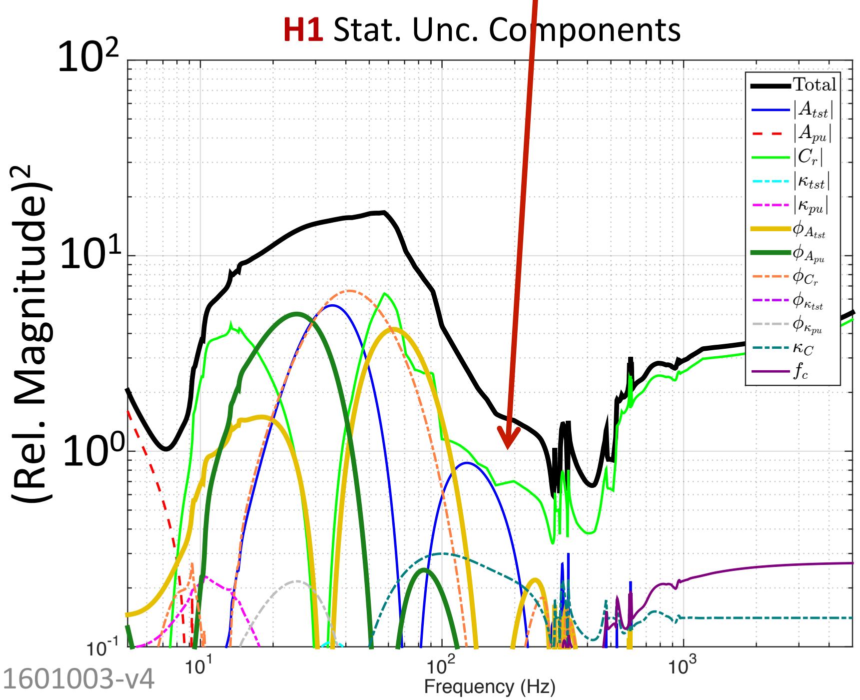
H1's sensing function uncertainty is estimated totally differently than L1

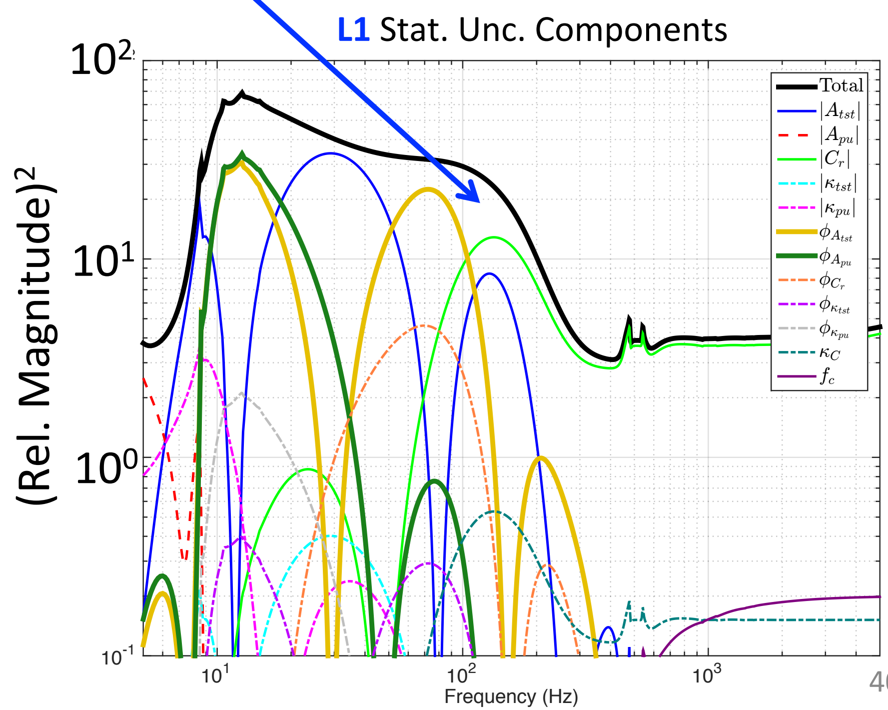


Sensing Summary:

- Differences in assessment of uncertainty are essential incomparable, unfortunate for L1

- Tough to say if H1 is under or L1 is over estimated





Why is L1's Statistical Uncertainty Larger than H1's?

Conclusions

- Kissel had claimed that **L1**'s statistical uncertainty suffered from too few measurements. **This is just not true** (I hear you now Shivaraj!)
- For Actuation **L1** took (PCAL Actuation) data out to higher frequency, which (for the PUM stage) includes violin modes.
- Revisiting the data, **L1's PUM** uncertainty is large because systematic errors (from PUM stage violin modes) were not consistent between data sets. Remaining systematic after subtracting out an overall fit to all measurements still left significant, non-Gaussian residual spread, miss-construed as large statistical fluctuation.
- **L1's TST** stage measurements have a few low-frequency outliers (remaining systematics?) that doubled the standard deviation w.r.t. **H1**, and includes 4 months worth of actuation strength change making measurement-to-measurement results non-Gaussian.
- Sensing function uncertainty is treated differently between sites. Could say **L1** is over estimated, or **H1** is underestimated.

For the future:

- Both sites *need* to measure out the the same frequency band (H1 should go higher!)
- If combining multiple measurements for TST, make sure clustered tightly in time or compensated for long-term drift.
- Need to look into a better treatment of residuals for determining statistical uncertainty that is less sensitive to systematic error fits, time-dependence and outliers.
- Need to treat the uncertainties the same between sites. If we see inconsistencies between measurements – investigate why!