Thank you for your thorough reading of this paper. Your comments were helpful, and led to several changes in the paper text and figures, listed below in response to your comments:

****
** My main comments which I think deserve to improve the text are
about the use of H_{PCAL}.

It is defined at the end of section II-C-1 as the photon calibrator
actuation strength. Then it is redefined in section II-C-2 as a
relative uncertainty in the strength.
--> The first definition of H_PCAL is not used later. Hence it can
probably be removed to keep only the relative uncertainty.

--> in this same paragraph about how H_PCAL modifies the measured C
and A, it is not clear what the arrow means, from which to which
measurement. Reading it, I have the feeling that the formulas are in
the wrong way, but probably it is only due to a missing explanation.
Something like:
- C_meas is measured assuming some PCAL actuation strength
- assume that in reality the actuation strength is larger
(i.e. H_PCAL > 1)
- then C_true = C_meas / H_PCAL
But it is written wrote: Cmeas -> H_PCAL* Cmeas
Similar logic for A_meas.
Can the authors check and possibly better explain this part?

- Moved discussion of H_PCAL in section II-C-2 to section II-C-1 for more clarity and motivation for including the correction factor, changed H_PCAL to H_PCAL(t) everywhere for clarity, and removed old Equations 19, 20, and 21 with the confusing arrow notation by including H_PCAL(t) in Equation 14, 15, and 17.

****
** Below are a few minor comments:

- update the date of the Observing Run O2 in the introduction.
  - Updated the dates for the start and ends of O1 and O2

- add somewhere in the section II the frequency range of the needed
models for sensing and actuation
  - Updated section II with detector GW frequency sensitivity range (10-5000 Hz)
  - Updated section II with frequency range the DARM model transfer functions are modeled and measured at (5-5000 Hz)

- in section II-A, at the very end, delta C^{GP}(f) is defined by GP
is not (it is later at the end of section II-B).
  - Rearranged the explanation for \delta C^{GP}(f) so it's clear that GP = Gaussian Process regression

- section II-B: can "small enough" to neglect quadratic term be
quantified?
  - Added to section II-B a better explanation of electrostatic drive linearity by explaining quadratic terms are not actually neglected. Changes in the linear response originate from charging on the test mass, which changes the relative voltage between the controller and test mass, a quadratic effect. We track changes in the linear response over time using kappa_T(t).

- fig. 3 and fig.4 : it may help to add the dates of the reference
measurements for H1 and L1 (as already done in tables 1 and 2).

- Put the reference measurement dates in Figure 3 and 4's captions.

- fig.4: if the UIM and PUM models are useful only at low frequency, where there are data, there model could be shown only in this frequency range (<100 Hz and <600 Hz for example). It could help reading the figure at high frequency in particular.
  - Changed Figure 4 plots cut off the phase at 300 Hz for UIM actuation and 400 Hz for PUM actuation.

- section III-B: in the first paragraph, a word is missing after "MCMC method"
  - Updated section III-B to fix "MCMC method" sentence.

- fig. 6: can LHO be replaced by H1?
  - Changed Figure 6 plot to make title say "H1" not "LHO"

- fig. 7: it can be useful for the reader to recall that it is the same plot at fig 3-left, but gathering all the collected data, not only the reference data. And maybe give the period of the data collection.
  - Changed Figure 7 plot to match color scheme from Figure 3, and altered text to reference Figure 3 at the end of section III-B.

- section III-C:
after the enumeration, it may be useful to add:
1 and 2 from the MCMC ensemble results --> + on the reference measurements
3 and 4 from the GP regressions on the residuals --> + to incorporate unknown syst. errors

- section III-D: stocastic (missing 'h')
  - In sections III-C and III-D, took suggested wording changes.

****
** A few questions, that may not require paper update:

- in fig.1, the electrostatic actuators are called High Range on X end test mass and Low Noise on Y end test mass.
Are there really different and not used symmetrically?

Each optic has the same hardware, but are operated differently. End Test Mass Y is switched from high range to low noise control during observing time. End Test Mass X could be used for this, but it is not necessary. The differential arm (DARM) degree of freedom only requires only one optic to be actively controlled to hold the interferometer in it's on-resonance, no-light-in-the-antisymmetric-port state. This raises the question: how do we keep both arm cavities on resonance if we push on only one optic? We also control the common arm (CARM) degree of freedom with a much higher bandwidth controller (~ 20 kHz, as compared to the DARM loop bandwidth of ~40 Hz). The CARM loop is controlled by changing the laser frequency: any time DARM is altered from ETMY actuation, CARM instantly compensates for DARM's changes, keeping both arms on resonance. For more info on the aLIGO control loops, see Chapter 2 of Evan Hall's thesis: https://thesis.library.caltech.edu/10031/.

- in section II-A, about the sensing time delay tau_C:
it includes the light travel time of the 4 km arm and some compensation for using a simplified arm cavity model.
The given time delay is 12 μs.
The travel in the arm being ~13 μs, I understand that the compensation is 1 μs

only?
Is it right? Up to which frequency is it valid?

I mistakenly reported the 12 us as the total sensing delay. The true sensing model time delay is ~77.6 microseconds. 13.3 us delay from the light travel time, 76.3 us delay from digital acquisition, and -12 us time advance added to compensate our single pole approximation of a Fabry Perot response. In LIGO DCC G1501316, we have found that the single pole approximation with the artificial time advance is good to within a percent at 3 kHz, and good to within 3% at 5 kHz, but falls off in accuracy quickly after that. The single pole approximation must be investigated more thoroughly if calibration uncertainties and errors are to be further reduced.

- section II-B: why is there only one facteur $k_{PU]$ for both upper-intermediate and penultimate stages time variations, can't they change independently?

There is only one factor of k_PU(t) because the electromagnetic actuation strength on the penultimate stage (PUM) and upper-intermediate (UIM) stage doesn't change significantly over the run. If needed, they could be separated, but this has not yet proved to be necessary.

- section II-B: "the electro-mechanical transfer functions for each stage are independently measured and included in the model with negligible uncertainty".
    --> in which frequency range are they measured? does it include some low frequency resonant frequencies? if yes, can't these response/frequencies vary with time and then is it taken into account somewhere?

We do have measurements and models of the LIGO suspensions, and have mapped out all their resonances and quality factors. The main pendulum resonance frequencies are 1 Hz and below. Our actuation measurements A_i(f) begin at 5 Hz, in Figure 4. We do not expect the mechanical response of the suspensions to change over time, nor the electronics actuating upon them, except for the electrostatic drive strength due to charge buildup on the test mass.

- section II-C-2: for the measurement of 1/(1+G), I suppose that what is really measured is G, adding some excitation 'post' in the loop just after d_err or d_ctrl probes and measured the transfer function d_err/d_err_post?

Yes. In LIGO we often take OLG transfer functions via a swept sine IN1/IN2 method. IN1 is our error signal, d_err, to which an excitation EXC is applied, giving an IN2 = IN1 + EXC. From some basic loop algebra we find IN1/IN2 = OLG.

All Changes to Uncertainty Methods paper for PRD:

- Moved discussion of H_PCAL in section II-C-2 to section II-C-1 for more clarity and motivation for including the correction factor, changed H_PCAL to H_PCAL(t) everywhere for clarity, and removed old Equations 19, 20, and 21 with the confusing arrow notation by including H_PCAL(t) in Equation 14, 15, and 17.
- Updated the dates for the start and ends of O1 and O2
- Updated section II with detector GW frequency sensitivity range (10-5000 Hz)
- Updated section II with frequency range the DARM model transfer functions are modeled and measured at (5-5000 Hz)
- Added to section II-B a better explanation of electrostatic drive linearity by explaining quadratic terms are not actually neglected. Changes in the linear response originate from charging on the test mass, which changes the relative voltage between the controller and test mass, a quadratic effect. We track changes in the linear response over time using kappa_T(t).

- Changed the sensing model time delay in section II-A from 12 microseconds to 77.6 microseconds. See the answer to question 2.
- Updated section III-B to fix “MCMC method” sentence.
- Rearranged the explanation for \delta C^{GP}(f) so it’s clear that GP = Gaussian Process regression
- Put the reference measurement dates in Figure 3 and 4’s captions.
- In the caption of Figure 7, emphasized the measurements come from the entirety of O2.
- In sections III-C and III-D, took suggested wording changes.
- Changed Figure 9 plots to include August calibrations in All of O2 C01 plots.
- Changed Figure 4 plots cut off the phase at 300 Hz for UIM actuation and 400 Hz for PUM actuation.
- Changed Figure 6 plot to make title say “H1" not “LHO”
- Changed Figure 7 plot to match color scheme from Figure 3, and altered text to reference Figure 3 at the end of section III-B.