

# Characterization of systematic error in Advanced LIGO calibration

Ling Sun<sup>1</sup>, Evan Goetz<sup>2</sup>, Jeffrey S. Kissel<sup>3</sup>, Joseph Betzwieser<sup>4</sup>,  
Sudarshan Karki<sup>5</sup>, Aaron Viets<sup>6</sup>, Madeline Wade<sup>7</sup>,  
Dripta Bhattacharjee<sup>5</sup>, Vladimir Bossilkov<sup>8</sup>, Pep B. Covas<sup>9</sup>,  
Laurence E. H. Datrier<sup>10</sup>, Rachel Gray<sup>10</sup>,  
Shivaraj Kandhasamy<sup>11</sup>, Yannick K. Lecoeuche<sup>3</sup>,  
Gregory Mendell<sup>3</sup>, Timesh Mistry<sup>12</sup>, Ethan Payne<sup>13</sup>,  
Richard L. Savage<sup>3</sup>, Alan J. Weinstein<sup>1</sup>, Stuart Aston<sup>4</sup>,  
Aaron Buikema<sup>14</sup>, Craig Cahillane<sup>1</sup>, Jenne C. Driggers<sup>3</sup>,  
Sheila E. Dwyer<sup>3</sup>, Rahul Kumar<sup>3</sup>, and Alexander Urban<sup>15</sup>

<sup>1</sup>LIGO, California Institute of Technology, Pasadena, CA 91125, USA

<sup>2</sup>University of British Columbia, Vancouver, BC V6T 1Z4, Canada

<sup>3</sup>LIGO Hanford Observatory, Richland, WA 99352, USA

<sup>4</sup>LIGO Livingston Observatory, Livingston, LA 70754, USA

<sup>5</sup>Institute of Multi-messenger Astrophysics and Cosmology, Missouri Institute of Science and Technology, Rolla, MO 65409, USA

<sup>6</sup>Concordia University Wisconsin, 12800 N Lake Shore Dr, Mequon, WI 53097, USA

<sup>7</sup>Kenyon College, Gambier, OH 43022, USA

<sup>8</sup>OzGrav, University of Western Australia, Crawley, Western Australia 6009, Australia

<sup>9</sup>Universitat de les Illes Balears, IAC3—IEEC, E-07122 Palma de Mallorca, Spain

<sup>10</sup>SUPA, University of Glasgow, Glasgow G12 8QQ, UK

<sup>11</sup>Inter-University Centre for Astronomy and Astrophysics, Pune 411007, India

<sup>12</sup>The University of Sheffield, Sheffield S10 2TN, UK

<sup>13</sup>OzGrav, School of Physics & Astronomy, Monash University, Clayton 3800, Victoria, Australia

<sup>14</sup>LIGO, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>15</sup>Louisiana State University, Baton Rouge, LA 70803, USA

E-mail: [ling.sun@ligo.org](mailto:ling.sun@ligo.org)

7 May 2020

**Abstract.** The raw outputs of the detectors within the Advanced Laser Interferometer Gravitational-Wave Observatory need to be calibrated in order to produce the estimate of the dimensionless strain used for astrophysical analyses. The two detectors have been upgraded since the second observing run and finished the year-long third observing run. Understanding, accounting, and/or compensating for the complex-valued response of each part of the upgraded detectors improves the overall accuracy of the estimated

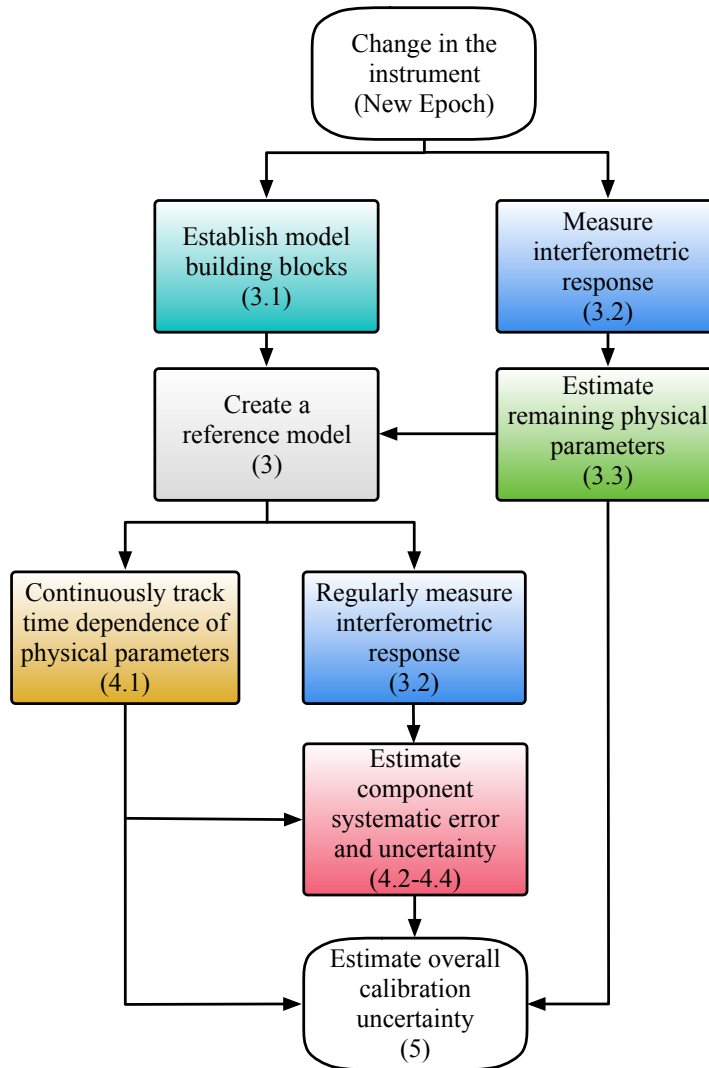
detector response to gravitational waves. We describe improved understanding and methods used to quantify the response of each detector, with a dedicated effort to define all places where systematic error plays a role. We use the detectors as they stand in the first half (six months) of the third observing run to demonstrate how each identified systematic error impacts the estimated strain and constrain the statistical uncertainty therein. For this time period, we estimate the upper limit on systematic error and associated uncertainty to be  $< 7\%$  in magnitude and  $< 4$  deg in phase (68% confidence interval) in the most sensitive frequency band 20–2000 Hz. The systematic error alone is estimated at levels of  $< 2\%$  in magnitude and  $< 2$  deg in phase.

## 1. Introduction

The Advanced Laser Interferometer Gravitational-Wave Observatory (Advanced LIGO) detectors [1] and the Virgo detector [2] have directly observed transient gravitational waves from multiple binary black hole coalescences and one binary neutron star merger in the first and second observing runs [3]. After a series of instrument upgrades to further improve the sensitivity, e.g., replacing test masses and optics, increasing laser power, and adding squeezed light [4], the two LIGO detectors started the third observing run (O3), together with Virgo, on April 1st, 2019, and ended the first half of O3 (O3A) on Oct 1st, 2019 [5, 6, 7, 8].

The time series of dimensionless strain,  $h$ , measured by each detector and used to determine the detection of a gravitational-wave (GW) signal and infer the properties of the astrophysical source, is reconstructed from digitized raw output of each detector. This reconstruction process, with an accurate and precise model of the detector’s response to  $h$ , is referred to as “calibration.” The accuracy and precision of  $h$  are important for detecting gravitational wave signals and crucial for the reconstruction of their astrophysical parameters [9, 10, 11].

We report the accuracy and precision of  $h$  by estimating the upper and lower 68% confidence interval bounds on the combination of systematic error and uncertainty for each detector response. The systematic error is quantified by propagating the error of each modeled component to the response of each detector. The statistical uncertainty arises from either the random noise in the measurements, repeated sampling of parameters from a random parent distribution, or the uncertainty in the unknown systematic error. The resulting upper and lower bounds on the error and uncertainty in  $h$  are thus complex-valued, frequency-dependent functions. Photon calibrators (Pcal), which independently use photon radiation pressure to produce strain within the detector [12, 13], are the primary absolute reference used to validate the estimates of  $h$  itself as well as the combined error and uncertainty. We expect the ratio of the estimated  $h$  to the strain produced by the Pcal systems to lie within these bounds 68% of the time [10, 14].



**Figure 1.** Flowchart of how the detector systematic error and uncertainty estimate is produced. Parenthetical numbers in each box guide the reader to the corresponding section in the paper.

In the first and second observing runs of Advanced LIGO (O1 and O2), we achieved a combined error and uncertainty limit of  $\lesssim 5\%$  in magnitude and  $\lesssim 3$  deg in phase in the most sensitive frequency band 20–2000 Hz. The method to determine those estimates for O1 and O2 is presented in [14]. Details of how a complete model is used to compute and produce the  $h$  data stream can be found in [11]. In this paper, we update the discussion of the methods in [14], use new studies of the upgraded O3A detectors to elucidate all sources of systematic error considered, and estimate the contribution of each source through measurements and Bayesian inference.

The structure of the paper is as follows. In section 2, we review the model components of the detector’s response to  $h$ , adding new qualifying details that are important to the O3 detectors, and discuss each component’s contribution to the detector’s response. In sections 3 and 4, we describe the procedure for creating a detector response model and estimating its error and uncertainty, following the workflow in figure 1: With a verified absolute calibration reference, a model of the detectors’ mechanical dynamics, and detailed measurement of the detectors’ electronics (section 3.1), we estimate remaining detector response parameters through Markov chain Monte Carlo (MCMC) analysis of interferometric measurements (sections 3.2 and 3.3). We then discuss how continuous time dependence in model parameters within a given observation period are tracked and accounted for (section 4.1), limitations of the O3A detector model components (section 4.2), estimation of residual frequency-dependent error and statistical uncertainty through Gaussian Process Regression (GPR) methods (section 4.3), and the negligible and/or unaccounted for systematic errors in each component (section 4.4). Finally, the overall uncertainty and the best estimated error in O3A are presented in section 5. We summarize and conclude in section 6.

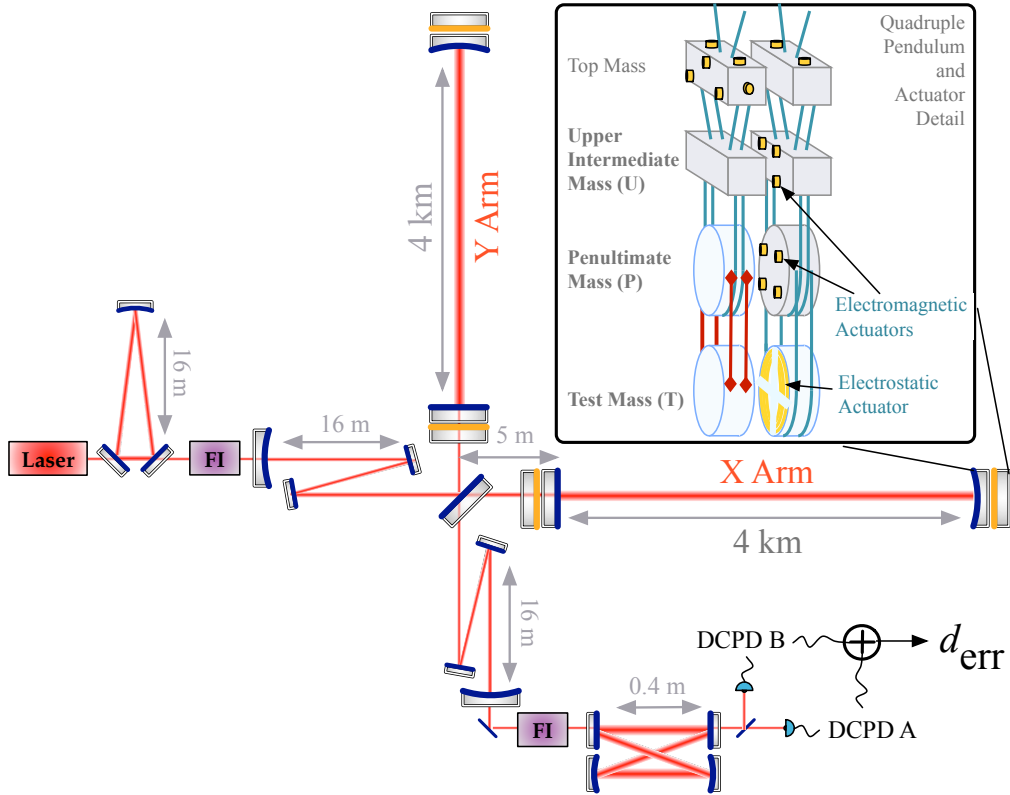
## 2. Model fundamentals

While the instrument has been upgraded between O2 and O3 [4], the conceptual design of the Advanced LIGO detectors has not changed fundamentally since the first observation of gravitational waves [15], as described in, e.g., [16]. The optical configuration of the two LIGO detectors remain dual-recycled, Michelson interferometers with 4-km-long Fabry-Pérot resonant cavities (figure 2). These detectors have been built to measure a dimensionless strain,  $h$ , incident upon them, defined by the differential changes in arm length (DARM length),  $\Delta L_{\text{free}}$ , divided by the average length of the arms  $L$ ,

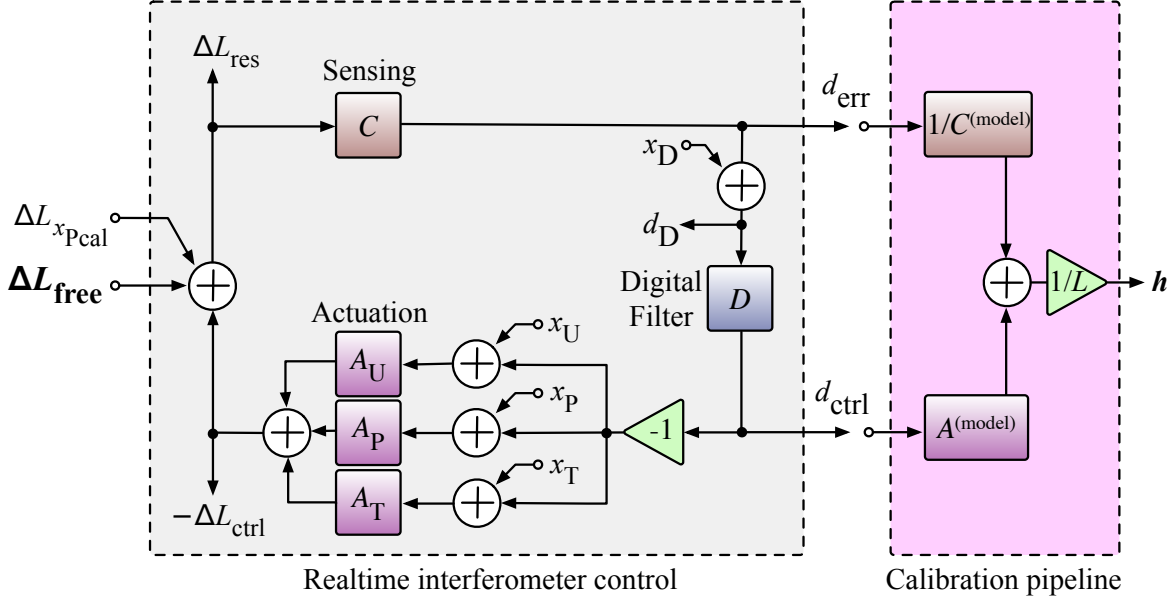
$$h = \frac{\Delta L_{\text{free}}}{L} = \frac{\Delta L_x - \Delta L_y}{L}, \quad (1)$$

where  $\Delta L_x$  and  $\Delta L_y$  are the displacements in the two orthogonal arms, X and Y, respectively. Due to the presence of noise and the desire to maintain the resonance condition of the optical cavities, the detectors do not directly measure  $\Delta L_{\text{free}}$ . Instead  $\Delta L_{\text{free}}$  is derived from the error and control signals of the DARM control loop, using methods described previously in [10, 11, 14, 17].

Figure 3 shows the interferometer DARM feedback control loop (gray box) and the calibration process (pink box) at a conceptual level. The loop contains the physical interferometer, analog electronics, analog-to-digital converters, a network of “front-end” computers, and digital-to-analog converters, as described in [18, 19]. The residual DARM displacement  $\Delta L_{\text{res}}$  is converted by the sensing function  $C$ , to produce the digital output  $d_{\text{err}}$ . The error signal is filtered through a set of digital filters  $D$ , creating the digital



**Figure 2.** Conceptual diagram of the optical configuration of the Advanced LIGO interferometers: dual-recycled, Fabry-Pérot Michelson. The X and Y arms are 4-km-long, Fabry-Pérot cavities formed by the highly reflective end test masses and partially transmissive input test masses. Pre-stabilized laser light enters the detector from the left, and is further stabilized using an input mode cleaner optical cavity. Cleaned light then enters the Power Recycling Cavity (formed by a partially transmissive input coupler and two high reflectors), is split by a 50/50 beamsplitter, and sent into the long arm cavities where the light interacts most with the potentially changing gravitational field. The light returning from the arm cavities interferes at the beamsplitter, and is then extracted from the beamsplitter’s anti-symmetric port by the Signal Recycling Cavity (SRC), similarly formed by two high reflectors and a partially transmissive output coupler. Finally, light exiting the SRC is cleaned with an additional resonant cavity, referred to as the “output mode cleaner”. Faraday Isolators (FI) are used for optical isolation of the main interferometer from the rest of the instrument. The transmitted light of the output mode cleaner is split onto two photodiodes, whose output current is turned to voltage, conditioned, digitized, de-conditioned digitally, and then linearly combined to form  $d_{\text{err}}$ . Inset: one of the full quadruple pendulum suspension systems and its actuators.



**Figure 3.** Advanced LIGO differential arm (DARM) length feedback control loop (gray box) and the generation of calibrated strain data (pink box). The sensing function  $C$  converts the residual DARM displacement  $\Delta L_{\text{res}}$  to the digital error signal  $d_{\text{err}}$ . The digital filter  $D$  processes  $d_{\text{err}}$  and produces the digital control signal  $d_{\text{ctrl}}$ . The actuation function  $A$  converts  $d_{\text{ctrl}}$  to the force allocated to the test masses, producing displacement  $-\Delta L_{\text{ctrl}}$  to suppress  $\Delta L_{\text{free}}$ . During the time dedicated to loop characterization (see section 3.2), DARM displacement excitations  $\Delta L_{x_{\text{Pcal}}}$  are added using the photon radiation pressure actuator system. Similarly,  $x_D$  and  $x_i$  ( $i = U, P, T$ ; refer to figure 2 for the definitions of  $U, P, T$ ) are added using the quadruple pendulum actuator via the digital control system. In the presence of  $x_D$ , the digital signal  $d_D = d_{\text{err}} + x_D$  may be used to characterize the DARM loop suppression. In the pink box,  $h$  is constructed using the sensing and actuation models,  $C^{(\text{model})}$  and  $A^{(\text{model})}$ . Note that the  $h$  here is the estimated DARM strain rather than the GW strain.

control signal,  $d_{\text{ctrl}}$  (i.e.,  $d_{\text{ctrl}} = Dd_{\text{err}}$ ). The actuation function  $A$  converts  $d_{\text{ctrl}}$  to the displacement  $\Delta L_{\text{ctrl}}$ , generated by the control force acting on the test masses that form the arm cavities. The displacement  $\Delta L_{\text{ctrl}}$  suppresses  $\Delta L_{\text{free}}$  caused by external stimuli, holding the optical cavities on resonance and leaving a small amount of  $\Delta L_{\text{res}}$  in the DARM loop. Conceptually,  $\Delta L_{\text{free}}$  is reconstructed with models of these functions as

$$\Delta L_{\text{free}} = \Delta L_{\text{res}} + \Delta L_{\text{ctrl}} = \frac{1}{C^{(\text{model})}} * d_{\text{err}} + A^{(\text{model})} * d_{\text{ctrl}}. \quad (2)$$

We can define a response function,  $R^{(\text{model})}$ , given by

$$\tilde{R}^{(\text{model})} = \frac{1 + \tilde{A}^{(\text{model})} \tilde{D} \tilde{C}^{(\text{model})}}{\tilde{C}^{(\text{model})}} = \frac{1 + \tilde{G}^{(\text{model})}}{\tilde{C}^{(\text{model})}}, \quad (3)$$

in frequency domain, where  $\tilde{G}^{(\text{model})} \equiv \tilde{A}^{(\text{model})} \tilde{D} \tilde{C}^{(\text{model})}$  is the DARM open loop gain,

such that

$$h = \frac{R^{(\text{model})} * d_{\text{err}}}{L}. \quad (4)$$

It is desirable to produce calibrated strain with low latency for quick electromagnetic follow-up. To fulfill this desire, a reasonably accurate, low-latency estimate of  $h$  is created in near-real-time. Later, a carefully-vetted, most-accurate estimate of  $h$  is delivered within a few months after the raw data are stored.

The low-latency estimate of  $h$  is produced from the model in two parts. In the first part, the models  $A^{(\text{model})}$  and  $C^{(\text{model})}$  are grossly reproduced by infinite impulse response (IIR) filters, which modify copies of  $d_{\text{err}}$  and  $d_{\text{ctrl}}$  in near real-time on a parallel computer within the network of the feedback control system. The resulting crude estimates of  $\Delta L_{\text{res}}$  and  $\Delta L_{\text{ctrl}}$ , summed to form a version of  $\Delta L_{\text{free}}$ , are all stored for later consumption. This “front-end” production of  $\Delta L_{\text{free}}$  is limited by causality and the finite sample rate of the system, but good enough to quickly assess the detector noise performance. However, the real-time limitations produce sufficient systematic errors in the front-end production of  $\Delta L_{\text{free}}$  that must be rectified before the data can be used in astrophysical analyses. As such, in the second part, the roughly calibrated  $\Delta L_{\text{res}}$  and  $\Delta L_{\text{ctrl}}$  are pulled from the front-end storage, and modified further with finite impulse response (FIR) filters with  $\sim 10$  seconds of latency. The extra bit of latency allows for acausal improvement to produce a low-latency data stream that faithfully estimates  $h$  with manageable systematic error.

This low-latency (online) estimate of  $h$  uses the best models of the detector at the time of recording. Over the course of any observing run, data dropouts due to computer failures, mistakes in modeling  $A^{(\text{model})}$  and  $C^{(\text{model})}$ , and unknown residual systematic errors are often identified. Further, methods may be developed at a later time to correct for systematic errors. Finally, there are known model components excluded from the IIR and FIR reproductions of  $A^{(\text{model})}$  and  $C^{(\text{model})}$  for expediency, which create further, albeit small, systematic error in all online estimated  $h$ .

Hence it motivates the creation of an additional high-latency (offline) estimate of  $h$ , allowing for improved accuracy, which uses the best models developed after the low-latency data are collected, stored, and understood. The offline estimate of  $h$  is created entirely with the FIR reproductions of  $A^{(\text{model})}$  and  $C^{(\text{model})}$ , starting “from scratch” with  $d_{\text{err}}$  and  $d_{\text{ctrl}}$ . Further details of the computational software and methods for producing these versions of  $h$  can be found in [11].

In this paper, we focus on the systematic error and uncertainty of the offline estimated  $h$ . The accuracy and precision of any estimate of  $h$  for a given detector is estimated by comparing a large collection of independent measurements of the detector response using the actuation excitation paths at  $x_{\text{Pcal}}$ ,  $x_{\text{D}}$ , and  $x_i$  ( $i = U, P, T$ ) against the model  $R^{(\text{model})}$ . In sections 2.1 and 2.2, we detail the components and parameters of the sensing

and actuation function models,  $C^{(\text{model})}$  and  $A^{(\text{model})}$ , respectively. Section 2.3 describes causes and impacts of slow time-variation of these frequency-dependent functions. The frequency-dependent contribution of  $C$  and each component in  $A$  to the response  $R^{(\text{model})}$  is discussed in section 2.4. The systematic error in  $R^{(\text{model})}$  and the impact from each component are defined in section 2.5.

### 2.1. Sensing function

The sensing function  $C(f)$  is the response of the filtered, digitized combination of photo-detector output signals, i.e.  $d_{\text{err}}$ , to the residual DARM displacement,  $\Delta L_{\text{res}}$ . This response is complex-valued (amplitude and phase), frequency-dependent, and slowly time-varying. It is comprised of a linear combination of several conceptually different parts: (1) the opto-mechanical, interferometric response to  $\Delta L_{\text{res}}$ , producing power (in units of watts) at the output of the signal recycling mirror, (2) the opto-electronic processing of that power into photo-current, including any optical loss on the path to and through the output mode cleaner, the final beamsplitter ratio as the transmitted light is sent to the readout photodiodes, and the photodiodes' response, (3) the analog signal processing electronics for the photodiodes and analog-to-digital conversion process which turn photo-current into digital counts, and (4) the conditioning, re-combination, and linearization of those digitized counts into a suitable, single error signal for the DARM control loop.

In O3, we retain the same frequency-domain model transfer function for this collection of conceptual parts (at a particular time) as in [14], analytically given by,

$$\tilde{C}^{(\text{model})}(f) = \left( \frac{H_C}{1 + i f f_{cc}^{-1}} \right) \left( \frac{f^2}{f^2 + f_s^2 - i f f_s Q^{-1}} \right) C_R(f) \exp(-2\pi i f \tau_C). \quad (5)$$

The overall gain of the sensing function,  $H_C$ , is the product of the scalar gains from each component in all four parts (in units of digital counts of  $d_{\text{err}}$  per meter of DARM length). Aside from  $H_C$ , the frequency dependence of part (1), represented by the first two parenthetical terms, defines the response of the coupled Fabry-Pérot arm cavities and signal recycling cavity (SRC) to  $\Delta L_{\text{free}}$ . In the first parenthetical term,  $f_{cc}$  is the differential coupled-cavity pole frequency. In the second term, the numerator represents two zeros at 0 Hz, and  $f_s$  and  $Q$  in the denominator are, respectively, the pole frequency and quality factor. Collectively the zeros and poles of the second parenthetical term represent the optical spring response created by any detuning present between the SRC and the arm cavities. The approximations and deficiencies within the first two terms are described in sections 4.2 and 4.4. The collective frequency response of the analog electronics described in parts (2)–(4) are addressed in two ways. Some portions of the response are paired with corresponding inverse digital filters, applied after the photodiode signal is digitized. Thus, they compensate the analog response within the DARM loop



itself (“in-loop”) and are not explicitly included in (5). The portions in parts (2)–(4) not compensated in-loop are collected within  $C_R$  (see further discussion in sections 3.1 and 4.4). The collection of analog and digital time delays from all four parts is denoted by  $\tau_C$  in the final term.

## 2.2. Actuation function

The actuation function,  $A$ , is the response of the control DARM displacement,  $\Delta L_{\text{ctrl}}$ , to the requested digital control signal,  $d_{\text{ctrl}}$ . Like the sensing function described above, it is composed of several components. We first qualify the O3 actuator model by extending the discussion in previous work [11, 14].

First, we consider the DARM control system only in the frequency band above 5 Hz. Below 5 Hz, actuation from absolute references (such as the Pcal) cannot be sufficiently resolved in the detector noise in  $\Delta L_{\text{free}}$ . Hence, any further allocation of  $\Delta L_{\text{ctrl}}$  to other actuators below 5 Hz, e.g., to the first, top-most stage of the quadruple suspension, is ignored.

Second, while  $\Delta L_{\text{ctrl}}$  may be induced by actuating on any stage of any of the four arm cavity optics quadruple pendulum systems [20, 21], we reduce complexity by only modeling the DARM control actuator as the bottom three stages of a single quadruple pendulum. In other words, if the upper two suspension stages of the X arm end test mass and the final test mass stage in the Y arm are used in combination to produce  $\Delta L_{\text{ctrl}}$ , each stage is measured and modeled independently, and the actuation from each stage is summed as though created by single quadruple pendulum.

Third, each detector has many other cavities length and angle degrees of freedom that must be controlled. Some of those control systems also use the quadruple suspension systems as actuators. These auxiliary control loops will only impact the DARM loop response if there is cross-coupling from  $\Delta L_{\text{ctrl}}$  to the auxiliary degree of freedom and from the auxiliary degree of freedom back to  $\Delta L_{\text{res}}$ . Reducing potential auxiliary cross-coupling to DARM is an essential element of the Advanced LIGO detectors collective control system design [1, 22]. Further, the O3 detectors only use three actuation stages among the six available lower stages of the end test mass suspensions to create  $\Delta L_{\text{ctrl}}$ . The actuation model does not include any cross-coupling with auxiliary degrees of freedom.

Finally, each quadruple pendulum system is actually a pair of closely adjacent quadruple suspensions, with the “main chain” holding the suspended test mass, and the “reaction chain” suspending equally isolated masses upon which the actuators are mounted (see the inset of figure 2 and [20, 21]). Among the lowest three stages of each quadruple suspension, the upper intermediate (UIM), and penultimate (PUM), are driven by magnetic coil actuators. The lowest stage of the suspension, named the test mass (TST) stage, is driven by an electrostatic actuator system. The force from the

actuators on the reaction chain is considered to be applied directly to the center of the mass at each stage of the main chain. For the purpose of estimating the displacement of the test mass, only the dynamic response of the main chain is modeled; it is not necessary to take into consideration the added complexity of the reaction chain.

With these qualifying remarks, the response of each actuator stage is modeled as (1) the digital distribution system which allocates  $d_{\text{ctrl}}$  (i.e., the filtered  $d_{\text{err}}$ ) to the computer that controls the three end test mass suspension stages, where subsequent digital filtering (i.e., the assignment of frequency-dependent control authority) and signal conditioning occurs, (2) the digital-to-analog converters and associated signal processing electronics that convert the conditioned digital signal into electrical signal suitable for that stage's actuator, (3) the mechanical pendulum dynamics of the stage's actuator itself, and (4) the mechanical, force-to-displacement response of quadruple pendulum suspension system in the DARM direction from the given stage to the optic.

Thus, the total actuation model (at a particular time) is similar to that in [10, 14, 11], with only slight modifications,

$$\begin{aligned} \tilde{A}^{(\text{model})}(f) = & [F_U(f)H_U\tilde{A}_U(f)\exp(-2\pi if\tau_U) \\ & + F_P(f)H_P\tilde{A}_P(f)\exp(-2\pi if\tau_P) \\ & + F_T(f)H_T\tilde{A}_T(f)\exp(-2\pi if\tau_T)], \end{aligned} \quad (6)$$

where  $U$ ,  $P$ , and  $T$  represent the UIM, PUM, and TST stages, respectively (see the inset of figure 2). For each stage ( $i = U, P, T$ ),  $F_i(f)$  is the digital, frequency-dependent filter which allocates  $d_{\text{ctrl}}$  to the appropriate stage,  $H_i$  is the overall gain, the product of the scalar gains of each component in all four parts (in units of meters of DARM length per digital count of  $d_{\text{ctrl}}$ ), and  $\tau_i$  is the total time delay in the digital-to-analog conversion. Similar to the sensing function, some portions of the analog electronics frequency response in part (2) are paired with inverse digital filters applied before converting the digitized signal to analog voltage, and thus not explicitly included in the model. Thus,  $\tilde{A}_i(f)$  includes the dynamical force-to-displacement frequency response of the quadruple pendulum and the residual response from any uncompensated analog electronics (for further discussion, again, see sections 3.1 and 4.4). Note that (6) differs from the equivalent expressions in [10, 11, 14] only in the generalization of  $\tau_i$  to be an arbitrary delay at each actuator stage, instead of a common delay for all stages. Limitations of this model are discussed in section 4.4.

### 2.3. Time dependence

The static, reference models described by (5) and (6) are constructed with parameters  $H_C$ ,  $f_{cc}$ ,  $f_s$ , and  $Q$  for  $\tilde{C}^{(\text{model})}$  and  $H_i$  ( $i = U, P, T$ ) for  $\tilde{A}^{(\text{model})}$  that are measured at a given time. Some parameters, however, are slowly varying over time due to various

physical mechanisms [23]. Sensing function parameters  $H_C$ ,  $f_{cc}$ ,  $f_s$ , and  $Q$  fluctuate on a time-scale of minutes due to the variations of optical alignment in the arm cavities, the relative alignment between the arm cavities and the SRC, and the laser power. The overall strength of the TST electrostatic actuator changes slowly on the time-scale of days to weeks due to the slow accumulation of static charges around the test mass and reaction mass. The overall strengths of the UIM and PUM magnetic coil actuators are expected to be static, but occasional changes in actuator electronics in the path often require compensation. The time-dependent sensing and actuation functions are virtually identical to those in [10, 14, 11, 17], and are summarized here:

$$\begin{aligned} \tilde{C}(f; t) = \kappa_C(t) & \left( \frac{H_C}{1 + i f f_{cc}^{-1}(t)} \right) \left( \frac{f^2}{f^2 + f_s^2(t) - i f f_s(t) Q^{-1}(t)} \right) \\ & \times C_R(f) \exp(-2\pi i f \tau_C), \end{aligned} \quad (7)$$

where  $\kappa_C(t)$  is a dimensionless, real-valued, scale factor characterizing the frequency-independent variations of  $H_C$ , and

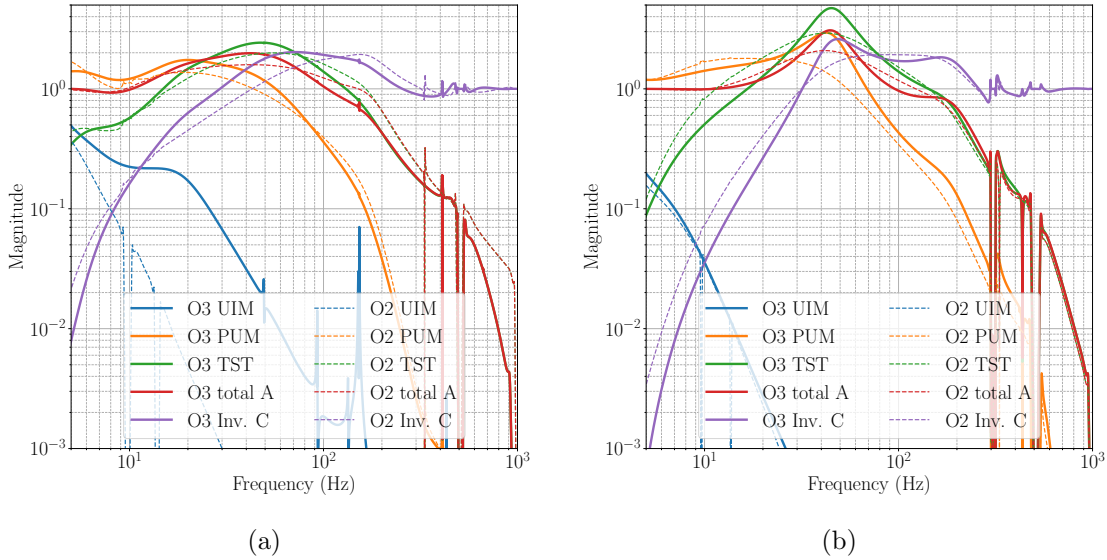
$$\begin{aligned} \tilde{A}(f; t) = [\kappa_U(t) F_U(f) H_U \tilde{A}_U(f) \exp(-2\pi i f \tau_U) \\ + \kappa_P(t) F_P(f) H_P \tilde{A}_P(f) \exp(-2\pi i f \tau_P) \\ + \kappa_T(t) F_T(f) H_T \tilde{A}_T(f) \exp(-2\pi i f \tau_T)], \end{aligned} \quad (8)$$

where  $\kappa_U(t)$ ,  $\kappa_P(t)$ , and  $\kappa_T(t)$  are similar dimensionless scale factors (though in this case complex) for the UIM, PUM, and TST stages, respectively, with the real parts varying about unity. In O1 and O2, the fluctuations in UIM and PUM stages were tracked with a combined factor  $\kappa_{PU}(t)$ . In O3,  $\kappa_{PU}(t)$  is replaced by separate scale factors  $\kappa_U(t)$  and  $\kappa_P(t)$  to provide more accurate tracking of temporal variation of the actuation functions. We refer to these time-dependent parameters in (7) and (8),  $\kappa_C(t)$ ,  $f_{cc}(t)$ ,  $f_s(t)$ ,  $Q(t)$ , and  $\kappa_i(t)$  ( $i = U, P, T$ ), as time-dependent correction factors (TDCFs). Additional details are provided in section 4.1.

#### 2.4. Contribution to the response function

The DARM loop response function is dominated by the actuation and sensing components at low and high frequencies, respectively. The exact frequency dependence is determined by choices made in the digital filtering, i.e., in the shape of  $D$  and  $F_i$  ( $i = U, P, T$ ), as well as the physical setup and state of the detector. It is important to quantify the frequency-dependent contributions to the response function from each component to determine how each component contributes to the uncertainty and systematic error. Figure 4 shows the magnitudes of these contributions, i.e.,  $F_i H_i \tilde{A}_i^{(\text{model})} \tilde{D} / \tilde{R}^{(\text{model})}$  ( $i = U, P, T$  for each individual suspension stage),  $\tilde{A}^{(\text{model})} \tilde{D} / \tilde{R}^{(\text{model})}$ , and  $1 / (\tilde{C}^{(\text{model})} \tilde{R}^{(\text{model})})$ . A similar figure for phase contributions is not shown here for brevity.

The solid curves in figure 4 are computed from the reference model used in September 2019, towards the end of O3A. As described in section 2.3, the strength of actuators and alignment of optical cavities can change, resulting in different contributions. In addition, different choices may be made and evolve with time to accommodate new detector parameters and/or to improve detector noise performance. For comparison, the O2 model values are shown as dashed curves, reflecting the impacts due to these changes. In particular, the contribution from the TST actuator at Livingston has increased in O3A compared to O2. The same amount of error and uncertainty in modeling  $\tilde{A}_T^{(\text{model})}$  in O3A as in O2 results in a larger overall calibration uncertainty around 50 Hz (see section 5).



**Figure 4.** Contributions of each stage of the actuators, the overall actuation, and the inverse sensing to the response function at (a) Hanford and (b) Livingston. The solid and dashed curves indicate the static, reference models used towards the end of O3A and O2, respectively.

### 2.5. Systematic error definition

The frequency-dependent systematic error of the response function, equivalent to the systematic error in estimated  $h$ , is defined by

$$\tilde{\eta}_R \equiv \frac{\tilde{R}}{\tilde{R}^{(\text{model})}} = \frac{\delta\tilde{R}}{\tilde{R}^{(\text{model})}} + 1, \quad (9)$$

where  $\delta\tilde{R}/\tilde{R}^{(\text{model})} = \tilde{\eta}_R - 1$  is the relative error in the response function as defined in [14]. By applying  $\tilde{\eta}_R$  to the model response function, we obtain the true response  $\tilde{\eta}_R \tilde{R}^{(\text{model})}$ .

A systematic error in  $\tilde{C}$ , defined by  $\tilde{\eta}_C \equiv \tilde{C}/\tilde{C}^{(\text{model})}$ , will impact the response function systematic error as

$$\tilde{\eta}_{R;C} = \frac{1}{\tilde{R}^{(\text{model})}} \left[ \frac{1}{\tilde{\eta}_C \tilde{C}^{(\text{model})}} + \tilde{A}^{(\text{model})} \tilde{D} \right]. \quad (10)$$

Similarly, a systematic error in  $\tilde{A}_i$  ( $i = U, P, T$ ), defined by  $\tilde{\eta}_{A_i} \equiv \tilde{A}_i/\tilde{A}_i^{(\text{model})}$ , will impact the response function systemic error as

$$\tilde{\eta}_{R;A_i} = \frac{1}{\tilde{R}^{(\text{model})}} \left[ \frac{1}{\tilde{C}^{(\text{model})}} + \left( \tilde{\eta}_{A_i} \tilde{A}_i^{(\text{model})} + \sum_{j \neq i} \tilde{A}_j^{(\text{model})} \right) \tilde{D} \right]. \quad (11)$$

These definitions are employed in sections 3.1.3, 4.3, and 5.

### 3. Construct a reference model

In this section, we describe the method and procedure of constructing a static reference DARM loop model. Section 3.1 discusses the tools prepared and measurements made before constructing the model: (a) the photon calibrator absolute reference, (b) a verified model of the quadruple pendulum mechanical dynamics, and (c) a characterization of all actuator and photodiode signal processing electronics present in  $A$  and  $C$ . Section 3.2 describes the measurements of the remaining model parameters, which are only measurable when the detectors are in their nominal low-noise configuration, and section 3.3 explains how these parameters and associated uncertainty are computed.

#### 3.1. Essential building blocks

The systematic error and uncertainty associated with the absolute reference of the DARM loop model and other essential “building blocks” are discussed in this subsection. Prior to O3A, the dynamics of the quadruple pendulum and response of signal processing electronics were included in the DARM loop model without considering their contributions to the uncertainty or systematic error in the detector’s response. However, as our understanding of the detectors improves, we now consider them to be a potential source of systematic error, and thus their fundamentals are described in more detail here.

*3.1.1. Photon calibrator absolute reference* The displacement fiducials upon which all estimates of  $h$  depend are generated by the Pcal systems [12, 24]. These systems employ power-modulated auxiliary lasers with beams reflecting from end test masses to displace the mirrors via photon radiation pressure. Pcal systems are deployed on both end test masses of each interferometer. Their functionality is summarized here along with system updates relevant to the O3A observing run.

Each Pcal system has a 2-watt laser operating at 1047-nm wavelength, housed in a transmitter module located outside the vacuum envelope. A feedback control loop that uses an acoustic-optic modulator to vary the laser power in response to a digital excitation signal,  $x_{\text{Pcal}}$ , generates a power-modulated output waveform that reproduces the excitation waveform. The modulated laser light is directed into the vacuum envelope and reflects from the surface of the end test mass, producing true DARM displacement,  $\Delta L_{x_{\text{Pcal}}}$  (see figure 3). The reflected light is directed to a laser power sensor that uses an integrating sphere and photodetector to generate a digital signal,  $d_{\text{Pcal}}$ , proportional to the received laser power. The bandwidth of this laser power control servo is approximately 100 kHz. The Pcal systems can thus produce arbitrary time-dependent forces resulting in  $\Delta L_{x_{\text{Pcal}}}$  similar to those that can be produced by the actuators of the quadruple suspension system. We estimate the induced DARM displacement from digitized photodiode signal  $d_{\text{Pcal}}$  with a bank of digital filters  $H_{\text{Pcal}}$ ,

$$\Delta L_{\text{Pcal}} = H_{\text{Pcal}} d_{\text{Pcal}}, \quad (12)$$

where  $\Delta L_{\text{Pcal}}$  is the estimate of the true DARM displacement  $\Delta L_{x_{\text{Pcal}}}$ . See (1) in [12] for details of  $H_{\text{Pcal}}$ .

Though the test mass displacement decreases as the square of the modulation frequency, with a maximum modulated power of approximately 1 W, the Pcal systems can generate  $\Delta L_{x_{\text{Pcal}}}$  that is orders of magnitude larger than the  $\Delta L_{\text{free}}$  noise floor across the most sensitive band of the detector. The digital control system allows for arbitrary excitation waveforms, but two specific waveforms for  $x_{\text{Pcal}}$  are typically used. The first is a sequence of monochromatic sinusoidal length modulations, referred to as a *swept-sine* excitation. The second is a colored random noise modulation used to probe more sensitive frequency regions with high frequency resolution.

The  $1\sigma$  uncertainties of  $H_{\text{Pcal}}$ , and thus  $\Delta L_{\text{Pcal}}$  as an estimate of  $\Delta L_{x_{\text{Pcal}}}$ , for all four LIGO Pcal systems during O3A are 0.54%. They are dominated by unknown systematic errors rather than by statistical variations in measured values [13]. At the end of O3A, the  $H_{\text{Pcal}}$  value was refined using system characterization measurements carried out during the six months of the run, as well as the correction of errors in the masses of the end mirrors. These updates are accounted for by a multiplicative correction factor,  $\eta_{\text{Pcal}}$ , applied to  $H_{\text{Pcal}}$  for each Pcal system. For the Hanford reference Pcal system on the Y arm end test mass, we have  $\eta_{\text{Pcal}} = 1.0043$ . For the Livingston reference system, also on the Y arm end test mass, we have  $\eta_{\text{Pcal}} = 1.0031$  [25]. Accounting for these systematic errors and uncertainties is discussed in section 5.

*3.1.2. Dynamics of the quadruple suspension* Preliminary models of the quadruple suspension system rigid-body, force-to-displacement transfer functions were developed from first principles well before the Advanced LIGO interferometers were installed [26, 27].

These preliminary models aided analysis and diagnostics of the early prototype quadruple suspension systems [28]. Refinements were added to the model in order to match them to the first production suspension system and improve their accuracy [29]. The refined model was later used to verify the function of all production quadruple suspensions installed in each LIGO detector.

The model parameters are kept up-to-date as the installed suspensions are modified (e.g., between O2 and O3, small, few-gram damping mechanisms were added to the test masses [30]). The change in model parameters can be typically quantified to high accuracy (e.g. each  $\sim 40$  kg test mass can be measured to an accuracy of  $\sim 10$  g), and the subsequent updated model parameters are revalidated to high precision through many local and interferometric measurements of the dynamical response. These models are used as the basis for the frequency dependence of force-to-displacement transfer functions in  $\tilde{A}_i$ . Beyond these rigid-body dynamics, we have found the need for additional, non-rigid body modifications to these transfer functions in order to improve the model accuracy. The impact of the additional modifications is discussed in sections 4.3 and 4.4.

*3.1.3. Signal processing analog electronics* The responses of all signal processing electronics are measured independently in advance and modeled as transfer functions with poles and zeros at well-determined frequencies. Within the sensing function, these electronics conditioning the current produced by the readout photodiodes are: the transimpedance amplifiers of the photodiodes, “whitening” filters (i.e., frequency-dependent, signal pre-amplification or noise reduction filters), and anti-aliasing filters. Within the actuation function, requested voltage at each stage is conditioned through anti-imaging filter electronics and sent to either the magnetic coil current drivers or electrostatic voltage drivers, depending on the actuator type of the given stage of the quadruple suspension. The actuator drivers also have frequency-dependent response for noise reduction much like the readout photodiode transimpedance amplifier and whitening filters.

The pole and zero frequencies for the responses of these electronic components range from as low as 0.5 Hz to as high as 50 kHz, all of which need to be included in  $A^{(\text{model})}$  and  $C^{(\text{model})}$  to produce accurate estimates of  $h$ . For example, to minimize the contribution to systematic error in the response function near the DARM loop unity gain frequency ( $\sim 100$  Hz), the phase of  $A^{(\text{model})}$  and  $C^{(\text{model})}$  must be accurate to a level of  $\lesssim 0.1$  deg. Such accuracy cannot be achieved if any of the poles or zeros, even those at  $\sim 50$  kHz, are excluded in the model. Measurements of the response of each electronic component from 0.1 Hz to 100 kHz are made using an analog spectrum analyzer. Pole and zero frequencies are determined by fitting the measured response of each electronic component to a model consisting of poles and zeros [31]. Only the poles and zeros frequency response is needed at this point and later the gain of each path is measured in

$H_C$  and  $H_i$  ( $i = U, P, T$ ) using techniques described in section 3.2.

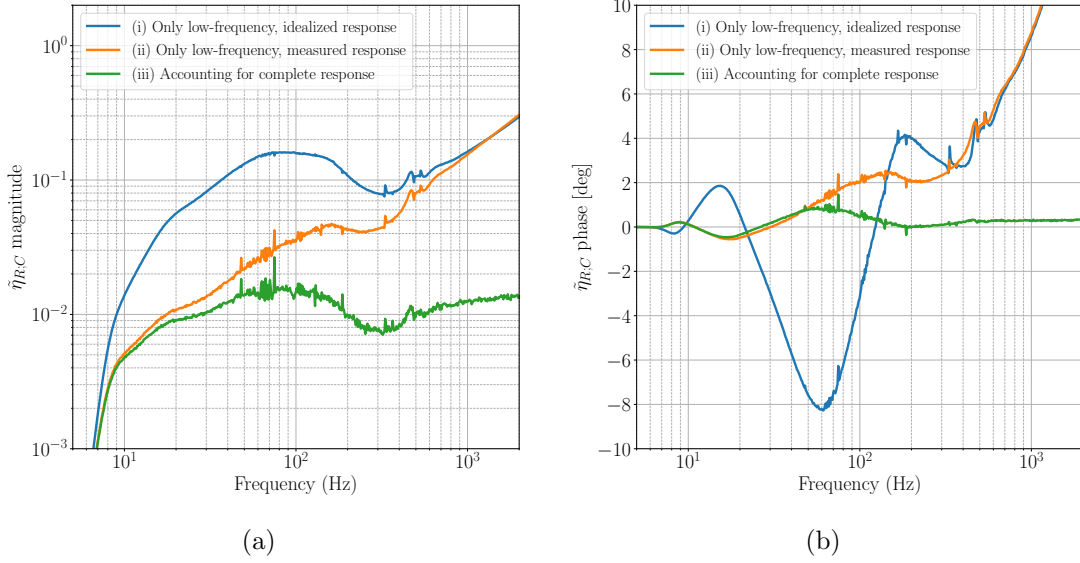
The pole and zero frequencies of the electronic components are used in different ways throughout the production of  $h$ . As described in sections 2.1 and 2.2, poles and zeros below  $\sim 500$  Hz are used to design digital IIR filters within the DARM loop that replicate the inverse response of the electronics. The DARM loop would be negatively impacted without including these poles and zeros. Each pair of analog response and compensating digital inverse response occurs before constructing  $d_{\text{err}}$  within  $C$  and after  $d_{\text{ctrl}}$  is distributed through  $A$ . Thus, the low-latency or offline estimate of  $h$  need only further include the higher-frequency poles and zeros. Poles and zeros above  $\sim 500$  Hz but below the Nyquist frequency of the real-time system ( $\sim 7000$  Hz) are included in the front-end IIR reproductions of  $A^{(\text{model})}$  and  $C^{(\text{model})}$  (outside the DARM loop) to produce the roughly calibrated  $\Delta L_{\text{free}}$ . Limitations of the front-end IIR filter construction result in growing systematic error approaching the Nyquist frequency and prevent the inclusion of any response above the Nyquist frequency. The second part of the low-latency pipeline repairs any distorted high-frequency response of IIR models of  $A^{(\text{model})}$  and  $C^{(\text{model})}$  and includes the response of super-Nyquist poles and zeros to form the low-latency estimate of  $h$ . The most accurate, offline estimate of  $h$  includes all poles and zeros above  $\sim 500$  Hz (i.e., those not compensated within the DARM loop) in the FIR reproductions of  $A^{(\text{model})}$  and  $C^{(\text{model})}$ .

The contribution to the systematic error in  $h$  from each electronic component in  $C$  and  $A$  is evaluated by  $\tilde{\eta}_{R;C}$  or  $\tilde{\eta}_{R;A}$  in equations 10 and 11, respectively. An example is given in figure 5 for the sensing function whitening filter electronics alone, assuming all other electronics are modeled perfectly. To emphasize the need for careful measurements, the blue curves indicate the resulting  $\tilde{\eta}_{R;C}$  if the in-loop compensation filters are designed using only pole and zero frequencies below  $\sim 500$  Hz as reported in the design specifications; the poles or zeros above  $\sim 500$  Hz are not included. The orange curves show the resulting  $\tilde{\eta}_{R;C}$  if in-loop compensation filters are designed using pole and zero frequencies obtained via fitting but do not account for poles or zeros above  $\sim 500$  Hz. Finally, the green curves correspond to  $\tilde{\eta}_{R;C}$  remaining in the low-latency and offline estimates of  $h$  if the in-loop compensation filters are designed with the measured poles and zeros below  $\sim 500$  Hz and include all other poles and zeroes above 500 Hz in the FIR reproductions of  $A^{(\text{model})}$  and  $C^{(\text{model})}$ .

### 3.2. Interferometric measurements

At frequencies much lower than the acoustic resonance frequencies of the test masses ( $\lesssim 1$  kHz), the excitations from the Pcal systems  $x_{\text{Pcal}}$  cause DARM displacement  $\Delta L_{x_{\text{Pcal}}}$ , equivalent to  $\Delta L_{\text{free}}$ . Thus, measuring the DARM loop error signal  $d_{\text{err}}$  in the presence of these Pcal excitations, while the detector is otherwise fully functional, is a direct





**Figure 5.** Response function error ( $\tilde{\eta}_{R;C}$ ) in magnitude (a) and phase (b) from the systematic error in sensing function whitening filter electronics alone. The colored curves indicate the resulting  $\tilde{\eta}_{R;C}$ , if (i) the in-loop compensation filters are informed only by design specifications, and high-frequency poles and zeros are excluded from  $A^{(\text{model})}$  and  $C^{(\text{model})}$  (blue), (ii) the in-loop compensation filters are informed by measurements, but high-frequency poles and zeros are still excluded (orange), and (iii) all measured pole and zero frequencies are included (green) as in the final estimate of  $h$ .

measure of the (inverse) response function,

$$\frac{\tilde{d}_{\text{err}}}{\widetilde{\Delta L}_{\text{Pcal}}} = \frac{1}{\tilde{R}^{(\text{meas})}} = \frac{\tilde{C}^{(\text{meas})}}{1 + \tilde{A}^{(\text{meas})} D \tilde{C}^{(\text{meas})}}, \quad (13)$$

where the superscript “(meas)” stands for the measurement. To measure the sensing function  $\tilde{C}^{(\text{meas})}$ , an additional separate measurement of the loop suppression is required (see figure 3),

$$\frac{\tilde{d}_{\text{D}}}{\tilde{x}_{\text{D}}} = \frac{1}{1 + \tilde{A}^{(\text{meas})} \tilde{D} \tilde{C}^{(\text{meas})}}, \quad (14)$$

where  $x_{\text{D}}$  indicates the displacement excitations added using the quadruple pendulum actuator system, and  $d_{\text{D}}$  is the sum of  $d_{\text{err}}$  and  $x_{\text{D}}$ . Measurements in (13) and (14) need to be taken sufficiently close in time, such that the time dependence of  $A$  and  $C$  can be ignored. Combining (13) and (14), we obtain the sensing function directly as

$$\tilde{C}^{(\text{meas})}(f) = \left( \frac{\tilde{d}_{\text{err}}(f)}{\widetilde{\Delta L}_{\text{Pcal}}(f)} \right) \left( \frac{\tilde{x}_{\text{D}}(f)}{\tilde{d}_{\text{D}}(f)} \right). \quad (15)$$

For frequencies above 1 kHz, where much longer averaging is required to obtain appreciable signal-to-noise (SNR) with respect to the detector’s noise floor, the sensing

function is measured by introducing discrete sinusoidal excitations from 1 kHz to 4 kHz, at 500 Hz intervals. Excitations at a single frequency are left on for 24 hours of “observation-ready” (i.e., nominal operating configuration) time before moving to the next frequency in the sequence. At frequencies above 1 kHz, where the detector’s response is determined by the sensing function, we have  $|\tilde{A}^{(\text{meas})}\tilde{D}\tilde{C}^{(\text{meas})}| \lesssim 10^{-4}$  and thus the independent measurement of the loop suppression in (15) is not necessary, i.e.,  $|\tilde{x}_D/\tilde{d}_D - 1| \lesssim 10^{-4}$ . Therefore, for these frequencies, the sensing function is well-approximated by

$$\tilde{C}^{(\text{meas})}(f) \approx \frac{\tilde{d}_{\text{err}}(f)}{\Delta\tilde{L}_{\text{Pcal}}(f)}, \quad (16)$$

using the average of two consecutive 30-minute fast Fourier transforms (FFTs) of the measurements. The mean value at each frequency is calculated using these 1-hour stretches over 24 hours. Due to the long duration of the measurements, the resulting sensing function must be corrected for time dependence using  $\kappa_C(t)$  and  $f_{cc}(t)$  (see section 4.1).

The Pcal systems are also used to determine the actuator strength,  $H_i$ , for each stage of the quadruple suspension. Measuring  $d_{\text{err}}$  caused by the excitations from the suspension actuators,  $\tilde{x}_i(f)$ , gives

$$\frac{\tilde{d}_{\text{err}}}{\tilde{x}_i} = \frac{\tilde{A}_i^{(\text{meas})}\tilde{C}^{(\text{meas})}}{1 + \tilde{A}^{(\text{meas})}D\tilde{C}^{(\text{meas})}}. \quad (17)$$

Combining (13) and (17), we can extract  $\tilde{A}_i^{(\text{meas})}$  ( $i = U, P, T$ ), as in [10, 14],

$$\begin{aligned} \tilde{A}_i^{(\text{meas})}(f) &= \left[ \tilde{F}_i(f)H_i\tilde{A}_i(f)\exp(-2\pi if\tau_i) \right]^{(\text{meas})} \\ &= \left( \frac{\Delta\tilde{L}_{\text{Pcal}}(f)}{\tilde{d}_{\text{err}}(f)} \right) \left( \frac{\tilde{d}_{\text{err}}(f)}{\tilde{x}_i(f)} \right). \end{aligned} \quad (18)$$

The relative magnitude uncertainty and absolute phase uncertainty of the measured transfer functions at each frequency point is given by [14, 32]

$$\sigma^{(\text{meas})}(f) = \sqrt{\frac{1 - \gamma^2(f)}{2N_{\text{avg}}\gamma^2(f)}}, \quad (19)$$

where  $N_{\text{avg}}$  is the number of values averaged. The coherence,  $\gamma^2(f)$ , between the excitation  $x$  and readout  $d$  at frequency  $f$  is calculated by [11]

$$\gamma^2(f) = \frac{|\langle \tilde{x}^*(f)\tilde{d}(f) \rangle|^2}{\langle |\tilde{x}(f)|^2 \rangle \langle |\tilde{d}(f)|^2 \rangle}, \quad (20)$$

where the angled brackets denote averaging, and the asterisk denotes complex conjugation.

Measurements described by (15) and (18) are repeated with a weekly cadence throughout the run. A single set of them is used to construct a reference model (see section 3.3). The collection of weekly measurements is used to assess static, frequency-dependent systematic errors (see sections 4.2 and 4.3).

### 3.3. Model parameter estimation

Remaining parameters in the DARM reference model are determined from one set of measurements (section 3.2) taken at the reference time, after dividing out all known frequency dependence from section 3.1.3, using MCMC fitting algorithms [33]. The remaining parameters are

$$\boldsymbol{\lambda}^C = [H_C, f_{cc}, f_s, Q, \delta\tau_C], \quad (21)$$

for the sensing function, and

$$\boldsymbol{\lambda}_i^A = [H_i, \delta\tau_i], \quad (22)$$

for each  $i$ th stage of the actuation function, where  $\delta\tau_C$  and  $\delta\tau_i$  are the residual time delays of  $\tau_C$  and  $\tau_i$ , respectively. We note that only measurement data at frequencies below 1 kHz are used for parameter estimation. The high-frequency measurements are used for studying the static, residual systematic error and statistical uncertainty above 1 kHz (see section 4.3). The MCMC method produces the posterior distributions of the multivariate parameters assuming normally distributed priors for  $H_C$ ,  $f_{cc}$ ,  $H_i$ , and flat (uniform) priors for  $f_s$ ,  $Q^{-1}$ ,  $\delta\tau_C$ , and  $\delta\tau_i$  ( $i = U, P, T$ ). The maximum a posteriori (MAP) values,  $\boldsymbol{\lambda}_{\text{MAP}}^C$  and  $\boldsymbol{\lambda}_{\text{MAP}}^A$ , are adopted to create the DARM response model,

$$\tilde{R}^{(\text{model})}(f) = \frac{1}{\tilde{C}^{(\text{model})}(\boldsymbol{\lambda}_{\text{MAP}}^C; f)} + \tilde{A}^{(\text{model})}(\boldsymbol{\lambda}_{\text{MAP}}^A; f)\tilde{D}(f). \quad (23)$$

When any physical change of the interferometer is too large to be corrected by the TDCFs, or a precursory component has changed, we create a new calibration “epoch.” It is likely that in any new epoch one or more parameters in the existing  $\boldsymbol{\lambda}_{\text{MAP}}^C$  and  $\boldsymbol{\lambda}_{\text{MAP}}^A$  (and hence  $\tilde{R}^{(\text{model})}$ ) are no longer valid. The MCMC parameter estimation process is repeated using new measurements,  $A^{(\text{meas})}$  and  $C^{(\text{meas})}$ , in order to create an updated reference model and account for the precursory changes. Table 1 lists the O3A epochs in both detectors and the main changes associated with each. We quantify the calibration systematic error and statistical uncertainty for each epoch in section 5.

**Table 1.** O3A calibration epochs and the main changes in each epoch.

| Hanford epoch     | Changes  |
|-------------------|--|
| (a) Mar 28–Jun 11 | Start of the run   |
| (b) Jun 11–Aug 28 | Input power increased; angular control system modified       |
| (c) Aug 28–Oct 1  | Added a microscopic length offset to SRC to relieve detuning |
| Livingston epoch  | Changes  |
| (a) Mar 28–Jun 11 | Start of the run   |
| (b) Jun 11–Oct 1  | Adjusted the gain in the TST actuator due to a 4% drift      |

An example of the MCMC fitting for the sensing function at Hanford is given in figures 6 and 7. A set of measurements is taken in the frequency band 5–1084 Hz, and passed to the MCMC algorithm. The five-dimensional fitting results are shown in figure 6. Posterior distributions of five parameters in  $\boldsymbol{\lambda}^C$  are shown in the diagonal panels. Contour curves show the covariance of each parameter pair, at  $1\sigma$ ,  $2\sigma$ , and  $3\sigma$  levels (from dark to light). The cyan (solid) and red (dashed) lines indicate the MAP and  $1\sigma$  values for each parameter, respectively.

The reference sensing model, created using the MAP parameters shown in figure 6, is then compared to the original measured data points, plotted in figure 7. The left column shows the magnitude and phase of both the reference model (grey curve) and measurement (red points). The right column displays the residual between the two. The units of the sensing function are shown in digital counts of  $d_{\text{err}}$  per meter change in the DARM length. The deviation between the measurement and the model below 20 Hz is due to a poorly-modeled effect (discussed in section 4.2), and hence the measurements below 20 Hz are not used to inform the MCMC fit (as denoted by the dashed vertical lines).

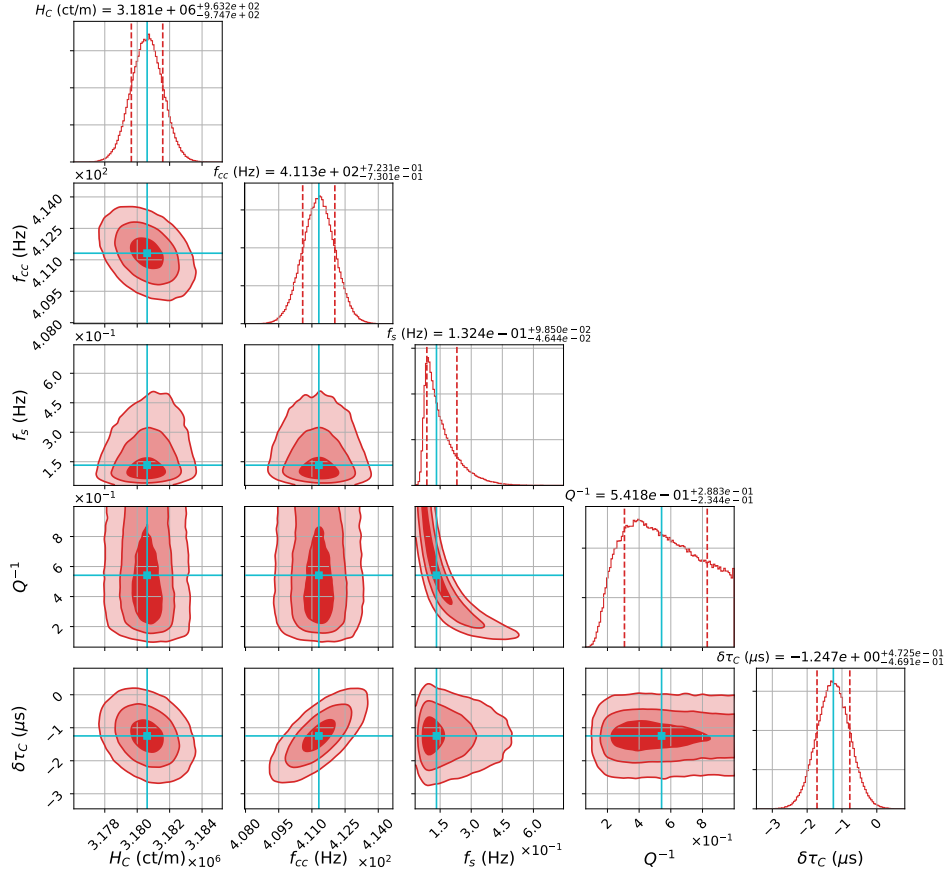
A similar procedure is repeated to generate the model of the three end test mass actuator stages. The Hanford detector produces  $\Delta L_{\text{ctrl}}$  with all three end test mass suspension stages on the X arm during O3A. The Livingston detector uses the TST stage on the Y arm, and the PUM and UIM stages on the X arm.‡ In this “split actuator” configuration for the Livingston detector, one must include information in the model reflecting that different computers and digital-to-analog converters are used to create  $\Delta L_{\text{ctrl}}$ . This is done by allowing for a different computational time delay in the model for each stage. Thus,  $\tau_i$  is fit independently and included in the model. After accounting for these different delays, we find that all remaining residuals of  $\tau_i$  are consistent with zero. The uncertainty of  $\tau_i$  is discussed in section 4.4.

## 4. Understanding of systematic errors

### 4.1. Time dependent systematic error

The TDCFs,  $\kappa_C$ ,  $f_{cc}$ ,  $f_s$ ,  $Q$ ,  $\kappa_U$ ,  $\kappa_P$ , and  $\kappa_T$  from (7) and (8), are monitored by a collection of monochromatic, high-SNR sinusoidal excitations (“calibration lines”) injected into the DARM control loop by both Pcal and suspension actuators. After demodulating the magnitude and phase of these calibration lines in  $d_{\text{err}}$ , the TDCFs are calculated from (15)–(18) (see complete derivation in [17, 11, 23]) and applied to the appropriate components in  $\Delta L_{\text{res}}$  and  $\Delta L_{\text{ctrl}}$ . The uncertainties for all TDCFs are computed using

‡ In O1 and O2,  $\Delta L_{\text{ctrl}}$  was produced entirely by the three end test mass actuator stages on the Y arm at Hanford, and the three stages on the X arm at Livingston.

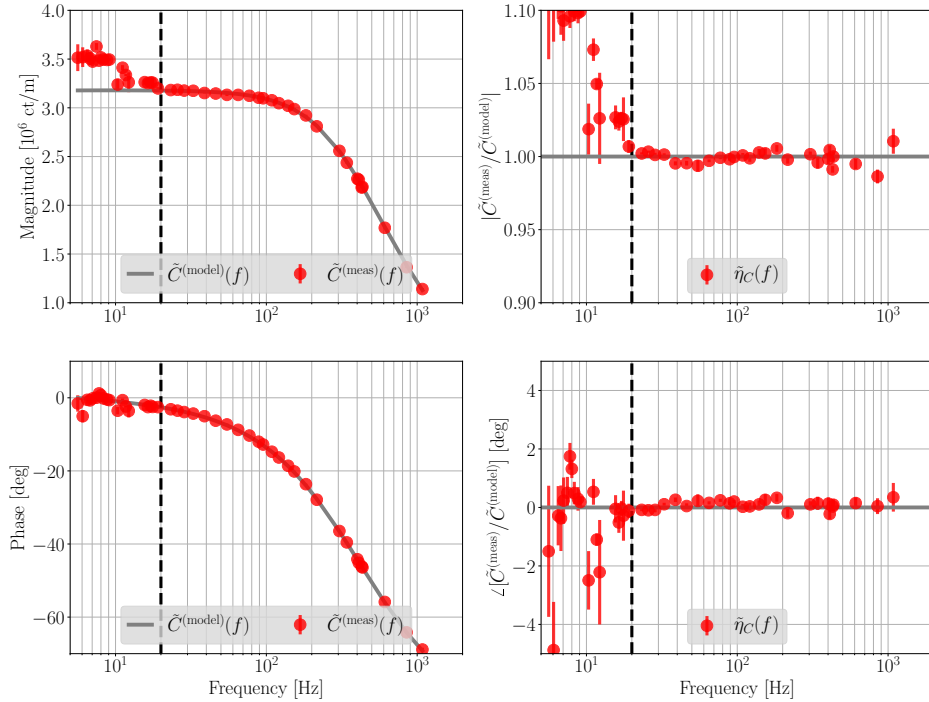


**Figure 6.** Corner plot of posterior distributions of the sensing parameters,  $\lambda^C$ , at Hanford. The one-dimensional histograms along the diagonal are the posterior distributions for the optical gain  $H_C$ , Fabry-Pérot coupled cavity pole frequency  $f_{cc}$ , SRC optical spring frequency  $f_s$ , inverse optical spring quality factor  $Q^{-1}$ , and residual time delay  $\delta\tau_C$ , from top left to bottom right. The off-diagonal two-dimensional histograms show the covariance of two parameters;  $1\sigma$ ,  $2\sigma$ , and  $3\sigma$  levels are delineated by contours (from dark to light). The cyan lines indicate the MAP values for each parameter. The dashed red lines in the 1-D histograms indicate the  $1\sigma$  values in the distribution.

(19).

Table 2 shows how time-dependent corrections are applied in O1, O2, and O3, for each of the three calibration pipelines. In O1 and O2, no TDCF was applied to correct for systematic errors in the front-end calibration, and only the scalar (i.e., frequency independent) factors were applied in the low-latency, online calibration. § The frequency-dependent factor  $f_{cc}$  was only applied in the high-latency, offline calibration, producing

§ Recall that the gains of the UIM and PUM suspension stages were tracked by the combined factor  $\kappa_{PU}$  in O1 and O2.



**Figure 7.** Sensing measurement  $\tilde{C}^{(\text{meas})}(f)$  and the reference model  $\tilde{C}^{(\text{model})}(f)$  created with the MAP parameters in figure 6 (left column), and the fractional residual between the two (right column). The gray curves and red markers indicate the reference model and the measurements, respectively. Vertical error bars indicate uncertainties of the measurements. The top and bottom rows display the magnitude and phase, respectively. The MAP values in figure 6 are inferred from the measurements above 20 Hz only (on the right side of the vertical dashed lines; see explanations in text).

the final, corrected strain data a couple of months after the data were acquired. In O3, both the scalar factors and  $f_{cc}$  are applied in all of the three calibration pipelines. The front-end pipeline computes and applies the TDCFs separately and independently from the online and offline pipelines for real-time detector performance assessment and consistency checks. The remaining two TDCFs,  $f_s$  and  $Q$ , are monitored but not applied to the data in any of the pipelines. Static reference values for these terms are used by the pipelines and taken from  $\lambda_{\text{MAP}}^C$ .

The impacted frequency bands and level of systematic errors from all TDCFs are slightly different at Hanford than at Livingston due to differences in the design of digital filters, as shown in the contribution curves in figure 4.

As an example to show the necessity of time-dependent corrections, we quantify the systematic errors that would be present in the O3A Hanford model, if the TDCFs were not applied. Similar to studies in [17], figure 8 shows the estimated systematic

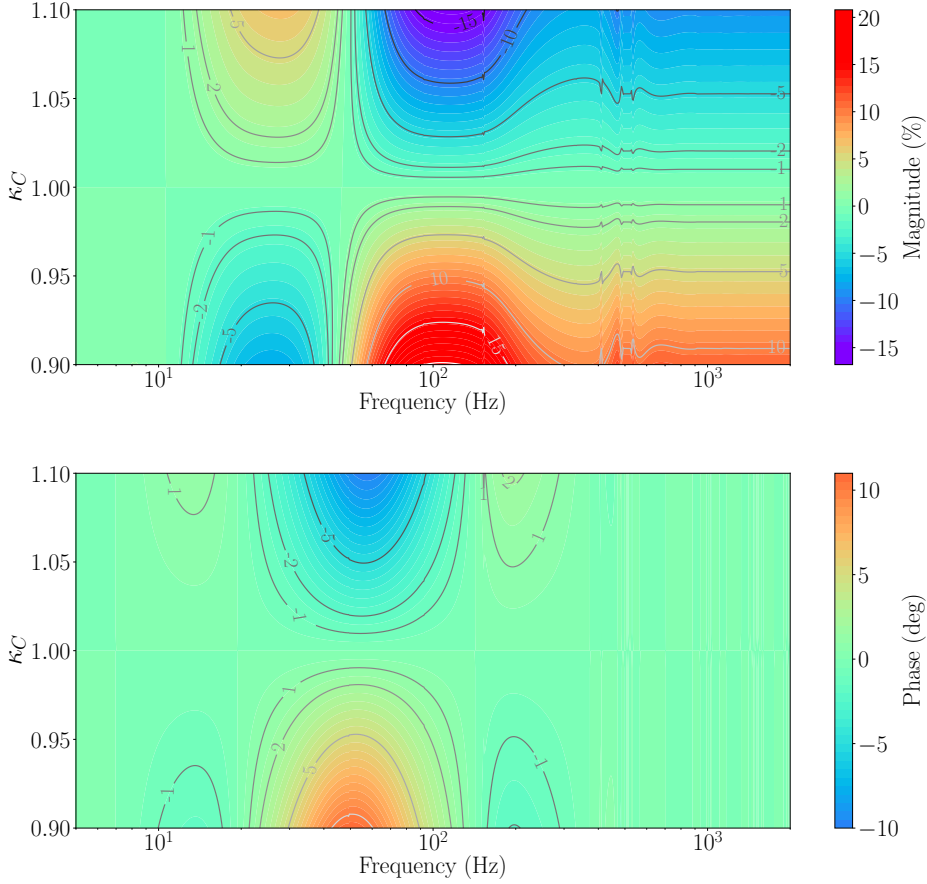
**Table 2.** Time-dependent correction factors applied in each of the calibration pipelines in the three observing runs.

| Calibration pipeline   | O1 and O2                                 | O3   |
|------------------------|---|--|
| Front-end              | None                                      | $\kappa_C, \kappa_U, \kappa_P, \kappa_T, f_{cc}$ |
| Low-latency (online)   | $\kappa_C, \kappa_{PU}, \kappa_T$         | $\kappa_C, \kappa_U, \kappa_P, \kappa_T, f_{cc}$ |
| High-latency (offline) | $\kappa_C, \kappa_{PU}, \kappa_T, f_{cc}$ | $\kappa_C, \kappa_U, \kappa_P, \kappa_T, f_{cc}$ |

errors (colored contours) introduced in  $\tilde{R}^{(\text{model})}$  (i.e.,  $\tilde{\eta}_{R;C} - 1$ ) if the time-variation of  $H_C$ , tracked by the factor  $\kappa_C$ , is not corrected. The top and bottom panels correspond to magnitude and phase of  $\tilde{\eta}_{R;C} - 1$ , respectively. The measured fractional variation of  $\kappa_C$  is typically at the level of 1%–2%, and can be as large as  $\sim 10\%$  in either detector. As shown in the figure, an uncorrected  $\sim 5\%$  change in  $\kappa_C$  will result in  $\sim 10\%$  systematic error in the magnitude of  $\tilde{R}^{(\text{model})}$ , near 100 Hz. Similar plots for impacts of uncorrected  $f_{cc}$ ,  $\kappa_T$ ,  $\kappa_P$ , and  $\kappa_U$  are shown in Appendix A.

The impact of TDCFs related to SRC detuning,  $f_s$  and  $Q$ , was not discussed in [17]. To study the impact from time-varying SRC detuning effect, we create an example with a perfectly tuned SRC reference model (i.e.,  $f_s = 0$  Hz) and vary only  $f_s$  ( $Q$  is fixed at 52.14). In (7), the time-varying  $f_s$  value is always positive but can be real or imaginary, corresponding to an anti-spring-like or a spring-like detuned optical response, respectively. We quantify the spring-like or anti-spring-like effect with  $f_s^2$  for simplicity; i.e.,  $f_s^2 < 0$  is a spring response, and  $f_s^2 > 0$  is an anti-spring response. Figure 9 shows  $\tilde{\eta}_{R;C} - 1$  in colored contours if the time-variation of  $f_s^2$  is not corrected. The variation of  $f_s^2$ , denoted by  $\Delta f_s^2$  on the vertical axis, covers both the anti-spring-like and spring-like detuned optical responses. For  $|\Delta f_s^2| \lesssim 25 \text{ Hz}^2$ , the impact is generally negligible. It has been found that occasionally we have  $|\Delta f_s^2| \gtrsim 50 \text{ Hz}^2$ , resulting in an error of  $\gtrsim 5\%$  in the magnitude of  $\tilde{R}^{(\text{model})}$ . See further discussion and treatment of the resulting systematic error in section 4.2.

An exceptional period occurred early in O3A due to two issues. (1) Prior to April 16, 2019, at Hanford,  $\kappa_P$  and  $\kappa_U$  were not applied because the estimates were computed using calibration lines separated by  $\sim 20$  Hz. Such large frequency separations invalidated approximations used to compute  $\kappa_P$  and  $\kappa_U$ . The systematic error introduced by this issue is accounted for when reporting the overall accuracy of the estimated  $h$  in section 5. (2) From April 1 to June 11, 2019 at Livingston and from April 16 to June 11, 2019 at Hanford, the complete complex values for all actuator TDCFs,  $\kappa_T$ ,  $\kappa_P$ , and  $\kappa_U$ , rather than only the real values were applied to  $h$ . Applying the complex actuator TDCFs was found to cause an overall increase in the systematic error in  $h$ . For all other time periods in O3A, only the real part of  $\kappa_T$ ,  $\kappa_P$ , and  $\kappa_U$ , were applied to  $h$ . The impact of (2) is discussed as follows.

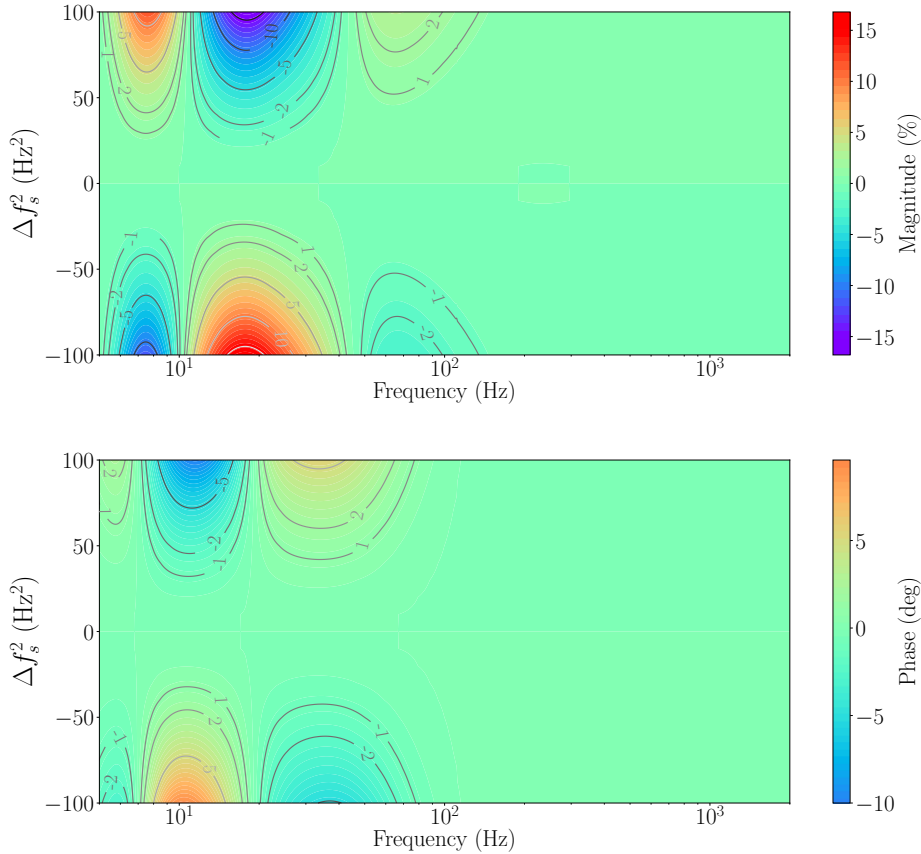


**Figure 8.** Magnitude (top) and phase (bottom) of the fractional error  $\tilde{\eta}_{R;C} - 1$  in the Hanford detector’s response [O3A Epoch (c)] as a function of frequency due to uncorrected gain variations in the sensing function, tracked by the scalar time-dependent factor,  $\kappa_C$ .

By design, all of the actuator TDCFs at the reference time equal  $1 + 0i$ . At other times, the actuator TDCFs can take different values, typically with the real term  $1 \pm 0.05$  and imaginary term  $(0 \pm 0.01)i$ . Non-zero imaginary terms may arise due to small physical effects that change during the observing period or because approximations used to estimate the TDCFs break down [34]. If the former is the case, then we expect that applying the full complex-valued TDCFs should reduce the measured systematic error in  $h$ . If the latter, then applying the full complex-valued TDCFs does not correctly compensate because there was no actual physical change so we expect the systematic error in  $h$  to increase. The application in early O3A described above was found to have typically increased measured systematic error, indicating that the latter is the problematic element during this period.

We characterize the response function systematic error in two cases: the full complex-





**Figure 9.** Magnitude (top) and phase (bottom) of the fractional error  $\tilde{\eta}_{R,C} - 1$  in the Hanford detector’s response [O3A Epoch (c)] as a function of frequency due to uncorrected time-dependent SRC optical spring frequency,  $f_s$  (with fixed  $Q=52.14$ ).

valued actuator TDCF’s are applied and only the real-valued actuator TDCF’s are applied. The method for computing the response function systematic error and associated uncertainty is described in section 5. For those results presented in section 5, we have only considered the actuator TDCF’s as real-valued. Figures showing side-by-side results from applying only real-valued and full complex-valued actuator TDCF’s are provided in Appendix B. The impact on the systematic error alone is  $\lesssim 1\%$  in magnitude and  $\lesssim 0.2$  deg in phase, and remains within the overall associated uncertainty. Work is underway to implement an improved method of computing the actuation TDCF’s that does not suffer this breakdown of approximations and the full complex-value may be trusted to reflect true physical effects [34].

#### 4.2. Deficiencies in the sensing function model at low frequencies

Understanding the results from weekly measurements of the Hanford sensing function below 20 Hz has posed a challenge unique in the advanced detector era. This section describes the results observed.

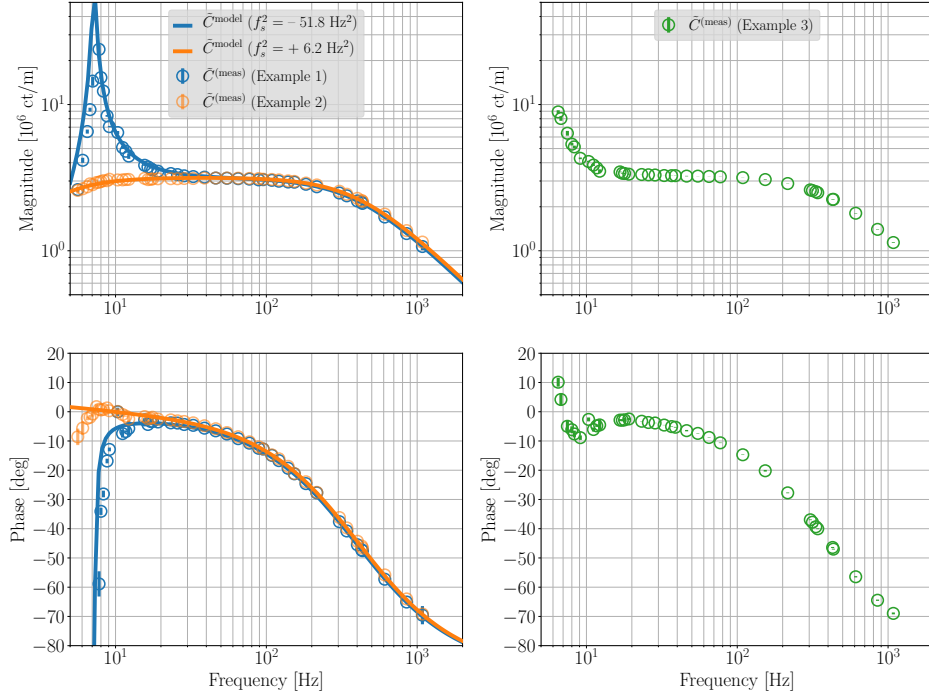
In the previous O1 and O2 observing runs, measurements at Hanford showed evidence for slight detuning of the SRC with respect to the arm cavities [14]. Detuning of the SRC can be caused by either misalignment or mode mismatch with the arm cavities; both misalignment and mode mismatch can change as the thermal lenses in the input test masses change [35]. To account for the impact on the sensing function, an invertible, phenomenological, analytic representation of an optical spring was included in the model, parameterized by  $f_s$  and  $Q$ , as in (5) (see derivation in [22]). A fixed, positive value of  $f_s^2$  was sufficient to describe the ensemble of sensing function measurements at Hanford throughout O1 and O2. The measurement ensemble of the Livingston detector showed no sign of detuning, and  $f_s$  was set to 0 Hz. In O3A, the Livingston detector continues to show no sign of detuning, and  $f_s$  remains set to 0 Hz. The Hanford detector measurements, however, now show clear evidence for detuning responses with both  $f_s^2 > 0$  and  $f_s^2 < 0$ .

There is also evidence for two-way cross-coupling between the DARM and angular control systems at Hanford in O3A, further modifying the response below 20 Hz. To avoid point defects [36], the Hanford detector alignment scheme has been modified to position the laser light impinging on arm cavity optics away from the center of the optics. Angular motion of the optics will therefore be sensed as DARM length change, and actuators used for angular control create DARM length change. When there is a second cross coupling from DARM length to the angular sensors, the angular control loop response impacts the measured sensing function. In that case,  $C^{(\text{meas})}$  shows a complex, low-frequency response inconsistent with detuning and (5).

While the Hanford sensing function is more complicated than (5) in O3A, we nevertheless use continuous measurements of  $f_s$  to monitor changes in the sensing function. These measurements show a consistent evolution of  $f_s^2$  from positive to negative over the first  $\sim 2$  hours after the detector achieves “observation-ready” performance but before reaching thermal equilibrium. Once thermal equilibrium is achieved, the continuously monitored value stabilizes and shows only small variations at the level of  $|\Delta f_s^2| \lesssim 1 \text{ Hz}^2$ .

Figure 10 shows several examples of sensing function measurements at Hanford in O3A. The left panels show the comparison between (5) and two response measurements dominated by detuning. The right panels show an example measurement in which the low frequency response is dominated by angular cross-coupling.

We conclude that the Hanford sensing function low-frequency response depends on the complicated interaction between detuning, cross-coupling, and the thermal state of



**Figure 10.** Examples of O3A weekly sensing function measurements at Hanford in magnitude (top) and phase (bottom). The left panels show two measurements that behave like (5), with positive and negative  $f_s^2$  values indicated in the legend. The right panels show a measurement that cannot be explained by (5). Vertical error bars crossing the markers indicate uncertainties of the measurements, most of which are too small to be seen by eye.

the detector as exemplified by the measurements presented in this section. Since these effects are not modeled and poorly monitored, we use the discrepancy between model and the collection of weekly measurements to represent this deficiency as an unknown systematic error below 20 Hz using GPR described in section 4.3. Further limitations of (5) at frequencies above 20 Hz are discussed in section 4.4.

#### 4.3. Accounting for unknown static frequency dependence

Unknown systematic errors are accounted for by computing the complex-valued residuals between the model and measurements of the sensing and actuation functions. Weekly measurements taken throughout each epoch, including data at frequencies above 1 kHz and over relevant frequency bands (discussed in further detail below), are taken into consideration. Each measurement of these interferometer components has all known loop sub-components and all known TDCFs applied such that only unidentified systematic and measurement statistical uncertainty remain, resulting in measures of  $\tilde{\eta}_C$  and  $\tilde{\eta}_{A_i}$ . The

complex-valued frequency dependence and uncertainty of these residuals are characterized using the GPR method [37]. The posterior results from the GPR are then used as part of the overall response function uncertainty calculation [14].

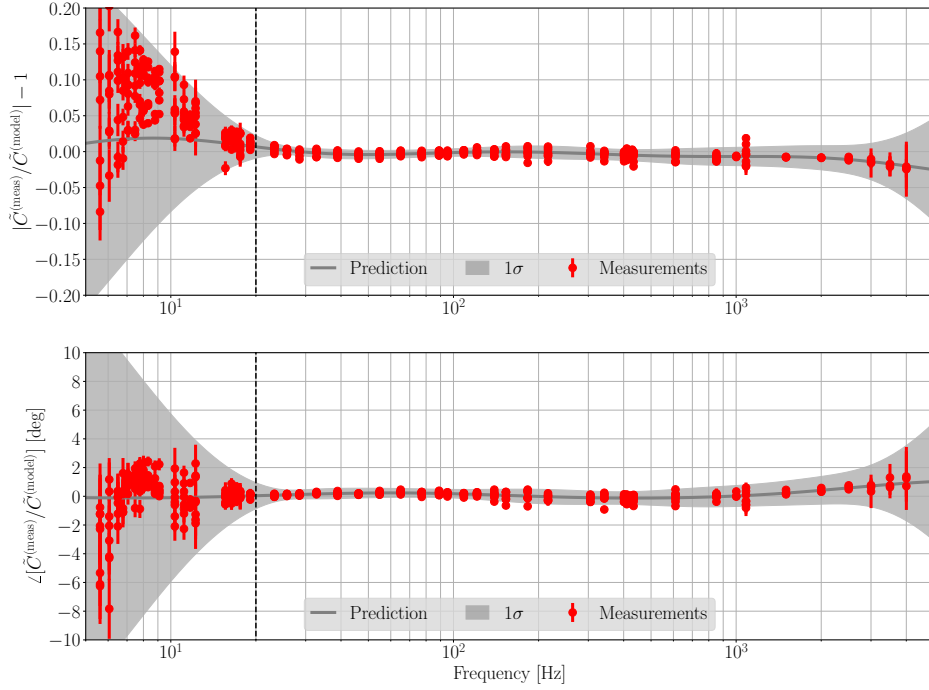
The GPR trains on the residual data using a physically motivated covariance kernel, defined as

$$k(\log(f), \log(f')) = \gamma_1^2 + \gamma_2^2 \exp\left(-\frac{(\log(f) - \log(f'))^2}{2\ell^2}\right), \quad (24)$$

where  $\{\gamma_1, \gamma_2, \ell\}$  are the hyperparameters of the covariance kernel with the following bounding values;  $\gamma_1 \in [0.9, 1.1]$ ,  $\gamma_2 \in [0.1, 2.0]$ , and  $\ell \in [0.5, 1.5]$ . Previous analysis in O1 and O2 determined the covariance kernel hyperparameters via GPR of the magnitude and phase residuals separately for each of the sensing and actuation functions [14]. For O3, we determined the covariance kernel for the complex-valued residuals such that correlations between magnitude and phase are preserved in the residuals for a given model. In addition, previous analysis provided more hyperparameters of the covariance kernel and did not restrict the parameter values away from unphysical regions of parameter space (e.g., covariance between complex-valued residuals of nearby frequency points should be preserved whereas residuals from widely spaced frequencies should have very small covariance). This updated kernel, hyperparameter ranges, and use of the complex-valued residuals addresses all of these issues.

Measurements of the Hanford sensing function during O3A have shown significant deviations from the reference model at frequencies  $\lesssim 20$  Hz (see section 4.2). It has proven difficult to model and track these changes a priori, so the variations are included as part of the residuals. Figure 11 shows the measured residuals together with the GPR posterior confidence intervals. Since residuals have larger variance at frequencies below 20 Hz, the posterior values of the covariance kernel hyperparameters are expected to have larger values at those frequencies, covering the range of measured residuals. For the Hanford sensing function residuals only, the hyperparameter fit interval is modified to be  $\ell \in [0.01, 0.5]$ , and the range of accepted frequency points is changed so that, by visual inspection, the uncertainty interval adequately covers the range of all measurement residuals (as shown in figure 11).

Although the Hanford sensing function residuals and, in turn, the posteriors from the GPR, are larger at frequencies below 20 Hz, the actual impact on the response function uncertainty is small. This is because the contribution to the response function by the sensing function is smaller at low frequencies than the contribution from actuation stages [i.e., below 20 Hz, the values of the purple curve in figure 4(a) are at least a factor of 2 times smaller than the total  $A$  curve]. The Hanford actuation measurements do not show such residual variations as those seen in the sensing measurements. No such variation is seen in any interferometric measurements of the Livingston detector.

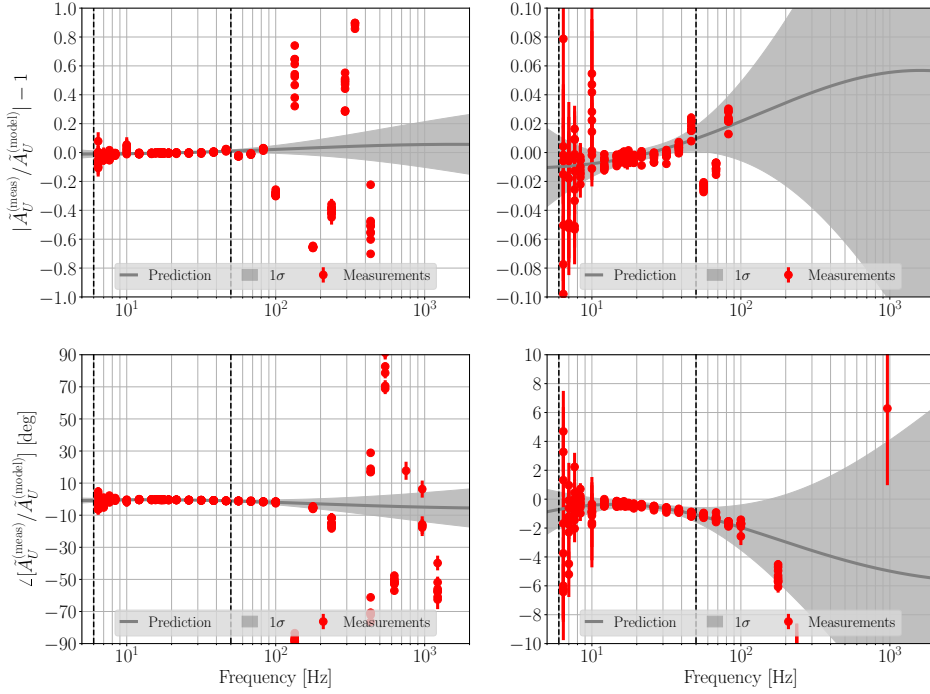


**Figure 11.** Gaussian Process Regression results of the systematic error in Hanford’s sensing model [O3A Epoch (c)]. The red markers are the residuals  $\tilde{\eta}_C$  between all the sensing measurements and the model in magnitude (top) and phase (bottom). The dark grey curve is the best prediction of the systematic error. The light grey shaded region indicates the  $1\sigma$  uncertainty on the systematic error.

In addition to the sensing function, it is instructive to consider the GPR for the UIM stage. Measurements for UIM actuation stages at both detectors are consistent with model expectations between 6 and 50 Hz, so only data within this frequency range are used by the regression (see, for example, Hanford results in figure 12). Similar to the sensing function below 20 Hz, the UIM contribution to the response function above 50 Hz is negligible (see figure 4). Outside this band, especially above 50 Hz, measurements do not agree with the model. Restricting the regression from 6 to 50 Hz may not accurately reflect the potentially large systematic error between the model and measurements above 50 Hz. The impact of neglecting this systematic effect above 50 Hz, however, is negligible because of the small UIM contribution. This is discussed further in section 4.4.

#### 4.4. Quantifying uncompensated systematic errors

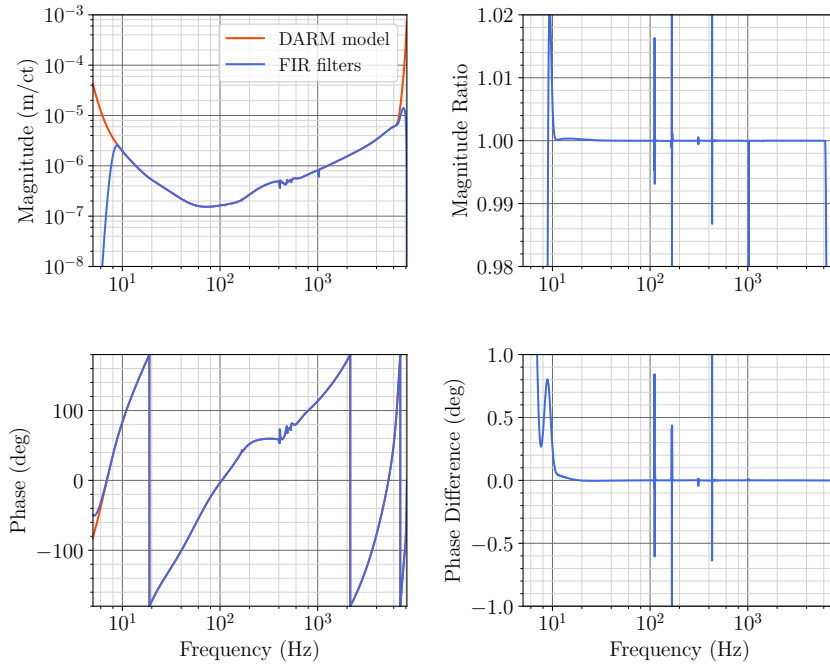
Some systematic errors in the calibrated strain data are known but uncompensated in O3A, even in the most-accurate, offline production of  $h$ . The sources of errors include: (1) FIR filters used to reproduce the offline  $h$  data stream, (2) intentionally applied low-pass



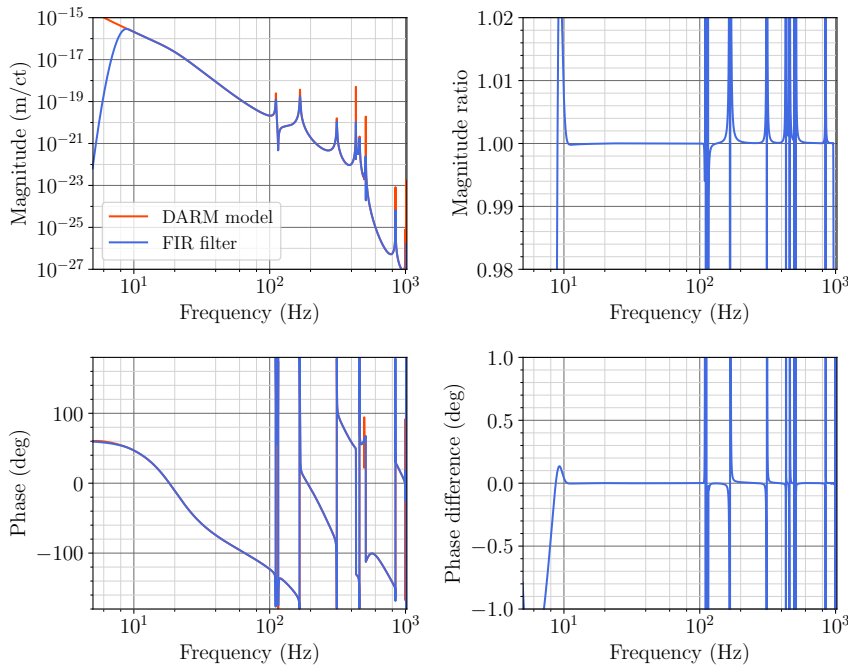
**Figure 12.** Gaussian Process Regression results of the systematic error in Hanford’s UIM actuation model [O3A Epoch (c)]. The panels on the right are zoomed in on the vertical axis to display the error and uncertainty in the band of interest. The red markers are the residuals  $\tilde{\eta}_{AU}$  between all the UIM actuator measurements and the model in magnitude (top) and phase (bottom). The dark grey curve is the best prediction of the systematic error. The light grey shaded region indicates the  $1\sigma$  uncertainty on the systematic error.

and high-pass filters for improved data handling, (3) the exclusion of cross-coupling with auxiliary degrees of freedom in the actuation or sensing models, (4) imperfect modeling of the UIM force-to-displacement transfer function, (5) unknown residual time delay between actuator stages, (6) impact of SRC detuning on the coupled-cavity pole frequency, (7) imperfect modeling of the response of a Fabry-Pérot cavity to length changes, (8) non-rigid-body displacement of the test masses in the presence of Pcal forces, and (9) uncertainty in the timing synchronization between multiple elements of the DARM control system. Sources (7)–(9), without being accounted for, result in errors mainly at frequencies above 1 kHz. These errors are typically small enough that they do not significantly contribute to the uncertainty and systematic error of the response function at frequencies between 20 and 2000 Hz, although some features do contribute appreciably in narrow frequency bands. We discuss and quantify each of these errors in this section for completeness, but do not include them in the final estimate of the uncertainty.

FIR filters cannot perfectly reproduce all details of the DARM loop model. In the frequency band from 10 Hz to 6 kHz, however, these errors are generally less than 0.1% in magnitude and 0.1 deg in phase. Figure 13 shows a comparison between the FIR filter implementation of  $\tilde{R}^{(\text{model})}(f)$  in the offline calibration pipeline and the frequency-domain DARM model. Some narrow-band systematic errors can be seen in the residual (right panels), caused by sharp spectral features that are difficult to resolve using FIR filters only a few seconds in duration. A Kaiser window is applied to the FIR filter in the time domain, resulting in a frequency resolution of  $\sim 3$  Hz. Such errors mostly originate from the filters that model the actuation system, especially the UIM stage at Hanford. Figure 14 displays the comparison between the frequency response of the UIM FIR filter and the frequency-domain model at Hanford. Narrow-band systematic errors caused by the limitation of short-duration FIR filters are left uncompensated. This is a compromise between the data-loss due to FIR impulse response settling and the accuracy of reproducing the sharp spectral features.



**Figure 13.** Comparison between the effective response function implemented by the FIR calibration filters (blue) and the frequency-domain model of the response function (red) at Hanford. The left panels show the response functions, and the right panels show the residuals. The top and bottom panels are for magnitude and phase, respectively. Note that the sharp, narrow-band feature from the UIM actuator dynamics (e.g., at  $\sim 150$  Hz) is difficult to model with short FIR filters (see detailed discussion in text).



**Figure 14.** Comparison between the frequency response of the UIM actuation filter (blue) and the frequency-domain UIM model (red) at Hanford. Panels on the left and right display the two UIM transfer functions and the fractional residual between them, respectively. The top and bottom panels are for magnitude and phase, respectively. The sharp features from  $\sim 100$ – $500$  Hz are difficult to resolve with short FIR filters.

Below 10 Hz and above 6 kHz, the detector sensitivity degrades rapidly, and the data become dominated by  $\Delta L_{\text{free}}$  that is not of astrophysical origin. For convenience in data handling, as well as the prevention of spectral leakage or aliasing, aggressive high-pass and low-pass FIR filters (with corner frequencies of 9 Hz and 6 kHz, respectively) are applied in post-processing in both the low-latency and high-latency calibration pipelines. The well-understood systematic error resulting from these additional FIR filters only impacts these extreme frequency regions.

The physical construction of the detectors minimizes the cross-coupling between auxiliary control loops and the DARM loop. The auxiliary loop control designs are adjusted to further reduce this cross-coupling as the detectors sensitivity is improved. Section 4.2 describes the first evidence of undesirable interactions between auxiliary control loops and the DARM loop during an observation period. While more sophisticated models that account for these interactions are academically interesting, future improvements to the detector hardware and control system parameters will render complex models unnecessary. Residual errors from cross-coupling effects are accounted



for with techniques described in section 4.3.

The UIM-to-TST, force-to-displacement (i.e., force from UIM stage to displacement at TST stage) transfer function shows a number of resonant features above 50 Hz (as discussed in section 4.3; see figure 12). In the mid-frequency band (50–250 Hz), the features result from the twisting and/or bending of the UIM stage vertical blade springs in the longitudinal direction as a result of the force producing  $\Delta L_{\text{ctrl}}$ . Between O2 and O3, damping mechanisms were modified on the UIM vertical isolation blade springs (see [38] for details). The increased weight of these improved dampers lowered the bending mode frequencies of the blades, changing the force-to-displacement transfer function for the UIM. While these improved dampers change the bending modes in similar ways for both Hanford and Livingston detectors, the impact is only significant on the Hanford response function because of different choices in  $F_U$  between two detectors [compare figures 4(a) and 4(b)]. These changes were not included in the UIM actuator model, resulting in an underestimate of the contribution from  $A_U$  to  $R$  around the frequency of the bending mode resonances. This results in three narrow, resonant features appearing in spectra of the Hanford calibrated data stream in the band 150–155 Hz. These narrow features have a maximum excursion ( $\sim 1$  Hz width) of  $\pm 3\%$  in magnitude and 3 deg in phase. Careful inspection of figure 15 (discussed later in section 5) at  $\sim 150$  Hz hints at this error but does not resolve it in the overall systematic error estimate. Recent investigations and efforts have resolved this error, but it remains an uncompensated systematic error in the Hanford O3A data.

The estimate of the overall residual time delay from each actuation stage is determined by the MCMC fit. However, as shown in figure 12 and discussed above, the data input to the MCMC may include discrepancy between the model and measurement unrelated to a time delay. In that case, fitting for only a scalar  $H_i$  and a delay  $\tau_i$  is incorrect. We thus, after accounting for all understood time delays, restrict the frequencies of the MCMC to a band where an actuator transfer function appears to be frequency-independent in order to determine  $H_i$ . Any remaining uncertainty in timing for each actuator stage ( $\tau_i$ ) or in the sensing function ( $\tau_C$ ) is determined via GPR as described in section 4.3.

Two effects left out of the opto-mechanical portion of the sensing function model  $C^{(\text{model})}$  may have potential impact. First, in (5) we assume the SRC detuning effect is small enough that the coupled-cavity, single-pole response of the coupled arm and SRC cavities (the first parenthetical term) and the detuned SRC response (the second parenthetical term) can be separated and parameterized independently by  $f_{cc}$  and  $f_s$ . The physical model from which (5) was derived [22, 39], however, suggests that the response at frequencies  $\sim 300$  Hz and above may no longer be described by  $f_{cc}$  alone when SRC detuning is sufficiently large. The amount of detuning is proportional to any modification between the GW signal phase and the phases determined by two physical

quantities: (a) the homodyne phase  $\zeta$  determined by the laser wavelength, and (b) the signal extraction phase  $\phi_{\text{SRC}}$  determined by the SRC cavity length. Both  $\zeta$  and  $\phi_{\text{SRC}}$  are nominally 90 deg. The residual between the physical model response and the approximate model response in (5) shows the systematic error is frequency dependent, but does not exceed 1% in magnitude or 1 deg in phase at frequencies below 1 kHz for the measured extremes of detuning,  $|f_s^2| \lesssim 75 \text{ Hz}^2$  (or equivalently,  $|\phi_{\text{SRC}} - 90^\circ| < 1^\circ$  and  $|\zeta - 90^\circ| < 3^\circ$ ). Second, this single-pole response is also an approximation to the complete response of the Fabry-Pérot arm cavities fluctuation in their lengths [40]. This approximation leads to errors in the sensing function at high frequencies above 1 kHz in both magnitude and phase (larger in phase). The resulting phase error is compensated by an artificial time delay of  $-11.7$  microseconds, included in  $\tau_C$  [the last term in (5)]. The magnitude error, increasing with frequency up to 4% at 5 kHz, is left uncompensated [41]. The systematic errors in magnitude and phase resulting from these two approximations of the detector full opto-mechanical response are accounted for in the uncertainty of unknown systematic error via GPR.

During Pcal excitation ( $x_{\text{Pcal}}$ ), the actuation forces deform the test masses in their natural bending modes, producing a deformation-induced, arm cavity length variation (not equivalent to the displacement of center of the mass) sensed by the interferometer. This phenomenon impacts the accuracy of  $\Delta L_{\text{Pcal}}$  at high frequencies ( $\gtrsim 1$  kHz) to a level depending on the positions on the test mass surface where the Pcal beams reflect [12, 24]. We estimate that the reflecting positions are within  $\pm 2$  mm of their optimal locations (close to the nodal circle of the dominant mode). The magnitude error in the estimate of  $H_{\text{Pcal}}$  due to the deformation is  $\lesssim 0.1\%$  below 1 kHz, increases with frequency, and reaches at most  $\sim 5\%$  at 5 kHz [24]. The phase error may also increase with frequency, but is expected to be less than 0.5 deg even at 5 kHz. This may limit the accuracy of the long-duration measurements used to characterize the sensing function. We see no evidence for this error exceeding all other known and unknown systematic errors above 1 kHz (e.g., see figure 11). As such, this effect has been excluded from  $H_{\text{Pcal}}$ .

Finally, within a given detector, the analog and digital components of the DARM loop are synchronized via a sophisticated timing system [42]. The uncertainty in synchronization of these components is less than 1 microsecond throughout O3A [43]. The frequency-dependent phase impact from timing uncertainty on an individual detector is believed to be within the bounds of the unknown systematic error,  $\tilde{\eta}_C$  or  $\tilde{\eta}_A$ , estimated via GPR, and hence is not explicitly accounted for. The GW detectors within the network are synchronized to each other via the GPS receivers of the timing systems. The network timing uncertainty, estimated to be at the level of 10 microseconds [44], is negligible compared to the uncertainty in estimates of the time-of-arrival for any GW event (typically at the level of 1 millisecond).

## 5. Combined error and uncertainty estimate

### 5.1. Estimate at a given time

We numerically estimate the combined estimate of systematic error and uncertainty,  $\tilde{\eta}_R(f; t)$ , in each detector's response function at a given time  $t$  as follows.

Ten thousand response functions,  $\tilde{R}_i(f; t)$ , are constructed with

$$\tilde{R}_i(f; t) = \eta_{\text{Pcal}_i} \left[ \frac{1}{\tilde{\eta}_{C_i}(f)\tilde{C}(\boldsymbol{\lambda}_i^C; f; t)} + \tilde{\eta}_{A_i}(f)\tilde{A}(\boldsymbol{\lambda}_i^A; f; t)\tilde{D}(f) \right]. \quad (25)$$

Here,  $i$  indexes each response function and all draws associated with it. The  $i$ th sensing and actuation functions  $\tilde{C}(\boldsymbol{\lambda}_i^C; f; t)$  and  $\tilde{A}(\boldsymbol{\lambda}_i^A; f; t)$  are constructed using (7) and (8), with the  $i$ th draw from the MCMC posterior distributions of the reference model parameters,  $\boldsymbol{\lambda}^C$  and  $\boldsymbol{\lambda}^A$  (section 3.3). Within the time-dependent  $\tilde{C}(\boldsymbol{\lambda}_i^C; f; t)$  and  $\tilde{A}(\boldsymbol{\lambda}_i^A; f; t)$ , TDCFs at time  $t$  are applied (section 4.1). To account for the uncertainties of the TDCFs, we draw TDCF samples from normal distributions centered at the values recorded at time  $t$  with  $1\sigma$  standard deviation calculated using (19). The complex-valued, fractional, frequency-dependent residual functions,  $\tilde{\eta}_{C_i}(f)$  and  $\tilde{\eta}_{A_i}(f)$ , are drawn from the sensing and actuation GPR posterior distributions, respectively (section 4.3). By drawing samples from the MCMC and GPR posterior distributions, the covariance between parameters in  $\boldsymbol{\lambda}^C$  or  $\boldsymbol{\lambda}^A$ , and covariance between frequency points of  $\tilde{\eta}_{C_i}(f)$  or  $\tilde{\eta}_{A_i}(f)$ , is preserved. Finally,  $\eta_{\text{Pcal}_i}$  is an overall multiplicative real-valued scale factor drawn from a normal distribution centered at  $\eta_{\text{Pcal}}$  for each detector, with  $1\sigma$  standard deviation equal to the Pcal system uncertainty (section 3.1). This factor accounts for the Pcal uncertainty and systematic error common to all interferometric measurements and TDCF computations for a given detector. Therefore it is convenient to apply  $\eta_{\text{Pcal}_i}$  to each  $\tilde{R}_i(f; t)$  rather than equivalently to  $H_{\text{Pcal}}$ , or to each interferometric transfer function and TDCF calculations that involve  $\Delta L_{\text{Pcal}}$ .

The time-dependent MAP response function  $\tilde{R}_{\text{MAP}}(f; t)$  is constructed with the MAP parameters of the sensing and actuation functions ( $\boldsymbol{\lambda}_{\text{MAP}}^C$  and  $\boldsymbol{\lambda}_{\text{MAP}}^A$ ), similar to (3), given by

$$\tilde{R}_{\text{MAP}}(f; t) = \frac{1}{\tilde{C}(\boldsymbol{\lambda}_{\text{MAP}}^C; f; t)} + \tilde{A}(\boldsymbol{\lambda}_{\text{MAP}}^A; f; t)\tilde{D}(f). \quad (26)$$

At time  $t$ , if the TDCFs are applied to the  $h$  data stream (i.e., the time-dependent systematic error is removed from the estimated  $h$ ), the same TDCFs recorded at that time are also applied to  $\tilde{C}(\boldsymbol{\lambda}_{\text{MAP}}^C; f; t)$  and  $\tilde{A}(\boldsymbol{\lambda}_{\text{MAP}}^A; f; t)$ . Otherwise the reference TDCF values are used to construct  $\tilde{R}_{\text{MAP}}$ , such that the same systematic error (due to the uncorrected TDCFs) as that left in the estimated  $h$  is kept in  $\tilde{R}_{\text{MAP}}$ .

We then divide each  $\tilde{R}_i(f; t)$  by  $\tilde{R}_{\text{MAP}}(f; t)$  to create the probability distribution of  $\tilde{\eta}_{R_i}(f; t)$ . At any given time  $t$  and frequency  $f$ , the median (50th percentile) value

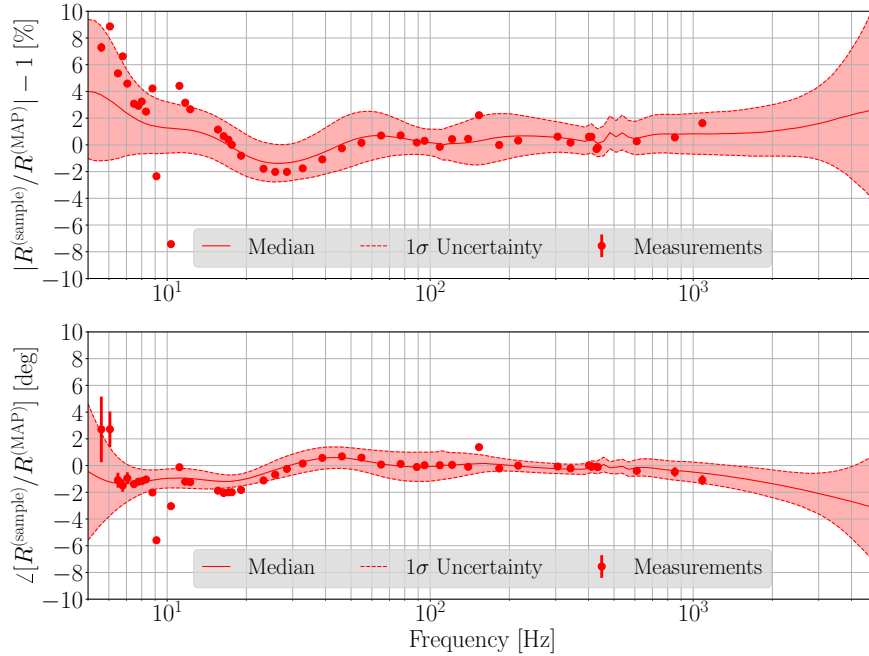
of the distribution  $\tilde{\eta}_{R_i}(f; t)$  represents the total systematic error in  $\tilde{R}^{(\text{model})}(f; t)$  at that time and frequency. The 16th and 84th percentiles represent the lower and upper bounds, respectively, of the combined systematic error and  $1\sigma$  statistical uncertainty in  $\tilde{R}^{(\text{model})}(f; t)$ . As such, these percentiles of  $\tilde{\eta}_{R_i}(f; t)$  represent the complex-valued, frequency-dependent, overall uncertainty and systematic error bounds of  $h$  at time  $t$ .

An example of the combined uncertainty and error estimate,  $\tilde{\eta}_R(f; t)$ , for the Hanford detector is shown in figure 15. The vertical axes indicate the excursions of  $\tilde{\eta}_R(f; t)$  from zero systematic error, i.e., unity magnitude (top panel) and zero phase (bottom panel). The solid curve shows the median value of  $\tilde{\eta}_R(f; t)$ , indicating the best estimate of the frequency-dependent systematic error in the response function at that time. The dashed curves bounding the shaded region represent the collection of  $1\sigma$  uncertainties, including that of the systematic error. Within the frequency band 20–1000 Hz, the red dots show a swept-sine measurement of  $hL/\Delta L_{\text{Pcal}}$  taken on September 16, 2019, that aligns with the estimate of  $\tilde{\eta}_R(f; t)$  at that time. Some measured data points at frequencies below 20 Hz deviate from the median curve and exceed the  $1\sigma$  uncertainty bounds. This is evidence of systematic error induced by, e.g. detuning between the SRC and the arm cavities (see section 4.2), or resonant modes of the quadrupole suspension actuator stages that are not sufficiently accounted for when estimating  $\tilde{\eta}_R(f; t)$ . The single outlying data point around 150 Hz is caused by the imperfect dynamical model of the UIM stage at Hanford (see section 4.4).

## 5.2. Estimate over time

Estimates of the combined systematic error and uncertainty over longer periods are generated using the collections of time-specific estimates described in section 5.1. To quantify the final calibration accuracy and precision in O3A, the entire duration is split into three epochs for Hanford, and two for Livingston (see table 1). Each epoch is defined by a physical configuration change in the detector. Within each epoch only TDCF’s vary. Previous shorter-duration observing runs did not require intra-run epochs, hence estimates of systematic error and associated uncertainty were constructed for those entire observing runs [14]. In O3A, the combined uncertainty and systematic error for each epoch is quantified using the collection of percentile curves of  $\tilde{\eta}_{R_i}(f; t)$  described above, and shown in figures 16 and 17.

The procedure of calculating these results is as follows. First, the distribution of  $\tilde{\eta}_{R_i}(f; t)$  is computed with a 1-hour cadence during observing periods. Second, we compute the 16th, 50th, and 84th percentile curves from each of these distributions representing the systematic error and  $1\sigma$  uncertainty bounds at that time (i.e., the lower, median, and upper curves shown in figure 15). Third, all hourly percentile curves within an epoch form a complex-valued, frequency-dependent “epoch distribution” of systematic



**Figure 15.** Combined error and uncertainty estimate at the reference time of Epoch 3 for the Hanford detector. The top and bottom panels show the frequency-dependent excursions of the response from unity magnitude and zero phase compared to  $\tilde{R}_{\text{MAP}}$ , respectively. The dashed curves indicate the 16th and 84th percentiles of the response distribution  $\tilde{R}_i$  compared to  $\tilde{R}_{\text{MAP}}$ . The solid curve is the median of the distribution, indicating the frequency-dependent systematic error in  $\tilde{R}^{(\text{model})}$ . The shaded region bounded by the dashed curves represents the  $1\sigma$  uncertainty bounds on the systematic error. The red dots show a set of validating measurement taken on September 16, 2019, which are generally consistent with the overall uncertainty estimate. Vertical error bars crossing the markers indicate uncertainties of the measurements, most of which are too small to be seen by eye.

error estimates (the medians and the upper and lower uncertainty bounds). Finally, the epoch distribution is used to determine the rate that the upper and lower bounds exceed a given value (frequency-dependent; in magnitude and phase) and the variability of those curves.

In the left panels of figures 16 and 17, the white curves, constructed from the 50th percentile of the hourly medians of the epoch distribution  $\tilde{\eta}_{R_i}(f; t)$ , show the frequency-dependent systematic error for each epoch. The 68%, 95%, and 99% percentile curves of the  $1\sigma$  uncertainty boundaries in the epoch distributions are shown as dark, moderate, and light shaded regions, respectively. These percentile curves quantify the time-dependent variation of the combined uncertainty and systematic error bounds over the entire epoch. Figures 16(b), 16(c), 17(a), and 17(b) show that the variation of the overall uncertainty bounds is generally negligible (i.e., the 68%, 95%, and 99%

percentile boundaries almost overlap in each epoch). Figure 16(a), however, shows that the variation of the uncertainty bounds during the first epoch of O3A for the Hanford detector is not negligible. The 95% and 99% percentiles deviate from the 68% percentile due to uncorrected  $\kappa_U(t)$  and  $\kappa_P(t)$  variations during the first 16 days of the first epoch at Hanford (see section 4.1).

In the right panels, we introduce a simplified presentation of the results. For brevity, the systematic error and uncertainty estimate for a given epoch across a given frequency band is quoted by calculating the maximum excursions from zero systematic error. The maximum excursion values are determined as follows. First, at all frequencies, the absolute values of the  $1\sigma$  upper and lower bounds of the epoch-distribution curves are computed in each epoch. Then, a frequency-dependent curve (dashed) is formed by taking the larger of the two absolute values at any given frequency. The solid curve in each of the right panels represents the absolute values of the white median curve on the left. Finally, the maximum value of these curves is determined within a frequency band, over which a given GW analysis is conducted. The star and dot markers indicate the maximum excursions in the frequency band 20–2000 Hz, corresponding to the 68% and 50% percentiles of the epoch distribution, respectively. The values of these markers in each epoch for each detector are listed in table 3. The maximum median values represent the best estimate of the systematic error bounds in the band 20–2000 Hz.

**Table 3.** O3A calibration epochs and the maximum  $1\sigma$  and median excursions of response from unity magnitude and zero phase compared to  $\tilde{R}_{\text{MAP}}$ , in the frequency band 20–2000 Hz. The maximum median values represent the best estimate of the systematic error bounds.

| Hanford epoch     | Max $1\sigma$<br>magnitude [%] | Max $1\sigma$<br>phase [deg] | Max median<br>magnitude [%] | Max median<br>phase [deg] |
|-------------------|--------------------------------|------------------------------|-----------------------------|---------------------------|
| (a) Mar 28–Jun 11 | 6.96                           | 3.79                         | 1.58                        | 0.86                      |
| (b) Jun 11–Aug 28 | 4.11                           | 2.34                         | 1.15                        | 0.92                      |
| (c) Aug 28–Oct 1  | 3.33                           | 1.53                         | 1.42                        | 1.00                      |
| Livingston epoch  | Max $1\sigma$<br>magnitude [%] | Max $1\sigma$<br>phase [deg] | Max median<br>magnitude [%] | Max median<br>phase [deg] |
| (a) Mar 28–Jun 11 | 6.37                           | 3.49                         | 1.13                        | 1.59                      |
| (b) Jun 11–Oct 1  | 5.99                           | 3.68                         | 1.09                        | 2.09                      |

As detectors have become more sensitive and more transient GW events are observed, it is desirable to frequently deliver offline-calibrated data and estimates of the systematic error and uncertainties to GW analyses. Therefore, in O3A (and for future observing periods), that data and the overall uncertainty for collections of epochs as described above are delivered in  $\sim 3$ -month intervals, the boundaries of which are coincidentally

aligned with those of the Hanford epochs. Balancing the requirements of (a) delivering high-quality data and uncertainty estimates quickly and (b) maintaining systematic error at a level that does not impact astrophysical parameter estimation requires that we do not intend to revise data or estimates of previously vetted intervals, unless circumstances are extraordinary.

### 5.3. Interpretation and discussion

Astrophysical parameter estimation for any GW event in O3A has used the most accurate, offline-calibrated data [5, 6, 7]. The calibration systematic error and uncertainty folded into the parameter estimation is informed by the 68% percentile boundaries of the single hourly  $\tilde{\eta}_{R_i}(f; t)$  distribution for the time closest to the event (e.g., figure 15). A five-point interpolation of the frequency dependent boundary curves is used as an approximation to the full  $\tilde{\eta}_{R_i}(f; t)$  distribution [45, 46]. In searches for persistent astrophysical signals, the offline calibrated data and the 68% percentile curve from each epoch distribution is used as representative of the uncertainty and systematic error estimate for the entire duration of the search (e.g., figures 16 and 17).

Throughout all epochs of O3A, the systematic error is less than 2% in magnitude and 2 deg in phase in the band 20–2000 Hz at both detectors (as indicated by the solid curves in the right panels of figures 16 and 17; see table 3 also). As discussed in section 5.2, we expect GW events within a given epoch will have the same estimated systematic error defined by the physical configuration of the detector. In the first epoch at Hanford [figure 16(a)], the uncertainty on the systematic error is larger than usual in the 1–4 kHz frequency band because no measurement had yet been made in that band. The uncertainty estimate is also larger in the first epoch at Livingston because measurements in the 1–4 kHz band were sparse [figure 17(a)]. Also at Livingston, the increase in the contribution of the TST actuator to the response function at  $\sim 50$  Hz that occurred between O2 and O3A [as shown in figure 4(b)] leads to the relatively higher uncertainty around 50 Hz, as shown in figure 17. We anticipate reduction of this uncertainty at  $\sim 50$  Hz in future observing runs.

The LIGO Scientific Collaboration and Virgo Collaboration use near-realtime analyses to quickly process data in search of transient GW sources, enabling multi-messenger astrophysics [47]. These analyses use the low-latency estimate of  $h$  for detection of GW events and preliminary parameter estimation. Low-latency data, however, occasionally contains increased systematic errors due to a variety of factors. The increased systematic errors are often reduced after a short period ( $\sim$ weeks). The maximum systematic error in the low-latency data does not exceed 6% in magnitude and 5 deg in phase at Hanford, and does not exceed 10% in magnitude and 6 deg in phase at Livingston across the frequency band 20–2000 Hz at any time during O3A.

This is verified by comparing the low-latency product  $hL$  to  $\Delta L_{\text{Pcal}}$  during various forms of Pcal measurements in the 5–1200 Hz band made periodically throughout O3A (e.g., as shown in figure 15). Although the systematic error and associated uncertainty is larger in the low-latency calibrated data than in the high-latency data, analyses using the low-latency data can nevertheless confidently detect GW events and make rapid astrophysical parameter estimates. Near-realtime analyses have been shown to be robust against calibration errors of this scale [48].

## 6. Conclusion

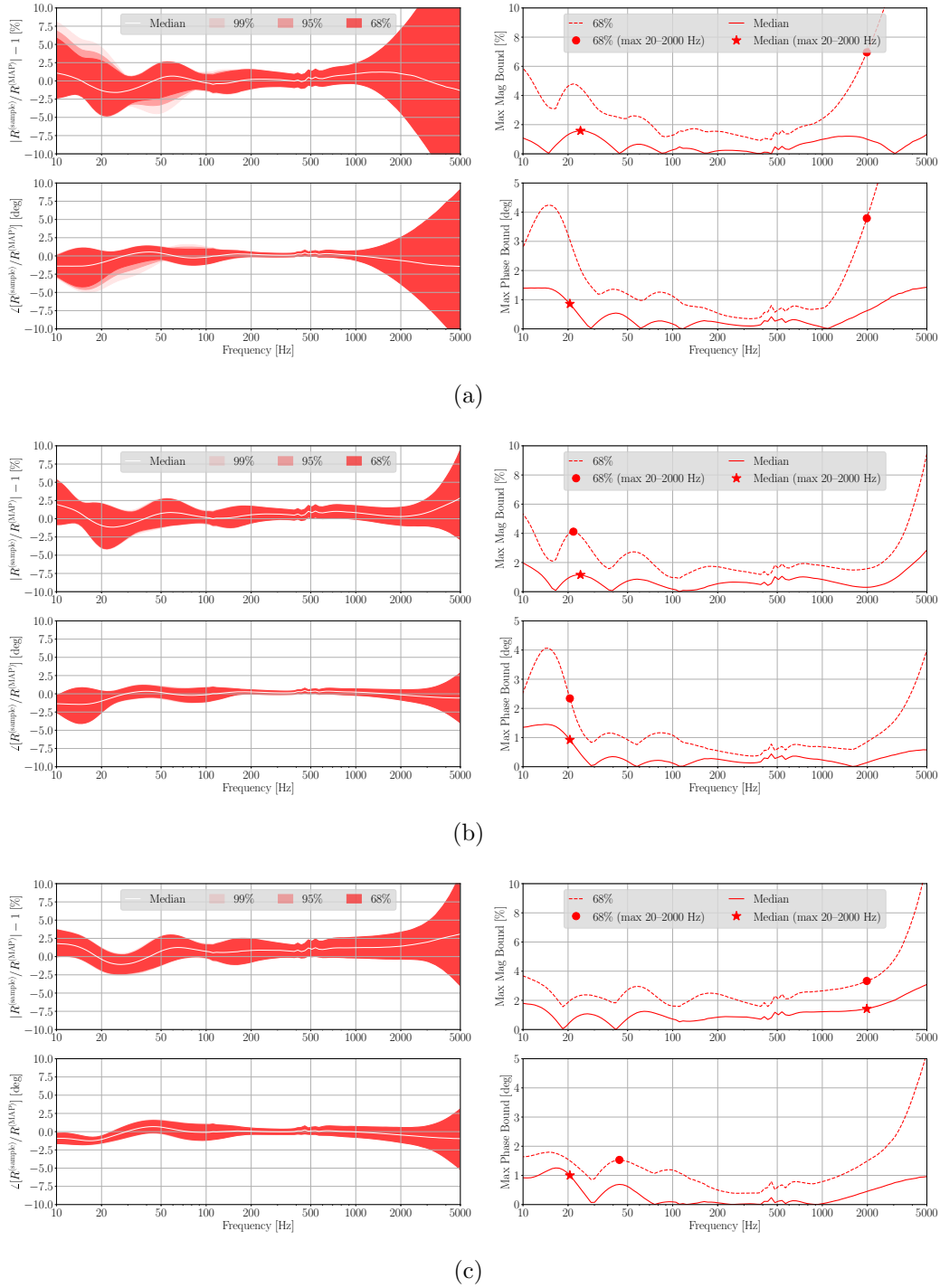
In this paper, we (1) review the procedure for creating a model of the DARM loop used to produce the calibrated data streams,  $h$ , for the Advanced LIGO detectors; (2) present the systematic errors incurred at each stage of that procedure; and (3) quantify the resulting overall accuracy and precision for the most accurate, offline version of  $h$  used for GW astrophysical parameter estimation in O3A. The discussion of systematic error includes all known sources, and, where possible, how they have been accounted for in  $h$  or in the overall systematic error and uncertainty estimate. In O3A, the overall, combined systematic error and associated uncertainty of the most accurate, offline-calibrated data is within 7% in magnitude and 4 deg in phase in the frequency band 20–2000 Hz. In this same band, the systematic error alone is estimated to be below 2% in magnitude and 2 deg in phase. This is similar to the accuracy and precision as achieved by LIGO in O2 [14]. Current detection of GW events and estimation of their astrophysical parameters are not yet limited by such levels of uncertainty and systematic error [7, 45].

As the global GW detector network sensitivity increases, however, detector calibration systematic error and uncertainty plays an increasingly important role. Limitations caused by calibration systematics on estimated GW source parameters, precision astrophysics, population studies, cosmology, and tests of general relativity are possible. For example, correlated systematic errors in the estimated luminosity distance of high-SNR GW events due to calibration systematic errors could bias estimates of the cosmological Hubble constant,  $H_0$ . Efforts to integrate the work presented in this paper into future GW event astrophysical parameter estimation are ongoing, including the use of the full, numerically evaluated, distribution of systematic error and uncertainty. These efforts will enable quantifying the impact of calibration systematics on individual GW events as well as studies that rely on a population of GW events. Additionally, these efforts guides research and development of new techniques currently underway to further reduce combined calibration systematic error and uncertainty below the 1% level, a key milestone towards minimizing impacts of calibration systematics on astrophysical and cosmological results.

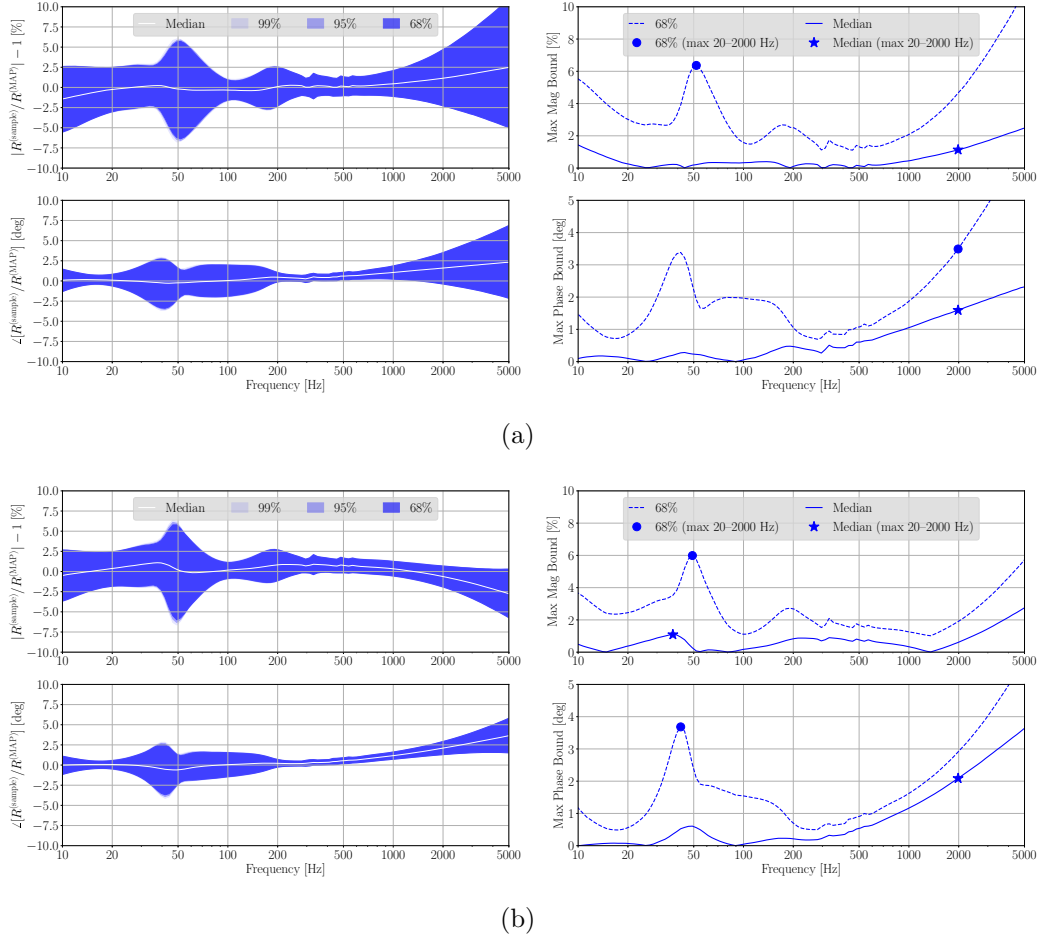


## 7. Acknowledgments

The authors gratefully acknowledge the operators, commissioners, and LSC fellows at Hanford and Livingston for their help in setting up the detector configurations and taking measurements needed for this work. The authors also gratefully acknowledge Jonathan Gair, Keita Kawabe and Loïc Rolland for the review and comments. LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the United States National Science Foundation (NSF), and operates under cooperative agreement PHY-1764464. Advanced LIGO was built under award PHY-0823459. The authors gratefully acknowledge the support of the United States NSF for the construction and operation of the LIGO Laboratory and Advanced LIGO as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. LS, DB, VB, PBC, LEHD, RG, TM, EP, AB and CC acknowledge the LSC Fellows program for supporting their research at LIGO sites. DB and SK are supported by NSF award PHY-1921006. AV is supported by NSF award PHY-1841480. MW is supported by NSF awards PHY-1607178 and PHY-1847350. PBC acknowledges the support of the Spanish Agencia Estatal de Investigación and Ministerio de Ciencia, Innovación y Universidades grants FPA2016-76821-P, the Vicepresidencia i Conselleria d'Innovació, Recerca i Turisme del Govern de les Illes Balears (grant FPI-CAIB FPI/2134/2018), the Fons Social Europeu 2014-2020 de les Illes Balears, the European Union FEDER funds, and the EU COST actions CA16104, CA16214, CA17137 and CA18108. EP is supported by ARC CE170100004. The authors would like to thank all of the essential workers who put their health at risk during the COVID-19 pandemic, without whom we would not have been able to complete this work. This paper carries LIGO Document Number LIGO-P1900245.



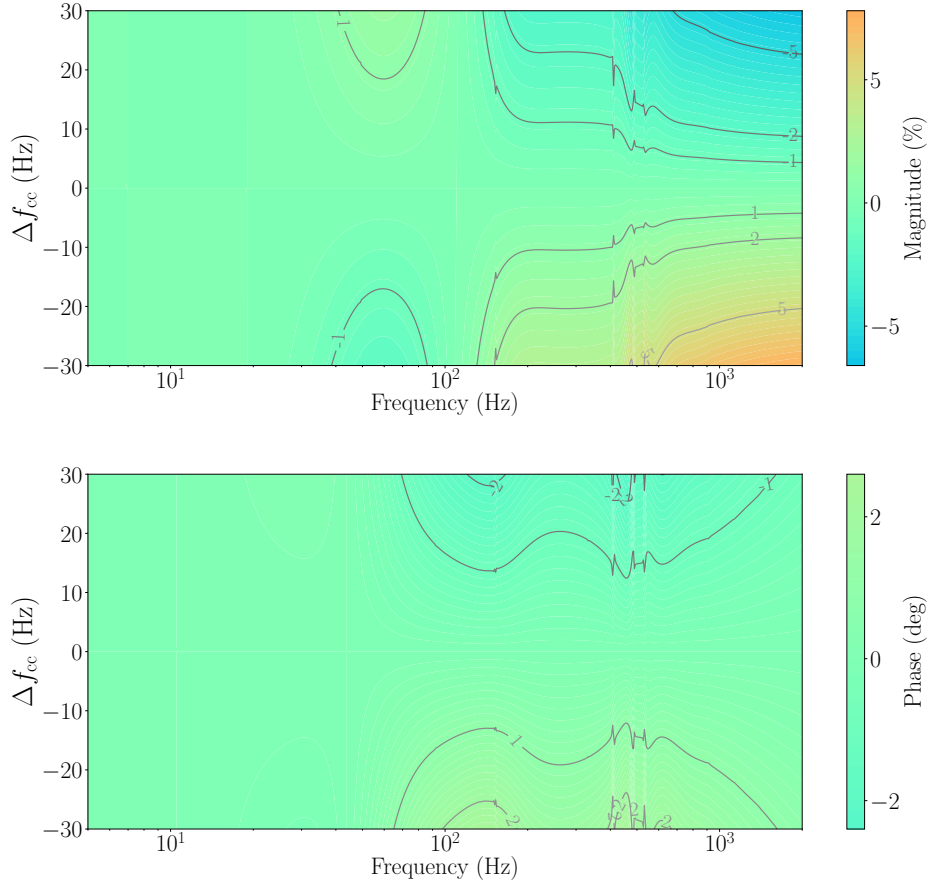
**Figure 16.** Uncertainty percentiles (left) and the maximum bounds (right) for Hanford. The three subfigures correspond to Hanford epoch (a)–(c) in table 3. The top and bottom panels of each subfigure show the frequency dependent excursions of response from unity magnitude and zero phase compared to  $\tilde{R}_{MAP}$ , respectively. The percentiles are obtained from all the hourly evaluated  $\tilde{\eta}_R(f; t)$  over each epoch. In the left panels, the colors represent  $1\sigma$  uncertainty for 68%, 95%, and 99% of the run time, as indicated in the legend. The white curve indicate the median excursion. The absolute values of the boundaries (median and 68%) in the left panels are plotted on the right. The star and dot markers indicate the median and  $1\sigma$  maximum excursions in the frequency band 20–2000 Hz, respectively.



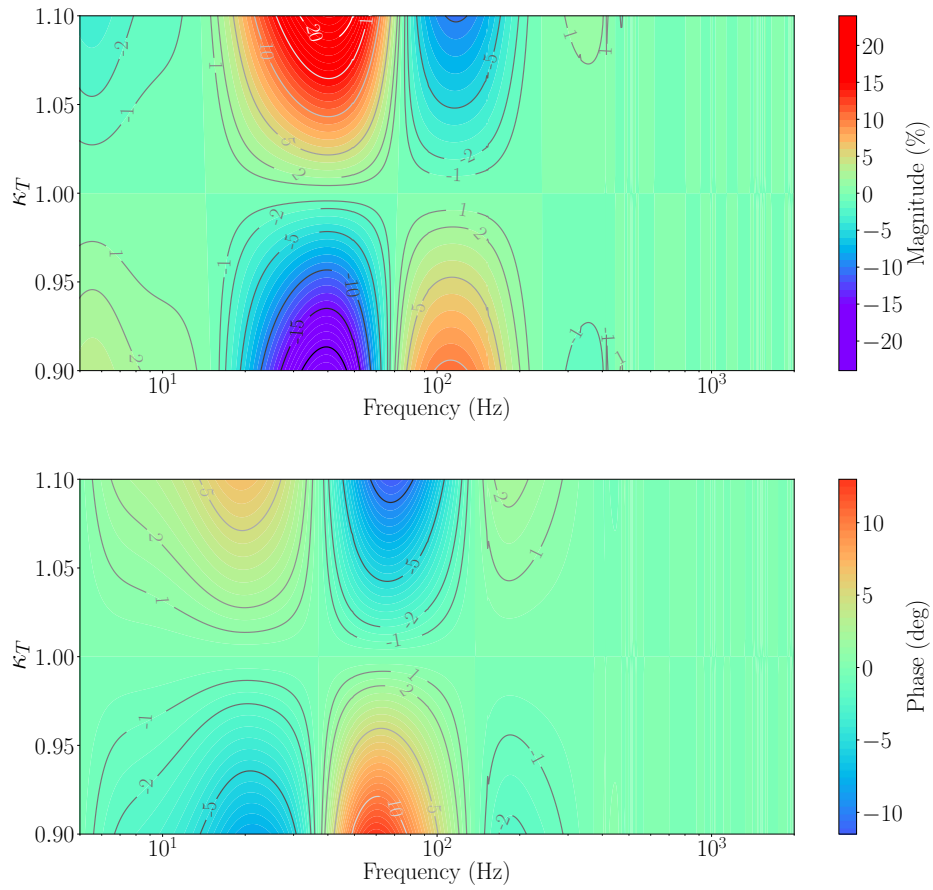
**Figure 17.** Uncertainty percentiles (left) and the maximum bounds (right) for Livingston. The two subfigures correspond to Livingston epoch (a)–(b) in table 3. The top and bottom panels of each subfigure show the frequency dependent excursions of response from unity magnitude and zero phase compared to  $\tilde{R}_{\text{MAP}}$ , respectively. The percentiles are obtained from all the hourly evaluated  $\tilde{\eta}_R(f; t)$  over each epoch. In the left panels, the colors represent  $1\sigma$  uncertainty for 68%, 95%, and 99% of the run time, as indicated in the legend. The white curve indicate the median excursion. The absolute values of the boundaries (median and 68%) in the left panels are plotted on the right. The star and dot markers indicate the median and  $1\sigma$  maximum excursions in the frequency band 20–2000 Hz, respectively.

## Appendix A. Impact of uncorrected TDCF's

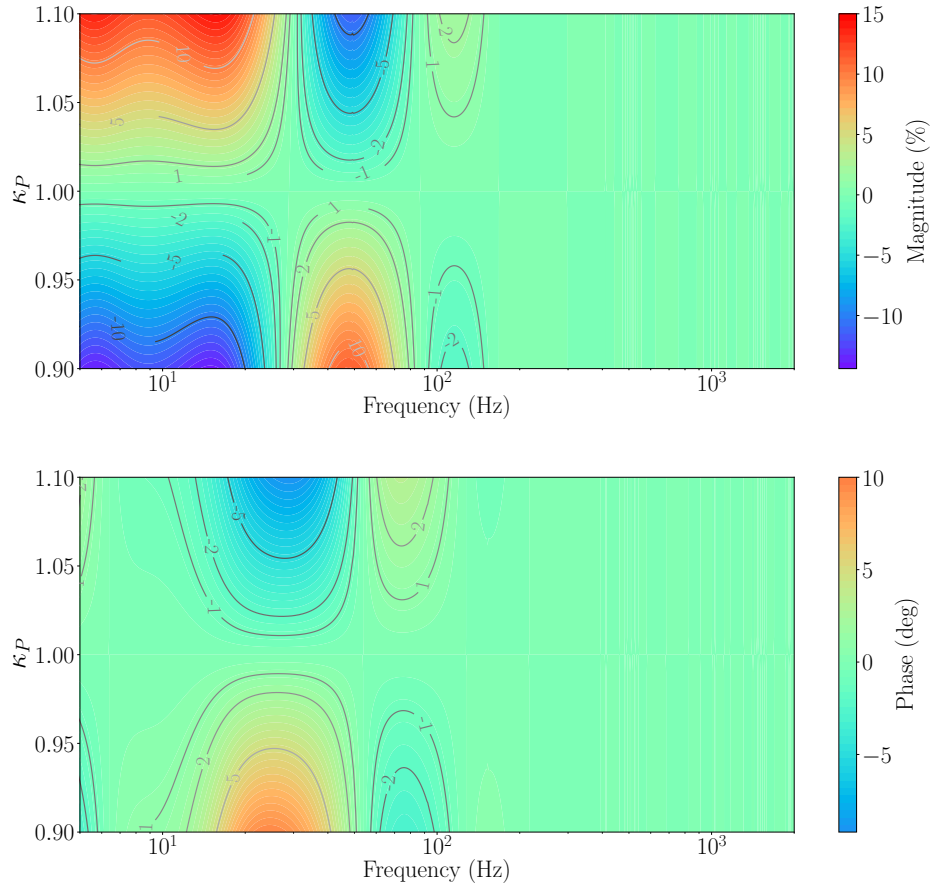
The figures in this appendix show the impact of uncorrected TDCF's,  $f_{cc}$ ,  $\kappa_T$ ,  $\kappa_P$ , and  $\kappa_U$ , on Hanford detector's response (see details in section 4.1).



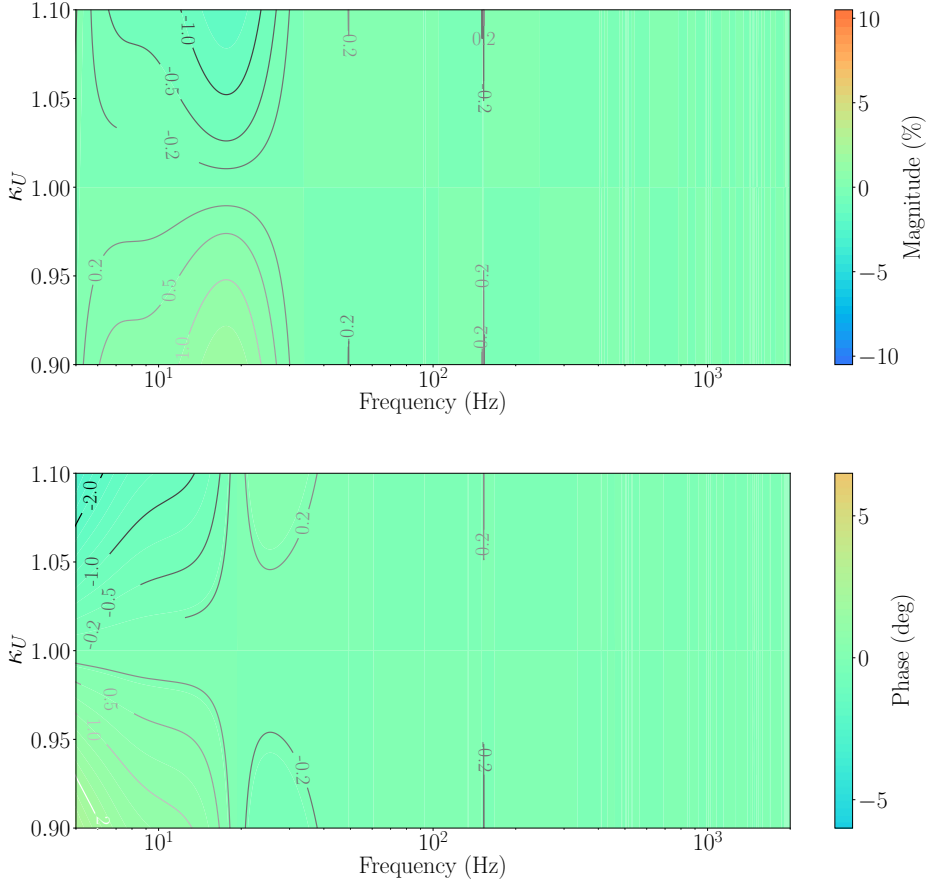
**Figure A1.** Magnitude (top) and phase (bottom) of the fractional error  $\tilde{\eta}_{R,C} - 1$  in the Hanford detector's response [O3A Epoch (c)] as a function of frequency due to uncorrected time-dependent coupled cavity pole frequency,  $f_{cc}$ . The reference value of  $f_{cc}$  is 410.6 Hz.



**Figure A2.** Magnitude (top) and phase (bottom) of the fractional error  $\tilde{\eta}_{R;A_T} - 1$  in the Hanford detector's response [O3A Epoch (c)] as a function of frequency due to uncorrected gain variations in the TST actuation stage, tracked by the scalar time-dependent factor,  $\kappa_T$ .



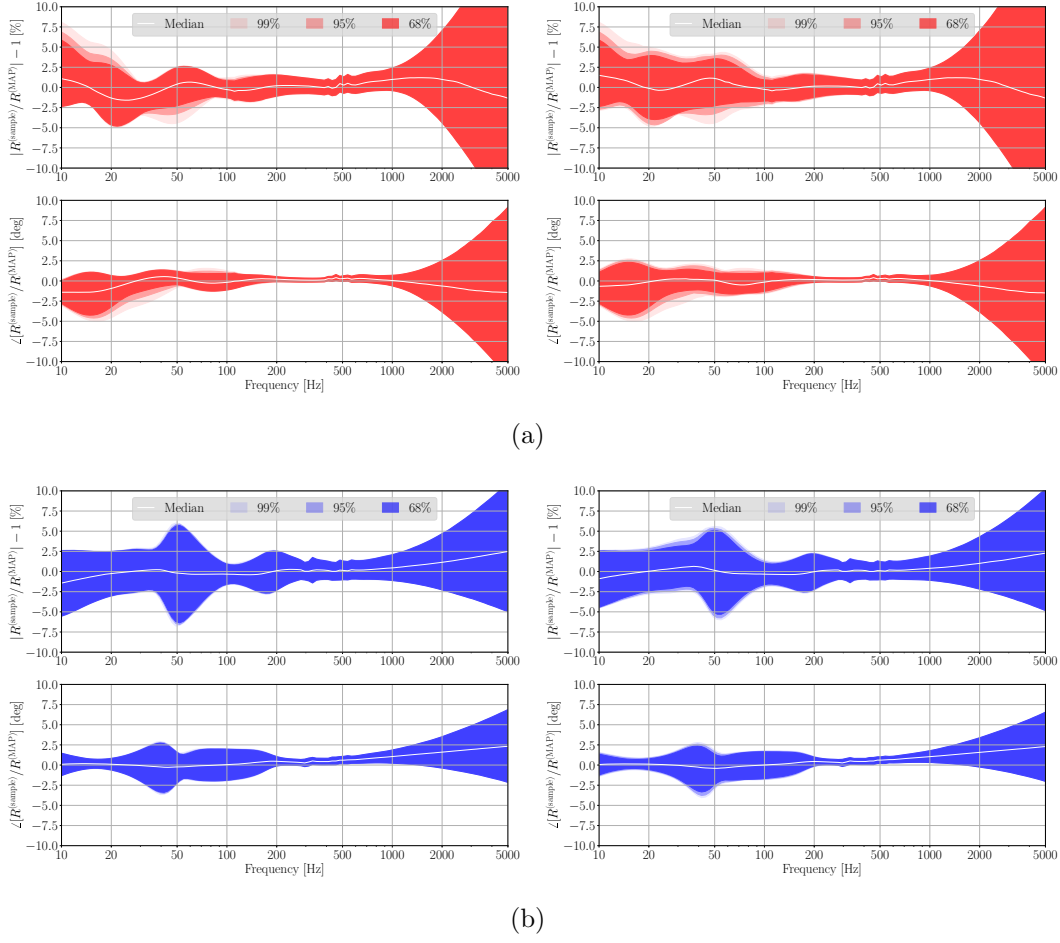
**Figure A3.** Magnitude (top) and phase (bottom) of the fractional error  $\tilde{\eta}_{R;A_P} - 1$  in the Hanford detector's response [O3A Epoch (c)] as a function of frequency due to uncorrected gain variations in the PUM actuation stage, tracked by the scalar time-dependent factor,  $\kappa_P$ .



**Figure A4.** Magnitude (top) and phase (bottom) of the fractional error  $\tilde{\eta}_{R_i;A_U} - 1$  in the Hanford detector's response [O3A Epoch (c)] as a function of frequency due to uncorrected gain variations in the UIM actuation stage, tracked by the scalar time-dependent factor,  $\kappa_U$ .

## Appendix B. Impact of complex-valued actuator TDCFs in early O3A

The figure in this appendix shows the estimates of systematic error and associated uncertainty using the collection of percentile curves of  $\tilde{\eta}_{R_i}(f; t)$  in the first epoch. The comparison of two cases are shown side-by-side: only the real-valued actuator TDCFs are applied (left) and the full complex-valued actuator TDCFs are applied from April 16 to June 11, 2019 at Hanford and from April 1 to June 11, 2019 at Livingston (right). See discussions in section 4.1.



**Figure B1.** Uncertainty percentiles for (a) Hanford and (b) Livingston in the first epoch in O3A. The top and bottom panels of each subfigure show the frequency dependent excursions of response from unity magnitude and zero phase compared to  $\tilde{R}_{\text{MAP}}$ , respectively. The percentiles are obtained from all the hourly evaluated  $\tilde{\eta}_R(f; t)$  over the first epoch. In the left panels, the results are obtained when only the real-valued actuator TDCFs are applied [equivalent to figures 16(a) and 17(a) in section 5.2]. In the right panels, the results are obtained when complex-valued actuator TDCFs are applied from April 16 to June 11, 2019 at Hanford and from April 1 to June 11, 2019 at Livingston. The colors represent  $1\sigma$  uncertainty for 68%, 95%, and 99% of the run time, as indicated in the legend. The white curve indicate the median excursion.



- [1] Aasi J *et al.* (LSC) 2015 *Classical and Quantum Gravity* **32** 074001
- [2] Acernese F *et al.* (Virgo) 2015 *Classical and Quantum Gravity* **32** 024001
- [3] Abbott B P *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) 2019 *Phys. Rev. X* **9**(3) 031040 URL <https://link.aps.org/doi/10.1103/PhysRevX.9.031040>
- [4] Buikema A, Cahillane C, Mansell G, Blair C *et al.* In preparation
- [5] Abbott B P *et al.* 2020 *The Astrophysical Journal* **892** L3 URL <https://doi.org/10.3847/2041-8213/ab75f5>
- [6] Abbott B P *et al.* 2020 (*Preprint* 2004.08342) URL <https://arxiv.org/abs/2004.08342>
- [7] Abbott B P *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) In preparation
- [8] GraceDB – Gravitational-Wave Candidate Event Database <https://gracedb.ligo.org/superevents/public/03/>
- [9] Lindblom L 2009 *Phys. Rev. D* **80**(4) 042005 URL <https://link.aps.org/doi/10.1103/PhysRevD.80.042005>
- [10] Abbott B P *et al.* (LIGO Scientific Collaboration) 2017 *Phys. Rev. D* **95**(6) 062003 URL <https://link.aps.org/doi/10.1103/PhysRevD.95.062003>
- [11] Viets A D, Wade M, Urban A L, Kandhasamy S, Betzwieser J, Brown D A, Burguet-Castell J, Cahillane C, Goetz E, Izumi K, Karki S, Kissel J S, Mendell G, Savage R L, Siemens X, Tuyenbayev D and Weinstein A J 2018 *Classical and Quantum Gravity* **35** 095015 URL <https://doi.org/10.1088/1361-6382/aab658>
- [12] Karki S, Tuyenbayev D, Kandhasamy S, Abbott B P, Abbott T D, Anders E H, Berliner J, Betzwieser J, Cahillane C, Canete L, Conley C, Daveloza H P, De Lillo N, Gleason J R, Goetz E, Izumi K, Kissel J S, Mendell G, Quetschke V, Rodruck M, Sachdev S, Sadecki T, Schwinberg P B, Sottile A, Wade M, Weinstein A J, West M and Savage R L 2016 *Review of Scientific Instruments* **87** 114503 (*Preprint* <https://doi.org/10.1063/1.4967303>) URL <https://doi.org/10.1063/1.4967303>
- [13] Bhattacharjee D, Savage R, Betzwieser J, Bossilkov V, Kandhasamy S, Karki S, Lecoecuche Y, Payne E *et al.* In preparation
- [14] Cahillane C, Betzwieser J, Brown D A, Goetz E, Hall E D, Izumi K, Kandhasamy S, Karki S, Kissel J S, Mendell G, Savage R L, Tuyenbayev D, Urban A, Viets A, Wade M and Weinstein A J 2017 *Phys. Rev. D* **96**(10) 102001 URL <https://link.aps.org/doi/10.1103/PhysRevD.96.102001>
- [15] Abbott B P *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) 2016 *Phys. Rev. Lett.* **116**(6) 061102 URL <https://link.aps.org/doi/10.1103/PhysRevLett.116.061102>
- [16] Abbott B P, Abbott R, Abbott T, Abernathy M, Acernese F, Ackley K, Adams C, Adams T, Addesso P, Adhikari R *et al.* 2016 *Physical review letters* **116** 131103
- [17] Tuyenbayev D, Karki S, Betzwieser J, Cahillane C, Goetz E, Izumi K, Kandhasamy S, Kissel J S, Mendell G, Wade M, Weinstein A J and Savage R L 2016 *Classical and Quantum Gravity* **34** 015002 URL <https://doi.org/10.1088/0264-9381/34/1/015002>
- [18] Bork R, Aronsson M, Barker D, Batch J, Heefner J, A Ivanov R M, Sandberg V and Thorne K 2011 New control and data acquisition system in the advanced ligo project Tech. Rep. ligo-p1100052 LIGO Laboratory URL <https://dcc.ligo.org/LIGO-P1100052/public>
- [19] Bork R, Hanks J, Barker D, Betzwieser J, Rollins J, Thorne K and von Reisb E 2020 (*Preprint* 2005.00219) URL <https://arxiv.org/abs/2005.00219>
- [20] Robertson N, Cagnoli G, Crooks D, Elliffe E, Faller J, Fritschel P, Goßler S, Grant A, Heptonstall A, Hough J *et al.* 2002 *Classical and Quantum Gravity* **19** 4043
- [21] Aston S, Barton M, Bell A, Beveridge N, Bland B, Brummitt A, Cagnoli G, Cantley C, Carbone L, Cumming A *et al.* 2012 *Classical and Quantum Gravity* **29** 235004
- [22] Hall E D 2017 *Long-baseline laser interferometry for the detection of binary black-hole mergers* Ph.D. thesis California Institute of Technology

- [23] Goetz E, Kandhasamy S, Kissel J, Viets A and Anand S 2019 Update to tracking temporal variations in darm loop model parameters: Individual actuation stage tracking, cancelled lines, and src detuning Tech. Rep. ligo-t1700106 LIGO Laboratory URL <https://dcc.ligo.org/LIGO-T1700106>
- [24] Karki S 2019 *Accurate and Precise Calibration of Advanced LIGO Detectors in the Era of Gravitational Wave Astronomy* Ph.D. thesis University of Oregon
- [25] Bhattacharjee D, Savage R L, Betzwieser J, Bossilkov V, Kandhasamy S, Karki S, Lecoecuche Y and Payne E 2019 Photon calibrator actuation coefficients and uncertainties update at the start of o3b Tech. Rep. ligo-g1902259 LIGO Laboratory URL <https://dcc.ligo.org/LIGO-G1902259>
- [26] Barton M 2010 Models of the advanced ligo suspensions in mathematica Tech. Rep. LIGO-T020205 LIGO Laboratory URL <https://dcc.ligo.org/LIGO-T020205/public>
- [27] Barton M 2011 Models of the advanced ligo suspensions in matlab Tech. Rep. LIGO-T080188 LIGO Laboratory URL <https://dcc.ligo.org/LIGO-T080188>
- [28] Shapiro B 2010 Fitting the quad noise prototype model to measured data Tech. Rep. ligo-t1000458 LIGO Laboratory URL <https://dcc.ligo.org/LIGO-T1000458/public>
- [29] Shapiro B, Barton M, Mavalvala N, Mittleman R and Youcef-Toumi K 2014 *IEEE/ASME Transactions on Mechatronics* **20** 13–23
- [30] Biscans S, Gras S, Blair C, Driggers J, Evans M, Fritschel P, Hardwick T and Mansell G 2019 *Physical Review D* **100** 122003
- [31] IIRrational: Rational Function fitting for System ID <https://iirrational.readthedocs.io/en/latest/>
- [32] Bendat J and Piersol A 2011 *Random Data: Analysis and Measurement Procedures* Wiley Series in Probability and Statistics (Wiley) ISBN 9781118210826 URL <https://books.google.com.au/books?id=qYSViFRNMLwC>
- [33] Foreman-Mackey D, Hogg D W, Lang D and Goodman J 2013 *Publications of the Astronomical Society of the Pacific* **125** 306–312
- [34] Viets A D 2019 *Theses and Dissertations* **2135** URL <https://dc.uwm.edu/etd/2135>
- [35] Miyakawa O, Ward R, Adhikari R, Evans M, Abbott B, Bork R, Busby D, Heefner J, Ivanov A, Smith M *et al.* 2006 *Physical Review D* **74** 022001
- [36] Brooks A *et al.* 2020 URL <https://dcc.ligo.org/LIGO-P1900287>
- [37] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M and Duchesnay E 2011 *Journal of Machine Learning Research* **12** 2825–2830 URL <https://scikit-learn.org>
- [38] Torrie C and Robertson N 2016 Recent results and conclusions from tests of the uim blade non-magnetic damper Tech. Rep. ligo-t1600046 LIGO Laboratory URL <https://dcc.ligo.org/LIGO-T1600046/public>
- [39] Hall E D, Cahillane C, Izumi K, Smith R J E and Adhikari R X 2019 *Classical and Quantum Gravity* **36** 205006 URL <https://doi.org/10.1088/1361-6382/ab368c>
- [40] Rakhmanov M, Savage R, Reitze D and Tanner D 2002 *Physics Letters A* **305** 239–244 ISSN 0375-9601 URL <http://www.sciencedirect.com/science/article/pii/S037596010201469X>
- [41] Izumi K 2015 Toward accurate darm length response modeling in aligo calibration model Tech. Rep. ligo-g1501316 LIGO Laboratory URL <https://dcc-llo.ligo.org/LIGO-G1501316>
- [42] Bartos I, Bork R, Factourovich M, Heefner J, Márka S, Márka Z, Raics Z, Schwinberg P and Sigg D 2010 *Classical and Quantum Gravity* **27** 084025
- [43] Asali Y, Countryman S and Marka Z 2019 aligo timing system checks during o3a Tech. Rep. ligo-t1900799 LIGO Laboratory URL <https://dcc.ligo.org/LIGO-T1900799>
- [44] U.S. Department of Defense, SPS and GPS 2008 *Global positioning system standard positioning service performance standard.* URL <http://www.gps.gov/technical/ps/>

2008-SPS-performance-standard.pdf

- [45] Farr W, Farr B and Littenberg T 2016 Modelling calibration errors in cbc waveforms Tech. Rep. ligo-t1400682 LIGO Laboratory URL <https://dcc.ligo.org/LIGO-T1400682>
- [46] Vitale S, Del Pozzo W, Li T G F, Van Den Broeck C, Mandel I, Aylott B and Veitch J 2012 *Phys. Rev. D* **85**(6) 064034 URL <https://link.aps.org/doi/10.1103/PhysRevD.85.064034>
- [47] Abbott B P *et al.* 2019 *The Astrophysical Journal* **875** 161 URL <https://doi.org/10.3847/1538-4357/ab0e8f>
- [48] Abbott B P, Abbott R, Abbott T, Abernathy M, Acernese F, Ackley K, Adams C, Adams T, Addesso P, Adhikari R *et al.* 2016 *Physical Review D* **93** 122003