

Calibration and Uncertainty Quantification of Gravity Wave Parameterization in an Intermediate Complexity Climate Model

Laura A Mansfield¹ and Aditi Sheshadri¹

¹Stanford University

November 26, 2022

Abstract

The drag due to breaking atmospheric gravity waves plays a leading order role in driving the middle atmosphere circulation, but as their horizontal wavelength ranges from tens to thousands of kilometers, part of their spectrum must be parameterized in climate models. Gravity wave parameterizations prescribe a source spectrum of waves in the lower atmosphere and allow these to propagate upwards until they either dissipate or break, where they deposit drag on the large-scale flow. These parameterizations are a source of uncertainty in climate modeling which is generally not quantified. Here, we explore the uncertainty associated with a non-orographic gravity wave parameterization in a global climate model of intermediate complexity, using the Calibrate, Emulate and Sample (CES) method. We first calibrate the uncertain parameters that define the gravity wave source spectrum in the tropics, to obtain climate model settings that are consistent with properties of the primary mode of tropical stratospheric variability, the Quasi-Biennial Oscillation (QBO). Then we use a Gaussian process emulator to sample the calibrated distribution of parameters and quantify the uncertainty of these parameter choices. We find that the resulting parametric uncertainties on the QBO period and amplitude are of a similar magnitude to the internal variability under a 2xCO₂ forcing.

Hosted file

mov01.gif available at <https://authorea.com/users/556165/articles/606347-calibration-and-uncertainty-quantification-of-gravity-wave-parameterization-in-an-intermediate-complexity-climate-model>

1 **Calibration and Uncertainty Quantification of Gravity Wave Parameterization in an**
2 **Intermediate Complexity Climate Model**

3 **L. A. Mansfield¹ and A. Sheshadri¹**

4 **¹Department of Earth System Science, Stanford University, Stanford, CA, USA**

5 Corresponding author: Laura A. Mansfield (lauraman@stanford.edu)

6 **Key Points:**

- 7 • We calibrate tropical parameters in a gravity wave parameterization to obtain selected
8 properties of the Quasi-Biennial Oscillation.
- 9 • We use a Gaussian process to emulate an intermediate complexity climate model and
10 then learn a distribution of gravity wave parameters.
- 11 • We explore the gravity wave parametric uncertainty of the Quasi-Biennial Oscillation
12 period and amplitude in a double CO₂ scenario.

13

14

15

16

17 Abstract

18 The drag due to breaking atmospheric gravity waves plays a leading order role in driving the
19 middle atmosphere circulation, but as their horizontal wavelength ranges from tens to thousands
20 of kilometers, part of their spectrum must be parameterized in climate models. Gravity wave
21 parameterizations prescribe a source spectrum of waves in the lower atmosphere and allow these
22 to propagate upwards until they either dissipate or break, where they deposit drag on the large-
23 scale flow. These parameterizations are a source of uncertainty in climate modeling which is
24 generally not quantified. Here, we explore the uncertainty associated with a non-orographic
25 gravity wave parameterization in a global climate model of intermediate complexity, using the
26 Calibrate, Emulate and Sample (CES) method. We first calibrate the uncertain parameters that
27 define the gravity wave source spectrum in the tropics, to obtain climate model settings that are
28 consistent with properties of the primary mode of tropical stratospheric variability, the Quasi-
29 Biennial Oscillation (QBO). Then we use a Gaussian process emulator to sample the calibrated
30 distribution of parameters and quantify the uncertainty of these parameter choices. We find that
31 the resulting parametric uncertainties on the QBO period and amplitude are of a similar
32 magnitude to the internal variability under a $2xCO_2$ forcing.

33

34 Plain Language Summary

35 Atmospheric gravity waves are created in the lower atmosphere by disturbances such as
36 mountains, convection and fronts. They travel upwards and break in the upper atmosphere,
37 which slows down the flow and has large effects on the circulation, including driving a tropical
38 oscillation. Gravity waves have a wide range of spatial scales and a large portion of these are
39 smaller than the grid size of a climate model. This means they cannot be resolved exactly and
40 instead, they are represented through approximations called “parameterizations”, which
41 introduce a source of uncertainty in climate model output. In this study, we tune a
42 parameterization so that the model produces a oscillation in the tropical middle atmosphere, with
43 a defined period and amplitude, which is one of the main features of the climate driven primarily
44 by gravity waves. We also explore uncertainties associated with the parameterization.

45

46

47

48 **1 Introduction**

49

50

50 **1.1 Atmospheric gravity waves**

51

52

53

54

55

56

57

58

59

60

Atmospheric gravity waves or buoyancy waves, which owe their existence to the restoring force of gravity in a stratified flow, play a substantial role in the exchange of momentum between the Earth's surface and the free atmosphere. They are forced by a range of processes including orography, convection and frontogenesis in the lower atmosphere. While propagating upwards through decreasing density, gravity waves gain amplitude and eventually break, depositing momentum. This influences the large-scale flow, and affects the circulation, temperature, structure, chemistry and composition of the middle and upper atmosphere (Alexander & Dunkerton, 1999).

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

The horizontal length scale of gravity waves ranges from tens to thousands of kilometers. While the larger scale gravity waves are resolved by the dynamics in climate models, waves smaller than $2\times$ the horizontal resolution cannot be resolved, leading to an underestimate of gravity wave drag from the dynamical core. At this time, current climate models designed for CMIP6 have resolutions of 1° - 2.8° , equivalent to ~ 100 - 250 km spacing at the equator (Priestley et al., 2020; Richter & Tokinaga, 2020). At these resolutions, the majority of gravity wave drag is not resolved and is instead represented through both orographic and non-orographic gravity wave parameterizations (e.g. Alexander & Dunkerton, 1999; Scinocca, 2003; Warner & McIntyre, 1999). These aim to describe the large-scale effect that sub-grid scale gravity waves have on the flow and are often necessary to obtain realistic circulation patterns, for example, to reduce model biases (e.g., Palmer et al., 1986) and to induce a spontaneous Quasi-Biennial Oscillation (QBO) (Bushell et al., 2020). Parameterized gravity waves are required even at the higher resolution end of the spectrum of models, for instance, HighResMIP, which have resolutions higher than 50 km but typically still include some parameterized sub-grid scale gravity waves (e.g. Kodama et al., 2021). Sub-grid scale parameterizations make several assumptions about the nature of gravity waves which becomes a source of uncertainty in climate models. Several recent studies harness machine learning methods to learn data-driven gravity wave parameterizations, which may be faster and/or more accurate (e.g. Chantry et al., 2021; Espinosa et al., 2022; Matsuoka et al., 2020). This study makes use of machine learning methods, but rather than replacing traditional parameterizations, we instead calibrate an existing gravity wave parameterization and quantify uncertainties associated with it.

83

84

83 **1.2 Gravity wave parameterizations and associated uncertainties**

84

85

86

87

88

89

90

91

92

A common type of parameterization is the Lindzen-type parameterization, based on Lindzen (1981), which assumes gravity waves are launched at a fixed source level in the troposphere and propagate in the vertical column until they reach saturation. This is called the critical level, at which it is assumed that breaking occurs, depositing gravity wave drag. These have been further developed into spectral parameterizations, in which a complete spectrum of waves is launched, leading to a spectrum of critical levels rather than a single level (Alexander & Dunkerton, 1999). In this type of parameterization, there are several parameter choices to be made, for instance, the phase speeds, amplitudes and location of launched gravity waves. These

93 all influence the magnitude and spatial structure of gravity wave drag deposited by the
94 parameterization.

95
96 The parameters should ideally be chosen so that the parameterization output (here the
97 unresolved gravity wave drag) is consistent with observations. However, obtaining observations
98 of gravity wave drag caused by unresolved gravity wave breaking is not trivial. Observations of
99 total gravity wave momentum flux are available, but it is not clear how to obtain the momentum
100 flux attributed to the subgrid-scale gravity waves. Importantly, the main goal of
101 parameterizations is to obtain climate model output consistent with the macrophysical climate
102 state (i.e., large-scale flow and circulation), rather than the microphysical (i.e., gravity wave
103 drag). Therefore, the typical approach is to tune the parameterization to obtain a consistent
104 climate state (e.g. Barton et al., 2019; Couvreur et al., 2021; Donner et al., 2011; Dunbar et al.,
105 2021; Scaife et al., 2002).

106
107 Calibration of parameters traditionally involves manual tuning of parameter values until a
108 reasonable output is obtained (e.g. Donner et al., 2011; Kodama et al., 2021), but in recent years
109 has been automated with statistical methods such as Bayesian optimization (Kennedy &
110 O’Hagan, 2001), iterative refocusing/history matching (Williamson et al., 2013) and ensemble
111 Kalman methods (Cleary et al., 2021). These methods typically calibrate the parameters by
112 minimizing a loss function that describes the difference between the climate model output and
113 the observations.

114
115 Even after calibration, sub-grid scale parameterizations are a substantial source of
116 uncertainty in climate model output that is generally not considered in model analysis.
117 Uncertainty quantification is a growing field for parameterizations including clouds (Pathak et
118 al., 2021), convection (Dunbar et al., 2021), aerosol microphysics (Lee et al., 2012) and ocean
119 processes (Souza et al., 2020), but has not yet been applied for gravity wave parameterizations.
120 In this paper, we combine calibration and uncertainty quantification methods to explore the
121 importance of parameter choices in a non-orographic gravity wave parameterization within an
122 idealized moist atmospheric model. Specifically, we use the Calibrate-Emulate-Sample
123 framework developed in Cleary et al. (2021) to first estimate the optimal parameters that give
124 model output consistent with observed properties of stratospheric phenomena and to further
125 assess the uncertainty of the output associated with the derived distribution of gravity wave
126 parameters.

127
128 In the remainder of this section, we describe the QBO, a large-scale oscillation in the
129 tropical stratosphere, realistic simulation of which has depended critically on the choices made in
130 gravity wave parameterization. Section 2 describes the model and gravity wave parameterization
131 used and Section 3 outlines the CES framework. The results of this are discussed in Section 4,
132 where we explore CES under the perfect model setting, assuming the “truth” to be a long
133 integration of our model. In Section 4.2, we explore the sensitivity of the QBO to gravity wave
134 parameters and in Section 4.3, we quantify uncertainties of the QBO due to the parameter
135 choices for a control climate and 2xCO₂ scenario. Section 5 contains a summary and discussion
136 of the work.

137
138

1.2 Quasi-Biennial Oscillation

The Quasi-Biennial oscillation (QBO) is the dominant mode of variability in the equatorial stratosphere, occurring in the vertical range of 5-100 hPa (Gray, 2010). The QBO consists of alternating westerly and easterly winds with a period of ~ 28 months, descending at ~ 1 km/month, as shown in Figure 1a, which shows a cross-section of the zonal mean zonal winds at the equator ($5^{\circ}\text{S} - 5^{\circ}\text{N}$) from global radiosonde observations (Freie Universität Berlin, 2007).

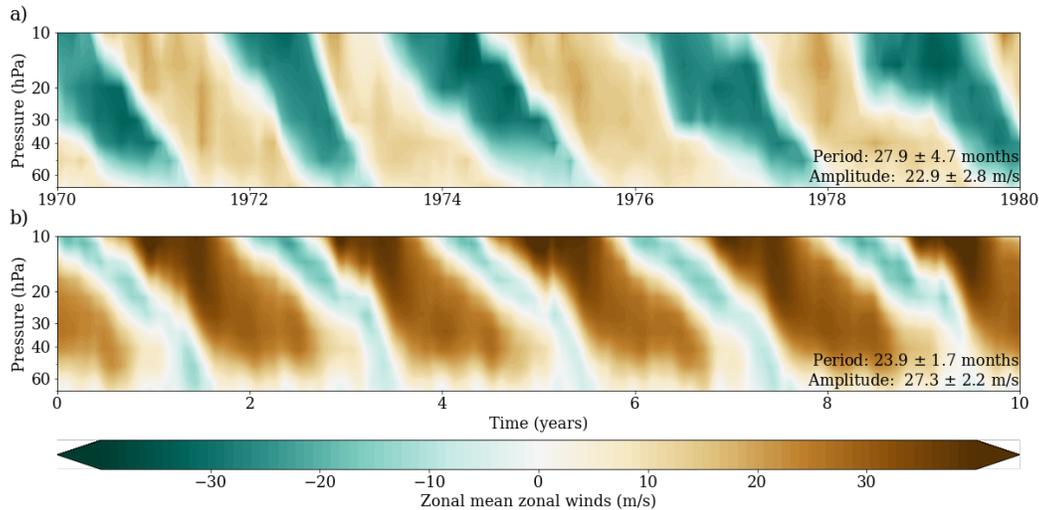


Figure 1. QBO zonal mean zonal winds at the equator ($5^{\circ}\text{S} - 5^{\circ}\text{N}$) over a 10-year segment from a) global radiosonde observations (Freie Universität Berlin, 2007) and b) the model used in this study (MiMAv2.0, (Garfinkel et al., 2020)). In the bottom right corner are the period and amplitudes, shown as the means and 1 standard deviation estimated from a) the 68 year period of observations and b) a 50- year control simulation of MiMA.

The QBO is driven by a broad spectrum of waves, including large-scale Kelvin and Rossby-gravity waves, mesoscale inertia-gravity and high frequency small-scale gravity waves (Baldwin et al., 2001; Lindzen & Holton, 1968). The latter are the gravity waves with zonal wavenumber > 40 , corresponding to zonal wavelengths between 10 and 1000 km, i.e., mostly sub-grid scale in climate models. These contribute significant forcing to the QBO, without which climate models cannot produce a spontaneous QBO. Specifically, only 10 out of 47 CMIP5 models included a non-orographic gravity wave parameterization and of these, only 5 displayed a QBO-like signal (Schenzinger et al., 2017). Based on more recent models that obtain a spontaneous QBO, at least half of the forcing required is contributed from non-orographic gravity wave parameterizations (Holt et al., 2020). This makes the QBO a sensible phenomenon to consider when calibrating the gravity wave parameterization (Anstey et al., 2016; Barton et al., 2019; Scaife et al., 2002).

Simulating a realistic QBO in climate models is important for not just accurately reproducing the tropical stratosphere, but also for tropical convection (Rao et al., 2020), the

169 subtropical jet (Garfinkel & Hartmann, 2011) and the stratospheric polar vortices. The QBO is
 170 known to strengthen the polar vortex during the westerly QBO phase and weaken it during the
 171 easterly QBO phase, leading to more sudden stratospheric warmings (SSWs) and hence colder
 172 surface temperatures in winter (the Holton-Tan relationship, Holton & Tan, 1980). Studies also
 173 indicate the QBO influences the transport of aerosols and other atmospheric constituents into and
 174 out of the polar vortex (Strahan et al., 2015).

175

176 The QBO is defined by a variety of metrics. The first order properties to consider are the
 177 period and amplitude of the QBO, which we consider at the reference level 10 hPa, where the
 178 QBO amplitude is generally a maximum (Bushell et al., 2020; Richter et al., 2020). The zonal
 179 mean zonal winds between $5^{\circ}S$ and $5^{\circ}N$ at 10 hPa, \bar{u}_{eq} , are first smoothed using a 5-month
 180 binomial filter to remove fast fluctuations. Following Schenzinger et al. (2017), a single QBO
 181 cycle is determined based on the times at which \bar{u}_{eq} transitions from westward to eastward. The
 182 period is defined as the time between subsequent transitions and the amplitude is defined as the
 183 maximum amplitude of the zonal mean zonal winds, i.e. $\max |\bar{u}_{eq}|$. This gives a period and
 184 amplitude for each cycle of the QBO, from which the mean and standard deviation can be
 185 estimated.

186

187

188 **2 Model Setup**

189 **2.1 Model**

190

191 In this study, we explore the uncertainty of a climate model with respect to the Lindzen-
 192 type spectral parameterization introduced in Alexander & Dunkerton (1999), hereafter AD99.
 193 We explore uncertainties related to AD99 parameters that describe the spectrum of launched
 194 gravity waves at the source level. For the climate model, we use the Model of an idealized Moist
 195 Atmosphere version 2.0 (MiMAv2.0) (Garfinkel et al., 2020; Jucker & Gerber, 2017). This is
 196 chosen because it is of intermediate complexity and results in reasonable atmospheric variability,
 197 including obtaining a realistic QBO and stratospheric polar vortex but at a lower computational
 198 cost than more complex coupled GCMs. We run MiMA at 2.8° resolution (or ~ 300 km at
 199 equator), which corresponds to T42 spectral resolution, i.e., resolving waves only with
 200 wavenumber smaller than 42. This leaves the small-scale gravity waves noted as influential for
 201 the formation of the QBO (wavenumber >40 (Baldwin et al., 2001)) to be parameterized. These
 202 gravity waves are instead captured by the AD99 parameterization, described below.

203

204 **2.2 Gravity wave parameterization**

205

206

207 AD99 is a gravity wave parameterization that does not separate the source of gravity
 208 waves and treats both orographic and non-orographic gravity waves in the same way. Instead, it
 209 launches gravity waves with a fixed phase speed for orographic waves and a spectrum of gravity
 210 waves for non-orographic gravity waves. We focus on the non-orographic gravity waves for this
 211 study.

212

213 **Gravity wave source**

214
215
216
217
218
219
220

The non-orographic component of AD99 launches a spectrum of gravity waves with discretized phase speeds centered at $c_0 = 0$ m/s from the source level. The width of this spectrum is defined by the half-width, c_w , which is chosen to be 35 m/s in the default setting, but is not easily constrained by observations. The spectrum of wave momentum flux at phase speed c is given by

$$B_0(c) = \frac{F_{P0}(c)}{\bar{\rho}_0} = \text{sign}(c - \bar{u}_0) B_m \exp\left[-\left(\frac{c - c_0}{c_w}\right)^2 \ln 2\right] \quad (1)$$

221
222

223
224
225
226
227
228
229

where $F_{P0}(c)$ is the momentum flux carried by a wave with phase speed c and $\bar{\rho}_0$ is the mean flow density at the source level. B_m is the amplitude of waves with zero phase speed and can be constrained by observed $\overline{u'w'}$ and $\overline{v'w'}$ local wave events. $B_0(c)$ is the momentum flux amplitude in active times and determines when the wave will break, along with the mean flow profile.

230
231
232
233

The total momentum flux depends not just on $B_0(c)$, but also on the intermittency of the gravity waves. With time, the intermittency reduces the total momentum flux compared to $B_0(c)$ (the momentum flux in active times) and is modeled in AD99 with an intermittency scaling factor,

$$\varepsilon = \frac{F_{S0} \Delta c}{\bar{\rho}_0 \sum_c |B_0(c)| \Delta c} \quad (2)$$

234
235

236
237
238
239
240

where F_{S0} is the total gravity wave stress at the source level, Δc is phase speed resolution of the spectrum and $\bar{\rho}_0$ is the mean density at the source level. This equation describes the ratio between the total time-averaged momentum flux to the total momentum flux averaged over all phase speeds of the spectrum.

241
242
243
244
245
246
247
248

Although long-term averages of observed $\overline{u'w'}$ and $\overline{v'w'}$, e.g., from superpressure balloons can be used to estimate the observed total momentum flux (Geller et al., 2013; Jewtoukoff et al., 2015), it is not necessarily optimal to constrain F_{S0} in this way. Climate models typically require the total momentum flux to be smaller than observed values by a factor of 3-5 in order to obtain realistic large-scale flow (Plougonven et al., 2020). This means F_{S0} is not easily constrained by observations and must instead be calibrated to obtain a realistic macrophysical climate state. This gives two uncertain parameters to be calibrated in this study: c_w and F_{S0} (highlighted in red in Equations (1) and (2) respectively).

249
250
251

Gravity wave breaking

252
253
254
255

Given these properties of gravity waves at the source level, the parameterization allows gravity waves to propagate upwards. At each level the parameterization checks if the intrinsic frequency magnitude is less than the reflection frequency, and if so, the waves undergo total internal reflection and are eliminated. A stability criterion is also checked at each level, for all

256 phase speeds. The portion of the wave spectrum with phase speeds that do not satisfy the
 257 stability criteria undergo breaking and are removed from the spectrum. On breaking, the mean-
 258 flow forcing and eddy diffusion coefficients are estimated and fed back into the large-scale flow.
 259 For waves that break, indexed by j , between level z_{n-1} and z_n , the forcing on the mean flow is:

$$260 \quad X(z_{n-1/2}) = \frac{\epsilon}{\bar{\rho}(z_{n-1/2})\Delta z} \sum_j F_{P0}(c_j)$$

261 and the eddy diffusion coefficient is:

$$262 \quad D(z_{n-1/2}) = \frac{\epsilon}{\bar{\rho}(z_{n-1/2})\Delta z} \frac{1}{N^2(z_{n-1/2})} \sum_j \left(c_j - \bar{u}(z_{n-1/2}) \right) F_{P0}(c_j)$$

263 where N is the Brunt-Väisälä frequency and $F_{P0}(c_j)$ is the discretized momentum flux carried
 264 by waves with phase speed c_j at the source level. Note this relates to F_{S0} , the total momentum
 265 flux at the source level, as $F_{S0} = \sum_{i=1}^{N_c} F_{P0}(c_i)$. The parameters that define the source spectrum
 266 affect the forcing and eddy diffusion coefficient through the intermittency scaling factor
 267 (Equation (2)) and any uncertainty in parameters such as c_w and F_{S0} propagate through to affect
 268 the mean flow.

269

270

271 **Latitude dependence of source terms**

272

273 Alexander & Dunkerton (1999) introduce this parameterization for a single vertical
 274 column with the intention that it could be applied to global climate models with one-dimensional
 275 calculations based on the wind and stability profiles at each geographic point in the model, i.e.,
 276 for each longitude and latitude. Alexander & Rosenlof (2003) find that gravity wave sources in
 277 the tropics can differ significantly from those in the extratropics in observations. This can be
 278 included in the parameterization by providing latitude-dependent source parameters for c_w and
 279 F_{S0} .

280

281 The AD99 implementation in MiMA allows c_w to be defined in the tropics ($10^\circ S$ to $10^\circ N$)
 282 independently of its value outside this region. This means tropical values of c_w can be varied,
 283 e.g., to explore its effects on the QBO (Garfinkel et al., 2022), while keeping the extratropical
 284 value of c_w fixed in order to maintain the stratospheric polar vortices. In this study, we only
 285 consider c_w in the tropics, with c_w in the extratropics kept fixed at 35 m/s.

286

287 F_{S0} is also latitude dependent. It is typical for GCMs to prescribe a peak in F_{S0} in the
 288 tropics due to tropical precipitation (e.g., the Canadian Middle Atmosphere Model (CMAM,
 289 Anstey et al. (2016) and MERRA reanalysis/Fortuna version of the Goddard Earth Observing
 290 System Mode (GEOS-5) (Molod et al., 2012))) and/or additional stress in extratropical storm
 291 track regions, in some cases with a larger value of F_{S0} in the northern hemisphere compared to
 292 the southern hemisphere to improve the simulation of the stratospheric polar vortices (e.g.,
 293 AM3/4, the atmospheric components of the global model from Geophysical Fluid Dynamics
 294 Laboratory (GFDL) (Donner et al., 2011; Zhao et al., 2018)). We include the latter, by setting a
 295 base of 0.0043 Pa in the extratropics, with an additional 0.0035 Pa in the northern hemisphere
 296 that appears to provide roughly the correct number of sudden stratospheric warmings (Equation
 297 A3 of Garfinkel et al., 2022). In the tropics ($10^\circ S$ to $10^\circ N$), we define $F_{S0} = B t_{eq}$ as the
 298 parameter of interest, responsible for modulating properties QBO. Table 1 shows the two

299 parameters calibrated and assessed in this study and their values chosen for the control run
 300 setting.

301

302 *Table 1. Description of the two parameters calibrated in this study*

Parameter	Description	Control value
c_w	Half-width of phase speed in tropics ($10^\circ S$ to $10^\circ N$)	35 m/s
Bt_e	Total gravity wave stress in tropics ($10^\circ S$ to $10^\circ N$)	0.0043 Pa

303

304 Garfinkel et al. (2022) assessed the sensitivity of the QBO in MiMA to c_w and Bt_{eq} .
 305 They found that the QBO amplitude is significantly more sensitive than the period. Increasing
 306 Bt_{eq} leads to a faster and stronger QBO. While increasing c_w also leads to a faster and stronger
 307 QBO, the period is not affected significantly when c_w is increased beyond 25m/s.
 308

309 **3 Calibrate, Emulate and Sample Method**

310

311 The goal of uncertainty quantification is to obtain a distribution of model outputs, given a
 312 distribution of model parameters. To do this, we need samples from the optimal distribution of
 313 model parameters that produce model outputs in agreement with an observed dataset. We employ
 314 the Calibrate, Emulate and Sample (CES) method (Cleary et al., 2021; Dunbar et al., 2021;
 315 Howland et al., 2022). This involves (a) calibration of model parameters so that the model output
 316 agrees with the observed dataset, (b) emulation of the expensive model given model parameters
 317 to allow for quick evaluations and (c) sampling from the calibrated distribution of model
 318 parameters with the emulator.

319

320

321 **3.1 Calibration**

322

323 The first step of CES is the calibration, for which we use Ensemble Kalman Inversion
 324 (EKI). Following Cleary et al. (2021), we define the inverse problem as

325

326

$$327 \quad \mathbf{y} = \mathcal{G}(\boldsymbol{\theta}) + \eta \quad (3)$$

328 where $\boldsymbol{\theta}$ are the unknown model parameters (in this case, parameters that define the gravity wave
 329 spectrum at the source level, c_w and Bt_{eq}); $\mathcal{G}(\boldsymbol{\theta})$ is the forward model (in this case, MiMA with
 330 the AD99 gravity wave parameterization); \mathbf{y} is the observable (in this case, long-term averages
 331 of stratospheric phenomena); and η is the internal noise on the system. For simplicity, this noise
 332 is assumed to be Gaussian, $\eta \sim N(0, \Gamma)$ (Cleary et al., 2021).
 333

334 The goal of the calibration step is to learn the optimal distribution of parameters given the
 335 observed data, $p(\boldsymbol{\theta} | \mathbf{y})$. This is linked to the likelihood, $p(\mathbf{y} | \boldsymbol{\theta})$, and the prior, $p(\boldsymbol{\theta})$, through
 336 Bayes' theorem:

$$337 \quad p(\boldsymbol{\theta} | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) \quad (4)$$

338
 339
 340 This optimal parameter distribution can be found by minimizing a misfit function which
 341 describes a distance between the data, \mathbf{y} , and the forward model, $\mathcal{G}(\boldsymbol{\theta})$. Following Dunbar et al.
 342 (2021), we define the misfit function to be:

$$343 \quad \Phi(\boldsymbol{\theta}, \mathbf{y}) = \frac{1}{2} \|\mathbf{y} - \mathcal{G}(\boldsymbol{\theta})\|_{\Gamma}^2 = \frac{1}{2} (\mathbf{y} - \mathcal{G}(\boldsymbol{\theta}))^T \Gamma^{-1} (\mathbf{y} - \mathcal{G}(\boldsymbol{\theta})) \quad (5)$$

344 where $\|\cdot\|_{\Gamma} = \sqrt{(\cdot)^T \Gamma^{-1} (\cdot)}$ is the Mahalanobis distance. This is the exponent of a Gaussian
 345 distribution and optimizing this equates to optimizing the log-likelihood when a Gaussian
 346 likelihood is chosen ($p(\mathbf{y} | \boldsymbol{\theta})$). Various optimization methods can be used to minimize $\Phi(\boldsymbol{\theta}, \mathbf{y})$.
 347 Here, we use EKI (Iglesias et al., 2013), which is a derivative-free optimization method, based
 348 on Ensemble Kalman filtering which is extensively used in numerical weather prediction to
 349 estimate a model state of atmospheric variables given observations. EKI uses the same concepts
 350 to solve the inverse problem (Equation (3)), but with two fundamental differences to Ensemble
 351 Kalman filtering used in data assimilation: (1) we aim to find the model parameters $\boldsymbol{\theta}$ given
 352 observations \mathbf{y} , removing dependence on the atmospheric state variable by integrating these out
 353 with long simulations, rather than finding atmospheric state variables) and (2) the inversion is
 354 done offline, without an update to the data at each iteration (i.e., no time dependence).
 355
 356

357
 358 In EKI, we take an ensemble of model parameters, labelled subscript $m = 1, \dots, M$,
 359 initially drawn from the prior, denoted $\boldsymbol{\theta}_m^{(0)} \sim p^{(0)}(\cdot)$. At each iteration, denoted superscript
 360 (n), the forward model gives $\mathcal{G}(\boldsymbol{\theta}_m^{(n)})$ which is used to update each ensemble member at the next
 361 iteration with

$$362 \quad \boldsymbol{\theta}_m^{(n+1)} = \boldsymbol{\theta}_m^{(n)} + C_{\theta\mathcal{G}}^{(n)} \left(\Gamma + C_{\mathcal{G}\mathcal{G}}^{(n)} \right)^{-1} \left(\mathbf{y} - \mathcal{G}(\boldsymbol{\theta}_m^{(n)}) \right)$$

363 where $C_{\mathcal{G}\mathcal{G}}^n$ is the covariance matrix of the ensemble output and $C_{\theta\mathcal{G}}$ is the cross-covariance
 364 matrix between the ensemble parameters and ensemble outputs. Note that $C_{\theta\mathcal{G}}^{(n)} \left(\Gamma + C_{\mathcal{G}\mathcal{G}}^{(n)} \right)^{-1}$ is
 365 the Kalman gain where $\left(\Gamma + C_{\mathcal{G}\mathcal{G}}^{(n)} \right)$ is the innovation covariance, describing the covariance
 366 matrix of the differences between \mathbf{y} and $\mathcal{G}(\boldsymbol{\theta}_m^{(n)})$.
 367

368 Parameters and Priors

369
 370 In this study, the model parameters are

$$371 \quad \boldsymbol{\theta} = (c_w, Bt_{eq},)$$

372 with units [m/s, Pa], described in Table 1, and the model outputs are
 373

$$\mathbf{y} = (T_{QBO}, A_{QBO},)$$

where T_{QBO} is the QBO period in months at 10 hPa and A_{QBO} is the QBO amplitude in m/s at 10 hPa.

When defining the priors on the model parameters, we first consider physical constraints that total gravity wave stress and the half-width of the phase speeds must be positive everywhere, i.e. $Bt_{eq} > 0$ and $c_w > 0$.

We enforce these hard constraints by imposing log-normal priors on all parameters, which equates to transforming the parameters to

$$\hat{\boldsymbol{\theta}} = (\exp(c_w), \exp(Bt_{eq}))$$

and carrying out the calibration on $\hat{\boldsymbol{\theta}}$ with normal priors. The mean and variance are calculated by transforming a normal distribution with means $\boldsymbol{\mu} = (35, 0.0043)$ and variances $\boldsymbol{\sigma}^2 = (10^2, 0.001^2)$ through the exponential map.

3.2 Emulation

The calibration step allows us to learn the distribution of optimal parameters given the observations. For uncertainty quantification of the model output, we would next sample from this distribution, e.g., with a Monte Carlo method such as MCMC. However, since this requires many expensive model evaluations, we build an emulator that can be evaluated cheaply. The emulator can be trained with the samples obtained through the EKI calibration step above. These samples are ideal as they cover the posterior distribution (particularly in the later iterations of EKI) and the prior distribution (in the early iterations of EKI).

The emulator we use here is a Gaussian process (GP) emulator, which is a popular Bayesian emulation tool in the calibration and uncertainty quantification community (e.g. Couvreur et al., 2021; Kennedy & O’Hagan, 2001; Williamson et al., 2016). This is because GPs model the distribution of functions that satisfies a given dataset, meaning they can produce a mean function and a measure of uncertainty around this (e.g., the standard deviation or confidence intervals). GPs use a Bayesian approach, where the user defines a prior GP which is combined with the dataset in Bayes’ theorem to derive a posterior GP that agrees with the data. Deriving the posterior GP is tractable because a GP assumes that any input values are linked through a multivariate Gaussian distribution. Following this assumption, the GP emulator can be evaluated at new unseen input values to obtain a distribution of possible outputs, i.e., a mean and a standard deviation. The Gaussian process emulator has the additional benefit that it smooths the output, leading to better convergence properties for the MCMC algorithm used in the sample step of CES (as it reduces the chance of the MCMC becoming “stuck” in local minima). Dunbar et al. (2021) note the Gaussian process as suitable for climate problems since we are approximating climate properties, defined on an infinite time horizon, with finite time averages. Here, we assume that finite time averaged data is a noisy approximation of the infinite time average, where the noise is assumed to be Gaussian given large enough timescales, due to the central limit theorem. A Gaussian process emulator can also learn this internal noise, as described below.

419 The Gaussian process approximates the output of MiMA given gravity wave parameters
 420 i.e.

$$421 \mathcal{G}(\boldsymbol{\theta}) \approx GP(m(\boldsymbol{\theta}), C(\boldsymbol{\theta}, \cdot))$$

422 where $m(\boldsymbol{\theta})$ is the mean function and $C(\boldsymbol{\theta}, \boldsymbol{\theta}')$ is the covariance function (or kernel) that
 423 describes the covariance between two parameter choices, $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$. We make choices for the
 424 prior mean function and prior covariance function which both control the structure of the
 425 Gaussian process emulator (Rasmussen & Williams, 2006). The prior mean function is often
 426 assumed to be zero so that all choices are determined by the covariance function, as done here.
 427 The covariance function defines the similarity of two inputs $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ and how this propagates
 428 through to the similarity of the outputs $\mathcal{G}(\boldsymbol{\theta})$ and $\mathcal{G}(\boldsymbol{\theta}')$. For this we use a squared exponential
 429 kernel and assume independent length scales for each parameter dimension (also known as
 430 automatic relevance determination), with an additive white noise kernel, which represents the
 431 internal variability, consistent across all values of $\boldsymbol{\theta}$. The length scale and variance
 432 hyperparameters are learned using type II maximum likelihood using Scikit-learn (Pedregosa et
 433 al., 2011). Note that prior to building the Gaussian process emulator, we remove correlations
 434 between the outputs by performing Singular Value Decomposition (SVD).
 435
 436
 437
 438

439 3.3 Sample

440
 441 With the GP emulator, we can now (approximately) evaluate $\mathcal{G}(\boldsymbol{\theta})$ rapidly. This means
 442 we can obtain the posterior distribution on $\boldsymbol{\theta}$ given the dataset \mathbf{y} by running a Markov Chain
 443 Monte Carlo (MCMC) simulation, which typically require $O(10^5)$ function evaluations. The
 444 posterior distribution is given by Equation (4) where $p(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood, assumed to be
 445 Gaussian, i.e.

$$446 p(\mathbf{y}|\boldsymbol{\theta}) = \frac{1}{\sqrt{\det(\boldsymbol{\Gamma})}} \exp\left(-\frac{1}{2}\left((\mathbf{y} - \mathcal{G}(\boldsymbol{\theta}))^T \boldsymbol{\Gamma}^{-1}(\mathbf{y} - \mathcal{G}(\boldsymbol{\theta}))\right)\right)$$

447
 448 We use the same priors defined for the calibration (Section 3.1). We run a Metropolis
 449 random walk MCMC for 10^5 iterations (after 1000 burn-in iterations) to obtain the posterior
 450 distribution (Metropolis et al., 1953). The random walk step size is determined to ensure an
 451 acceptance rate close to 25% (Roberts & Rosenthal, 2004). Note that the MCMC is carried out in
 452 the decorrelated space, after performing SVD. All results are presented after transforming back
 453 into the original parameter space.
 454
 455
 456

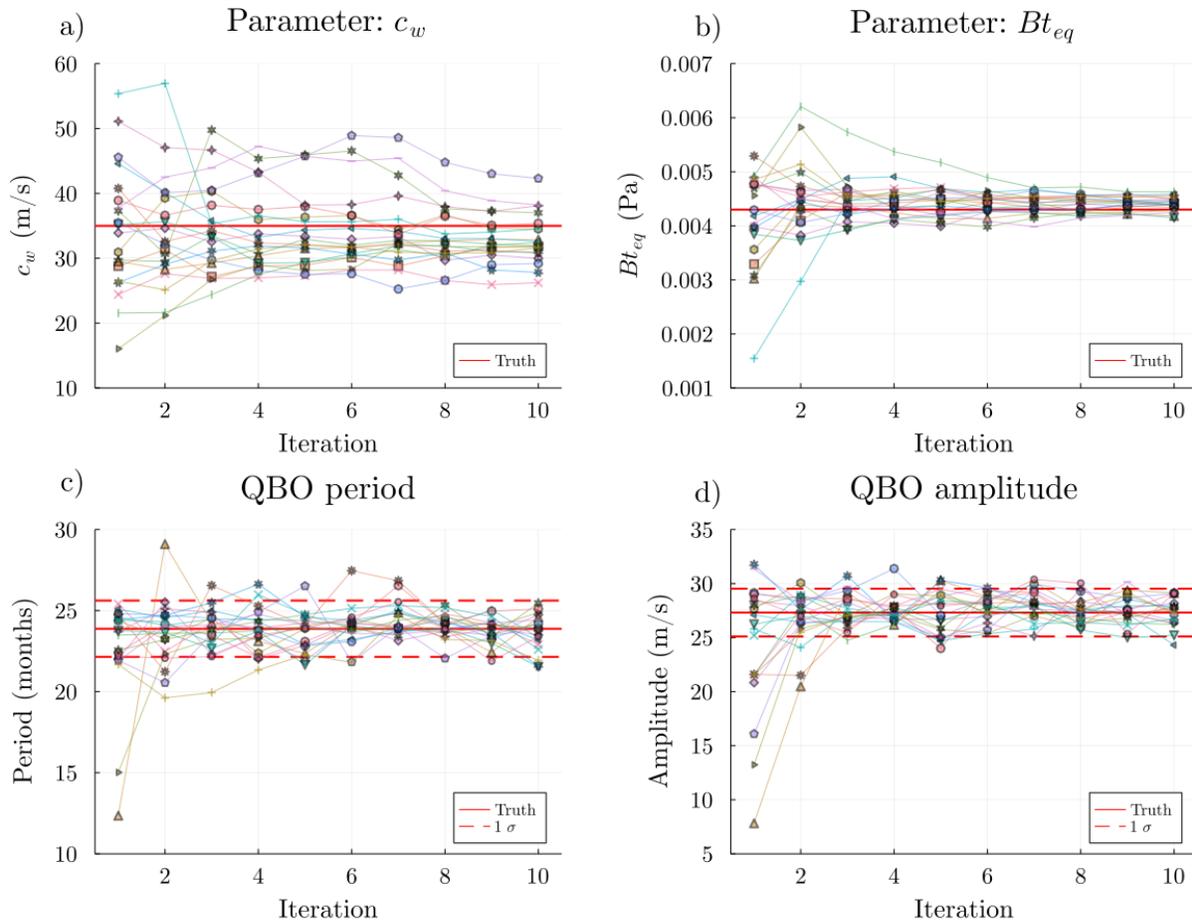
457 4 Results

458 4.1 Calibrate, Emulate and Sample in the perfect model setting

459
 460

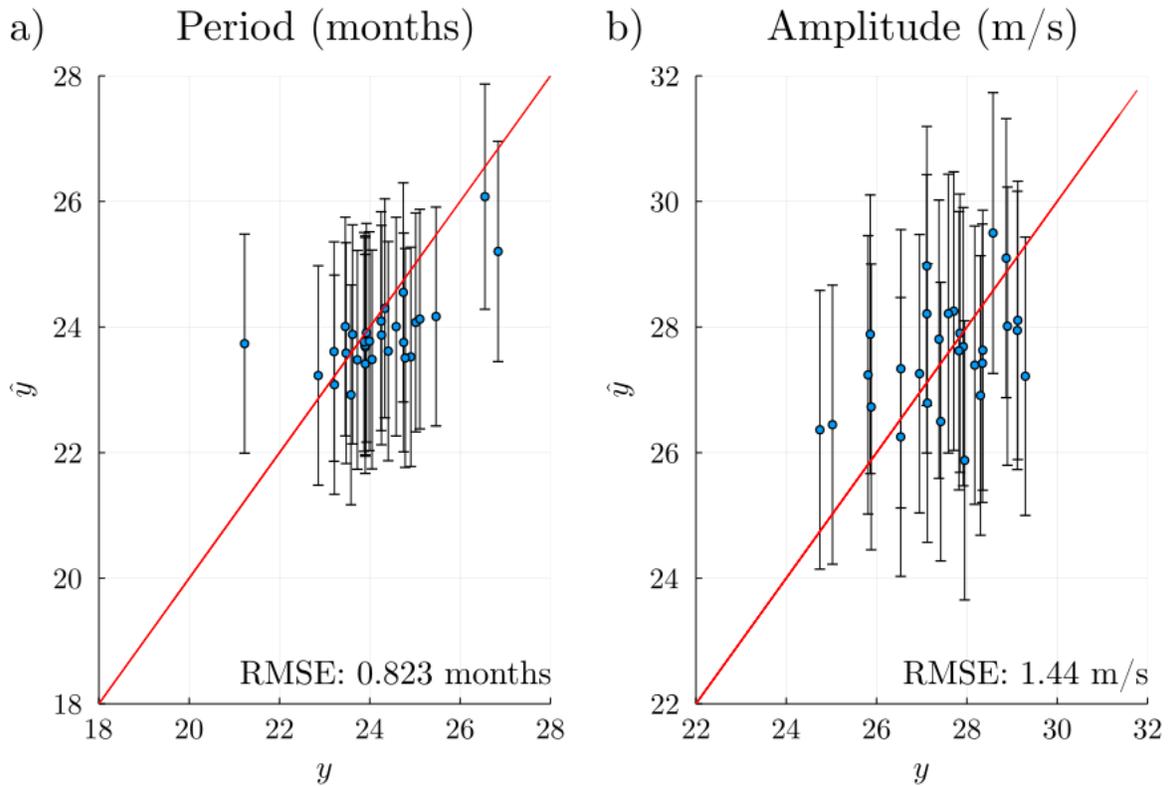
461 We explore the results of CES with the “perfect model” setting, as done in Dunbar et al.
 462 (2021), where we define the “truth” to be a long 50-year integration of MiMA, with known
 463 model parameters, here $c_w = 35$ m/s and $Bt_{eq} = 0.0043$ Pa. The long simulation gives a QBO
 464 period of 23.9 ± 1.7 years and amplitude 27.3 ± 2.2 m/s (shown in Figure 1b), where the
 465 uncertainties here are 1 standard deviation across all QBO cycles in the 50-year integration. The
 466 calibration step learns the posterior distribution of parameter values that gives a QBO consistent
 467 with this. It allows us to test the method on a simpler problem while developing an
 468 understanding of how the model parameters relate to each other.

470 The first step of CES is to calibrate c_w and Bt_{eq} to the QBO metrics for period and
 471 amplitude. EKI is run with an $M = 20$ ensemble. Figure 2 shows the EKI for 10 iterations,
 472 where the top two panels show the gravity wave parameters c_w and Bt_{eq} and the bottom two
 473 panels show the model output. The parameters appear to move closer to convergence after
 474 around 6-8 iterations.



475
 476 *Figure 2. (a-b) Parameter and (c-d) model output values for all iterations of EKI for the perfect*
 477 *model setting, where iteration 1 consists of parameter values drawn from the prior. Each*
 478 *line/marker represents a single ensemble member. The red line denotes in (a-b) the “truth” i.e.,*
 479 *the known parameter values (Table 1) and in (c-d) the model output obtained in one long MiMA*
 480 *simulation with these parameter values, with the dashed red line showing 1 standard deviation*
 481 *across the simulation.*

482 Considering each ensemble member at each iteration, EKI gives a total of 200 input-
 483 output pairs. These data are used to train the Gaussian process emulator in the emulation stage of
 484 CES. First, the validity of the emulator is tested by training the GP emulator on 170 input-output
 485 pairs, which include all data from the first three iterations and the rest selected at random from
 486 the last seven iterations. This leaves aside 30 samples for testing, randomly selected from the last
 487 seven iterations (to avoid testing involving extrapolation to regions of the parameter space
 488 outside of the posterior distribution). Figure 3 shows this test data, y , against the Gaussian
 489 process prediction \hat{y} , where a perfect prediction would be these points lying on the $\hat{y} = y$ line
 490 shown in red. The error bars indicate the 1σ uncertainty predicted by the Gaussian process
 491 emulator. The $\hat{y} = y$ line falls within 1σ of the Gaussian process prediction for the majority of
 492 test data points, as required for an accurate emulator.
 493



494
 495 *Figure 3. Plots of emulator performance on example test data points, selected at random from*
 496 *the last 8 iterations of EKI for a) period and b) amplitude of the QBO. The test data values are*
 497 *plotted on the x-axis (y) and the Gaussian process emulator predictions are plotted on the y-axis*
 498 *(\hat{y}), where the error bars indicate the Gaussian process 1σ levels. The red line shows where $\hat{y} =$*
 499 *y , indicating a perfect prediction.*

500
 501

502 To maximize accuracy, the final emulator used is trained on all 200 samples. A sweep
 503 across the parameter space is carried out by varying c_w from 10 to 70 m/s and Bt_{eq} from 0.002
 504 to 0.007 Pa. Figure 4 shows contour plots of a) the QBO period and b) the QBO amplitude for
 505 this parameter sweep across c_w and Bt_{eq} . The points indicate the training data values, showing
 506 an agreement with the GP emulator. Note that the training points are fairly crowded within the

507 region where the misfit function is minimized ($25 \lesssim c_w \lesssim 40$ m/s and $0.004 \lesssim Bt_{eq} \lesssim 0.005$
 508 Pa). Outside this region, the GP emulator is extrapolating to new regions of the parameter space
 509 and therefore is less trustworthy. The 1σ level predicted by the GP emulator also highlights this
 510 in Figure 4c-d for the period and amplitude respectively.

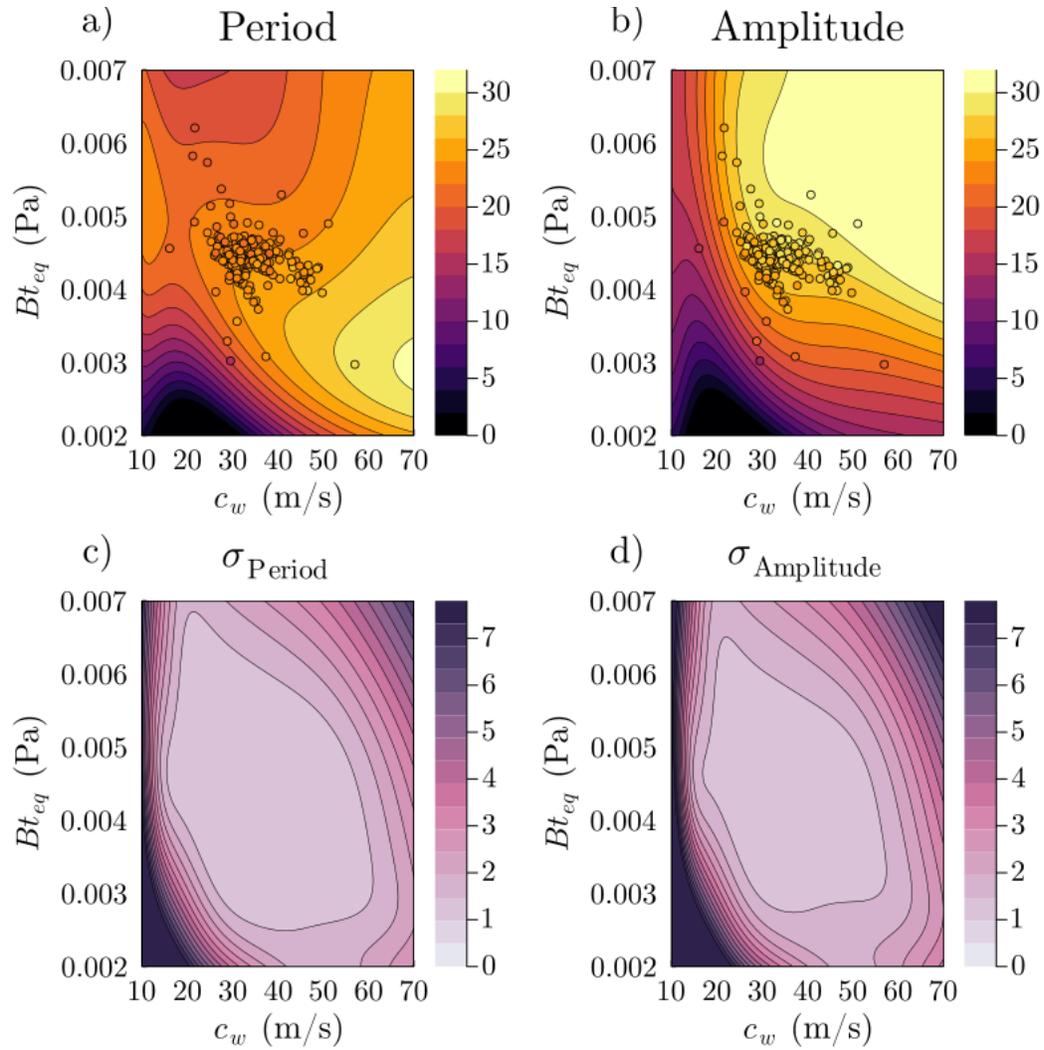
511

512 The contour plot in Figure 4a estimates a maximum in QBO period for relatively high c_w
 513 (50-70 m/s) when Bt_{eq} is chosen to be relatively low (0.002-0.003 Pa). Increasing Bt_{eq} from
 514 here leads to a faster QBO, consistent with the idealized models of Holton & Lindzen (1972) and
 515 Plumb (1977), since increased gravity wave stress leads to increased deceleration of winds and
 516 therefore more rapidly descending westerly/easterly shear zones (Dunkerton, 1997; Schirber et
 517 al., 2015). Decreasing c_w also leads to a slightly faster period, consistent with Garfinkel et al.
 518 (2022), possibly due to the weaker QBO present under slower phase speeds.

519

520 Figure 4b shows a peak in QBO amplitude when both c_w and Bt_{eq} are relatively high.
 521 Increasing c_w increases the QBO amplitude since the higher phase speeds contribute to the faster
 522 westerlies and easterlies in the QBO (Holton & Lindzen, 1972; Plumb, 1977; Schirber et al.,
 523 2015) but only up until c_w reaches around 30 m/s. Beyond this, increasing c_w has minimal effect,
 524 also seen in Garfinkel et al. (2022). This could be because phase speeds much faster than the
 525 easterlies/westerlies do not reach a critical level in the stratosphere where $c = u$, and instead
 526 continue propagating upwards, without depositing drag until reaching the sponge layer. For $c_w \gtrsim$
 527 30 m/s, the amplitude is more sensitive to Bt_{eq} , where increasing the gravity wave stress will
 528 increase the drag deposited and therefore lead to a stronger QBO.

529



530

531

532

533

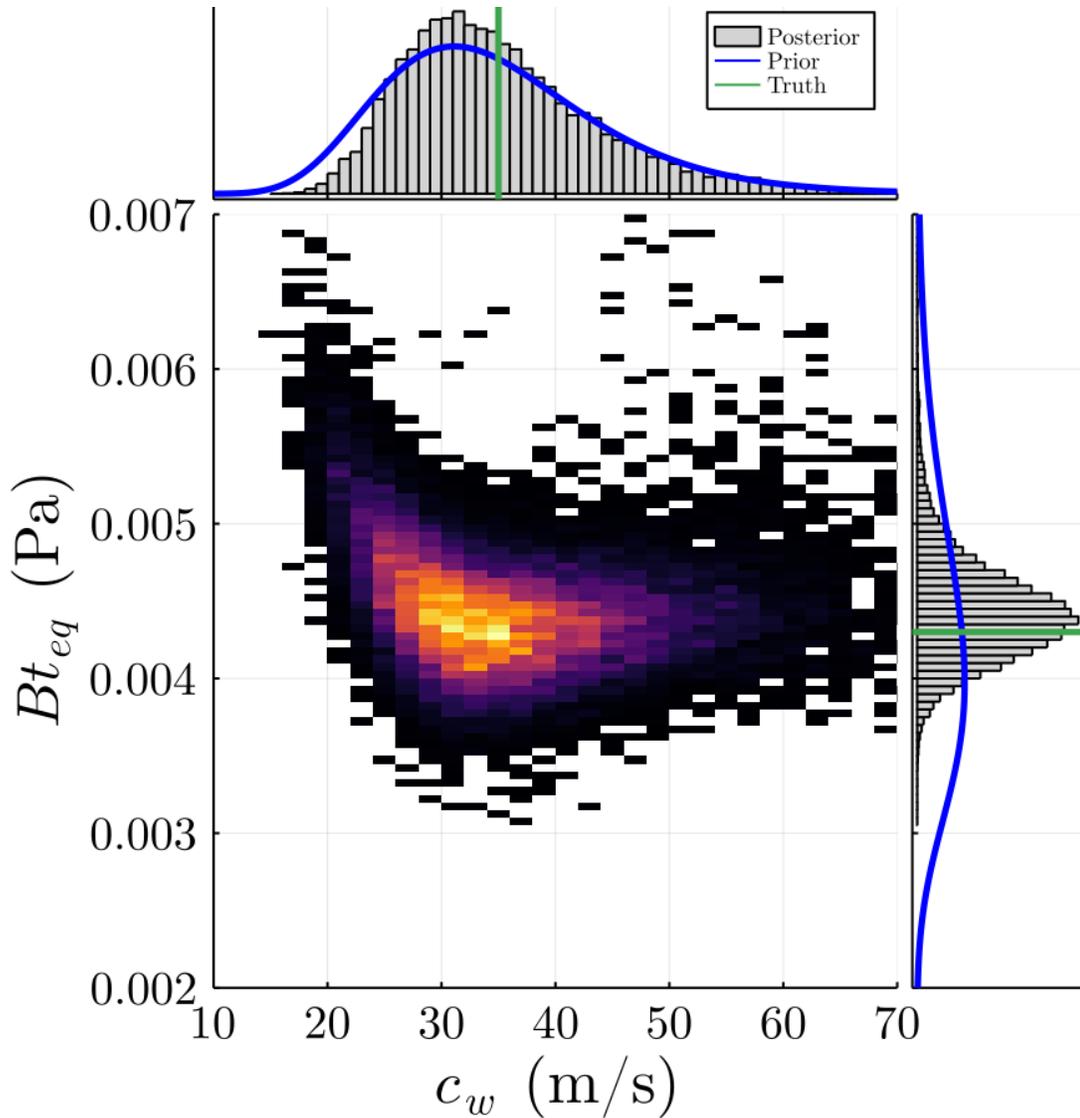
534

535

536

Figure 4. Gaussian process emulator predictions over a sweep across parameter values ($c_w = 10 - 70$ m/s, $Bt_{eq} = 0.002 - 0.007$ Pa) learned from the EKI in the perfect model setting for a) QBO period and b) QBO amplitude. The scatter points indicate the training data from MiMA simulations obtained through EKI. The 1σ uncertainty associated with these predictions are shown in c) for the period and d) for the amplitude.

536



537
 538 *Figure 5. Samples from the posterior distribution of c_w and Bt_{eq} generated by the MCMC in the*
 539 *final stage of CES. The marginal distributions are shown on the corresponding axis, with the*
 540 *prior distributions shown in blue and the known “truth” in green.*

541
 542 In the last stage of CES, we sample from the posterior distribution using an MCMC (see
 543 Supporting Movie S1). After removing 10000 iterations for burn-in, 80000 samples from the
 544 posterior distribution are shown in Figure 5, where the 2D histogram is shown in the center with
 545 the marginal posterior distributions for c_w and Bt_{eq} shown on the corresponding axis. The prior
 546 distribution is also shown in blue, with the known truth in green. The 2D histogram shows a
 547 correlation between c_w and Bt_{eq} , indicating that a sample with a larger value of c_w can still
 548 produce a QBO with a realistic period and amplitude if Bt_{eq} is decreased appropriately. The
 549 narrower posterior distribution for Bt_{eq} indicates this is more crucial for obtaining a correct
 550 QBO, while the posterior distribution for c_w more closely follows the prior distribution chosen.

551 Sampling the parameters from this histogram gives a QBO consistent with the “truth” selected
 552 here.

553
 554
 555
 556
 557
 558

4.3 Global Sensitivity Analysis

559 We carry out Global Sensitivity Analysis (GSA) to measure the sensitivity of the climate
 560 model output to the gravity wave parameters through variance-based sensitivity indices that
 561 describe how much of the variance in the output can be attributed to the variance in each input
 562 parameter for a given input parameter distribution (Saltelli et al., 2007). This method averages
 563 over all possible values for all other parameters (‘global’ sensitivity analysis) rather than keeping
 564 them fixed at the default values (‘local’ sensitivity analysis). This requires a large number of
 565 samples of the model, so the availability of the emulator to obtain inexpensive samples is crucial
 566 for this analysis.

567
 568
 569

The first order sensitivity index describes the variance in an output variable, Y , due to a
 single parameter, θ_i , and is given by

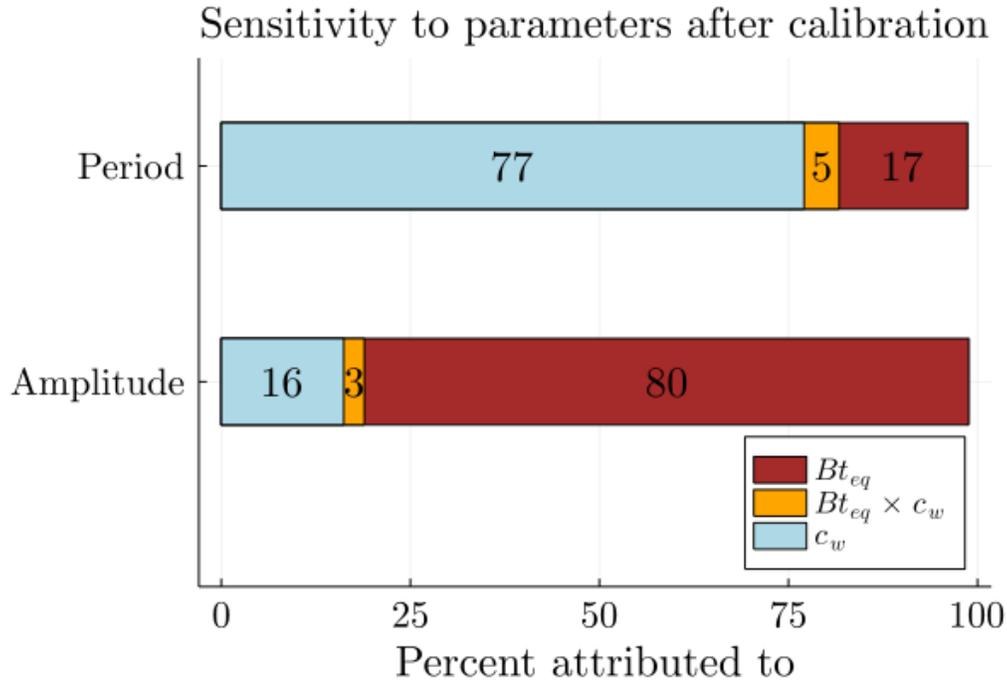
$$570 \quad SI_i = \frac{Var(\theta_i)(E_{\theta_{\sim i}}(Y|\theta_i))}{Var(Y)}$$

571 where $Y|\theta_i$ denotes the estimated output due to parameter θ_i and $E_{\theta_{\sim i}}(\cdot)$ indicates the average
 572 over all other parameters except for θ_i . The Sobol’ method (Sobol’, 2001) approximates this by
 573 estimating $Var(\theta_i)$ (see Saltelli et al., 2010). Higher order sensitivity indices can be estimated
 574 to attribute the interaction between multiple parameter values.

575
 576

We estimate first order sensitivity indices in the decorrelated space (applying SVD to
 577 remove correlations between c_w and Bt_{eq}). After transforming these back into the real space, the
 578 sensitivity indices in percentages of the QBO period and amplitude are shown in

579 Figure 6. The QBO period is most sensitive to c_w , while the QBO amplitude is most
 580 sensitive to Bt_{eq} . This is in agreement with the contour plots in Figure 4 in the region of the
 581 calibration. We expect that the QBO period is primarily controlled by Bt_{eq} and therefore after
 582 calibration, the remaining uncertainties are due to uncertainties in c_w . The QBO amplitude is
 583 mostly governed by c_w , which pushes QBO wind speeds towards the phase speeds. During the
 584 calibration stage, c_w is constrained so that remaining uncertainties in the QBO amplitude are
 585 caused mostly by Bt_{eq} . Note that the interaction terms are small, since the analysis is carried out
 586 in the decorrelated space.



587

588 *Figure 6 Sensitivity indices as a percentage, describing the proportion of variance in the QBO*
 589 *period and amplitude attributed to the variance in the parameters, c_w and Bt_{eq} .*

590

591

592

593

594

4.3 Uncertainty Quantification in New Scenario

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

Understanding the uncertainty in climate model output due to the gravity wave parameterization is one of the main motivations for this analysis. In this section, we explore the parametric uncertainty in a climate change projection, meaning the uncertainty in model output that is due to the possible values that c_w and Bt_{eq} could take. This can be assessed through a perturbed parameter ensemble, where an ensemble of simulations is run with parameter values sampled from their distribution in Figure 5 (Murphy et al., 2014). Here we run a perturbed parameter ensemble for a $2xCO_2$ integration. We use this ensemble of simulations to quantify parametric uncertainty for both scenarios.

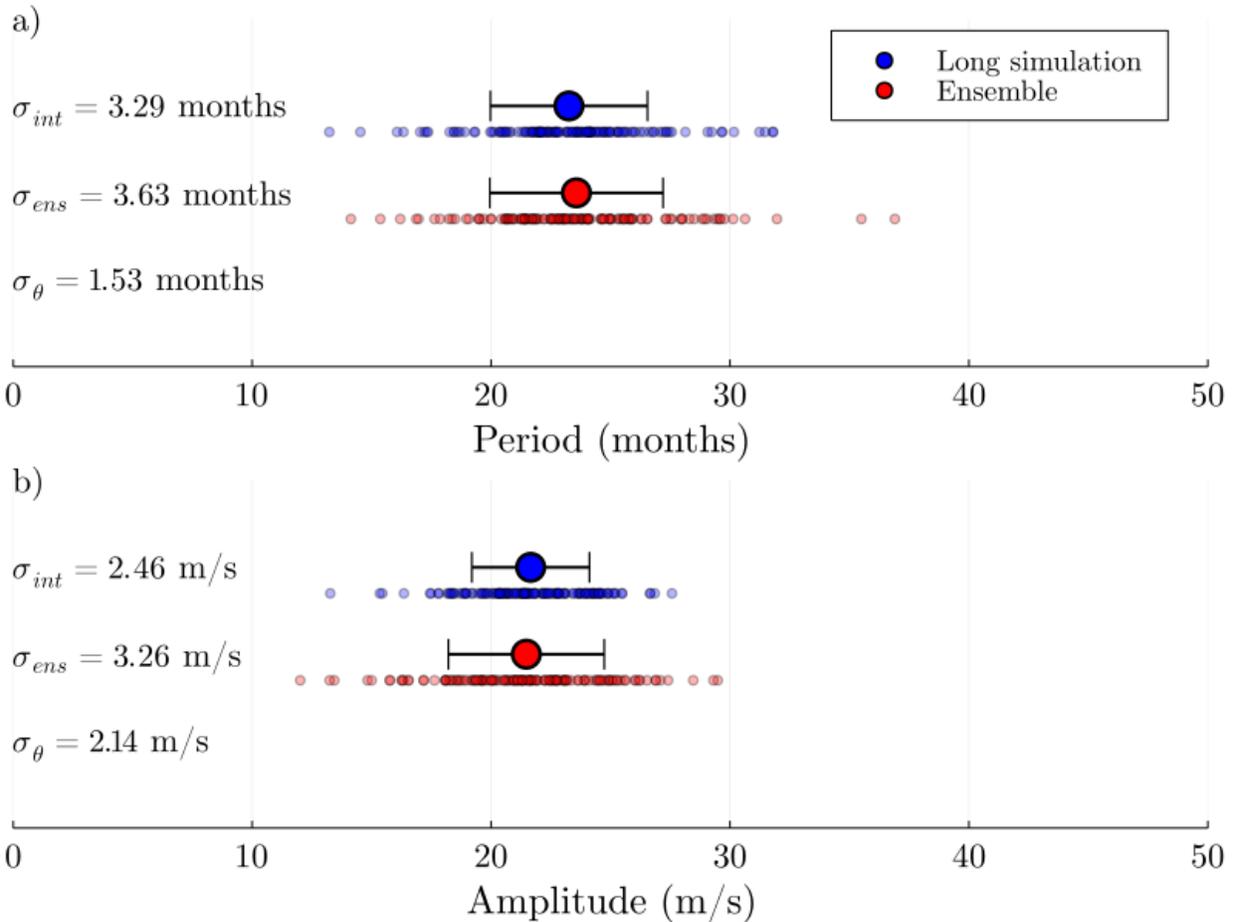
We run a perturbed parameter ensemble of 50 simulations for 10 years each, initialized with a spun-up climate (Wan et al., 2014), obtained through a 200 year $2xCO_2$ integration with fixed model parameters. Each 10-year simulation provides around 4-5 QBO cycles per ensemble member, after allowing 1 year for spin-up (a total of 140 QBO cycles). The QBO period and amplitudes are plotted in red in Figure 7 and compared against a single long simulation in blue, which was run for 300 years to giving roughly the same number of QBO cycles (142 cycles). Note that several QBO disruptions occurred in both the long simulation and the ensembles, so

612 these were removed before analysis. All QBO cycles for both the long simulation and the
 613 ensemble members are shown in Supplementary Figures S1-2.

614
 615 The larger variance in the ensembles (red) in Figure 7 compared to the long simulation
 616 (blue) is due to the uncertainty in parameter values. The internal variability can be estimated as
 617 the standard deviation across the 300-year simulation, denoted σ_{int} in Figure 7. The difference
 618 between the standard deviation in the ensemble, σ_{ens} , and the internal variability can be used to
 619 estimate the parametric uncertainty, σ_{θ} , by assuming a Gaussian distribution of QBO periods and
 620 amplitudes across all cycles so that $\sigma_{ens}^2 = \sigma_{int}^2 + \sigma_{\theta}^2$.

621
 622 This gives parametric uncertainty estimates in the period of 1.53 months and in the
 623 amplitude of 2.14 m/s under 2xCO₂ forcing, when the parameter values are sampled from the
 624 distribution in Figure 5. Here we have tuned the parameters to a long integration of a present-day
 625 climate, but the natural extension would be to calibrate parameters to observations, which would
 626 introduce further uncertainties. Therefore we may expect the parametric uncertainties presented
 627 here to be a lower bound on uncertainties associated with the gravity wave parameterization.

628
 629
 630



631
 632 *Figure 7. Range of values of QBO a) period and b) amplitude for a 2xCO₂ scenario for a long*
 633 *simulation of 300 years in blue, where parameter values are fixed at $c_w = 35$ m/s, $Bt_{eq} =$*

634 0.0043 Pa, compared against an ensemble in red (50 simulations, each of 10 years) where
 635 parameter values are drawn from the distribution in Figure 5. The large markers show the mean
 636 across the long simulation/ensemble and the error bars show 1 standard deviation. The smaller
 637 markers show the period and amplitude for all QBO cycles. Note that QBO disruptions are
 638 removed before analysis. The internal variability estimated from the long simulation is shown as
 639 σ_{int} , the ensemble variability is σ_{ens} , and the parametric uncertainty is σ_{θ} .

640
 641
 642

643 5 Discussion

644

645 This study demonstrates how the Calibrate, Emulate and Sample (CES) method can be
 646 applied to tune parameters and quantify uncertainties associated with a gravity wave
 647 parameterization within an intermediate complexity climate model. We have explored the
 648 application of CES under the perfect model setting, where we prescribe the “truth” as a long
 649 model simulation with known parameter values. However, in future studies this will be extended
 650 to a more realistic setting, using observational data from global radiosonde measurements as the
 651 “truth” (Freie Universität Berlin, 2007).

652 The CES method allows us to learn the optimal distribution of parameter values for the
 653 half-width of the phase speeds, c_w , and the total gravity wave stress, Bt_{eq} , both of which define
 654 the gravity wave spectrum at the source level. We find that these parameters have an anti-
 655 correlated distribution, i.e. a higher value of Bt_{eq} can be compensated with a lower value of c_w
 656 to achieve the same QBO period and amplitude.

657 A global sensitivity analysis highlighted that after calibration the QBO period is most
 658 sensitive to c_w , since it has been constrained mainly by Bt_{eq} , which directly influences the
 659 deceleration of easterly/westerly winds. Similarly, the QBO amplitude is more sensitive to Bt_{eq} ,
 660 as wind speeds are constrained predominantly by gravity wave phase speeds c_w (Dunkerton,
 661 1997; Lindzen & Holton, 1968).

662 We have quantified parametric uncertainties associated with the gravity wave
 663 parameterization under a $2\times\text{CO}_2$ forcing as 1.53 months for the QBO period and 2.14 m/s for the
 664 amplitude. We expect these to be a lower bound on the parametric uncertainty, since we
 665 calibrated the parameters to a long model integration, in the absence of realistic QBO variability
 666 and measurement error. These are of a similar order of magnitude to the internal variability,
 667 highlighting their relevance to climate change projections. Note that parametric uncertainty does
 668 not account for uncertainty in the structure of the parameterization itself, rather the uncertainty in
 669 the parameter values of c_w and Bt_{eq} alone. Here, the parameter values are tuned based on the
 670 QBO in the present day climate, isolating the effects the gravity wave parameters from any
 671 changes in the source, such as convection, which is likely to change under a warming climate.

672 In this study, we calibrated to the QBO period and amplitude at 10 hPa, since these are
 673 the first order properties of the QBO. Further extensions of this would be to explore other
 674 properties of the QBO such as the period and amplitudes at different levels of the stratosphere or
 675 the westerly and easterly amplitudes (e.g. to reduce the westerly bias in MiMA in Figure 1). This
 676 may be more complicated as Giorgetta et al. (2006) find that both the QBO in the lower

677 stratosphere and the westerly phase of the QBO are controlled more by resolved waves, rather
 678 than sub-grid scale parameterizations.

679 Calibrating the gravity wave parameterization to obtain a realistic QBO can potentially
 680 lead to compensating model errors at higher latitudes (Anstey et al., 2016). It is known that non-
 681 orographic gravity waves contribute to the breakdown of the polar vortices, influencing the
 682 frequency and properties of Sudden Stratospheric Warmings (SSWs) (Siskind et al., 2007, 2010;
 683 Wright et al., 2010) and the timing of the Spring final warming (Gupta et al., 2021). The effect of
 684 varying extratropical gravity wave parameters has not yet been explored in MiMA. Calibrating
 685 extratropical gravity wave parameters to properties of the stratospheric polar vortex in both
 686 hemispheres is a topic of future research.

687 The introduction of automated methods such as Ensemble Kalman Inversion allows us to
 688 calibrate sub-grid scale parameterizations in GCMs, as far fewer climate model integrations are
 689 required ($O(100)$ compared to $O(10^5)$). However, for high complexity GCMs, even running 100
 690 model integrations is highly costly, which is why these are typically tuned crudely (e.g. Kodama
 691 et al., 2021). Learning the optimal gravity wave parameters of intermediate complexity climate
 692 models, such as MiMA, is a potential step forward for estimating gravity wave parameters in
 693 higher complexity models.

694
 695

696 **Acknowledgments**

697 This research was made possible by Schmidt Futures, a philanthropic initiative founded by Eric
 698 and Wendy Schmidt, as part of the Virtual Earth System Research Institute (VESRI). AS
 699 acknowledges support from the National Science Foundation through grant OAC-2004492. We
 700 thank Oliver Dunbar and Tapio Schneider for useful discussions. The authors have no conflicts
 701 of interest.

702

703 **Open Research**

704 The code used in this analysis, including scripts to run MiMA and reproduce all results
 705 presented here can be found at 10.5281/zenodo.6629730. The codebase for Calibrate, Emulate,
 706 Sample and Ensemble Kalman Inversion are both maintained by the Climate Modeling Alliance
 707 (Clima) group at Caltech and can be found at
 708 <https://github.com/CliMA/CalibrateEmulateSample.jl> and
 709 <https://github.com/CliMA/EnsembleKalmanProcesses.jl>. The Model of an idealized Moist
 710 Atmosphere (MiMA) (Garfinkel et al., 2020; Jucker & Gerber, 2017) is available at
 711 <https://github.com/mjucker/MiMA>.

712

713 **References**

- 714 Alexander, M. J., & Dunkerton, T. J. (1999). A Spectral Parameterization of Mean-Flow Forcing due to Breaking
 715 Gravity Waves. *Journal of the Atmospheric Sciences*, 56(24), 4167–4182. [https://doi.org/10.1175/1520-0469\(1999\)056<4167:ASPOMF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1999)056<4167:ASPOMF>2.0.CO;2)
 716 Alexander, M. J., & Rosenlof, K. H. (2003). Gravity-wave forcing in the stratosphere: Observational constraints
 717 from the Upper Atmosphere Research Satellite and implications for parameterization in global models.
 718 *Journal of Geophysical Research (Atmospheres)*, 108, 4597. <https://doi.org/10.1029/2003JD003373>
 719 Anstey, J. A., Scinocca, J. F., & Keller, M. (2016). Simulating the QBO in an Atmospheric General Circulation
 720 Model: Sensitivity to Resolved and Parameterized Forcing. *Journal of the Atmospheric Sciences*, 73(4),
 721 1649–1665. <https://doi.org/10.1175/JAS-D-15-0099.1>
 722

- 723 Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander,
724 M. J., Hirota, I., Horinouchi, T., Jones, D. B. A., Kinnersley, J. S., Marquardt, C., Sato, K., & Takahashi,
725 M. (2001). The quasi-biennial oscillation. *Reviews of Geophysics*, *39*(2), 179–229.
726 <https://doi.org/10.1029/1999RG000073>
- 727 Barton, C. A., McCormack, J. P., Eckermann, S. D., & Hoppel, K. W. (2019). Optimization of Gravity Wave Source
728 Parameters for Improved Seasonal Prediction of the Quasi-Biennial Oscillation. *Journal of the Atmospheric*
729 *Sciences*, *76*(9), 2941–2962. <https://doi.org/10.1175/JAS-D-19-0077.1>
- 730 Bushell, A. C., Anstey, J. A., Butchart, N., Kawatani, Y., Osprey, S. M., Richter, J. H., Serva, F., Braesicke, P.,
731 Cagnazzo, C., Chen, C.-C., Chun, H.-Y., Garcia, R. R., Gray, L. J., Hamilton, K., Kerzenmacher, T., Kim,
732 Y.-H., Lott, F., McLandress, C., Naoe, H., ... Yukimoto, S. (2020). Evaluation of the Quasi-Biennial
733 Oscillation in global climate models for the SPARC QBO-initiative. *Quarterly Journal of the Royal*
734 *Meteorological Society*, *n/a*(n/a). <https://doi.org/10.1002/qj.3765>
- 735 Chantry, M., Hatfield, S., Dueben, P., Polichtchouk, I., & Palmer, T. (2021). Machine Learning Emulation of
736 Gravity Wave Drag in Numerical Weather Forecasting. *Journal of Advances in Modeling Earth Systems*,
737 *13*(7), e2021MS002477. <https://doi.org/10.1029/2021MS002477>
- 738 Cleary, E., Garbuno-Inigo, A., Lan, S., Schneider, T., & Stuart, A. M. (2021). Calibrate, emulate, sample. *Journal of*
739 *Computational Physics*, *424*, 109716. <https://doi.org/10.1016/j.jcp.2020.109716>
- 740 Couvreur, F., Hourdin, F., Williamson, D., Roechrig, R., Volodina, V., Villefranque, N., Rio, C., Audouin, O.,
741 Salter, J., Bazile, E., Brient, F., Favot, F., Honnert, R., Lefebvre, M.-P., Madeleine, J.-B., Rodier, Q., & Xu,
742 W. (2021). Process-Based Climate Model Development Harnessing Machine Learning: I. A Calibration
743 Tool for Parameterization Improvement. *Journal of Advances in Modeling Earth Systems*, *13*(3),
744 e2020MS002217. <https://doi.org/10.1029/2020MS002217>
- 745 Donner, L. J., Wyman, B. L., Hemler, R. S., Horowitz, L. W., Ming, Y., Zhao, M., Golaz, J.-C., Ginoux, P., Lin, S.-
746 J., Schwarzkopf, M. D., Austin, J., Alaka, G., Cooke, W. F., Delworth, T. L., Freidenreich, S. M., Gordon,
747 C. T., Griffies, S. M., Held, I. M., Hurlin, W. J., ... Zeng, F. (2011). The Dynamical Core, Physical
748 Parameterizations, and Basic Simulation Characteristics of the Atmospheric Component AM3 of the GFDL
749 Global Coupled Model CM3. *Journal of Climate*, *24*(13), 3484–3519.
750 <https://doi.org/10.1175/2011JCLI3955.1>
- 751 Dunbar, O. R. A., Garbuno-Inigo, A., Schneider, T., & Stuart, A. M. (2021). Calibration and Uncertainty
752 Quantification of Convective Parameters in an Idealized GCM. *Journal of Advances in Modeling Earth*
753 *Systems*, *13*(9), e2020MS002454. <https://doi.org/10.1029/2020MS002454>
- 754 Dunkerton, T. J. (1997). The role of gravity waves in the quasi-biennial oscillation. *Journal of Geophysical*
755 *Research: Atmospheres*, *102*(D22), 26053–26076. <https://doi.org/10.1029/96JD02999>
- 756 Espinosa, Z. I., Sheshadri, A., Cain, G. R., Gerber, E. P., & DallaSanta, K. J. (2022). Machine Learning Gravity
757 Wave Parameterization Generalizes to Capture the QBO and Response to Increased CO₂. *Geophysical*
758 *Research Letters*, *49*(8), e2022GL098174. <https://doi.org/10.1029/2022GL098174>
- 759 Freie Universität Berlin. (2007, October 25). *The Quasi-Biennial-Oscillation (QBO) Data Serie*. The Quasi-
760 Biennial-Oscillation (QBO) Data Serie. [https://www.geo.fu-](https://www.geo.fu-berlin.de/en/met/ag/strat/produkte/qbo/index.html)
761 [berlin.de/en/met/ag/strat/produkte/qbo/index.html](https://www.geo.fu-berlin.de/en/met/ag/strat/produkte/qbo/index.html)
- 762 Garfinkel, C. I., Gerber, E. P., Shamir, O., Rao, J., Jucker, M., White, I., & Paldor, N. (2022). A QBO Cookbook:
763 Sensitivity of the Quasi-Biennial Oscillation to Resolution, Resolved Waves, and Parameterized Gravity
764 Waves. *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002568.
765 <https://doi.org/10.1029/2021MS002568>
- 766 Garfinkel, C. I., & Hartmann, D. L. (2011). The Influence of the Quasi-Biennial Oscillation on the Troposphere in
767 Winter in a Hierarchy of Models. Part I: Simplified Dry GCMs. *Journal of the Atmospheric Sciences*,
768 *68*(6), 1273–1289. <https://doi.org/10.1175/2011JAS3665.1>
- 769 Garfinkel, C. I., White, I., Gerber, E. P., Jucker, M., & Erez, M. (2020). The Building Blocks of Northern
770 Hemisphere Wintertime Stationary Waves. *Journal of Climate*, *33*(13), 5611–5633.
771 <https://doi.org/10.1175/JCLI-D-19-0181.1>
- 772 Geller, M. A., Alexander, M. J., Love, P. T., Bacmeister, J., Ern, M., Hertzog, A., Manzini, E., Preusse, P., Sato, K.,
773 Scaife, A. A., & Zhou, T. (2013). A Comparison between Gravity Wave Momentum Fluxes in
774 Observations and Climate Models. *Journal of Climate*, *26*(17), 6383–6405. [https://doi.org/10.1175/JCLI-D-](https://doi.org/10.1175/JCLI-D-12-00545.1)
775 [12-00545.1](https://doi.org/10.1175/JCLI-D-12-00545.1)
- 776 Giorgetta, M. A., Manzini, E., Roeckner, E., Esch, M., & Bengtsson, L. (2006). Climatology and Forcing of the
777 Quasi-Biennial Oscillation in the MAECHAM5 Model. *Journal of Climate*, *19*(16), 3882–3901.
778 <https://doi.org/10.1175/JCLI3830.1>

- 779 Gray, L. J. (2010). Stratospheric Equatorial Dynamics. In *The Stratosphere: Dynamics, Transport, and Chemistry*
780 (pp. 93–107). American Geophysical Union (AGU). <https://doi.org/10.1002/9781118666630.ch5>
- 781 Gupta, A., Birner, T., Dörnbrack, A., & Polichtchouk, I. (2021). Importance of Gravity Wave Forcing for
782 Springtime Southern Polar Vortex Breakdown as Revealed by ERA5. *Geophysical Research Letters*,
783 *48*(10), e2021GL092762. <https://doi.org/10.1029/2021GL092762>
- 784 Holt, L. A., Lott, F., Garcia, R. R., Kiladis, G. N., Cheng, Y.-M., Anstey, J. A., Braesicke, P., Bushell, A. C.,
785 Butchart, N., Cagnazzo, C., Chen, C.-C., Chun, H.-Y., Kawatani, Y., Kerzenmacher, T., Kim, Y.-H.,
786 McLandress, C., Naoe, H., Osprey, S., Richter, J. H., ... Yukimoto, S. (2020). An evaluation of tropical
787 waves and wave forcing of the QBO in the QBOi models. *Quarterly Journal of the Royal Meteorological*
788 *Society*, *n/a*(*n/a*). <https://doi.org/10.1002/qj.3827>
- 789 Holton, J. R., & Lindzen, R. S. (1972). An Updated Theory for the Quasi-Biennial Cycle of the Tropical
790 Stratosphere. *Journal of the Atmospheric Sciences*, *29*(6), 1076–1080. [https://doi.org/10.1175/1520-0469\(1972\)029<1076:AUTFTQ>2.0.CO;2](https://doi.org/10.1175/1520-0469(1972)029<1076:AUTFTQ>2.0.CO;2)
- 791 Holton, J. R., & Tan, H.-C. (1980). The Influence of the Equatorial Quasi-Biennial Oscillation on the Global
792 Circulation at 50 mb. *Journal of the Atmospheric Sciences*, *37*(10), 2200–2208.
793 [https://doi.org/10.1175/1520-0469\(1980\)037<2200:TIOTEQ>2.0.CO;2](https://doi.org/10.1175/1520-0469(1980)037<2200:TIOTEQ>2.0.CO;2)
- 794 Howland, M. F., Dunbar, O. R. A., & Schneider, T. (2022). Parameter Uncertainty Quantification in an Idealized
795 GCM With a Seasonal Cycle. *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002735.
796 <https://doi.org/10.1029/2021MS002735>
- 797 Iglesias, M. A., Law, K. J. H., & Stuart, A. M. (2013). Ensemble Kalman methods for inverse problems. *Inverse*
798 *Problems*, *29*(4), 045001. <https://doi.org/10.1088/0266-5611/29/4/045001>
- 800 Jewtoukoff, V., Hertzog, A., Plougonven, R., Cámara, A. de la, & Lott, F. (2015). Comparison of Gravity Waves in
801 the Southern Hemisphere Derived from Balloon Observations and the ECMWF Analyses. *Journal of the*
802 *Atmospheric Sciences*, *72*(9), 3449–3468. <https://doi.org/10.1175/JAS-D-14-0324.1>
- 803 Jucker, M., & Gerber, E. P. (2017). Untangling the Annual Cycle of the Tropical Tropopause Layer with an
804 Idealized Moist Model. *Journal of Climate*, *30*(18), 7339–7358. <https://doi.org/10.1175/JCLI-D-17-0127.1>
- 805 Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical*
806 *Society: Series B (Statistical Methodology)*, *63*(3), 425–464. <https://doi.org/10.1111/1467-9868.00294>
- 807 Kodama, C., Ohno, T., Seiki, T., Yashiro, H., Noda, A. T., Nakano, M., Yamada, Y., Roh, W., Satoh, M., Nitta, T.,
808 Goto, D., Miura, H., Nasuno, T., Miyakawa, T., Chen, Y.-W., & Sugi, M. (2021). The Nonhydrostatic
809 ICosahedral Atmospheric Model for CMIP6 HighResMIP simulations (NICAM16-S): Experimental
810 design, model description, and impacts of model updates. *Geoscientific Model Development*, *14*(2), 795–
811 820. <https://doi.org/10.5194/gmd-14-795-2021>
- 812 Lee, L. A., Carslaw, K. S., Pringle, K. J., & Mann, G. W. (2012). Mapping the uncertainty in global CCN using
813 emulation. *Atmospheric Chemistry and Physics*, *12*(20), 9739–9751. <https://doi.org/10.5194/acp-12-9739-2012>
- 814 Lindzen, R. S. (1981). Turbulence and stress owing to gravity wave and tidal breakdown. *Journal of Geophysical*
815 *Research*, *86*(C10), 9707. <https://doi.org/10.1029/JC086iC10p09707>
- 816 Lindzen, R. S., & Holton, J. R. (1968). A Theory of the Quasi-Biennial Oscillation. *Journal of the Atmospheric*
817 *Sciences*, *25*(6), 1095–1107. [https://doi.org/10.1175/1520-0469\(1968\)025<1095:ATOTQB>2.0.CO;2](https://doi.org/10.1175/1520-0469(1968)025<1095:ATOTQB>2.0.CO;2)
- 818 Matsuoka, D., Watanabe, S., Sato, K., Kawazoe, S., Yu, W., & Easterbrook, S. (2020). Application of Deep
819 Learning to Estimate Atmospheric Gravity Wave Parameters in Reanalysis Data Sets. *Geophysical*
820 *Research Letters*, *47*(19), e2020GL089436. <https://doi.org/10.1029/2020GL089436>
- 821 Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of State
822 Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, *21*(6), 1087–1092.
823 <https://doi.org/10.1063/1.1699114>
- 824 Molod, A., Takacs, L., Suarez, M., Bacmeister, J., Song, I.-S., & Eichmann, A. (2012). *The GEOS-5 Atmospheric*
825 *General Circulation Model: Mean Climate and Development from MERRA to Fortuna*
826 (GSFC.TM.01153.2012). <https://ntrs.nasa.gov/citations/20120011790>
- 827 Murphy, J. M., Booth, B. B. B., Boulton, C. A., Clark, R. T., Harris, G. R., Lowe, J. A., & Sexton, D. M. H. (2014).
828 Transient climate changes in a perturbed parameter ensemble of emissions-driven earth system model
829 simulations. *Climate Dynamics*, *43*(9–10), 2855–2885. <https://doi.org/10.1007/s00382-014-2097-5>
- 830 Pathak, R., Dasari, H. P., El Mohtar, S., Subramanian, A. C., Sahany, S., Mishra, S. K., Knio, O., & Hoteit, I.
831 (2021). Uncertainty Quantification and Bayesian Inference of Cloud Parameterization in the NCAR Single
832 Column Community Atmosphere Model (SCAM6). *Frontiers in Climate*, *3*.
833 <https://www.frontiersin.org/article/10.3389/fclim.2021.670740>
- 834

- 835 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P.,
836 Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine
837 Learning in Python. *MACHINE LEARNING IN PYTHON*, 6.
- 838 Plougonven, R., de la Cámara, A., Hertzog, A., & Lott, F. (2020). How does knowledge of atmospheric gravity
839 waves guide their parameterizations? *Quarterly Journal of the Royal Meteorological Society*, 146(728),
840 1529–1543. <https://doi.org/10.1002/qj.3732>
- 841 Plumb, R. A. (1977). The Interaction of Two Internal Waves with the Mean Flow: Implications for the Theory of the
842 Quasi-Biennial Oscillation. *Journal of the Atmospheric Sciences*, 34(12), 1847–1858.
843 [https://doi.org/10.1175/1520-0469\(1977\)034<1847:TIOIWI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1977)034<1847:TIOIWI>2.0.CO;2)
- 844 Priestley, M. D. K., Ackerley, D., Catto, J. L., Hodges, K. I., McDonald, R. E., & Lee, R. W. (2020). An Overview
845 of the Extratropical Storm Tracks in CMIP6 Historical Simulations. *Journal of Climate*, 33(15), 6315–
846 6343. <https://doi.org/10.1175/JCLI-D-19-0928.1>
- 847 Rao, J., Garfinkel, C. I., & White, I. P. (2020). How Does the Quasi-Biennial Oscillation Affect the Boreal Winter
848 Tropospheric Circulation in CMIP5/6 Models? *Journal of Climate*, 33(20), 8975–8996.
849 <https://doi.org/10.1175/JCLI-D-20-0024.1>
- 850 Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press.
- 851 Richter, I., & Tokinaga, H. (2020). An overview of the performance of CMIP6 models in the tropical Atlantic: Mean
852 state, variability, and remote impacts. *Climate Dynamics*, 55(9), 2579–2601.
853 <https://doi.org/10.1007/s00382-020-05409-w>
- 854 Richter, J. H., Anstey, J. A., Butchart, N., Kawatani, Y., Meehl, G. A., Osprey, S., & Simpson, I. R. (2020).
855 Progress in Simulating the Quasi-Biennial Oscillation in CMIP Models. *Journal of Geophysical Research:*
856 *Atmospheres*, 125(8), e2019JD032362. <https://doi.org/10.1029/2019JD032362>
- 857 Roberts, G. O., & Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probability*
858 *Surveys*, 1(none), 20–71. <https://doi.org/10.1214/154957804100000024>
- 859 Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance based sensitivity
860 analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics*
861 *Communications*, 181(2), 259–270. <https://doi.org/10.1016/j.cpc.2009.09.018>
- 862 Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2007).
863 *Global Sensitivity Analysis. The Primer*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470725184>
- 864 Scaife, A. A., Butchart, N., Warner, C. D., & Swinbank, R. (2002). Impact of a Spectral Gravity Wave
865 Parameterization on the Stratosphere in the Met Office Unified Model. *Journal of the Atmospheric*
866 *Sciences*, 59(9), 1473–1489. [https://doi.org/10.1175/1520-0469\(2002\)059<1473:IOASGW>2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059<1473:IOASGW>2.0.CO;2)
- 867 Schenzinger, V., Osprey, S., Gray, L., & Butchart, N. (2017). Defining metrics of the Quasi-Biennial Oscillation in
868 global climate models. *Geoscientific Model Development*, 10(6), 2157–2168. <https://doi.org/10.5194/gmd-10-2157-2017>
- 870 Schirber, S., Manzini, E., Krismer, T., & Giorgetta, M. (2015). The quasi-biennial oscillation in a warmer climate:
871 Sensitivity to different gravity wave parameterizations. *Climate Dynamics*, 45(3), 825–836.
872 <https://doi.org/10.1007/s00382-014-2314-2>
- 873 Scinocca, J. F. (2003). An Accurate Spectral Nonorographic Gravity Wave Drag Parameterization for General
874 Circulation Models. *Journal of the Atmospheric Sciences*, 60(4), 667–682. [https://doi.org/10.1175/1520-0469\(2003\)060<0667:AASNGW>2.0.CO;2](https://doi.org/10.1175/1520-0469(2003)060<0667:AASNGW>2.0.CO;2)
- 876 Siskind, D., Eckermann, S. D., Coy, L., McCormack, J. P., & Randall, C. E. (2007). On recent interannual
877 variability of the Arctic winter mesosphere: Implications for tracer descent: MESOSPHERIC
878 INTERANNUAL VARIABILITY. *Geophysical Research Letters*, 34(9).
879 <https://doi.org/10.1029/2007GL029293>
- 880 Siskind, D., Eckermann, S., McCormack, J., Coy, L., Hoppel, K., & Baker, N. (2010). Case studies of the
881 mesospheric response to recent minor, major, and extended stratospheric warmings. *J. Geophys. Res.*, 115,
882 0–3. <https://doi.org/10.1029/2010JD014114>
- 883 Sobol', I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates.
884 *Mathematics and Computers in Simulation*, 55(1–3), 271–280. [https://doi.org/10.1016/S0378-4754\(00\)00270-6](https://doi.org/10.1016/S0378-4754(00)00270-6)
- 886 Souza, A. N., Wagner, G. L., Ramadhan, A., Allen, B., Churavy, V., Schloss, J., Campin, J., Hill, C., Edelman, A.,
887 Marshall, J., Flierl, G., & Ferrari, R. (2020). Uncertainty Quantification of Ocean Parameterizations:
888 Application to the K-Profile-Parameterization for Penetrative Convection. *Journal of Advances in*
889 *Modeling Earth Systems*, 12(12), e2020MS002108. <https://doi.org/10.1029/2020MS002108>

- 890 Strahan, S. E., Oman, L. D., Douglass, A. R., & Coy, L. (2015). Modulation of Antarctic vortex composition by the
891 quasi-biennial oscillation. *Geophysical Research Letters*, *42*(10), 4216–4223.
892 <https://doi.org/10.1002/2015GL063759>
- 893 Wan, H., Rasch, P. J., Zhang, K., Qian, Y., Yan, H., & Zhao, C. (2014). Short ensembles: An efficient method for
894 discerning climate-relevant sensitivities in atmospheric general circulation models. *Geoscientific Model*
895 *Development*, *7*(5), 1961–1977. <https://doi.org/10.5194/gmd-7-1961-2014>
- 896 Warner, C. D., & McIntyre, M. E. (1999). Toward an ultra-simple spectral gravity wave parameterization for
897 general circulation models. *Earth, Planets and Space*, *51*(7), 475–484. <https://doi.org/10.1186/BF03353209>
- 898 Williamson, D., Blaker, A. T., & Sinha, B. (2016). *Tuning without over-tuning: Parametric uncertainty*
899 *quantification for the NEMO ocean model* [Preprint]. *Climate and Earth System Modeling*.
900 <https://doi.org/10.5194/gmd-2016-185>
- 901 Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L., & Yamazaki, K. (2013). History
902 matching for exploring and reducing climate model parameter space using observations and a large
903 perturbed physics ensemble. *Climate Dynamics*, *41*(7–8), 1703–1729. [https://doi.org/10.1007/s00382-013-](https://doi.org/10.1007/s00382-013-1896-4)
904 [1896-4](https://doi.org/10.1007/s00382-013-1896-4)
- 905 Wright, C. J., Osprey, S. M., Barnett, J. J., Gray, L. J., & Gille, J. C. (2010). High Resolution Dynamics Limb
906 Sounder measurements of gravity wave activity in the 2006 Arctic stratosphere. *Journal of Geophysical*
907 *Research: Atmospheres*, *115*(D2). <https://doi.org/10.1029/2009JD011858>
- 908 Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., Chen, J.-H., Chen, X., Donner, L. J., Dunne, J.
909 P., Dunne, K., Durachta, J., Fan, S.-M., Freidenreich, S. M., Garner, S. T., Ginoux, P., Harris, L. M.,
910 Horowitz, L. W., Krasting, J. P., ... Xiang, B. (2018). The GFDL Global Atmosphere and Land Model
911 AM4.0/LM4.0: 2. Model Description, Sensitivity Studies, and Tuning Strategies. *Journal of Advances in*
912 *Modeling Earth Systems*, *10*(3), 735–769. <https://doi.org/10.1002/2017MS001209>
- 913

**Calibration and Uncertainty Quantification of Gravity Wave
Parameterization in an Intermediate Complexity Climate Model**

L. A. Mansfield¹ and A. Sheshadri¹

¹Department of Earth System Science, Stanford University, Stanford, CA, USA

Corresponding author: Laura A. Mansfield (lauraman@stanford.edu)

Contents of this file

Figures S1 to S2

Additional Supporting Information (Files uploaded separately)

Captions for Movies S1 to S2

Introduction

The Supporting Information includes Supplementary Figures showing the Quasi-Biennial Oscillation in used to estimate parametric uncertainty in this study and a Supplementary Movie showing the Markov chain Monte Carlo sampler exploring the posterior distribution.

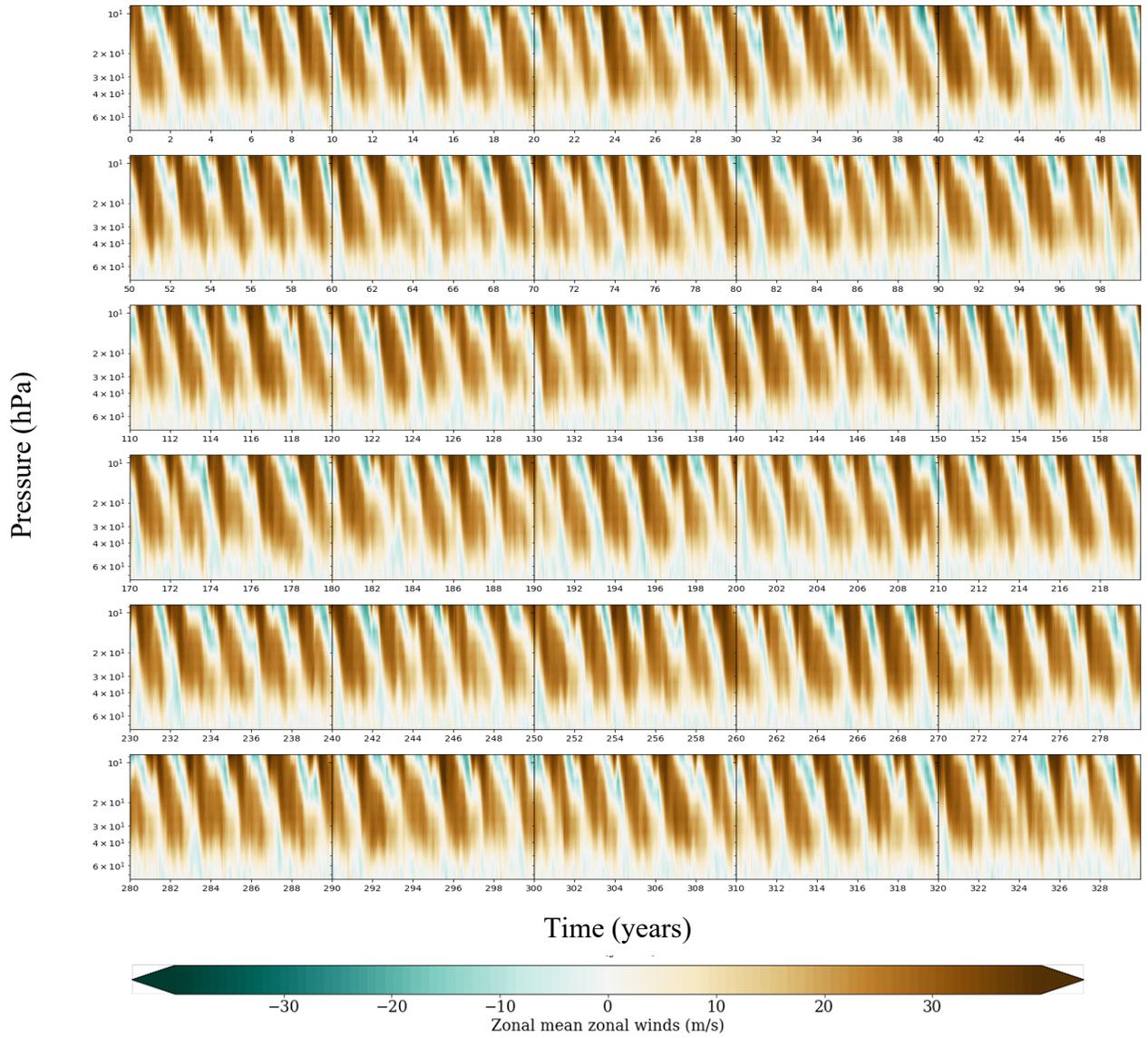


Figure S1. Time-height cross section of zonal mean zonal winds between 5°S and 5°N showing the QBO under a $2\times\text{CO}_2$ forcing for a 300 year simulation, where gravity wave parameters are set to their control values, $c_w = 35$ m/s and $Bt_{eq} = 0.0043$ Pa. The simulation is initialized from a spun-up state.

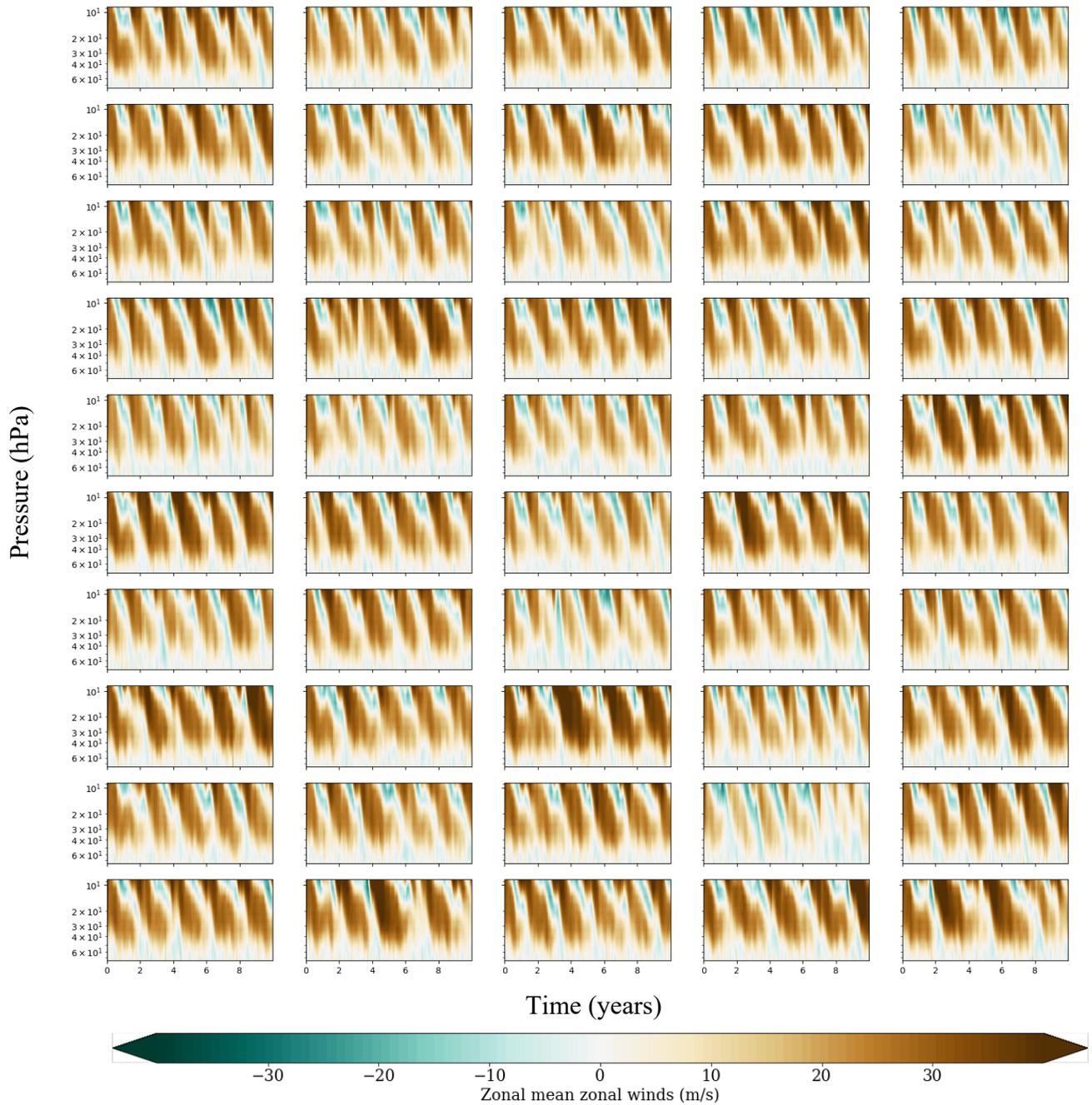


Figure S2. Time-height cross section of zonal mean zonal winds between 5°S and 5°N showing the QBO under a 2xCO₂ forcing for 50 ensemble members, where gravity wave parameters are sampled from the posterior distribution in Figure 6. Each ensemble member is initialized from a spun-up state and run for 10 years. Periods and amplitudes of each cycle within all ensemble members are used to generate Figure 7 of main text and to estimate the parametric uncertainty.

Additional Supporting Information (Files uploaded separately)

mov01.gif

Movie S1. Animation of MCMC samples from posterior distribution overlaid on Gaussian process predictions over parameter range (Figure 4 of main text) for perfect model setting.