Emergent Constraints on CMIP6 Climate Warming Projections: Contrasting Cloud- and Surface Temperature–Based Constraints

YONGXIAO LIANG, a NATHAN P. GILLETTE, b AND ADAM H. MONAHAN a

a School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia, Canada
b Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, Victoria, British Columbia, Canada

(Manuscript received 23 June 2021, in final form 26 November 2021)

ABSTRACT: The latest Coupled Model Intercomparison Project (CMIP6) multimodel ensemble shows a broader range of projected warming than the previous-generation CMIP5 ensemble. We show that the projected warming is well correlated with tropical and subtropical low-level cloud properties. These physically meaningful relations enable us to use observed cloud properties to constrain future climate warming. We develop multivariate linear regression models with metrics selected from a set of potential constraints based on a stepwise selection approach. The resulting linear regression model using two low-cloud metrics shows better cross-validated results than regression models that use single metrics as constraints. Application of a regression model using the low-cloud metrics to climate projections results in similar estimates of the mean, but substantially narrower uncertainty ranges, of projected twenty-first-century warming when compared with unconstrained simulations. The resulting projected global-mean warming in 2081–2100 relative to 1995–2014 is 2.84–5.12 K (5%–95% range) for Shared Socioeconomic Pathway (SSP) 5–8.5 compared with a range of 2.34–5.18 K for unconstrained projections, and 0.60–1.70 K for SSP1–2.6 compared to an unconstrained range of 0.38–2.04 K. We provide evidence for a higher lower bound of the projected warming range than that obtained from constrained projections based on the past global-mean temperature trend. Consideration of the impact of the sea surface temperature pattern effect on the recent observed warming trend, which is not well captured in the CMIP6 ensemble, indicates that the relatively low projected warming resulting from the global-mean temperature trend constraint may not be reliable and provides further justification for the use of climatologically based cloud metrics to constrain projections.

KEYWORDS: Climate change; Climate models; Climate variability; Cloud radiative effects

1. Introduction

While multimodel mean global warming projections for comparable scenarios are similar for CMIP6 and CMIP5, some climate models in the current-generation CMIP6 ensemble project large twenty-first-century warming and exhibit high climate sensitivity (Gettelman et al. 2019; Sellar et al. 2019; Swart et al. 2019; Voldoire et al. 2019) that is outside of the range of comparable CMIP5 projections (Meehl et al. 2020). Due to the emergence of more and higher climate sensitivity models, the CMIP6 multimodel ensemble simulates a larger spread of projected warming than CMIP5 (Nijsse et al. 2020; Tokarska et al. 2020). High levels of projected twenty-first-century warming in CMIP6 models are associated with high equilibrium climate sensitivity (ECS), although there is some scatter in the relationship (Tokarska et al. 2020). Model differences in simulated changes of shortwave reflection by low-level clouds (particularly in the tropics and midlatitudes) in response to climate change have been found to dominate the uncertainties in global warming projections in past intercomparisons (Brient and Schneider 2016; Vial et al. 2013). For example, Brient and Schneider (2016) point out that most of the variance of CMIP5 models’ ECS can be explained by model differences in low-cloud–induced shortwave reflection over tropical oceans. Recent studies provide evidence that changes to the representation of cloud process in CMIP6 models result in simulations in better agreement with satellite datasets in the midlatitudes than prior generations of models (Myers et al. 2021). However, there is a broad spread in cloud feedbacks in CMIP6 resulting in a broad range of climate sensitivity (Schlund et al. 2020; Zelinka et al. 2020). Statistical methods combining observations and model simulations are an effective approach to constraining projected warming and climate sensitivity. For example, Sherwood et al. (2020) consider multiple lines of evidence, including historical warming and the paleoclimate record as well as process understanding of feedbacks, to estimate the effective climate sensitivity using a Bayesian framework. Ribes et al. (2021) use climate models to provide a prior on the forced response and use the observational record to obtain its posterior distribution. Hattab et al. (2019) use a principal component regression method and discuss the selection of a robust set of observable predictors to estimate climate sensitivity.

The application of emergent constraints is an effective way to narrow the projected warming spread (Bretherton and Caldwell 2020; Nijsse et al. 2020; Tokarska et al. 2020). An emergent relationship between an observable quantity in the historical or present period and a quantity related to the
future climate (e.g., twenty-first-century warming) can serve as the basis to constrain projections. The emergent relationship is usually motivated by physical understanding of a process driving climate feedbacks, and observational estimates must exist to distinguish models with a realistic representation of the process from those with a less realistic representation (Bretherton and Caldwell 2020; Meehl et al. 2020; Tokarska et al. 2020). For an emergent constraint to be robust, a clear physical mechanism is required: statistical analyses in Caldwell et al. (2014) show that large correlations across models between past climate variables and future projections can occur by chance. Considering 11 metrics to constrain ECS from both CMIP5 and CMIP6 simulations, Schlund et al. (2020) find that most cloud metrics first identified using CMIP5 data show a weaker emergent constraint in CMIP6. Other lines of evidence further demonstrate that not all of these metrics are robust. In particular, Caldwell et al. (2018) use a feedback decomposition analysis to evaluate the performance of several emergent constraints on ECS and show that only 4 of 19 cloud constraints on ECS proposed in the literature can be considered credible.

Emergent constraints using certain cloud diagnostics have resulted in higher climate sensitivity than those obtained from unconstrained ensembles. Zhai et al. (2015) constrain ECS using the sensitivity of extratropical low cloud fraction to the seasonal cycle of sea surface temperature (SST), showing that the relatively high climate sensitivity models in CMIP3 and CMIP5 are more consistent with observed values of this metric than models with relatively low climate sensitivity. By combining information from several cloud-related metrics, Bretherton and Caldwell (2020) predict a larger constrained mean of ECS in CMIP5 than that of the raw ensemble.

Although higher or medium climate sensitivity models in CMIP5 and CMIP6 are generally found to be in better agreement with observational constraints using cloud-related metrics, recent analyses using the historical global mean surface air temperature warming trend as a constraint find the high-sensitivity models to be less consistent and suggest that they might overestimate the future warming trend (Liang et al. 2020; Nijsse et al. 2020; Tokarska et al. 2020). The preferential weighting of low warming projections when using the recent warming trend as a constraint is at least partially result of the so-called sea surface temperature pattern effect (associated with warming in the western equatorial Pacific Ocean and cooling in the eastern equatorial Pacific Ocean since about 1980), which has been shown to have a strong effect on the observed global mean near-surface air temperature (GSAT) trend in recent decades (Andrews et al. 2018; Dong et al. 2020; Gregory et al. 2020; Zhou et al. 2016, 2021). An increase in the zonal gradient of SST across the low latitudes of the Pacific Ocean has resulted in more low cloud coverage over the eastern part of the basin and a global-scale cooling effect due to the increased reflection of incoming shortwave radiation. As this pattern effect of east–west tropical Pacific SST gradient is more likely a result of internal variability rather than a long-term warming response, this cooling pattern is not expected to persist (Forster et al. 2021; Watanabe et al. 2021). The recent Assessment Report of the Intergovernmental Panel on Climate Change assigns a medium confidence of observed changes in the pattern effect resulting from internal variability (Forster et al. 2021). For the future changes of this pattern effect, there is medium confidence that the observed strengthening of the east–west SST gradient is temporary and will transition to a weakening of the SST gradient on centennial time scales. This SST pattern is not captured well by the CMIP5 or CMIP6 ensembles (Olonscheck et al. 2020). Compared with CMIP5, Olonscheck et al. (2020) found that the use of much larger initial condition ensembles of CMIP6 models does capture the observed cooling pattern, indicating that the absence of the pattern effect in the CMIP5 and CMIP6 ensembles can be partly interpreted as a sampling bias. Therefore, the relatively low observed trend potentially induced by internal variability will favor low climate sensitivity models and may result in spuriously low warming projections when applying the observed GSAT trend as a constraint.

While past studies have focused on the use of cloud metrics to constrain ECS (Bretherton and Caldwell 2020; Brient and Schneider 2016; Brient et al. 2016; Caldwell et al. 2018; Qu et al. 2014; Zhai et al. 2015), it is unclear how effectively cloud metrics constrain the transient projected climate warming, which is most relevant for adaptation and mitigation planning. This fact motivates us to investigate the use of cloud metrics to constrain CMIP6 projections of twenty-first-century warming under various SSP scenarios.

A number of issues regarding the development and implementation of emergent constraints remain outstanding. Although previous studies provide evidence of clear relationships between physically based cloud metrics and projected warming, the relative performance of cloud metrics as constraints compared with historical warming as a constraint is unclear. In particular, the relative impacts of the internal variability on the uncertainty range of constrained projections are unclear for different categories of metrics. Furthermore, as models in the CMIP6 archive share components the simulations they produce may not be statistically independent. The impact of this potential dependence needs to be addressed when assessing the performance of the metrics. Finally, different methods for applying constraints have been proposed. For example, some studies weight models based on their performance compared to observations (Brumner et al. 2019, 2020; Knutti et al. 2017; Lorenz et al. 2018; Sanderson et al. 2015a,b, 2017) while others use linear regression–based approaches (Cox et al. 2018; Hall et al. 2019; Nijsse et al. 2020; Schlund et al. 2020; Thackeray and Hall 2019; Tokarska et al. 2020). A direct comparison of these two approaches is needed.

To address these questions, in this paper we contrast the application of cloud metrics with the application of the global surface air temperature trend as emergent constraints on CMIP6 projected warming, and apply a multivariate linear regression model to make observationally constrained projections of twenty-first-century warming. We first apply a stepwise approach to select the most effective linear regression model from a subset of physically based metrics. The selected linear model is then evaluated in a cross-validated imperfect model test. We then use the linear regression model to constrain projected twenty-first-century GSAT changes from
<table>
<thead>
<tr>
<th>Model name</th>
<th>Historical</th>
<th>SSP1–2.6</th>
<th>SSP5–8.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCESS-CM2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ACCESS-ESM1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BCC-CSM2-MR</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CAMS-CSM1-0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>CanESM5</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>CESM2</td>
<td>6</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>CESM2-WACCM</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CNRM-CM6-1</td>
<td>10</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>CNRM-CM6-1-HR</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CNRM-ESM2-1</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>EC-Earth3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>FGOALS-e3-L</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FGOALS-g3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GISS-E2-1-G</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>HadGEM3-GC31-LL</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IPSL-CM6a-LR</td>
<td>16</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>KACE-1-0-G</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>MIROC-ES2L</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MIROC6</td>
<td>9</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>MPI-ESM1-2-HR</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MPI-ESM1-2-LR</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>MRI-ESM2-0</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>NEMO3</td>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>NorESM2-LM</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>NorESM2-MM</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>UKESM1-0-LL</td>
<td>8</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

CMIP6 simulations. Potential model dependence is taken into account during both processes of metric selection and warming projection constraint, and different approaches to applying the constraints (weighting vs linear regression) are directly compared.

2. Data and methods

a. Model simulations

We use output from 26 global climate models participating in CMIP6 (Table 1). We calculate the various metrics considered (section 2b) using historical simulations and then use observations with the emergent constraint approach to predict GSAT changes in 2081–2100 relative to 1995–2014 under SSP5-8.5 and SSP1-2.6 scenarios (O’Neill et al. 2016; Riahi et al. 2017). For those models which contributed initial condition ensembles, we consider all ensemble members individually (we sample one random realization per model 10,000 times, as outlined in section 2d) in order to assess the contribution of internal variability, and to not bias results toward models with particularly large ensembles.

b. Emergent constraint metrics considered

We consider five potential metrics for constraining GSAT projections: four cloud-related diagnostics and the historical GSAT trend (see appendix D). Caldwell et al. (2018) considered 19 cloud-based metrics proposed in the literature, and found that four have both clear physical connections with cloud feedbacks and significant correlations with ECS in CMIP5. Here, we consider these four cloud-related metrics as potential emergent constraints on twenty-first-century CMIP6 warming projections. Descriptions of these cloud metrics follow.

1) MARINE BOUNDARY LAYER CLOUD METRIC

The sensitivity of monthly marine boundary layer cloud (MBLC) fraction to the seasonal cycle of SST between 20° and 40° latitude in the Southern and Northern Hemispheres is well correlated with ECS across CMIP3 and CMIP5 models (Zhai et al. 2015). The fraction of MBLC is defined as the low cloud coverage (below 700 hPa) in subsidence regions (indicated by monthly climatologies of 500-hPa vertical velocity) over oceans between 20° and 40° latitude in both hemispheres with a random-overlap assumption (Manabe and Strickler 1964; Ramanathan et al. 1983; Stephens 1984). The MBLC metric is defined as the regression slope in

\[
\text{MBLC} = \alpha \left( \text{SST} \right) + \beta, \quad \text{mon} = 1, 2, \ldots, 12.
\]  

The overbar and angle brackets in Eq. (1) represent the monthly climatology and spatial average, respectively. This definition of the metric is similar that of Qu et al. (2014) but uses a region that is farther poleward, and focuses on monthly climatologies rather than interannual variations to reduce the uncertainty from internal variability. With future climate warming, an enhanced vertical humidity gradient between the marine boundary layer (MBL) and the free troposphere in subsidence regions can lead to greater in-cloud buoyancy and stronger production of turbulent kinetic energy in the MBL. This above process results in a weaker temperature inversion in the MBL, which can decrease low cloud formation (Bretherton et al. 2004; Rieck et al. 2012; Sherwood et al. 2014; Vial et al. 2016). As described in Zhai et al. (2015), the sensitivity of MBLC fraction to seasonal SST changes is similar to its sensitivity to centennial SST changes. The different model sensitivities of MBLC fraction to SST changes on centennial time scales account for much of the intermodel variation of climate sensitivity. Furthermore, as mentioned in section 1, the larger shortwave cloud feedback in the CMIP6 ensemble relative to the CMIP5 ensemble on average can be attributed to the changes in the simulation of midlatitude, mixed-phase clouds (especially in the Southern Hemisphere; Zelinka et al. 2020). These facts provide the motivation for applying the MBLC metric as an observational constraint. For the observed value of the MBLC metric, we use the value reported by Zhai et al. (2015) of \(-1.28% \pm 0.19% K^{-1}\) (mean \(\pm 1\) standard deviation), estimated over the period 2006–10.

2) BRIENT CLOUD ALBEDO METRIC

The Brient cloud albedo (BCA) metric measures the sensitivity of the anomalies of shortwave cloud albedo to SST changes over the tropical oceans (Brient and Schneider 2016).
The regions considered are the monthly-varying driest quartile of ocean grid cells between 30°S and 30°N, based on 500-hPa relative humidity. In climate models, the sensitivity of the tropical low cloud albedo to the underlying surface temperature in the present-day climate correlates with the strength of the shortwave tropical low cloud feedback and with future projected warming. Thus, the sensitivity of the variation of tropical low cloud reflection to changes of SST in the present climate can be used as an emergent constraint on future GSAT changes (Brient and Schneider 2016). A significant correlation between this metric and ECS has been found in CMIP6 simulations (Schlund et al. 2020).

3) LOWER-TROPOSPHERIC MIXING INDEX

Sherwood et al. (2014) developed three metrics to measure lower-tropospheric mixing: the Sherwood S metric, the Sherwood D metric, and the lower-tropospheric mixing index (LTMI), all of which are correlated with CMIP5 projected warming. The Sherwood S metric quantifies climatological small-scale mixing in the tropical lower free troposphere. The Sherwood D metric quantifies the large-scale mixing over the tropical lower troposphere. LTMI is defined as the sum of Sherwood D and S metrics (Sherwood et al. 2014). Climate models with stronger vertical moisture mixing in the lower troposphere tend to have a larger increase of moisture mixing with climate warming, which could decrease the boundary layer clouds because of the stronger convective drying under climate change (Sherwood et al. 2014). Sherwood et al. (2014) demonstrated that the S, D, and LTMI metrics are correlated with equilibrium climate sensitivity across climate models because of this link to low cloud feedbacks. Schlund et al. (2020) consider all three metrics as potential constraints, and show that only LTMI is significantly correlated with ECS in both CMIP5 and CMIP6 simulations. Therefore, in our study, we take LTMI as a potential constraint.

4) BRIENT CLOUD SHALLOWNESS METRIC

Brient et al. (2016) introduced the Brient cloud shallowness (BCS) metric of cloud shallowness:

$$\gamma = \frac{CF_{950}}{CF_{850} + CF_{950}},$$

which is defined in terms of the cloud fractions below 900 hPa ($CF_{950}$) and below 800 hPa ($CF_{850}$) over weakly subsiding tropical ocean regions (vertical velocity between 10 and 30 hPa day$^{-1}$). Following Brient et al. (2016), $CF_{950}$ is obtained by the mass-weighted cloud fractions between 1000 and 900 hPa (between 900 and 800 hPa for $CF_{850}$).

Models that have shallower clouds over weakly subsiding tropical regions (a large BCS) in the historical period tend to have more influence by convective drying of the planetary boundary layer relative to turbulent moistening under climate warming, which further decreases low cloud cover and leads to a larger positive low cloud feedback. The detailed physical mechanism is rather complicated, involving different partially cancelling effects, and is described in detail in Brient et al. (2016). Further evidence of the relation of BCS to cloud feedbacks comes from a sensitivity test considering the lateral entrainment rate of shallow convection in MPI-ESM (Mauritsen and Roeckner 2020). Mauritsen and Roeckner (2020) found that weak lateral entrainment rates lead to more stratiform clouds within the boundary layer (larger BCS), and that convection-induced drying leads to a stronger reduction of low cloud fraction under climate warming and a higher equilibrium climate sensitivity. As our observational BCS value, we use the value based on CALIPSO/GOCCP (GCM Oriented Cloud CALIPSO Product) data reported by Bretherton and Caldwell (2020): 45 ± 3% over 2006–12.

5) GSAT TREND METRIC

Past warming simulated by climate models is well correlated with future climate warming, especially for the 1970–2014 period (Brunner et al. 2020; Jiménez-de-la-Cuesta and Mauritsen 2019; Liang et al. 2020; Nijssse et al. 2020; Tokarska et al. 2020). The high correlation between the GSAT trend (GT) over 1970–2014 and projected warming emerges because of the small change of aerosol forcing relative to the dominant contribution from greenhouse gases over this period. In our analysis, we use HadCruT5 (Morice et al. 2021), which is spatially infilled, to compute the observational temperature trend (0.018 ± 0.001 K yr$^{-1}$, estimated over the period 1970–2014. The quoted observational uncertainty is
the 5%–95% range across the 200-member HadCRUT5 ensemble).

c. Linear regression and stepwise metric selection

Our aim is to integrate information from all the metrics we consider constraining projected warming based on CMIP6 models. Our primary method for doing this uses multivariate linear regression with stepwise selection (Senftleben et al. 2020), described in detail in appendix A. To avoid using multiple metrics describing related processes in a linear regression model, we use a stepwise regression method to select a subset of metrics. A risk of using all possible metrics in a single linear regression model is the possibility of overfitting resulting from spurious relations between historical metrics and future projections (Bracegirdle and Stephenson 2012). We apply an iterative approach combining forward selection with backward elimination to build a multiple regression model that can best project future GSAT changes (von Storch and Zwiers 1999). Specifically, this approach adds and eliminates variables iteratively to the linear regression model, stopping when the explained sum of squares (ESS) does not change significantly based on an F-test (with significance level $p = 0.1$). An alternative model weighting approach (Brunner et al. 2019, 2020; Knutti et al. 2017; Lorenz et al. 2018; Sanderson et al. 2015a,b, 2017) to constraining the ensemble is presented in appendix B.

d. Sampling from initial condition ensembles and uncertain observational values

About half of the models participating in CMIP6 provide ensembles of multiple initial-condition realizations. While all available ensemble members could be used together to derive an observationally constrained distribution, this approach

---

**Fig. 2.** A flowchart of the stepwise regression procedure using cloud metrics. For each step, the corresponding statistics are shown in Fig. 3.
would give us no information about the influence of intra-ensemble variability on our results. As well, if we were to follow this approach, those models with larger initial condition ensembles would have a stronger influence on our results. Therefore, when applying the linear regression model or weighting approach, we generate the constrained projections using one randomly selected ensemble member per model. To evaluate the influence of internal variability in the historical simulations on constrained projections, we repeat this process 10,000 times. This sampling approach is applied to all aspects of this study, including the imperfect model test and probabilistic validation (both described in section 2e), and observationally constrained projections. For observationally constrained projections, along with random sampling from the initial condition ensembles, we also sample the observed quantities \( X_0 \) in Eq. (A2) within their uncertainty ranges (assuming Gaussian distributions with means and standard deviations quoted above). For each random selection of ensemble members and each realization of the observations, we calculate a PDF of future GSAT change. We then average these PDFs together to calculate an overall PDF, from which we calculate a 5–95% range of future warming (appendix A).

The random selection of individual model realizations is important for our analysis because internal variability contributes considerable uncertainty to observational constraints, especially for metrics based on trends (Liang et al. 2020). The CMIP6 archive provides much larger initial condition ensembles than CMIP5, allowing us to better estimate the
contribution of internal variability to projected uncertainty. Note that this sampling process can only partly account for internal variability because approximately half the models used only have a single ensemble member. To estimate the effect of undersampling internal variability, we artificially reduce the number of models with multiple ensemble members in observational constraints to test whether projection statistics are convergent as more multiple ensemble models are included. We find that for our main analysis using the full CMIP6 ensemble, the effects of internal variability may be underestimated in observational constraints using the GSAT trend metric, but this is not the case when applying climatologically based cloud metrics (see Text S3 and Fig. S5 in the online supplemental material).

e. Imperfect model test

We apply a cross-validated imperfect model test to assess the performance of the emergent constraint approaches, comparing the linear regression and weighting approaches, as well as the metrics used in these approaches (Brunner et al. 2020; Liang et al. 2020). We first choose one model to act as pseudo-observations (the "truth" in the imperfect model test context), and then apply emergent constraint approaches with all remaining models to predict this truth. This procedure is repeated taking each model in turn as truth. We use probabilistic validation of the imperfect model test to assess the uncertainty estimates resulting from the emergent constraints. Across all models, we note the relative frequency with which pseudo-observations lie in each quintile of the imperfect model constrained projection (0%–20%, 20%–40%, etc.). If the constraining approach provides well-calibrated uncertainty estimates, the relative frequency within each quintile should be close to 0.2 (Liang et al. 2020). We also calculate the coverage frequency, defined as the percentage of pseudo-observations lying in the constrained 5%–95% predicted uncertainty range. If the uncertainty estimates work well, close to 90% of pseudo-observations should fall in the 5%–95% predicted uncertainty range.

There is an implicit assumption underlying our approach that cloud feedbacks not directly constrained in our study are not systematically biased in CMIP6 models. As will be demonstrated in section 3d, the agreement of constrained historical GSAT temperature evolution with observations provides evidence that the effects of any such systematic bias are limited.

3. Results

a. Metric performance

Based on the ECS results of previous studies discussed in section 2b, we take the BCS, BCA, LTMI, and MBLC metrics as our potential cloud-related constraints. Since the GSAT trend metric has been widely considered as an observational constraint on projections, we also compare constrained projections obtained solely from the cloud constraints with those using both the cloud constraints and the GT. We now evaluate how well these observable metrics are correlated with twenty-first-century warming using CMIP6 simulations. As shown in Fig. 1, projected warming is significantly correlated with BCA, MBLC, and GT, such that the $p$ values are smaller than 0.05 across the internal variability sampling distribution (section 2d). In contrast, the LTMI and BCS correlations are weaker, with $p$ values larger than $p = 0.05$. These results are similar to those of Schlund et al. (2020), who found that LTMI does not provide a strong emergent constraint for ECS in CMIP6 (in contrast to CMIP5). The spread of correlation coefficients is narrower for BCS, LTMI, and MBLC than BCA or GT metrics. The larger spread in BCA and GT reflects a larger contribution of internal variability in contrast to the other metrics.

b. Stepwise regression

We now investigate which combination of metrics produces the most effective emergent constraint. We first construct a
regression model using only cloud metrics. The stepwise regression approach we use adds (or removes) the most (or the least) important term in the linear regression model at each step, based on the results of $F$ tests (Figs. 2 and 3). For each step that requires calculation of an $F$ statistic (Figs. 3b–d), we account for the effects of model dependence by using 20 as the effective number of independent models (appendix C). Following the flowchart outlining the stepwise procedure (Fig. 2), we first build a single variable linear model with the MBLC metric because of all cloud metrics it produces the largest ESS value (Fig. 3a). In step 2, we build a two-variable linear model adding the BCS metric because this is the only cloud metric that results in a significant increase in ESS relative to the MBLC metric regression model. As shown in Fig. 3b, the lower 5th percentile of the $F$ statistic range for MBLC + BCS versus MBLC is greater than the critical $F$ value, which indicates a significant improvement to the linear regression model using the MBLC and BCS metrics relative to the MBLC metric only. For the other two choices, MBLC + BCA versus MBLC and MBLC + LTMI versus MBLC, the upper 95th percentile of the $F$ statistic is smaller than the critical $F$ value, indicating that these two combinations should not be considered further.

In step 3 of the procedure (Fig. 3c), we carry out a backward selection step by removing the MBLC metric. Estimates of the resulting $F$ statistic are always significantly larger than the critical $F$ values (irrespective of internal variability or changes in the effective number of degrees of freedom; appendix C), indicating that the model with both MBLC and BCS metrics has significantly larger ESS compared with the linear model including BCS only. Therefore, the MBLC and BCS metrics are retained in the linear model in step 3.

Three-variable linear models are considered in step 4. There is no significant improvement in the fit of the combination MBLC + BCS + LTMI or MBLC + BCS + BCA relative to MBLC + BCS. The upper 95th percentile of the $F$ statistic in MBLC + BCS + LTMI versus MBLC + BCS and the upper 90th percentile of the $F$ statistic in MBLC + BCS + BCA versus MBLC + BCS are smaller than the critical $F$ value. While the internal variability induced range of the MBLC + BCS + BCA versus MBLC + BCS $F$ statistic crosses the critical $F$ value, the median value is much smaller. Hence, all three-variable linear models fail to increase the ESS significantly compared with the model using MBLC and BCS metrics (Fig. 3d). Results from a sensitivity test addressing potential model dependence by varying the degrees of freedom to used to compute the $F$ statistic (appendix C; see Fig. S3) obtain the same set of stepwise selected metrics. Therefore, we use the MBLC and BCS metrics as constraints in our cloud metric based multiple diagnostic linear regression model to predict the GSAT changes in 2081–2100 under SSP5–8.5. We also carried out the stepwise selection on SSP1–2.6 (not shown) and get the same selected metrics as for SSP5–8.5.

Since the GT metric has been widely applied as an observational constraint in previous studies, we repeat the previous analysis including the GT metric in our stepwise regression (Fig. S4). The resulting regression model uses the MBLC and GT metrics as constraints. A schematic of the cloud metrics entering our final regression model using only cloud metrics. The stepwise regression approach we use adds (or removes) the most (or the least) important term in the linear regression model at each step, based on the results of $F$ tests (Figs. 2 and 3). For each step that requires calculation of an $F$ statistic (Figs. 3b–d), we account for the effects of model dependence by using 20 as the effective number of independent models (appendix C). Following the flowchart outlining the stepwise procedure (Fig. 2), we first build a single variable linear model with the MBLC metric because of all cloud metrics it produces the largest ESS value (Fig. 3a). In step 2, we build a two-variable linear model adding the BCS metric because this is the only cloud metric that results in a significant increase in ESS relative to the MBLC metric regression model. As shown in Fig. 3b, the lower 5th percentile of the $F$ statistic range for MBLC + BCS versus MBLC is greater than the critical $F$ value, which indicates a significant improvement to the linear regression model using the MBLC and BCS metrics relative to the MBLC metric only. For the other two choices, MBLC + BCA versus MBLC and MBLC + LTMI versus MBLC, the upper 95th percentile of the $F$ statistic is smaller than the critical $F$ value, indicating that these two combinations should not be considered further.

In step 3 of the procedure (Fig. 3c), we carry out a backward selection step by removing the MBLC metric. Estimates of the resulting $F$ statistic are always significantly larger than the critical $F$ values (irrespective of internal variability or changes in the effective number of degrees of freedom; appendix C), indicating that the model with both MBLC and BCS metrics has significantly larger ESS compared with the linear model including BCS only. Therefore, the MBLC and BCS metrics are retained in the linear model in step 3.

Three-variable linear models are considered in step 4. There is no significant improvement in the fit of the combination MBLC + BCS + LTMI or MBLC + BCS + BCA relative to MBLC + BCS. The upper 95th percentile of the $F$ statistic in MBLC + BCS + LTMI versus MBLC + BCS and the upper 90th percentile of the $F$ statistic in MBLC + BCS + BCA versus MBLC + BCS are smaller than the critical $F$ value. While the internal variability induced range of the MBLC + BCS + BCA versus MBLC + BCS $F$ statistic crosses the critical $F$ value, the median value is much smaller. Hence, all three-variable linear models fail to increase the ESS significantly compared with the model using MBLC and BCS metrics (Fig. 3d). Results from a sensitivity test addressing potential model dependence by varying the degrees of freedom to used to compute the $F$ statistic (appendix C; see Fig. S3) obtain the same set of stepwise selected metrics. Therefore, we use the MBLC and BCS metrics as constraints in our cloud metric based multiple diagnostic linear regression model to predict the GSAT changes in 2081–2100 under SSP5–8.5. We also carried out the stepwise selection on SSP1–2.6 (not shown) and get the same selected metrics as for SSP5–8.5.

Since the GT metric has been widely applied as an observational constraint in previous studies, we repeat the previous analysis including the GT metric in our stepwise regression (Fig. S4). The resulting regression model uses the MBLC and GT metrics as constraints. A schematic of the cloud metrics entering our final

FIG. 5. Scatterplots showing relationships between selected constraints and projected warming: the (a) GT, (b) MBLC (x axis reversed), and (c) BCS metrics. For illustration, one ensemble member per model is used. The correlation coefficients and $p$ values (relative to a null hypothesis of no correlation) are reported in the bottom right corner of each panel. The vertical lines show the observational values with means in solid and standard deviation in shadow. The dashed lines in each panel show the 66% confidence interval of the linear regression model [Eqs. (A3)–(A5)].
Although we do consider the use of GT, our main focus in the subsequent is on the use of the two metrics MBLC and BCS in our constraining process, since these two metrics are less influenced by internal variability than GSAT (also see section 3c) and result in the linear regression model with best predictive power.

To further illustrate the statistical relationship between the metrics considered and projected GSAT changes in the CMIP6 simulations, Fig. 5 shows scatterplots of each of the GT, MBLC, and BCS metrics with late twenty-first-century warming. These scatterplots each use only a single, randomly determined ensemble member for each model to calculate the metric and the projected warming. Figure 5 also presents the observed values of the metrics (with uncertainty ranges). For the MBLC metric, models near the center of the range of simulated GSAT changes are in best agreement with observations. These results suggest that applying the MBLC metric will not shift the center of the distribution much relative to the unconstrained ensemble, consistent with previous studies of constraints on ECS (Bretherton and Caldwell 2020; Caldwell et al. 2014; Schlund et al. 2020; Sherwood et al. 2014). In contrast, observed values of the GT and BCS metrics are toward the low end of the range of simulated values. This fact is also in agreement with previous studies (Bretherton and Caldwell 2020; Brient et al. 2016; Nijssse et al. 2020; Tokarska et al. 2020).

c. Imperfect model evaluation of constrained warming

Before presenting results of projected warming constrained by observations, we evaluate the performance of the emergent constraint approach in an imperfect model setting (section 2e), based on values of the root-mean-square error (RMSE) improvement (relative to the unconstrained ensemble) and correlation coefficient ($r$), both calculated using the pseudo-observations and the means of the constrained imperfect model ensemble (Fig. 6). We present linear regression results for each of the metrics separately, as well as linear regression and weighting results for the two pairs of metrics resulting from the stepwise selection procedure. The linear regression approach performs better than the weighting approach, resulting in larger values of both the correlation coefficient and the RMSE reduction (Fig. 6).

Consistent with the results of the stepwise selection procedure, the linear regression models using only a single metric do not result in constraints as effective as those using two metrics (with the exception of GT, which has median correlation and RMSE reduction values similar to the combination of MBLC and BCS metrics). On average, the constrained projections based on the combination of MBLC and GSAT metrics performs slightly better than those based on the combined MBLC and BCS metrics. However, the
uncertainty ranges (from the initial condition sampling) of both correlation coefficient and RMSE improvement are much wider for the combined MBLC and GSAT metrics than for the combined MBLC and BCS metrics, reflecting the substantially larger internal variability in GSAT estimates. As the slight improvement in performance of the combination of the MBLC and GSAT metrics comes with a substantial increase in uncertainty and risk of biases from using the GSAT metric due to poor representation of the pattern effect in CMIP6 models, these results suggest that the more robust constraints based on the cloud-based MBLC and BCS metrics should be preferred. Similar to SSP5–8.5 described above, the robustness of the MBLC and BCS metrics was determined for SSP1–2.6 using an imperfect model test (not shown).

To test whether the linear regression method with the selected metrics provides reliable uncertainty estimates, we conduct a probabilistic validation (section 2e). Figure 7a shows the results of this analysis for 2081–2100 under SSP5–8.5 with the uncertainty estimated by assuming Gaussian regression residuals (appendix A). Also shown in Fig. 7a are the results for the unconstrained ensemble. Similar to the unconstrained projection, the MBLC and BCS metric linear regression model produces relative frequencies of approximately 0.2 in each quintile (Fig. 7a). As shown in Fig. 7b, the average fraction of pseudo-observations lying within the 5%–95% uncertainty range predicted by linear regression with the assumption of Gaussian residuals is close to 90%, and the range of this fraction due to sampling internal variability is narrow. These results indicate that the assumption of Gaussian regression residuals results in an accurate coverage probability.

d. Observational constraints

Based on the imperfect model analyses presented above, we now apply the observed metrics to constrain warming projections using the linear regression approach assuming Gaussian residuals. As described in section 2d, we account for internal variability and observational uncertainty by constructing the regression models using one randomly selected ensemble member per model (Fig. 8) with one random realization of the observations sampled from the estimated distribution and repeating this process 10,000 times. The undersampling of internal variability resulting from the fact that only 12 models of SSP 5–8.5 have multiple ensemble members available is potentially important for GT but negligible for climatologically based cloud metrics (section 2d). Throughout these calculations we use the value 20 as the effective number of independent models in the CMIP6 ensemble (appendix C).

Before applying the observational constraint to future climate change, we investigate how the metrics we are considering constrain historical warming in the CMIP6 models (Fig. 9). Both the GT metric and the MBLC–BCS metric pair result in constrained historical warming consistent with observations, such that the observed GSAT always falls within the constrained 5%–95% uncertainty range. This result demonstrates that the linear regression approach with the selected metrics can capture the historical warming, and increases confidence in the constrained future warming results. This is further evidence that the constrained distribution using the MBLC and BCS metrics is the absence of a systematic bias relative to observations.

Constraining projections of twenty-first-century warming under SSP5–8.5 using the GT metric alone in the linear
The regression model produces lower values of the mean and 5th percentile of the warming distribution than either the unconstrained estimate or the constrained estimate using cloud metrics (Fig. 10 and Table 2; Liang et al. 2020; Nijsse et al. 2020; Tokarska et al. 2020). The 5%–95% uncertainty ranges of both sets of constrained projections are narrower than that of the unconstrained ensemble.

Relative to the unconstrained ensemble, the constrained projections using the cloud metrics result in an increase in the 5th percentile of warming and a decrease in the 95th percentile, with little effect on the mean. Specifically, with the cloud metric weighting, the unconstrained 5%–95% uncertainty range of warming in 2081–2100 relative to 1995–2014 under SSP5–8.5 of 2.34–5.81 K is narrowed to 2.84–5.12 K. Assuming that the models are independent slightly decreases this range (see Text S1, Table 2, and Table S1 in the online supplemental material).

Similar results are found for the projection of twenty-first-century warming under the SSP1–2.6 scenario (Table 2; see also Table S1). Weighting by the cloud metrics reduces the 5%–95% uncertainty range from 0.38–2.04 to 0.60–1.70 K, and reduces the mean from 1.30 to 1.15 K.

4. Summary and discussion

Previous studies have demonstrated the existence of physically meaningful relationships between equilibrium climate sensitivity and low-level cloud metrics across CMIP5 and CMIP6 models (Brient and Schneider 2016; Brient et al. 2016; Caldwell et al. 2016, 2018; Schlund et al. 2020; Zelinka et al. 2020; Zhai et al. 2015). These relationships enable us to constrain future warming using observed values of these metrics in a complementary approach to the use of the observed GSAT trend as a constraint. We have applied two cloud metrics, a marine boundary layer cloud (MBLC) metric (Zhai et al. 2015) and Brient cloud shallowness (Brient et al. 2016), obtained from a larger set of physically motivated cloud metrics using a stepwise selection process, to constrain future warming under two different scenarios, SSP1–2.6 and SSP5–8.5. In developing the emergent constraints, we have compared two different approaches: linear regression (Cox et al. 2018; Karpechko et al. 2013; Nijsse et al. 2020; Senftleben et al. 2020; Tokarska et al. 2020) and a weighting method (Brunner et al. 2019, 2020; Knutti et al. 2017; Lorenz et al. 2018; Sanderson et al. 2015a,b, 2017). Using a cross-validated imperfect model test across available CMIP6 models, we find that for the problem considered the linear regression approach produces more effective constraints than the weighting approach.

The cloud metrics we use to constrain future warming have less uncertainty, which might be due to the reduced effect of internal variability than the GSAT trend often used to constrain warming projections (Liang et al. 2020; Nijsse et al. 2020; Tokarska et al. 2020). As a result, the robustness of constrained projections based on cloud metrics is improved compared to those based on the GSAT trend metric. Furthermore, the SST pattern effect (Andrews et al. 2018; Watanabe et al. 2021; Zhou et al. 2016, 2021) is known to affect climate sensitivity but is not captured well in the CMIP6 archive, perhaps due to undersampling of internal variability (Forster et al. 2021; Olonscheck et al. 2020).
These considerations support the use of the cloud metric rather than the GSAT trend for constraining future warming.

We account for observational uncertainty and internal variability in our analysis by sampling from the estimated distribution of observational uncertainty and sampling individual members from initial condition ensembles when constructing observationally constrained projections. Applying the multiple observed cloud metrics as constraints to future GSAT changes, we find that for both the SSP1–2.6 and SSP5–8.5 scenarios the projected warming uncertainty ranges are considerably narrower relative to unconstrained simulations, with little change in mean warming (Fig. 10; see also Fig. S7). We also find that observationally constrained projections using climatological cloud metrics have substantially reduced prediction uncertainty associated with internal variability in historical simulations relative to constrained projections using the GSAT trend. Furthermore, our study provides evidence for increasing the lower bound of the warming range of CMIP6 projections, as well as lowering the upper bound. This result differs from constrained projections based on the GSAT trend alone, which exhibit a substantial decrease in the upper bound and the mean of the projection range, but little change in the lower bound (Brunner et al. 2020; Caldwell et al. 2018; Liang et al. 2020; Nijssen et al. 2020; Tokarska et al. 2020).

Our study provides a framework to apply multiple metrics to constrain future warming that is also appropriate for constrained projections of equilibrium climate sensitivity. Our results imply that the mean climate sensitivity of the CMIP6 ensemble may not in fact be biased high as some studies have suggested, and that uncertainties in projected warming can be considerably narrowed using physically reasonable cloud constraints.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Projected warming (5%–95%; units: K)</th>
<th>SSP1–2.6</th>
<th>SSP5–8.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconstrained</td>
<td>4.08 (2.34, 5.81)</td>
<td>1.30 (0.38, 2.04)</td>
<td></td>
</tr>
<tr>
<td>GT</td>
<td>3.76 (2.42, 5.11)</td>
<td>1.09 (0.33, 1.85)</td>
<td></td>
</tr>
<tr>
<td>MBLC+BCS</td>
<td>3.99 (2.84, 5.12)</td>
<td>1.15 (0.60, 1.70)</td>
<td></td>
</tr>
</tbody>
</table>
Acknowledgments. We acknowledge the World Climate Research Programme’s Working Group on the modeling of Sixth Coupled Model Intercomparison Project listing in Table 1. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC; Grant RGPIN-2017-04043) and the China Scholarship Council (CSC; Grant 202008180022).

Data availability statement. The data used in this study are openly available. The CMIP6 simulation data are available at https://esgf-node.llnl.gov/projects/cmip6/, and the HadCRUT5 dataset that we compute the observational temperature trend is available at https://www.metoffice.gov.uk/hadobs/hadcrut5/.

APPENDIX A

Linear Regression Method

The regression model used to implement the emergent constraints of future GSAT changes is

\[ y = \alpha + X^T \beta, \] (A1)

where \( y \) is the vector of modeled future GSAT change (each element corresponding to a different model), the vector \( \alpha \) and the vector \( \beta \) are the multiple regression parameters, and the metrics form the matrix \( X \) (\( X^T \) is the transpose of \( X \), the matrix that includes all metrics for all models). The number of rows in \( X^T \) is the number of climate models we use; the number of columns is the number of metrics we select. We estimate \( \alpha \) and \( \beta \) as \( \hat{\alpha} \) and \( \hat{\beta} \) using ordinary least squares regression, with \( X \) and \( y \) taken from CMIP6 models (using single randomly drawn members from each model with an initial condition ensemble; section 2d). Applying observational estimates \( X_0 \) of the corresponding selected metrics (sampled from the observational uncertainty) with the linear regression model yields a multidiagnostic constraint \( \hat{y}_0 \) (Karpechko et al. 2013; Senftleben et al. 2020):

\[ \hat{y}_0 = \hat{\alpha} + X_0^T \hat{\beta}. \] (A2)

If we assume Gaussian residuals as in previous studies (Schlund et al. 2020; Tokarska et al. 2020), the probability density function (PDF) for projected GSAT changes (\( y \)) given observation \( X_0 \) is

\[ p(y|X_0) = \frac{1}{\sqrt{2\pi\sigma^2_y}} \exp \left[ -\frac{(y - \hat{y}_0)^2}{2\sigma^2_y} \right]. \] (A3)

where

\[ \sigma^2_y = s^2 \left[ 1 + X_0^T (X^T X)^{-1} X_0 \right] \] (A4)

\[ s^2 = \frac{1}{n-p-n} \sum_{m=1}^M (y_m - \hat{y}_m)^2. \] (A5)

The derivation of Eq. (A4) is provided in Hooper and Zellner (1961). In Eq. (A5), \( M \) is the number of models and \( p \) denotes the number of metrics included in the regression; \( n \) is equal to the value of 26 when we take the full number of models; otherwise, \( n \) is equal to the value of 20 by accounting for model dependence (see appendix C).

As described in section 2d, we assess the effect of internal variability by taking one random selection of one ensemble member per model and simultaneously drawing one random choice of each observed metric from our estimate of the observed distribution. We repeat this process 10,000 times, and derive 10,000 estimates of the PDF of projected warming (shown as shading in Fig. 10). We next demonstrate that the sample mean of these PDFs (shown as solid lines in Fig. 10) is an estimate of the population estimate of the PDF of projected warming.

When our study samples over initial condition ensembles, each sample gives us a distribution of projected GSAT change [noted as \( x \) in Eq. (A6)] with a particular mean (\( \mu \)) and standard deviation (\( \sigma \)). Hence, multiplying the joint distribution of these two statistics \( f(\mu, \sigma) \), by a conditional distribution \( f(x|\mu, \sigma) \), and then integrating over \( \mu \) and \( \sigma \), gives us a population estimate \( p(x) \) of the marginal PDF of projected GSAT changes:

\[ p(x) = \int \int f(x|\mu, \sigma) f(\mu, \sigma) \, d\mu \, d\sigma. \] (A6)

Based on our sampling strategy, each derived PDF with its corresponding value of \( \mu \) and \( \sigma \) is equally probable. Sampling \( \mu \) and \( \sigma \) from their joint distribution and then averaging the resulting conditional distributions of \( x \) gives a sample estimate of this population mean distribution (solid curves of constrained PDFs).

APPENDIX B

Weighting Method

To relate our results to those of Liang et al. (2020), we consider a weighting method based on Sanderson et al. (2015a,b) and Knutti et al. (2017) as an alternative approach for deriving observationally constrained projections. We first define

\[ w_i = \frac{e^{-D_i/\sigma^2_i}}{1 + \sum_{j=1}^M e^{-S_{ij}/\sigma^2_i}}. \] (B1)

We then obtain \( W \), a vector of weights for each model, by normalizing \( w_i \) by the sum of \( w_i \). In Eq. (B1), \( M \) is the number of models considered, \( D_i \) is the difference between the simulated value of the metric from model \( i \) and the observed value, and \( S_{ij} \) is the difference between models \( i \) and \( j \) for the selected metric. When we weight models with multiple metrics, we calculate \( D_i \) and \( S_{ij} \) giving equal weights to each metric after normalizing each metric by the median across models. The parameter \( \sigma \) regulates the degree to which model similarities are penalized in the weights, and \( \sigma_0 \) regulates the effect of model performance on the weights (Knutti et al. 2017; Lorenz et al. 2018).
We calculate $\sigma_t$ and $\sigma_D$ based on approaches proposed by Knutti et al. (2017), Lorenz et al. (2018), and Brunner et al. (2019) (see Texts S1 and S2 in the online supplemental material). In contrast to the linear regression approach, this approach weights model projections based not only on their goodness of fit to historical observations but also based on similarity between models. The model weights obtained using Eq. (B1) not only yield the constrained multimodel mean, but also are used to weight models to obtain a multimodel distribution for the constrained uncertainty range (section 2e).

**APPENDIX C**

**Model Dependence**

Models from different institutions may share model components, and some individual modeling centers produce multiple closely related model versions. As a result, the individual members of the multimodel ensemble are not expected to be independent. This interdependence is accounted for in the weighting approach by downweighting models that are similar in their historical simulations. For the linear regression approach, the number of independent models must be estimated in order to determine the number of degrees of freedom in the regression. To this end, we assess model dependence using the “model genealogy” method (Masson and Knutti 2011). To build up the CMIP6 model genealogy, we apply a hierarchical clustering method (Brunner et al. 2020; Knutti et al. 2013; Masson and Knutti 2011) implemented by the linkage function in Python SciPy. The intermodel distance is defined as the area weighted mean square difference in near-surface air temperature averaged over 1970–2014 on a 1° grid between two models. We determine the number of independent models by comparing intermodel distances with inter-ensemble member distances in a model genealogy approach (Text S1, Figs. S1 and S2). Based on the results of this analysis, in sections 3b–3d the number of independent models considered in the linear regression approach is set to 20 rather than the full number of models. The sensitivity of the stepwise selection procedure to the number of statistical degrees of freedom used is presented in Figs. S3 and S4.

**APPENDIX D**

**Abbreviations for Trends and Metrics**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSAT</td>
<td>Global-mean near-surface air temperature</td>
</tr>
<tr>
<td>MBLC</td>
<td>The sensitivity of monthly marine boundary layer cloud fraction to SST changes between 20° and 40° latitude in the Southern and Northern Hemispheres</td>
</tr>
<tr>
<td>BCA</td>
<td>Brient cloud albedo metric, defined as the sensitivity of the deseasonalized shortwave cloud albedo to SST changes over the tropical oceans</td>
</tr>
<tr>
<td>LTM1</td>
<td>Lower-tropospheric mixing index, which measures the strength of lower-tropospheric mixing in ascending regions of the tropics</td>
</tr>
<tr>
<td>BCS</td>
<td>Brient cloud shallowness, defined in terms of the cloud fractions below 850 hPa (CF850) and below 950 hPa (CF950) over weakly subsiding tropical ocean regions</td>
</tr>
<tr>
<td>GT</td>
<td>Global-mean near-surface air temperature trend (GSAT trend)</td>
</tr>
</tbody>
</table>

**REFERENCES**


