The Impact of “Hot Models” on a CMIP6 Ensemble Used by Climate Service Providers in Canada: Do Global Constraints Lead to Appreciable Differences in Regional Projections?

ALEX J. CANNON

Climate Research Division, Environment and Climate Change Canada, Victoria, British Columbia, Canada

(Manuscript received 2 August 2023, in final form 19 December 2023, accepted 22 January 2024)

ABSTRACT: Canadian climate service providers offer projections from the Coupled Model Intercomparison Project (CMIP6) to help inform climate change mitigation and adaptation decisions. CMIP6 includes several “hot” climate models whose sensitivity to greenhouse gas forcings exceeds the likely range inferred from multiple lines of evidence. Global warming estimates assessed in the Sixth Assessment Report (AR6) of the Intergovernmental Panel on Climate Change (IPCC) were reduced by applying observational constraints on the historical rate of warming to the CMIP6 ensemble. This study assesses whether globally constrained CMIP6 projections for Canada are appreciably different from unconstrained projections. Two constraints are considered: one that removes models whose transient climate response lies outside the AR6 assessed range (TCRlikely), and the other that weights models to match the assessed distribution of equilibrium climate sensitivity (ECSall). Both constraints lead to appreciably cooler and drier projections than the unconstrained ensemble, with the strongest reductions seen in the upper end of the ensemble range, high-emissions scenario, end-of-century time period, and northern regions of Canada. In this case, constrained projections of annual mean temperature are 2–3°C cooler than the unconstrained projections, whereas projections of annual total precipitation are typically 20%–40% drier. Appreciable differences are also detected in the ensemble median of temperature extreme indices. Based on these results, it is recommended that a constrained ensemble be considered for regional projections to avoid the “hot model” problem. Alternatively, projections can be communicated conditional on a specified level of global warming, with global constraints then used to inform the timing of the warming level exceedance.

KEYWORDS: North America; Climate models; Ensembles; Internal variability; Trends; Climate services

1. Introduction

The mandate of Environment and Climate Change Canada’s (ECCC’s) Canadian Centre for Climate Services (CCCS) is to help Canadians increase their resilience to climate change. In collaboration with a group of regional climate service providers, including CLIMAtlantic, Ouranos, the Pacific Climate Impacts Consortium (PCIC), and the Prairie Climate Centre, CCCS supports delivery of future climate scenarios to Canadians through the ClimateData.ca online data portal. At the beginning of 2023, projections of climate variables for Canada were updated to include outputs from the latest phase of the Coupled Model Intercomparison Project (CMIP6). Data on the portal includes daily outputs from an ensemble of 26 climate models (Table 1) for each of three shared socioeconomic pathways (ScenarioMIP experiments for SSP1–2.6, low emissions; SSP2–4.5, moderate emissions; and SSP5–8.5, high emissions) that have been statistically bias adjusted and downscaled (Cannon et al. 2015; Werner and Cannon 2016; Hiebert et al. 2018) by PCIC, in collaboration with ECCC, onto an ~10 km grid over Canadian land areas. Downscaled data are used to calculate temperature and precipitation climate indices recommended by the Expert Team on Climate Change Detection and Indices (ETCCDI) (Zhang et al. 2011); a subset of these indices are summarized on the portal in terms of ensemble statistics (i.e., 10th, 50th, and 90th percentiles). As an example, the results of a query for a grid cell near Victoria, British Columbia, are shown in Fig. 1.

CMIP6 models form an “ensemble of opportunity,” which makes quantifying uncertainty in terms of ensemble statistics challenging (Tebaldi and Knutti 2007). Model results are contributed by modeling centers who are willing to participate. Hence, the ensemble is not designed to sample the known behavior and uncertainty of the climate system. This is exacerbated by systematic model biases in the historical mean state, which are removed on ClimateData.ca using statistical bias adjustment, as well as in the forced climate change signal. The latter bias is primarily due to unrealistic climate sensitivity (i.e., the warming response to external greenhouse gas forcing). Unlike the previous CMIP5 generation of models, CMIP6 includes several “hot models” (Hausfather et al. 2022) whose equilibrium climate sensitivities (ECSs) lie above the assessed likely range of Earth’s climate system (Sherwood et al. 2020; Meehl et al. 2020; Zelinka et al. 2020). Sobie et al. (2021) compared CMIP5 and CMIP6 projections of temperature and precipitation indices in regions of Canada. In general, CMIP6 models project larger changes in most temperature indices than CMIP5.
models, whereas greater precipitation changes in CMIP6 were found for extreme precipitation indices.

Recognizing the hot model problem, projections of global mean temperature change in the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) (Lee et al. 2021) were constrained using multiple lines of evidence, including observational constraints on the CMIP6 ensemble based on the rate of historical warming (Liang et al. 2020; Tokarska et al. 2020; Ribes et al. 2021), as well as assessed values of Earth’s ECS and transient climate response

<table>
<thead>
<tr>
<th>No.</th>
<th>Model</th>
<th>ECS (°C)</th>
<th>TCR (°C)</th>
<th>No.</th>
<th>Model</th>
<th>ECS (°C)</th>
<th>TCR (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>INM-CM4-8</td>
<td>1.83</td>
<td>–</td>
<td>14</td>
<td>CMCC-ESM2</td>
<td>3.58</td>
<td>1.92</td>
</tr>
<tr>
<td>2</td>
<td>INM-CM5-0</td>
<td>1.92</td>
<td>–</td>
<td>15</td>
<td>ACCESS-ESM1-5</td>
<td>3.88</td>
<td>1.97</td>
</tr>
<tr>
<td>3</td>
<td>NorESM2-MM</td>
<td>2.49</td>
<td>–</td>
<td>16</td>
<td>EC-Earth3</td>
<td>4.26</td>
<td>2.30</td>
</tr>
<tr>
<td>4</td>
<td>NorESM2-LM</td>
<td>2.56</td>
<td>1.49</td>
<td>17</td>
<td>EC-Earth3-Veg</td>
<td>4.33</td>
<td>2.66</td>
</tr>
<tr>
<td>5</td>
<td>MIROC6</td>
<td>2.60</td>
<td>1.55</td>
<td>18</td>
<td>TaoESM1</td>
<td>4.36</td>
<td>2.27</td>
</tr>
<tr>
<td>6</td>
<td>GFDL-ESM4</td>
<td>2.65</td>
<td>1.63</td>
<td>19</td>
<td>ACCESS-CM2</td>
<td>4.66</td>
<td>1.96</td>
</tr>
<tr>
<td>7</td>
<td>MIROC-ES2L</td>
<td>2.66</td>
<td>1.49</td>
<td>20</td>
<td>IPSL-CM6A-LR</td>
<td>4.70</td>
<td>2.35</td>
</tr>
<tr>
<td>8</td>
<td>FGOALS-g3</td>
<td>2.87</td>
<td>1.50</td>
<td>21</td>
<td>KACE-1.0-G</td>
<td>4.75</td>
<td>2.04</td>
</tr>
<tr>
<td>9</td>
<td>MPI-ESM1-2-HR</td>
<td>2.98</td>
<td>1.64</td>
<td>22</td>
<td>CNRM-ESM2-1</td>
<td>4.79</td>
<td>1.83</td>
</tr>
<tr>
<td>10</td>
<td>BCC-CSM2-MR</td>
<td>3.02</td>
<td>1.55</td>
<td>23</td>
<td>CNRM-CM6-1</td>
<td>4.90</td>
<td>2.22</td>
</tr>
<tr>
<td>11</td>
<td>MPI-ESM1-2-LR</td>
<td>3.03</td>
<td>1.82</td>
<td>24</td>
<td>UKEsim1-0-LL</td>
<td>5.36</td>
<td>2.77</td>
</tr>
<tr>
<td>12</td>
<td>MRI-ESM2-0</td>
<td>3.13</td>
<td>1.67</td>
<td>25</td>
<td>HadGEM3-GC31-LL</td>
<td>5.55</td>
<td>2.49</td>
</tr>
<tr>
<td>13</td>
<td>KIOST-ESM</td>
<td>3.36</td>
<td>1.90</td>
<td>26</td>
<td>CanESM5</td>
<td>5.64</td>
<td>2.71</td>
</tr>
</tbody>
</table>

Fig. 1. Results from a ClimateData.ca query for downscaled CMIP6 projections of mean temperature at a grid cell near Victoria, British Columbia, Canada. The lower and upper values of the range for each experiment are the 10th and 90th percentiles of the CanDCS-U6 ensemble, respectively.
(TCR). As a result, assessed ranges of global surface air temperature (GSAT) change in IPCC AR6 are lower than unconstrained ranges due to reduced influence of high-sensitivity models. The results on the global scale and the findings of Sobie et al. (2021) for Canada raise questions about how regional projections from the CMIP6 ensemble should be presented on ClimateData.ca. Are unconstrained projections robust and defensible? Should model projections instead be constrained or weighted? If so, how should constraints or weights be defined and applied?

Different viewpoints on these questions have been discussed in the literature. Hausfather et al. (2022) discourage the use of an ensemble of raw CMIP6 model outputs and recommend using a “warming level” approach (where analyses focus on the climate response to fixed levels of GSAT changes) or, in cases where the trajectory of warming matters, instead focusing on a constrained ensemble that removes models with a TCR that lies outside the assessed likely range. Because there is still an assessed 5% chance of ECS exceeding 5°C (Sherwood et al. 2020), one can then choose to focus on a separate subset of high-sensitivity models (i.e., those that lie in the low probability “hot tail” of the distribution). A counterpoint is provided by Bloch-Johnson et al. (2022), who argue that “models should be excluded only if compromised by physical errors. Because impact studies are strongly affected by the upper end of possibility, we call for guidance based on research into the physical reliability of ‘hot’ models.” The focus here is on the model constraint approach.

This study addresses two pragmatic questions related to these points. First, are there appreciable differences between unconstrained and constrained CMIP6 projections over Canada? Second, do different approaches to constraining projections lead to different results? To address these two questions, unconstrained projections in Canada, which are described in section 2, are compared against constrained projections based on two weighting schemes. As proposed by Hausfather et al. (2022), the first applies equal weights to models whose TCR lies within the assessed likely range; all other models are removed from the ensemble. In the second, model weights are optimized such that weighted quantiles of model ECS match quantiles of the assessed distribution of climate system ECS from Sherwood et al. (2020). Full details are provided in section 3. In section 4, mean temperature, total precipitation, and ETCCDI extreme indices are compared on national, regional, and local scales to see whether there are appreciable differences—those outside the range of internal variability—between the constrained and unconstrained projections. Issues related to equifinality and robustness of small ensembles are addressed in section 5. Finally, a summary and conclusions are presented in section 6.

2. Climate data

The ensemble of statistically downscaled CMIP6 climate projections used on ClimateData.ca was developed by PCIC with the support of ECCC. The underlying dataset, known as the Canadian Downscaled Climate Scenarios-Univariate method from CMIP6 (CanDCS-U6), is available from PCIC and ECCC. The 26 models included in the ensemble (Table 1) were selected from the entire CMIP6 ensemble based on practical considerations: projections of daily minimum temperature, maximum temperature, and precipitation had to be available for the historical experiment and three ScenarioMIP experiments (SSP1–2.6, low; SSP2–4.5, moderate; and SSP5–8.5, high) for the combined 1950–2100 period. A single realization of each climate model is available on ClimateData.ca. To allow the influence of internal variability on results to be quantified, an additional nine realizations of CanESM5 are also available from ECCC. Furthermore, ECCC provides ETCCDI climate extreme indices computed from CanDCS-U6 (ECCC 2022). The full set of 38 indices used in this study is listed in Table S1 in the online supplemental material.

Statistical downscaling of the raw climate model outputs—minimum temperature, maximum temperature, and precipitation—was performed using the Bias Correction/Constructed Analogs with Quantile delta mapping reordering (BCCAQv2) method (Werner and Cannon 2016; Hiebert et al. 2018). At the distributional level, the method adjusts climate model outputs so that biases in all quantiles of the empirical distribution, relative to the observational reference dataset of McKenney et al. (2011), are removed in the historical period. Following Cannon et al. (2015), the same adjustments are made to future projections subject to the added constraint that projected changes in quantiles are preserved. [Note: At the time of writing, a second set of downscaled CMIP6 outputs called CanDCS-M6, which is based on the multivariate algorithm of Cannon (2018), is being added to ClimateData.ca. Results here should apply equally to the new data, as the univariate distributions and change signals are the same in both CanDCS-U6 and CanDCS-M6.]

CanDCS-U6 is designed to provide data that are suitable for use as input to regional impact models or for the calculation of climate extreme indices. Systematic biases in the historical period are removed, which means that threshold-based indices or process representations that depend on absolute values better match historical observations. Furthermore, preservation of the forced response means that downscaled outputs are consistent with projected climate model anomalies at the large scale. However, if the forced response is unreliable (e.g., if the climate sensitivity is too low or high), then the resulting downscaled outputs will also be unreliable. The ECS and TCR values for the CanDCS-U6 models are shown in Table 1. Of the 26 models, 16 fall within the assessed likely range of TCR and just 12 within the likely ranges of both ECS and TCR. Figure 2a compares the empirical distribution of climate model ECS with the assessed probability density function of ECS from Sherwood et al. (2020). Note that Sherwood et al. (2020) provide an assessment of effective climate sensitivity, which is a linear approximation to the equilibrium warming defined by traditional measures of ECS; the two quantities are considered equivalent for the purposes of this study. In particular, there is an underrepresentation of climate models with an ECS around the mode of the assessed distribution (2–4°C) and an overrepresentation of models with high sensitivity (>4°C). With the assumption of exchangeability between model simulations and the real world, models weighted by assessed climate sensitivity allow us to obtain constrained future projections.
3. Model weighting

To mitigate the hot model problem, Hausfather et al. (2022) recommend constraining CMIP6 projections by calculating summary statistics from a reduced subset of climate models rather than the entire CMIP6 ensemble. Specifically, equal weights are given to models that fall within the assessed likely (66% likelihood range) of TCR (1.4°–2.2°C). All other models are assigned a weight of zero and removed from the ensemble. The resulting constrained ensemble, which includes 16 models, each with an equal weight of $1/16 = 0.0625$, is referred to here as TCRlikely. The impact of applying the TCRlikely constraint, which screens out one CanDCS-U6 model with low sensitivity and nine models with high sensitivity (see Table 1), on GSAT projections is shown in Fig. 2. The spread in GSAT is substantially reduced, especially at the upper end of the projection range. However, note that the empirical distribution of ECS in the constrained ensemble (Fig. 2b) is still quite different from the assessed distribution.

Other researchers have constrained GSAT model projections based on their agreement with observational constraints on the recent portion of the historical record. For example, the warming assessed in IPCC AR6 is based, in part, on three approaches that weight or constrain the models based on the realism of their rates of historical warming (Liang et al. 2020; Tokarska et al. 2020; Ribes et al. 2021). Alternatively, Gillett (2015) weighted models so that the resulting empirical distribution of TCR more closely matched an observationally constrained distribution from a detection and attribution analysis. Additional considerations, such as model dependence or other measures of model performance, can also be used to inform the choice of model weights (Brunner et al. 2020).

Although TCR, which focuses on the transient rather than equilibrium response of the climate system to external forcings, is thought to be more relevant than ECS for constraining decadal to centennial projections, some evidence suggests that ECS is a better predictor of climate model warming trends in the twenty-first century (Grose et al. 2018). To complement the TCRlikely screening approach, a second ECS-based method is used here to assign weights to climate models. The method is similar in principle to that of Gillett (2015), except that the goal is to match the assessed distribution of ECS rather than TCR.

Distributional matching is conducted by assigning weights to climate models so that weighted quantiles of the ensemble

---

**Fig. 2.** (a) The assessed probability density function of effective climate sensitivity (black line) from Sherwood et al. (2020) and the histogram of estimated ECS values (gray bars) for climate models in the unconstrained CanDCS-U6 ensemble used on ClimateData.ca. Individual model values are marked by asterisks (*) along the lower horizontal axis. Dark (light) orange shading shows the IPCC AR6 assessed likely (very likely) range of climate system ECS. (b) As in (a), except that the histogram is now based on equal weighting of models whose TCR lies within the IPCC AR6 assessed likely range. (c) Unconstrained ensemble medians (thin line) and 5th–95th-percentile ranges (thin lines) of 20-yr smoothed GSAT anomalies for historical (black), SSP1–2.6 (blue), SSP2–4.5 (orange), and SSP5–8.5 (red) experiment simulations. Anomalies are relative to 1850–1900 and match the IPCC AR6 definition of 0.85°C warming from preindustrial to 1995–2014. Error bars show the median and 5th–95th-percentile range at the end of the twenty-first century. (d) As in (c), but showing time series of GSAT anomalies for the constrained TCRlikely ensemble. (e) TCRlikely model weights.
of ECS values are as close as possible to quantiles from the assessed ECS distribution of Sherwood et al. (2020). Once weights have been optimized, they can be used to estimate other ensemble statistics, for example, specified percentiles of GSAT or regional climate extreme indices. An iterative quantile matching estimator (e.g., Delignette-Muller and Dutang 2015) is used here. The cost function $C$ to minimize is

$$C = \frac{1}{N_T} \sum_{t=1}^{N_T} (q_t - \hat{q}_t)^2 + \lambda \frac{1}{N} \sum_{i=1}^{N} \frac{w_i^2}{w_0^2},$$  \hspace{1cm} (1)$$

where $q_t$ is the $t$-quantile of the assessed ECS distribution, $T = \{0.025, 0.05, \ldots, 0.975\}$, $N_T$ is the number of elements in $T$, $\hat{q}_t$ is the weighted sample $t$-quantile of ECS values for the climate model ensemble, $w_i$ is the weight assigned to the $i$th of $N$ climate models, and $\lambda \geq 0$ and $w_0 = N^{-1}$ are hyperparameters that control the penalty assigned to nonzero weights; weights are constrained to be nonnegative and to sum to one via the softmax function.

A quasi-Newton optimization algorithm is used to adjust the model weights to minimize Eq. (1). The first term is a standard quantile matching cost function (Delignette-Muller and Dutang 2015), whereas the second is a penalty term that, for nonzero $\lambda$, exerts pressure to minimize the cardinality (i.e., the number of nonzero elements) of the set of ensemble weights (Weigend et al. 1991). The role of the penalty term in constructing alternative sets of weights for small subensembles of climate models will be discussed later in section 5. For now, $\lambda$ is set to zero and results are based on quantile matching without any explicit attempts to reduce the size of the ensemble.

This approach requires a consistent method for estimating unweighted and weighted sample quantiles. Unweighted quantiles reported on ClimateData.ca are calculated using the "type 7" plotting position algorithm of Hyndman and Fan (1996) as implemented in Bourgault et al. (2023); this method is adopted here. Similarly, type 7 weighted quantiles are estimated following Akinshin (2023). Notably, this estimator is consistent with the unweighted one and is stable (i.e., changes in the estimation are small when changes in the weights are small), which facilitates its use with a gradient-based iterative optimization algorithm.

Weights based on quantile matching, called ECSall, are shown in Fig. 3. The final set of ECSall weights is obtained by averaging values over 50 runs of the optimization algorithm, each starting from a different set of initial weights. Because there is uncertainty in the estimation of ECS from climate model simulations (Dunne et al. 2020), a small amount of Gaussian noise $\epsilon \sim N(0, 0.1)$ is applied to the ECS values at the beginning of each run. Multiple runs with random initialization and noise injection allow models with similar climate sensitivity to be weighted more equally; in

![Fig. 3](image-url)
any given trial, one model may be assigned more weight and the others less, but this uneven weighting is smoothed when averaging over multiple trials.

When ECSall weights are used to constrain the future evolution of GSAT anomalies, the projection range is, as expected, reduced, again most notably at the upper end (Fig. 3d). This reduction in warming is mainly driven by the reduced weighting of the three highest ECS models (HadGEM3-GC31-LL, UKESM1-0-LL, and CanESM5), as well as the group of models with ECS between 4°C and 5°C. By design, the empirical distribution of ECS in the constrained ensemble now closely matches the assessed distribution (Fig. 3b). Although it is challenging to evaluate the performance of the weighting methods, a comparison with temperature observations from Rohde and Hausfather (2020) suggests that the weighting of TCRlikely and ECSall improves the reliability of the raw CMIP6 ensemble over the historical period (Figs. S1 and S2). The evaluation shows that the weighting methods lead to narrower uncertainty range (Fig. S1), resulting in improved probabilistic calibration of the ensembles (Fig. S2). Constrained end-of-century projections of GSAT for TCRlikely, ECSall, and the constraints used in IPCC AR6 (Liang et al. 2020; Ribes et al. 2021; Tokarska et al. 2020) are compared in Fig. 4. When differences in the respective unconstrained ensembles are taken into consideration, the results are largely consistent between the constrained ensembles. The results from ECSall match more closely those from Liang et al. (2020), while the spread in the GSAT projections from TCRlikely, which completely omits information from the highest sensitivity models, is narrower than the others for the moderate (SSP2–4.5) and high (SSP5–8.5) emissions scenarios.

4. Constrained projections for Canada

TCRlikely and ECSall constraints reduce the impact of high-sensitivity CMIP6 climate models on projections of GSAT. Given that changes in many regional climate variables, including those in Canada (Sobie et al. 2021), scale with global warming (i.e., pattern scaling holds; Tebaldi and Arblaster 2014), it is hypothesized that the application of constraints inferred based on TCR and ECS will also lead to changes in Canadian climate scenarios. Pattern scaling provides the physical basis for the use of global metrics, such as ECS and TCR, to constrain the regional climate response. However, differences between constrained and unconstrained projections may be obscured by internal variability, since the relative magnitude of the forced signal tends to shrink with spatial scale (Fischer et al. 2013).

To test this hypothesis, constraints are applied to mean temperature, total precipitation, and ETCCDI extreme indices (Table S1) on national, regional, and grid cell scales, and differences are compared between constrained and unconstrained projections. The study domain is divided into five regions (Fig. 5) that are approximately the same size as the IPCC climate reference regions (Iturbide et al. 2020) but have been shifted northward to better distinguish the major climate zones of Canada. In particular, the three southern regions, which terminate at 50°N in IPCC AR6, are extended to 60°N; the two northern regions then cover latitudes north of 60°N. This division more closely reflects the climate regions used by Shephard et al. (2014) and other studies on the climate of Canada.

Noting the difficulty in assessing statistical significance from an ensemble of opportunity (von Storch and Zwiers 2012), determining whether differences are appreciable is instead based on their magnitude relative to internal variability.
Similar to Scinocca et al. (2016), the standardized difference $z$ between the constrained $\hat{x}_c$ and unconstrained $\hat{x}_u$ changes of a climate variable is given by

$$z = \frac{\hat{x}_c - \hat{x}_u}{\hat{s}},$$

where $\hat{s}$ is a measure of the spread in projected change due to internal variability. For CanDCS-U6, $\hat{s}$ is taken to be the standard deviation of projected changes in the 10 available downscaled realizations of CanESM5. It is assumed that CanESM5 accurately simulates natural variability and that its internal variability is equivalent to that of the other models. Under this definition, appreciable differences between constrained and unconstrained projections are more likely to be present when $|z| \gg 1$. For simplicity, two thresholds are used here: $|z| > 1.68$ (i.e., exceeding the 95th percentile of a standard normal distribution) and $|z| > 1.96$ (i.e., exceeding the 97.5th percentile).

Constrained and unconstrained projections of annual mean temperature and annual total precipitation for Canada as a whole are shown in Fig. 6; standardized differences are summarized in Fig. 7. For consistency with ClimateData.ca (Fig. 1), standardized differences are calculated for the 10th, 50th, and 90th percentiles of the ensemble projections; these values are reported for three time periods—near-term, midterm, and end of century—for low-, moderate-, and high-emissions scenarios.

For the 90th percentile of the temperature ensemble, TCRlikely projects appreciably cooler conditions ($z < -1.96$) than the unconstrained projections for eight of the nine combinations of scenario and time period. For ECSall, this drops to five of nine combinations, with appreciable differences seen in midterm and end-of-century projections for the high-emissions scenario only. At the end of century, the constrained ensemble median is also appreciably cooler for both TCRlikely and ECSall under the high-emissions scenario. For precipitation, appreciable differences are confined to the ensemble 90th percentile. Projections from ECSall and TCRlikely are appreciably drier at the end of century under the high ($z < -1.96$) and moderate emissions ($z < -1.68$) scenarios; the same pattern of differences extends to
TCRlikely for the midterm period, while the differences at this time for ECSall are limited to the moderate scenario ($z < -1.68$). Consistent with findings for GSAT, differences in Canada-wide projections of temperature and precipitation are more detectable for TCRlikely, which omits models with TCR $> 2.2^\circ$C, than for ECSall.

Figure 8 shows maps of the magnitude of differences in the ensemble 90th percentile between constrained and unconstrained projections under the strongest forcing (high-emissions scenario for end of century). For temperature, differences are appreciable across Canada, spatially uniform (constrained projections are typically 2$^\circ$–3$^\circ$C cooler than unconstrained), and consistent between methods, although TCRlikely is, on average, 0.2$^\circ$C cooler than ECSall. The spatial distribution of the precipitation differences is more heterogeneous. While constrained projections are typically drier (median relative reductions of 20%–40%) than unconstrained projections, with the exception of southeastern Canada (SEC), appreciable differences are mostly confined to northwestern Canada (NWC) and northeastern Canada (NEC) ($>70^\circ$N for ECSall and $>60^\circ$N for TCRlikely). The spatial patterns shown in Fig. 8 are reflected in mean results for the five regions (Figs. 9 and 10). Notably, appreciable differences between constrained and unconstrained temperature projections are more likely to be found in NWC and NEC than in the southern regions, especially for the ensemble median. This north–south asymmetry is also apparent for precipitation projections, where appreciable differences in the ensemble median for NWC and NEC can be found even under the moderate-emissions scenario for the midterm period. As noted before, differences are stronger for the TCRlikely constraint than for ECSall.

Results reported so far have focused on annual temperature and precipitation. Figure 11 extends the analysis to Canada-wide differences for the ETCCDI climate extreme indices (Table S1). In general, cooler and drier projections for the mean values are also reflected in the extreme indices. For example, the low end of the ensemble range for the number of icing days (id) and frost days (fd) increases in the cooler constrained projections, while the high end for the number of tropical nights (tr) and summer days (su) decreases. Similarly, the magnitude of precipitation extremes, such as annual total precipitation on days when daily precipitation is greater than the 99th percentile ($r_{99p}$), tends to be lower in the constrained projections; appreciable differences are more prevalent for the 90th percentile of the ensemble. Regionally, appreciable differences in precipitation extremes are more likely to be found in NWC and NEC than in the southern regions (Figs. S3–S7); unlike regional mean precipitation, however, appreciable differences in regional extremes are not detected for the ensemble median, likely due to increased internal variability at smaller spatial scales.

5. Reducing ensemble size

The TCRlikely and ECSall approaches to constraining the CanDCS-U6 ensemble differ in the way in which weights are assigned to climate models. TCRlikely assigns a weight of zero to some models, which leads to a reduced 16-member ensemble, while ECSall assigns nonzero weights to all members of the ensemble. From a practical perspective, a smaller ensemble may be desirable for impacts and adaptation studies that require substantial computational or time investments to complete. In
In many cases, reducing the size of the ensemble is a requirement when climate models are selected in an applied research context (Cannon 2015).

Although projections are similar for Canada, TCRlikely, which excludes the three highest ECS models, is somewhat cooler and drier than ECSall. If one were to construct an optimized ensemble of smaller size using the ECS matching approach (section 3), would the results be different from those for ECSall? This question raises the issue of equifinality. The goal of minimizing Eq. (1) is to match the assessed ECS distribution. Because independent climate models can have the same ECS, different sets of weights can ultimately lead to the same weighted ECS distribution. However, the projections of the weighted ensembles may differ due to structural differences between models that share the same ECS.

To investigate this possibility, three additional sets of weights are constructed using ECS matching but with the penalty term in Eq. (1) now exerting influence on the number of nonzero weights. An ensemble size of 10 is chosen to be on the same order as recent efforts aimed at the climate change impacts and adaptation community (Mahony et al. 2022). Optimization is carried out without added noise, and the values of $\lambda$ are adjusted so that 10 climate models contribute to each of the smaller ensembles (referred to as ECS10). The ECS10 variants differ primarily in terms of which high ECS model is included in the subset of 10 models (ECS10-CanESM5, ECS10-HadGEM3-CG31-LL, or ECS10-UKESM1-0-LL). The weighted distribution of ECS is effectively identical in each case (Table 1) and, on the global scale, the weighted projections of GSAT for ECS10 are similar to those for ECSall (see Figs. S1 and S2).

To assess whether this similarity extends to regional responses in Canada, standardized differences between constrained and unconstrained projections of the ETCCDI extreme indices are calculated for the ECS10 ensembles. As before, this is done for each region, percentile, SSP experiment, and time period. The sets of standardized differences $z$ for the three ECS10 variants are correlated with each other and with ECSall and TCRlikely. Results are shown in Fig. 12. As expected, very high correspondence (correlation $r = 0.95$) is found between ECS10-HadGEM3-CG31-LL, ECS10-UKESM1-0-LL, and ECSall. The other pairs of ECS10 ensembles are also highly correlated ($r > 0.9$), and the differences with ECSall are generally less than a few degrees Celsius.
CG31-LL and ECS10-UKESM1-0-LL; these two ensembles include models developed by the same center and share much of the same code (Kuma et al. 2023). The correlation between these variants and the ECS10 variant with CanESM5, which is from an independent family of climate models, is slightly lower ($r \approx 0.91$ and 0.92). ECS10-HadGEM3-CG31-LL and ECS10-UKESM1-0-LL also show greater consistency with ECSall ($r \approx 0.86$) than ECS10-CanESM5 ($r \approx 0.78$). In the latter case, the level of correspondence is slightly lower than that between ECSall and TCRlikely ($r \approx 0.8$).

Depending on the index, these differences can be regionally important. For example, Fig. S8 shows maps of the difference between constrained and unconstrained projections of the annual minimum value of the daily minimum temperature ($t_{mn}$) for ECS10-CanESM5 and ECS10-HadGEM3-CG31-LL at the end-of-century (90th percentile and high-emissions scenario). ECS10-HadGEM3-CG31-LL is appreciably cooler than the unconstrained ensemble across Canada (91% of grid cells), while this is the case in only half of grid cells for ECS10-CanESM5. In particular, ECS10-CanESM5 is consistent with the unconstrained ensemble in much of the NEC region.

6. Summary and conclusions

Global temperature changes assessed in IPCC AR6 were reduced by applying observational constraints to the CMIP6 ensemble to moderate the influence of “hot models.” It has been recommended that such constraints be applied to CMIP6 ensembles used by regional climate service providers (Hausfather et al. 2022), for example the CanDCS-U6 ensemble featured on ClimateData.ca. This study first assesses whether there are appreciable differences between unconstrained and constrained projections of temperature and precipitation over Canada; and second, whether different approaches to constraining projections lead to different results. Two constraints are considered: TCRlikely, which removes models whose TCR lies outside the AR6 assessed...
range, and ECSall, which weights models to match the assessed distribution of ECS. Temperature and precipitation indices are compared on national, regional, and local scales to see if there are appreciable differences, those outside the range of internal variability, between constrained and unconstrained projections.

The strongest appreciable differences, typically cooler and drier conditions in constrained ensembles, are found at the upper end of the ensemble range, in the high-emission scenario, in the end-of-century time period, and in northern regions of Canada. Depending on the variable, appreciable differences are also detectable in the ensemble median. TCRlikely, which completely omits information from models with the highest sensitivity, tends to be cooler and drier than ECSall. Furthermore, when the ECSall ensemble is reduced in size to 10 members, the results are sensitive to which high ECS model is included in the subset. For example, the degree of difference between ECS10-CanESM5 and ECSall is approximately the same as the difference between ECSall and TCRlikely. This highlights the role of structural model uncertainty in constrained projections.

While the TCRlikely and ECSall constraints explored in this study lead to broadly similar projections for Canada as a whole, remaining differences at the national and regional scales, as well as practical considerations, must also be taken into account. For example, the equal weight/subset approach of TCRlikely is attractive for its simplicity, but it involves excluding information from the low (but nonzero) probability hot tail of the CMIP6 distribution. On the other hand, ECSall and the reduced ECS10 variants are more consistent with the assessed distribution of ECS, but they require the use of weighted quantiles that add complexity to the delivery of climate services.

For regional variables other than temperature and precipitation, especially those related to dynamical processes with weaker links to climate sensitivity, a potential pitfall of global constraints is that their application may lead to an artificial
reduction of ensemble spread. This loss of information must be carefully weighed against the ability to mitigate known biases in ensemble statistics for temperature variables and precipitation extremes.

From a technical point of view, the quantile matching approach used to develop the ECSall and ECS10 constrained ensembles can be extended to include additional terms in its cost function. For example, other observational constraints, measures of global or regional performance, and measures of climate model independence could all be applied, although this would come at the expense of extra complexity. The physical plausibility, mechanism of operation, and robustness of any additional constraints or performance measures must be first carefully evaluated (Hall et al. 2019).

This study only considers global constraints. Although efforts have been made to select CMIP6 models based on historical performance in Canada (Mahony et al. 2022; Jeong and Cannon 2023), it is unclear whether regional constraints provide added value relative to simpler global constraints on climate sensitivity or historical rates of global warming. Recent work suggests that a careful combination of global and regional constraints can lead to better performance relative to the application of global observational constraints alone (Qasmi and Ribes 2022). However, Goldenson et al. (2023) found that some high-ECS climate models are among the best performers in terms of their ability to simulate historical circulation fields over the Northern Hemisphere. Similarly, Palmer et al. (2023) found that the removal of the models that performed the worst for Europe led to greater regional warming as the high-sensitivity models remained. More research on regional constraints is warranted.

In general, it is recommended that results from the CMIP6 ensemble be accompanied by documentation that clearly states the impact of the “hot model” problem on regional projections. Furthermore, the use of a constrained ensemble should be considered. Given that changes in temperature and precipitation

**FIG. 11.** Summary of appreciable differences between constrained and unconstrained projections of Canada-wide ETCCDI extreme indices for (a) ECSall and (b) TCRlikely constraints. Differences that lie outside ±1.68 and 1.96 standard deviations of internal variability are shaded.

**FIG. 12.** Correlation matrix summarizing the level of correspondence in differences between constrained and unconstrained projections of ETCCDI extreme indices across five weighting schemes (ECSall, TCRlikely, and three variants of ECS10) applied to the CanDCS-U6 ensemble. Correlation coefficients are calculated between standardized differences for all combinations of region, index, percentile, SSP experiment, and time period. The ECS10 variants differ primarily in terms of which high ECS model is included in the subset of 10 models (CanESM5, HadGEM3-CG31-LL, or UKESM1-0-LL).
variables in Canada are strongly correlated with global warming (Sobie et al. 2021), global constraints on climate sensitivity or historical warming provide a simple and robust path forward for these variables. This should also be true for other combinations of global regions and variables for which pattern scaling holds (Tebaldi and Arablaster 2014). To complement the constrained ensemble, unconstrained projections can also be communicated for specified levels of global warming, with global constraints then used to inform the timing of warming level exceedance.

Acknowledgments. The initial motivation for this article came from correspondence with Elaine Barrow at CCCS. Comments on the draft manuscript by Yongxiao Liang are appreciated. The provision of data is due to the efforts of PCIC, the Climate Data and Products Section of ECCC, and the collaborative partners responsible for ClimateData.ca.


REFERENCES