Received: 15 June 2021 | Revised: 1 December 2021 | Accepted: 2 December 2021
DOI: 10.1002/gdj3.142

DATA PAPER

Geoscience Data Journal RMetS WILEY

Canadian Large Ensembles Adjusted Dataset version 1 (CanLEADv1): Multivariate bias-corrected climate model outputs for terrestrial modelling and attribution studies in North America

Alex J. Cannon[1] | **Hunter Alford**[1] | **Rajesh R. Shrestha**[2] | **Megan C. Kirchmeier-Young**[3] | **Mohammad Reza Najafi**[4]

[1]Climate Research Division, Environment and Climate Change Canada, Victoria, BC, Canada

[2]Watershed Hydrology and Ecology Research Division, Environment and Climate Change Canada, Victoria, BC, Canada

[3]Climate Research Division, Environment and Climate Change Canada, Toronto, ON, Canada

[4]Department of Civil and Environmental Engineering, Western University, London, ON, Canada

Correspondence
Alex J. Cannon, Climate Research Division, Environment and Climate Change Canada, 2472 Arbutus Road, 2nd Floor, Victoria, British Columbia V8P 5C2, Canada.
Email: alex.cannon@ec.gc.ca

Abstract

The Canadian Large Ensembles Adjusted Dataset version 1 (CanLEADv1) contains 50-member ensembles of bias-adjusted near-surface global and regional climate model variables on a 0.5° grid over North America for historical and future scenarios (1950–2100). Canadian Earth System Model Large Ensembles (CanESM2 LE) and Canadian Regional Climate Model Large Ensemble (CanRCM4 LE) datasets are bias-corrected using a multivariate quantile-mapping algorithm for statistical consistency – in terms of marginal distributions and multivariate dependence structure – with two observationally constrained historical meteorological forcing datasets. For each observational dataset, bias-adjusted variables are provided for two sets of 50-member initial-condition CanESM2 ensembles (historical plus RCP8.5 scenarios, 1950–2005 and 2006–2100, respectively; and historicalNAT scenario, 1950–2020, which excludes anthropogenic forcings), and one 50-member CanRCM4 ensemble (historical plus RCP8.5). The archive includes daily minimum temperature, maximum temperature, precipitation, relative humidity, surface pressure, wind speed, incoming shortwave radiation and

Dataset

Identifier: https://open.canada.ca/data/en/dataset/a97edbc1-7fda-4ebc-b135-691505d9a595

Creator: Environment and Climate Change Canada.

Dataset correspondence: open-overt@tbs-sct.gc.ca.

Title: Canadian Large Ensembles Adjusted Dataset (CanLEADv1).

Publisher: Government of Canada.

Publication year: 2020.

Resource type: Data Entity.

Version: 1.

Geoscience Data Journal RMetS

incoming longwave radiation. Intended uses include hydrological and land surface impact modelling, as well as related event attribution studies.

KEYWORDS

bias correction, climate scenarios, counterfactual, downscaling, event attribution, hydrology, land surface, large ensemble, North America, regional climate model

1 | INTRODUCTION

There is growing interest in quantitative and physically consistent assessments of the potential impacts of future climate change on multiple affected sectors at a range of spatiotemporal scales. Sectoral impacts include those on water supply and demand, ecosystems, agriculture, forests, health and energy supply and demand, among others. To support these types of studies, efforts like the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) (Warszawski et al., 2013) provide impact modellers with observationally constrained global meteorological forcing data (Lange, 2018, 2019a, 2019c) and future projections from global climate models run under different greenhouse gas concentration scenarios. To minimize the influence of climate model biases on impact model results (e.g. Meyer et al., 2019; Zscheischler et al., 2019), climate model simulations are bias-adjusted towards the meteorological forcing data used for impact model calibration before dissemination (Lange, 2019b).

In addition to exploring future inter-sectoral impacts, studies are now considering the role of anthropogenic climate change on events that have already occurred. Recent examples include assessments of human-induced climate change contributions to area burned in the 2017 fire season in western Canada (Kirchmeier-Young et al., 2019) and to riverine flooding in Bangladesh in 2017 (Philip et al., 2019). In these types of event attribution studies, it is common to compare extreme events estimated using outputs from climate models run with both natural and anthropogenic forcings (so called ALL forcings) to those run with only natural forcings (NAT forcings). With access to large ensembles of ALL simulations – those representing the real world – and NAT simulations – those reflecting a world without human influence on the climate system – one can estimate probabilities of event occurrence in both factual and counterfactual worlds. The ratio of the two probabilities then quantifies the change in risk due to anthropogenic climate change. Counterfactual climate data is a recent addition to the set of ISIMIP simulations (Mengel et al., 2021).

The Canadian Large Ensembles Adjusted Dataset version 1 (CanLEADv1) is designed to support both inter-sectoral climate change impacts studies and event attribution studies. The main characteristics of the dataset are summarized in Tables 1 and 2. Unlike ISIMIP, which is global in scope, the focus of CanLEADv1 is on North America, with a further emphasis on terrestrial land surface and hydrological modelling users. To address these communities, the dataset includes 8 daily surface or near-surface meteorological variables commonly used for water and energy balance simulations: (a) minimum temperature (tasmin), (b) maximum temperature (tasmax), (c) precipitation (pr), (d) relative humidity (hurs), (e) surface pressure (ps), (f) wind speed (sfcWind), (g) incoming shortwave radiation (rsds) and (h) incoming longwave radiation (rlds). Furthermore, the regional focus permits the use of dynamically downscaled outputs from a regional climate model, as well as outputs from its parent global model.

Section 2 summarizes the main features of the dataset, technical details on its development and a brief evaluation of performance. As described in Section 2.1, CanLEADv1 is based on initial-condition ensembles of simulations from the Canadian Earth System Model – Canadian Regional Climate Model (CanESM2-CanRCM4) system, which is developed and run by Environment and Climate Change Canada's (ECCC's) Canadian Centre for Climate Modelling and Analysis (CCCma). Global and regional climate model outputs are bias-adjusted towards two observationally constrained 0.5° meteorological forcing datasets over North America. These observational forcing datasets are described in Section 2.2, and the combinations of model simulations and observational datasets that make up CanLEADv1 are given in Section 2.3. Section 2.4 provides technical details on the bias-adjustment algorithm, which is used to adjust the climate model outputs so that they match the multivariate dependence structure of the observational datasets. Section 2.5 evaluates the CanLEADv1 dataset in its ability to reproduce the extremes in multivariate hazard indices over the historical period. Finally, Section 2.6 places the global warming levels simulated by CanLEADv1's climate models in broader context.

As described in Section 3, CanLEADv1 is provided in self-describing, binary netCDF files under Canada's Open Government Licence. Potential dataset uses are outlined in Section 4. The large ensembles can be used to investigate the externally forced response, internal

TABLE 1 CanLEADv1 meteorological forcing variables

Name	Original variable	Adjusted variable	Units	Notes
Daily maximum near-surface (2 m) air temperature	tasmax	tasmax Adjust	K	Tasmax Adjust and tasmin Adjust derived from bias-adjustment of (tasmin + tasmax)/2 and (tasmax − tasmin)
Daily minimum near-surface (2 m) air temperature	tasmin	Tasmin Adjust	K	
Daily mean precipitation rate	pr	Pr Adjust	kg m^{-2} s^{-1}	
Daily mean near-surface relative humidity	hurs	Hurs Adjust	%	hurs (with respect to liquid water) derived from specific humidity, temperature and ps (Bolton, 1980)
Daily mean surface air pressure	ps	Ps Adjust	Pa	
Daily mean near-surface (10 metre) wind speed	sfcWind	sfcWind Adjust	m s^{-1}	
Daily mean surface downwelling shortwave radiation	rsds	Rsds Adjust	W m^{-2}	
Daily mean surface downwelling longwave radiation	rlds	Rlds Adjust	W m^{-2}	

TABLE 2 CanLEADv1 dataset characteristics

Ensemble	Members	Period	External forcing	Climate model	Output domain	Observational Target
CanESM2_ALL-EWEMBI-MBCn	50	1950–2100	historical/RCP8.5	CanESM2[a]	NAM-44i 0.5°	EWEMBI
CanESM2_NAT-EWEMBI-MBCn	”	1950–2020	historicalNat	”	”	”
CanESM2_ALL-S14FD-MBCn	”	1950–2100	historical/RCP8.5	”	”	S14FD
CanESM2_NAT-S14FD-MBCn	”	1950–2020	historicalNat	”	”	”
CanRCM4-EWEMBI-MBCn	”	1950–2100	historical/RCP8.5	CanRCM4[b]	”	EWEMBI
CanRCM4-S14FD-MBCn	”	”	”	”	”	S14FD

[a]Global model with native ~2.8° atmosphere.

[b]Regional model on native NAM-44 0.44° grid.

variability – which is an intrinsic property of a climate model and largely irreducible – and the relative role of external forcing and internal variability on the climate system. Large ensembles of ALL and NAT simulations can be compared in event attribution studies. Availability of consistently bias-adjusted outputs from the global and regional model components of the CanESM2-CanRCM4 system can be used to investigate the added value of dynamical downscaling, with multiple observational datasets partly accounting for observational uncertainty. Limitations, complementary datasets and potential future updates are summarized in Sections 5 and 6.

2 | DATA DESCRIPTION AND DEVELOPMENT

2.1 | Climate models and scenarios

The CanLEADv1 dataset is based on archived climate model simulations in the Canadian Earth System Model Large Ensembles (CanESM2 LE) (ECCC, 2017) and Canadian Regional Climate Model Large Ensemble (CanRCM4 LE) (ECCC, 2018) datasets. These two datasets consist of dynamically downscaled regional simulations over North America by version 4 of CCCma's

RCM, CanRCM4, along with global simulations from CanRCM4's parent model, CanESM2, which is CCCma's second generation Earth System Model. Scinocca et al. (2016) provided an overview and technical details of the coordinated global and regional climate modelling effort used to develop the CanESM2-CanRCM4 system.

CanESM2 includes interactive, coupled atmosphere, ocean, sea ice, land and carbon cycle components; the atmospheric model is configured to run globally at ~2.8° horizontal spacing (Arora et al., 2011). A large 50-member perturbed initial-condition ensemble, referred to as CanESM2 LE (Fyfe et al., 2017), was randomly initialized starting on 1 January 1950 from the 5 historical CanESM2 ensemble members contributed by CCCma to the Coupled Model Intercomparison Project phase 5 (CMIP5). Random perturbations to the initial atmospheric state at the beginning of 1950 were introduced via one of the cloud physics parameterisations. This parameterisation relies on a random number generator with a pre-set seed; the 10 individual simulations split from each of the 5 original members are based on different seeds. In this way, different realizations of historical and projected future climate were produced without any change to the model dynamics, physics or structure. The only differences are due to internal variability, which is a manifestation of natural climate variability simulated by the model. The historical ALL simulations used observed estimates of historical changes in solar, volcanic, greenhouse gas, aerosol, ozone and land-use forcings, whereas the counterfactual NAT simulations used solar and volcanic forcings only. To obtain future projections, historical ALL simulations, which end on 31 December 2005, were extended from January 2006 to the end of 2100 using the RCP8.5 emissions scenario.

Regional simulations that dynamically downscale global outputs from CanESM2 LE onto a 0.44° grid over North America were performed with CanRCM4 (Scinocca et al., 2016). CanRCM4 shares the same dynamical core as the Global Environmental Multiscale (GEM) model, which is the integrated weather forecasting and data assimilation system used by ECCC for operational Numerical Weather Prediction (NWP) (Côté et al., 1998). Unlike the operational forecasting system, however, CanRCM4 uses the same package of physical parameterisations as version 4 of CCCma's Canadian Atmospheric Global Climate Model (CanAM4) (von Salzen et al., 2013), which forms the atmospheric component of CanESM2. For CanRCM4 LE, NAT simulations were not run; hence, outputs are only available for the ALL runs (1950–2100).

CanESM2 contributions to CMIP5 have been evaluated alongside with other CMIP5 climate models; a summary of results, focusing on the global climate, is included in Flato et al. (2013). From a regional perspective, when assessed in terms of its ability to reproduce historical surface temperature and precipitation indices over extratropical Northern Hemisphere land areas, CanESM2's performance lay within the middle tercile of 31 CMIP5 models (Sillmann et al., 2013). Similarly, focusing on the historical frequency and persistence of daily atmospheric circulation patterns over North America, Cannon (2020) found that CanESM2 fell within the middle tercile of 30 CMIP5 and CMIP6 models in terms of frequency performance, and within the top tercile for persistence performance. Najafi et al. (2015) found good agreement between observed and CanESM2-simulated Arctic land temperature changes over the historical period.

Using North American Regional Climate Change Assessment Programme (NARCCAP) experimental protocols and evaluation measures (Mearns et al., 2012), Scinocca et al. (2016) compared the performance of CanRCM4 against other NARCCAP regional climate models. CanRCM4's performance fell within that of the other regional models assessed by Mearns et al. (2012). A growing body of literature has since evaluated the ability of the CanESM2-CanRCM4 system to simulate historical climate conditions in North America, including extremes of temperature and precipitation (e.g. Whan & Zwiers, 2015, 2016, among others). In general, performance of CanRCM4 is competitive with other regional models that have contributed simulations to NARCCAP and the North American Coordinated Regional Downscaling Experiment (NA-CORDEX) programme (McGinnis & Mearns, 2021).

2.2 | Observational datasets

Two observationally constrained historical meteorological forcing datasets, both based on third-generation global atmospheric reanalyses that have been further corrected towards observations on 0.5° global grids, are used as targets in the climate model bias-adjustments of daily tasmin, tasmax, pr, hurs, ps, sfcWind, rlds and rsds variables.

The S14 global meteorological forcing dataset (S14FD) (Iizumi et al., 2017) uses surface variables from the Japanese 55-year Reanalysis (JRA-55) (Kobayashi et al., 2015), whereas the EartH2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI) dataset (Lange, 2018, 2019a) takes variables from the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Reanalysis (ERA-Interim) (Dee et al., 2011). In both cases, reanalysis outputs are corrected so that climatological statistics over land match those of observational gridded datasets. Values over ocean grid cells are those of the raw JRA-55 and ERA-Interim reanalyses.

The use of different reanalyses in S14FD and EWEMBI, which include different NWP models and data assimilation systems, as well as different observational targets and observational bias removal methods, means that CanLEADv1 outputs sample both observational uncertainty and internal variability of the climate system. Iizumi et al. (2017) assessed the relative influence of uncertainties due to the use of different emission scenarios, climate models and observational bias-adjustment target datasets on near and distant future projections of temperature and precipitation indices over 22 sub-continental regions of the globe. They found that observational uncertainty exceeded scenario and climate model uncertainty over most combinations of precipitation indices and regions for both near and distant future.

When daily variables were evaluated against surface stations and independent flux tower observations, S14FD and WFDEI, the constituent dataset of EWEMBI that provides all variables except rlds and rsds, were consistently among the best performing datasets in an intercomparison of five global meteorological forcing datasets (Iizumi et al., 2017). Further, when assessed against the HadEX2 dataset (Donat et al., 2013) in terms of correlation (root mean squared error) of temperature and precipitation extremes indices, S14FD and EWEMBI were the top two datasets, performing best in 37.7% (39.7%) and 31.8% (22.7%) index/region combinations, respectively. Similar strong performance has been noted for assessment of S14FD precipitation (Singh & Najafi, 2020) and WFDEI precipitation (Wong et al., 2017) in Canada.

2.3 | Ensemble sizes

For CanESM2 LE, there are 2 scenarios (ALL and NAT), 2 observationally constrained target datasets for bias-adjustment (S14FD and EWEMBI), and 50 ensemble members, which gives a total of $2 \times 2 \times 50 = 200$ sets of outputs (22,200 simulated years). For CanRCM4 LE, which only has simulations for the ALL scenario, there are $2 \times 50 = 100$ sets of outputs (15,100 simulated years).

In both cases, CanLEADv1 provides variables on the CORDEX NAM-44i 0.5° grid over North America. CanESM2 outputs (~2.8° grid) and CanRCM4 outputs (0.44° grid) are bilinearly interpolated onto the NAM-44i grid before bias-adjustment (Figure 1).

2.4 | Multivariate bias-adjustment

Daily CanESM2 LE and CanRCM4 LE outputs on the NAM-44i grid are bias-adjusted so that they are statistically consistent with the historical S14FD and EWEMBI meteorological forcing datasets. Here, statistical consistency refers specifically to the multivariate distribution of the 8 meteorological forcing variables. As described by Cannon (2018), a multivariate version of quantile mapping – multivariate bias correction via N-dimensional probability distribution transfer, referred to as MBCn – is used to adjust the historical distribution of each simulated variable, as well as the statistical dependence between variables, so that these properties match those of the target observational dataset. Multivariate bias-adjustment is performed on a grid cell-by-grid cell basis.

Specifically, climate model outputs at each grid cell are adjusted using the 1981–2010 observational period for calibration, with multivariate bias correction applied over 30-year sliding windows from 1950–2100 (ALL) or 1950–2020 (NAT). In each window, the central 10 years are replaced, the window is slid 10 years, etc. until the end of the projection period is reached. To ensure an unbiased seasonal cycle, adjustments are applied to data pooled over 33-day-of-year sliding blocks – the central 11 days are replaced, the block is slid 11 days, etc. Following Kirchmeier-Young et al. (2017), each of the 50 ALL or NAT ensemble members is corrected separately using a different ALL member as the historical reference for calibration. This ensures that internal variability is not artificially suppressed (e.g. if each member were to be calibrated against itself).

To ensure that corrected values of tasmax exceed tasmin on all days, tasmin and tasmax are not corrected directly (Thrasher et al., 2012). Instead, multivariate bias-adjustment is applied to the diurnal temperature range (tasmax-tasmin) and approximate mean temperature [(tasmin + tasmax)/2] variables. To avoid physical inconsistencies in corrected humidity variables (Grenier, 2018), hurs values (evaluated with respect to water) are first computed from raw model and observational target values of specific humidity, mean temperature and ps (Bolton, 1980). Following Cannon (2018), the resulting hurs values are mapped from the unit interval onto the real line using a logit transform; the transformed values are then bias-adjusted.

Outside of the 1981–2010 calibration period, changes in corrected quantiles are constrained to match those in the raw climate model simulations (i.e. adjustments are change-preserving on a quantile-by-quantile basis) (Cannon et al., 2015). For ALL simulations, the climate change signal simulated by the climate model outside of the historical calibration period is preserved. For NAT simulations, the same approach is used to preserve differences from the ALL simulations resulting from the absence of anthropogenic forcings. Variables pr, rsds, sfcWind and diurnal temperature range are treated as ratio variables (i.e. relative changes in quantiles are preserved); all others are treated as interval variables (i.e. absolute changes are preserved). After bias-adjustment, transformed variables (logit transformed hurs, diurnal temperature range and

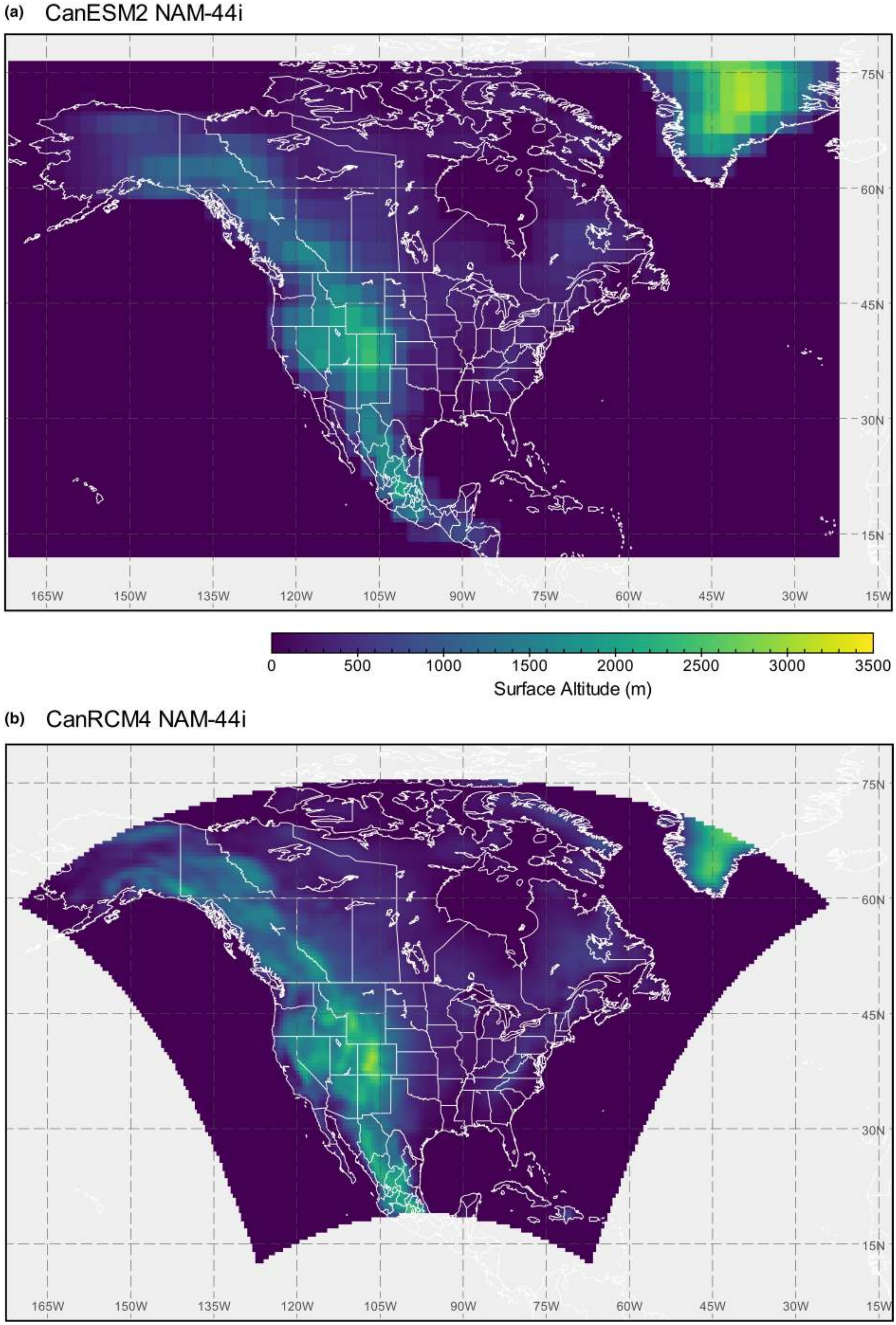

FIGURE 1 Spatial extent of the CORDEX NAM-44i 0.5° grid with interpolated (a) CanESM2 and (b) CanRCM4 surface topography for reference. CanRCM4 runs natively on a 0.44° rotated pole grid that only covers a portion of the regular NAM-44i grid

approximate mean temperature) are back-transformed to their original units.

The multivariate bias-adjustment algorithm is iterative, alternating between application of a random orthogonal rotation to the data and application of univariate quantile mapping to the rotated variables' marginal distributions. As pointed out by Cannon (2018) and François et al. (2020), the number of iterations affects both convergence of the bias-adjusted multivariate distribution to the observed distribution in the calibration period, as well as performance on non-calibration (i.e. projection) samples. Too few iterations can lead to underfitting and too many iterations can lead to overfitting. The optimal number of iterations for CanLEADv1 was set to 20 based on results from a reanalysis-driven CanRCM4 cross-validation experiment similar to that described in Section 5 of Cannon (2018), except with EWEMBI serving as the observational reference product. Figure 2 shows cross-validated measures of multivariate bias-adjustment performance, indicating that 20 iterations leads to optimal non-calibration performance. The same sequence of random rotations is used at all grid cells to limit the appearance of spatial artefacts at adjacent locations.

2.5 | Evaluation

To verify success of the multivariate bias-adjustments, performance over North American land grid cells is evaluated for the first member of each CanLEADv1 ensemble with respect to observationally constrained data from the 2011–2019 period; these observations lie outside the 1981–2010 period used to calibrate the bias-adjustments. The W5E5 global meteorological forcing data processed for ISIMIP (Lange, 2019c), a successor to EWEMBI based on the ERA5 reanalysis, is used as the observational reference. Note that CanESM2 and CanRCM4 simulations used in the evaluation period follow RCP8.5 greenhouse gas emissions rather than those observed. However, observed and RCP8.5 cumulative carbon emissions lie within 1% over this period (Schwalm et al., 2020) and hence differences in external forcing are unlikely to be of practical significance.

Following Zscheischler et al. (2019), multivariate hazard indices (Table 3) – the Canadian fire weather index, humidity index, wind chill and wet bulb globe temperature – are computed from the raw climate model outputs, CanLEADv1 bias-adjusted outputs and W5E5 observations. Each index is computed as a nonlinear function of two or more daily meteorological forcing variables. Importantly, the derived indices are not bias-adjusted directly. Hence, proper reproduction of their statistical characteristics in the evaluation period depends on correction of their component variables and their inter-dependence. Long-term median values over the 2011–2019 period are calculated for annual maxima of all indices except wind chill for which minima are calculated; values from CanLEADv1 and raw climate model outputs are then compared across the spatial domain with those calculated based on observations.

Figure 3 shows a Taylor diagram summarizing performance in terms of centred root mean squared error, spatial pattern correlation and spatial standard deviation. For illustration, maps of the median annual maximum

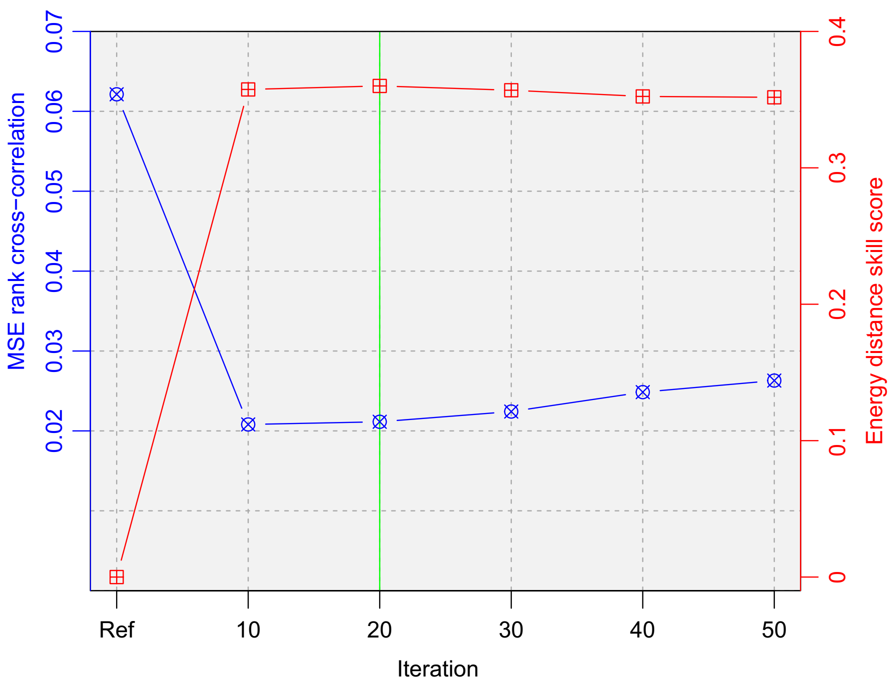

FIGURE 2 Cross-validated measures of multivariate bias-adjustment performance as a function of number of multivariate bias-adjustment iterations: (left axis) mean squared error (MSE) of inter-variable rank cross-correlations (blue, lower equals better); (right axis) energy distance skill score (red, higher equals better). The reference (Ref) values are for raw CanRCM4 simulations. The vertical green line indicates the number of iterations used for CanLEADv1

Geoscience Data Journal RMetS

TABLE 3 Multivariate hazard indices used to evaluate bias-adjustment performance in Section 2.5

Index	Hazard	Input variables	Reference
Canadian fire weather index	General index of fire danger optimized for forested areas in Canada	tasmax, pr, hurs, sfcWind	Van Wagner (1987)
Humidity index	Human heat stress due to combined effects of heat and humidity	(tasmin + tasmax)/2, hurs	Masterton and Richardson (1979)
Wind chill	Lowering of body temperature due to passing flow of cold air	(tasmin + tasmax)/2, sfcWind	Osczevski and Bluestein (2005)
Wet bulb globe temperature	Human heat stress in direct sunlight due to combined effects of temperature, humidity, wind speed and solar radiation	(tasmin + tasmax)/2, hurs, sfcWind, rsds	Liljegren et al. (2008)

fire weather index – the least well simulated hazard index – are shown in Figure 4 for W5E5 observations and raw/bias-adjusted CanRCM4 simulations. Because the role of internal variability is not taken into consideration, performance is best reported in terms of relative improvements

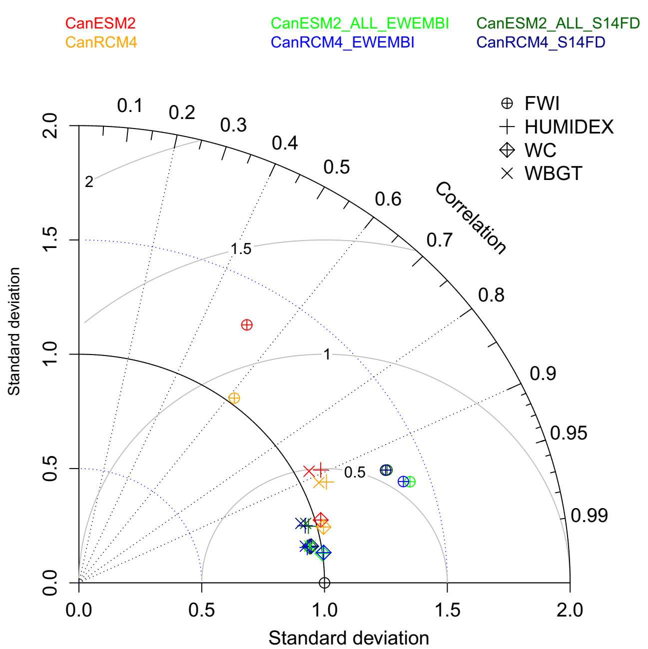

FIGURE 3 Taylor diagram showing spatial pattern correlations, standard deviations and centred root mean squared errors (grey arcs) of scaled 2011–2019 median multivariate hazard indices (Table 2) for bias-adjusted CanLEADv1 outputs – CanESM2_ALL-EWEMBI-MBCn (light green), CanESM2_ALL-S14FD-MBCn (dark green), CanRCM4-EWEMBI-MBCn (blue), and CanRCM4-S14FD-MBCn (dark blue) – and raw CanESM2 (red) and CanRCM4 (orange) climate model outputs. The first ensemble member is used in all cases. Observational reference values are 2011–2019 median W5E5 values scaled to have unit standard deviation; climate model indices are expressed in terms of W5E5 standard deviation units

between the bias-adjusted and raw climate model outputs for the ensemble member. For all indices, centred root mean squared error is reduced for bias-adjusted CanLEADv1 outputs relative to raw climate model outputs (Figure 3), with magnitudes ranging from reductions of 39%–42% for the fire weather index to 82%–85% for wind chill. Similarly, spatial pattern correlations improve (Figure 3), with increases in explained variance ranging from 3%–5% for WC to 50%–61% for the fire weather index. More generally, the ability of the multivariate bias-adjustment algorithm to correct biases in climate and weather model simulations has been evaluated for multiple end uses, including hydrological modelling (Meyer et al., 2019; Shrestha et al., 2019; Singh & Najafi, 2020; Su et al., 2020), calculation of multivariate heat and fire indices (Whan et al., 2021; Zscheischler et al., 2019), and estimation of compound extremes (Hao & Singh, 2020).

2.6 | Scenarios and global warming levels

The ALL forcing simulations of CanESM2 in CanLEADv1 are run using historical and RCP8.5 external forcing scenarios from 1950–2100. The amount of global warming, and hence the impact on the climate of North America, over this period depends on the strength of the response of CanESM2 to the prescribed anthropogenic forcings in these experiments. CanESM2 has an equilibrium climate sensitivity of 3.7°C and a transient climate response of 2.4°C (Vial et al., 2013). These values are on the high end of the assessed likely range of 1.5–4.5°C for equilibrium climate sensitivity and 1–2.5°C for transient climate response (Collins et al., 2013). Further, the RCP8.5 scenario represents the high end of possible baseline greenhouse gas emissions scenarios in a no-policy world. However, given current population and gross domestic product projections, emissions of CO_2 and projected global

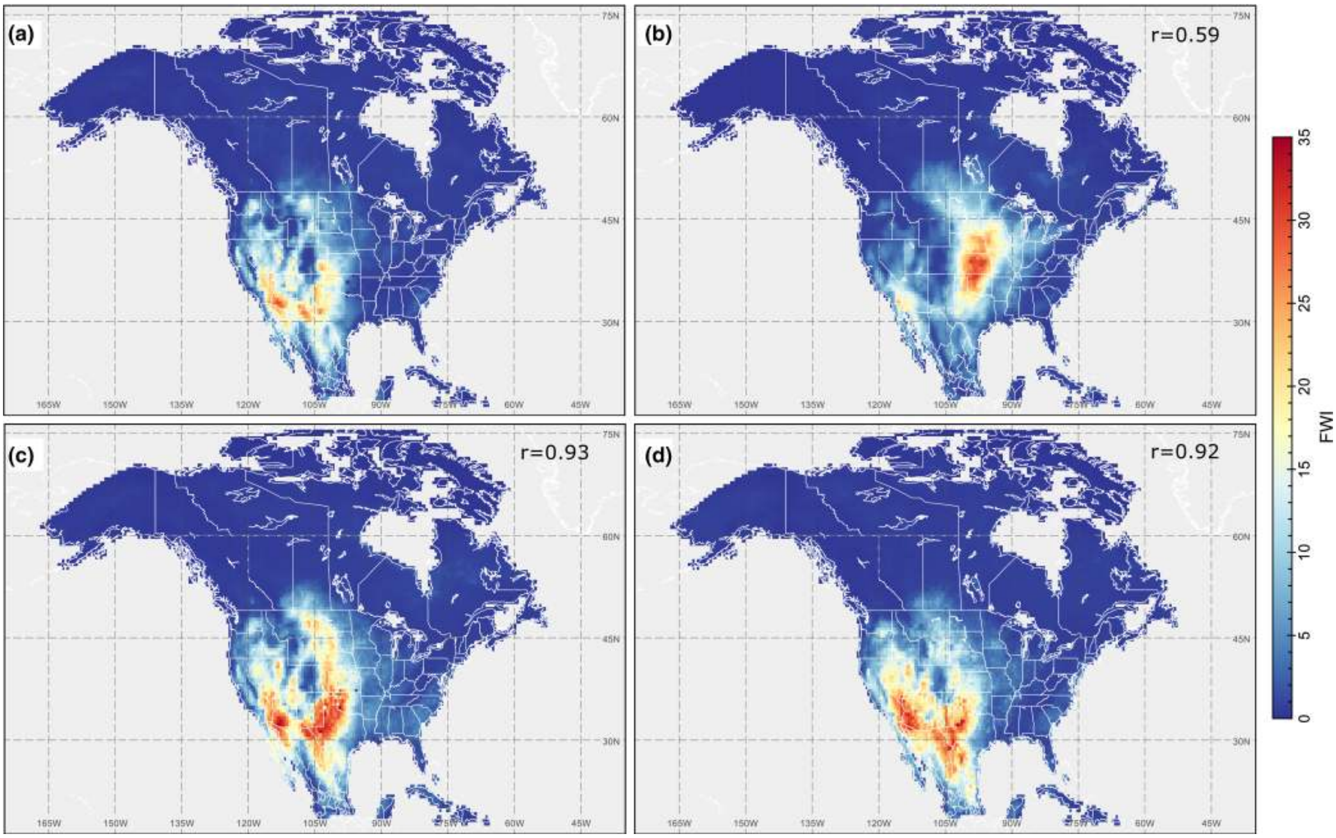

FIGURE 4 Median values of annual maximum FWI over the 2011–2019 period for (a) W5E5 observations, (b) raw CanRCM4 simulations, (c) bias-adjusted CanRCM4-EWEMBI-MBCn outputs and (d) bias-adjusted CanRCM4-S14FD-MBCn outputs. Pattern correlations between each of the CanRCM4 datasets and observations are shown in the top right corner. Differences between (c) and (d) are due exclusively to observational uncertainty; the first CanRCM4 member is used in all cases

warming that are as high as those under RCP8.5 are unlikely (Raftery et al., 2017).

Taken together, this means that the climate projections in CanLEADv1, which are based on CanESM2 simulations, warm faster than those based on most other combinations of climate model and emissions scenario. For example, CanESM2 projects global warming to reach ~5.9°C by the end of the 21st century under the RCP8.5 scenario (Figure 3), whereas the mean end-of-century warming for the full CMIP5 ensemble is ~4.3°C for RCP8.5 and ~1.6°C for RCP2.6 (Flato et al., 2019).

Rather than presenting results for a fixed period of simulation years (e.g. the 2050s or 2080s), which will depend on warming of CanESM2 under RCP8.5, one can instead present results for a specified level of global warming since the preindustrial (PI) period (e.g. +2°C or +3°C). Framing results in this manner relies on being able to identify the simulated time period that corresponds to the specified warming level. To this end, CanLEADv1 includes a comma-separated values file with the global mean annual surface air temperature anomaly time series presented in Figure 5. Assumptions (e.g. that the regional response to a given amount of global warming is independent of emissions scenario and climate model), caveats and examples of this approach to presenting regional climate scenarios are provided in Seneviratne et al. (2016), King et al. (2019) and Cannon et al. (2020). Similarly, the ensemble of national climate scenarios for Switzerland is, in part, constructed using a 'time-shift pattern scaling method' that fills in missing regional simulations using the relationship between time-of-warming in global models and the regional response for that level of warming under available RCP8.5 simulations (Sørland et al., 2020). This basic approach is extensible to new scenarios, for example the Shared Socioeconomic Pathways (SSPs) used in CMIP6.

3 | DATASET ACCESS AND FILE FORMATS

The Canadian Large Ensembles Adjusted Dataset version 1 (CanLEADv1) is available for download from the Government of Canada's Open data portal at https://open.canada.ca/data/en/dataset/a97edbc1-7fda-4ebc-b135-691505d9a595. Data are in self-describing, binary netCDF files with file names and metadata following the

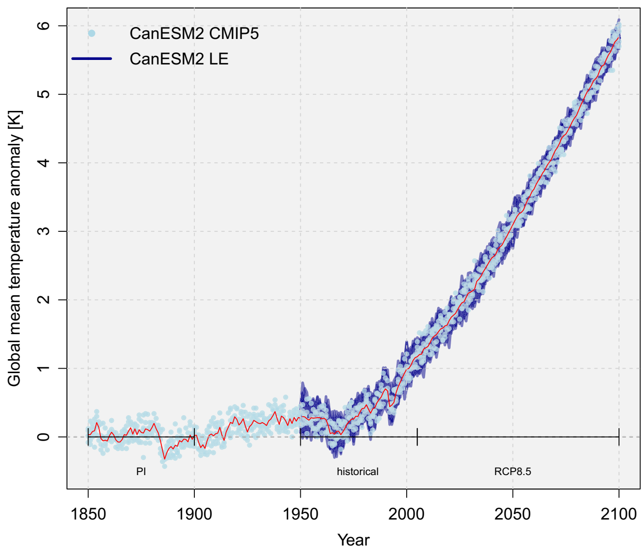

FIGURE 5 Time series of simulated global mean annual surface air temperature anomalies taken with respect to a preindustrial (PI) period (defined as 1850–1900) in the 5-member CanESM2 CMIP5 ensemble (1850–1950) and the 50-member CanESM2 LE under historical (1950–2005) and RCP8.5 (2006–2100) forcings. The red line shows the 5-member and 50-member ensemble mean values over the 1850–1950 and 1950–2100 periods, respectively

recommended Data Reference Syntax for bias-adjusted CORDEX simulations (Nikulin & Legutke, 2016). The dataset is provided under Canada's Open Government Licence, which grants users a worldwide, royalty-free, perpetual, non-exclusive licence to use the data, including for commercial purposes, subject to terms listed at https://open.canada.ca/en/open-government-licence-canada.

4 | POTENTIAL DATASET USE

4.1 | Hydrologic and land surface modelling

The CanLEADv1 dataset includes required meteorological forcings for hydrologic/land surface models, such as Variable Infiltration Capacity (Liang et al., 1994), Modelisation Environmentale Communautaire – Surface and Hydrology (Pietroniro et al., 2007), Soil and Water Assessment Tool (Arnold et al., 1998) and Canadian Land Surface Scheme including Biogeochemical Cycles (Melton et al., 2020), among others. With the physically consistent multivariable forcings from CanLEADv1, there is no need to employ empirical methods (Bennett et al., 2020) to derive additional meteorological forcings (e.g. radiation fluxes), which is the case for most statistically downscaled datasets that include only temperature and precipitation. Temporal disaggregation (e.g. from a daily to sub-daily time step) may, however, still be required, but can be accomplished using off-the-shelf software packages (Bennett et al., 2020; Förster et al., 2016).

Uncertainties in hydrologic fluxes, such as evapotranspiration and runoff timing and magnitude (Bohn et al., 2013; Mizukami et al., 2014), associated with the empirical estimates of the additional climate variables can thus be minimized. Furthermore, given that hydrological processes such as rain-on-snow events, spring freshet, etc. are dependent on interactions between multiple climate variables, the multivariate downscaled products, by virtue of maintaining realistic relationships between variables, could potentially lead to improved hydrologic and hydraulic simulations (Meyer et al., 2019; Shrestha et al., 2019; Singh & Najafi, 2020). Potential applications of CanLEADv1 include regional to continental scale hydrologic impact studies and/or investigation of the role of climate model's internal variability on hydrologic projections (Giuntoli et al., 2018). Furthermore, CanLEADv1 bias-adjusted outputs could be used as covariates for statistical modelling and analyses of nonstationary hydroclimatic extremes (e.g. Shrestha et al., 2017) and observationally constrained future runoff sensitivities (e.g. Lehner et al., 2019).

The CanLEADv1 dataset could also be used in regional to continental assessments of future snowpack changes in North America. Currently, due to the lack of bias-adjusted simulations of snowpack, there is still a need to rely on the raw climate model outputs for projections of future changes in the region (e.g. Fyfe et al., 2017; Jeong & Sushama, 2018; Mudryk et al., 2018; Shrestha et al., 2021). To this end, an offline energy balance snow model (e.g. Magnusson et al., 2015; Walter et al., 2005) forced with the bias-adjusted CanLEADv1 datasets could complement the climate model snowpack simulations.

4.2 | Event attribution

The CanLEADv1 dataset has many benefits for event attribution, which is a field of climate science that aims to quantify how anthropogenic climate change has influenced the frequency or magnitude of certain types of extreme events. Typically, this involves a comparison of the characteristics or likelihood of an extreme event between a climate model scenario that accounts for human emissions and a scenario without human influence on the climate. First, the models used in an event attribution study need to accurately represent the event in question, which may be improved by the bias-adjustment procedure used in CanLEADv1. Additionally, the multivariate nature of the bias-adjustment allows for the calculation of multivariate extremes or indices characterizing high-impact events such as drought or wildfires (Kirchmeier-Young et al., 2017, 2019; Zscheischler et al., 2019).

Second, large ensembles are needed to ensure a large enough sample size for estimating the likelihood of extreme events. Third, multiple forcing scenarios (ALL and NAT) enable a direct comparison between scenarios with and without anthropogenic forcing. In particular, there has previously been a lack of high-resolution datasets driven by NAT forcing. Finally, the scale of the CanLEADv1 dataset is finer than most global models and allows for the possibility of assessing smaller-scale events.

CanLEADv1 is also useful for investigating the changes in the observed hydrological processes and determining the degree of attributable risk of individual and compound extremes to climate change. It provides the opportunity to extend the existing methods that are developed to perform long-term trend attribution (Najafi et al., 2017), attribute individual extreme events and characterize the role of internal climate variability (Bellprat et al., 2019).

4.3 | Compound extremes

CanLEADv1 supports analyses of compound extremes – multiple extreme events that can interact in space and time – which can pose significant risks to societies, infrastructure, environments and various economic sectors (Zscheischler et al., 2020). Understanding the risks that are driven by cascading and compounding extremes requires an accurate representation of the underlying statistically dependent climate variables (Leonard et al., 2014). CanLEADv1, which is based on a multivariate bias-adjustment method, is, therefore, suitable for assessing the intensity and frequency of individual as well as compound extreme events at regional scales under climate change.

In addition to maintaining the dependence structure, the large ensemble simulations provide the opportunity to characterize nonstationary interdependencies between multiple drivers that can influence future changes in compound events including temperature and precipitation (e.g. heatwaves and droughts or warm-wet events), extreme wind and intense rainfall, among others (Singh et al., 2021). These interdependencies can be represented through nonstationary multivariate extreme value theory (e.g. via time-varying marginal distributions and dependence structures) applied to the CanLEADv1 bias-adjusted large ensemble simulations, which can lead to robust estimation of future joint return periods of compound events (Zscheischler et al., 2018). Further, the large ensemble can be used to quantify the contributions of internal climate variability and anthropogenic forcing to individual and compound extreme events, for example, compound floods, droughts, wildfires and heatwaves.

5 | LIMITATIONS AND COMPLEMENTARY DATASETS

The 50-member CanESM2/CanRCM4 ensembles were created by perturbing initial conditions of a 5-member set of historical CanESM2 simulations at the start of the 1950 simulation year. Due to the chaotic nature of the climate system, each of the 10 simulations start to diverge from their 5 parent members at this point. Depending on the use case (e.g. analyses of atmospheric versus land surface state), it may be necessary to exclude an initial spin-up period before the 50 simulations can be considered as independent realisations of the climate system.

CanLEADv1 samples observational uncertainty through the use of 2 observationally constrained reference datasets, and internal variability of the climate system through the use of 50-member initial-condition climate model ensembles. Other sources of uncertainty, including model uncertainty (only one climate modelling system and bias-adjustment method are used) and scenario uncertainty (future projections are based exclusively on RCP8.5) are not taken into account. Furthermore, CanRCM4 is run on a 0.44° grid, which limits its utility for high-resolution impact modelling, for example, hydrological modelling in small basins.

Structural model uncertainty, scenario uncertainty and additional observational uncertainty can be assessed using complementary datasets for North America. In particular, the NA-CORDEX programme provides bias-adjusted regional climate model outputs (McGinnis & Mearns, 2021) for daily tasmin, tasmax and surface wind components over the conterminous United States portion of the NAM-44i grid (using gridMET as the observational reference; Abatzoglou, 2013), and bias-adjusted daily tasmin, tasmax, pr, hurs and rsds over the NAM-44i grid (using Daymet as the observational reference; Thornton et al., 2020). In both cases, the same multivariate bias-adjustment algorithm is used as here, with some simulations available on the higher-resolution NAM-22i 0.25° grid. Outputs include a subset of combinations of single runs of 7 regional climate models driven by 9 global models under the RCP2.6, RCP4.5 and RCP8.5 scenarios. This includes a single realization from the CanESM2-driven CanRCM4 model used in CanLEADv1.

Few bias-adjusted climate datasets feature outputs that are consistent with both ALL and counterfactual non-warming forcings. Two notable examples that include the same variables at the same resolution as CanLEADv1 are the bias-corrected d4PDF historical and non-warming climate dataset (Iizumi, 2018) and the ISIMIP ATTRICI counterfactual climate for impact attribution dataset (Mengel et al., 2021). Unlike CanLEADv1, d4PDF and ATTRICI are global rather than regional in scope.

The d4PDF dataset covers the 1951-2010 historical period and contains the same daily variables as CanLEADv1. Bias-adjustment of 100-member ensembles of the Meteorological Research Institute Atmospheric Global Climate Model, version 3.2 (MRI-AGCM3.2) (Mizuta et al., 2017), one based on detrended historical boundary conditions and greenhouse gas and aerosol emissions set to PI values, is performed using the univariate quantile mapping methodology described by Iizumi et al. (2017) with the global S14FD dataset serving as the observational target. The ATTRICI dataset, unlike CanLEADv1 and d4PDF, which both rely on climate model simulations, constructs counterfactual climate data by statistically removing the climate change signal from two global observational datasets, including the W5E5 successor to EWEMBI (Lange, 2019c). This limits the ability to make attribution statements about the human influence on the climate system through emissions of greenhouse gases and aerosols.

6 | CONCLUSION AND FUTURE UPDATES

CanLEADv1 provides large ensembles of bias-adjusted CanESM2 and CanRCM4 climate model outputs on a 0.5° grid over North America. The dataset is designed to be useful for hydrological and land surface impact modelling, as well as related event attribution studies. Simulations include those based on ALL (1950–2100) and NAT (1950–2020) forcings from CanESM2 and ALL forcings from CanRCM4; anthropogenic forcings follow historical and RCP8.5 scenarios from CMIP5. Bias-adjustment is towards two observationally constrained historical datasets (S14FD and EWEMBI).

An update of CANLEAD is currently in the planning phase. Anticipated improvements include: (a) incorporation of higher-resolution simulations from CanRCM and other regional climate models (e.g. following CMIP6 CORDEX2 protocols; Gutowski et al., 2016); (b) use of multiple large global climate model ensembles and shared socioeconomic pathways; (c) adoption of higher-resolution observationally constrained historical datasets for bias-adjustment (e.g. Gasset et al., 2021; Muñoz-Sabater et al., 2021); (d) use of multiple bias-adjustment algorithms (e.g. François et al., 2020); and (e) outputs that include an expanded set of daily climate variables (e.g. including daily minimum and maximum relative humidity and wind components).

ACKNOWLEDGEMENT

The authors are grateful to the members of the Canadian Centre for Climate Modelling and Analysis for developing, running and archiving the CanESM2 and CanRCM4 LE simulations.

CONFLICT OF INTERESTS

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Alex J. Cannon: Conceptualization (lead); Investigation (lead); Methodology (lead); Software (lead); Supervision (lead); Validation (lead); Writing – original draft (lead); Writing – review & editing (lead). **Hunter Alford:** Methodology (supporting); Software (supporting); Validation (supporting); Writing – review & editing (supporting). **Rajesh R. Shrestha:** Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Megan C. Kirchmeier-Young:** Writing – original draft (equal); Writing – review & editing (equal). **Mohammad Reza Najafi:** Writing – original draft (equal); Writing – review & editing (equal).

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally shareable data necessary to reproduce the reported results. The data is available at https://open.canada.ca/data/en/dataset/a97edbc1-7fda-4ebc-b135-691505d9a595. Learn more about the Open Practices badges from the Center for Open Science: https://osf.io/tvyxz/wiki.

ORCID

Alex J. Cannon https://orcid.org/0000-0002-8025-3790
Rajesh R. Shrestha https://orcid.org/0000-0001-7781-6495

REFERENCES

Abatzoglou, J.T. (2013) Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, 33(1), 121–131. https://doi.org/10.1002/joc.3413

Arnold, J.G., Srinivasan, R., Muttiah, R.S. & Williams, J.R. (1998) Large area hydrologic modeling and assessment. Part I: Model development. *Journal of the American Water Resources Association*, 34(1), 73–89. https://doi.org/10.1111/j.1752-1688.1998.tb05961.x

Arora, V.K., Scinocca, J.F., Boer, G.J., Christian, J.R., Denman, K.L., Flato, G.M. *et al.* (2011) Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophysical Research Letters*, 38(L05805). 1–6. https://doi.org/10.1029/2010gl046270

Bellprat, O., Guemas, V., Doblas-Reyes, F. & Donat, M.G. (2019) Towards reliable extreme weather and climate event attribution. *Nature Communications*, 10(1732). 1–7. https://doi.org/10.1038/s41467-019-09729-2

Bennett, A., Hamman, J. & Nijssen, B. (2020) MetSim: A python package for estimation and disaggregation of meteorological data. *Journal of Open Source Software*, 5(47), 2042. https://doi.org/10.21105/joss.02042

Bohn, T.J., Livneh, B., Oyler, J.W., Running, S.W., Nijssen, B. & Lettenmaier, D.P. (2013) Global evaluation of MTCLIM and related algorithms for forcing of ecological and hydrological models. *Agricultural and Forest Meteorology*, 176, 38–49. https://doi.org/10.1016/j.agrformet.2013.03.003

Bolton, D. (1980) The computation of equivalent potential temperature. *Monthly Weather Review*, 108(7), 1046–1053. https://doi.org/10.1175/1520-0493(1980)108<1046:tcoept>2.0.co;2

Cannon, A.J. (2018) Multivariate quantile mapping bias correction: An N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dynamics*, 50(1–2), 31–49. https://doi.org/10.1007/s00382-017-3580-6

Cannon, A.J. (2020) Reductions in daily continental-scale atmospheric circulation biases between generations of global climate models: CMIP5 to CMIP6. *Environmental Research Letters*, 15(6), 064006. https://doi.org/10.1088/1748-9326/ab7e4f

Cannon, A.J., Jeong, D.I., Zhang, X. & Zwiers, F.W. (2020) *Climate-resilient Buildings and Core Public Infrastructure: An Assessment of the Impact of Climate Change on Climatic Design Data in Canada*. Gatineau, QC, Canada: Environment and Climate Change Canada. Available at: http://publications.gc.ca/pub?id=9.893021&sl=0 Accessed 28 July 2021.

Cannon, A.J., Sobie, S.R. & Murdock, T.Q. (2015) Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*, 28(17), 6938–6959. https://doi.org/10.1175/jcli-d-14-00754.1

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P. *et al.* (2013) Long-term climate change: Projections commitments and irreversibility. In: T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex & P.M. Midgley, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press. pp. 1029–1136.

Côté, J., Gravel, S., Méthot, A., Patoine, A., Roch, M. & Staniforth, A. (1998) The operational CMCMRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation. *Monthly Weather Review*, 126(6), 1373–1395. https://doi.org/10.1175/1520-0493(1998)126<1373:tocmge>2.0.co;2

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S. *et al.* (2011) The ERA-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553–597. https://doi.org/10.1002/qj.828

Donat, M.G., Alexander, L.V., Yang, H., Durre, I., Vose, R., Dunn, R.J.H. *et al.* (2013) Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *Journal of Geophysical Research: Atmospheres*, 118(5), 2098–2118. https://doi.org/10.1002/jgrd.50150

ECCC (2017) *Environment and Climate Change Canada, The Canadian Earth System Model Large Ensembles*. Government of Canada Open Data Portal. Available at: https://open.canada.ca/data/en/dataset/aa7b6823-fd1e-49ff-a6fb-68076a4a477c Accessed 28 July 2021.

ECCC (2018) *Environment and Climate Change Canada, The Canadian Regional Climate Model Large Ensemble*. Gatineau, QC, Canada: Government of Canada Open Data Portal. Available at: https://open.canada.ca/data/en/dataset/83aa1b18-6616-405e-9bce-af7ef8c2031c Accessed 28 July 2021.

Flato, G., Gillett, N., Arora, V., Cannon, A. & Anstey, J. (2019) Modelling future climate change. In: Bush, E. & Lemmen, D.S. (Eds.) *Chapter 3, Canada's Changing Climate Report*. Gatineau, QC, Canada: Government of Canada.

Flato, G., Marotzke, J., Abiodun, B., Pascale Braconnot, S., Chou, C., Collins, W. *et al.* (2013) Evaluation of climate models. In: T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex & P.M. Midgley, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press. pp. 741–866.

Förster, K., Hanzer, F., Winter, B., Marke, T. & Strasser, U. (2016) An open-source MEteoroLOgical observation time series DISaggregation Tool (MELODIST v0.1.1). *Geoscientific Model Development*, 9(7), 2315–2333. https://doi.org/10.5194/gmd-9-2315-2016

François, B., Vrac, M., Cannon, A.J., Robin, Y. & Allard, D. (2020) Multivariate bias corrections of climate simulations: Which benefits for which losses? *Earth System Dynamics*, 11(2), 537–562. https://doi.org/10.5194/esd-11-537-2020

Fyfe, J.C., Derksen, C., Mudryk, L., Flato, G.M., Santer, B.D., Swart, N.C. *et al.* (2017) Large near-term projected snowpack loss over the western United States. *Nature Communications*, 8(14996). 1–7. https://doi.org/10.1038/ncomms14996

Gasset, N., Fortin, V., Dimitrijevic, M., Carrera, M., Bilodeau, B., Muncaster, R. *et al.* (2021) A 10 km North American precipitation and land surface reanalysis based on the GEM atmospheric model. *Hydrology and Earth System Sciences*, 25(9), 4917–4945. https://doi.org/10.5194/hess-2021-41

Giuntoli, I., Villarini, G., Prudhomme, C. & Hannah, D.M. (2018) Uncertainties in projected runoff over the conterminous United States. *Climatic Change*, 150(3–4), 149–162. https://doi.org/10.1007/s10584-018-2280-5

Grenier, P. (2018) Two types of physical inconsistency to avoid with univariate quantile mapping: A case study over North America concerning relative humidity and its parent variables. *Journal of Applied Meteorology and Climatology*, 57(2), 347–364. https://doi.org/10.1175/jamc-d-17-0177.1

Gutowski Jr., W.J., Giorgi, F., Timbal, B., Frigon, A., Jacob, D., Kang, H.-S. *et al.* (2016) WCRP COordinated Regional Downscaling EXperiment (CORDEX): A diagnostic MIP for CMIP6. *Geoscientific Model Development*, 9(11), 4087–4095. https://doi.org/10.5194/gmd-9-4087-2016

Hao, Z. & Singh, V.P. (2020) Compound events under global warming: A dependence perspective. *Journal of Hydrologic Engineering*, 25(9), 03120001. https://doi.org/10.1061/(asce)he.1943-5584.0001991

Iizumi, T. (2018) *Bias-corrected d4PDF historical and non-warming climate data*. Kanagawa, Japan: Data Integration and Analysis System (DIAS). https://doi.org/10.20783/DIAS.544 Accessed 28 July 2021.

Iizumi, T., Takikawa, H., Hirabayashi, Y., Hanasaki, N. & Nishimori, M. (2017) Contributions of different bias-correction methods and reference meteorological forcing data sets to uncertainty in projected temperature and precipitation extremes. *Journal of Geophysical Research: Atmospheres*, 122(15), 7800–7819. https://doi.org/10.1002/2017jd026613

Jeong, D.I. & Sushama, L. (2018) Projected changes to extreme wind and snow environmental loads for buildings and infrastructure across Canada. *Sustainable Cities and Society*, 36, 225–236. https://doi.org/10.1016/j.scs.2017.10.004

King, A.D., Lane, T.P., Henley, B.J. & Brown, J.R. (2019) Global and regional impacts differ between transient and equilibrium warmer worlds. *Nature Climate Change*, 10(1), 42–47. https://doi.org/10.1038/s41558-019-0658-7

Kirchmeier-Young, M.C., Gillett, N.P., Zwiers, F.W., Cannon, A.J. & Anslow, F.S. (2019) Attribution of the influence of human-induced climate change on an extreme fire season. *Earth's Future*, 7, 2–10. https://doi.org/10.1029/2018ef001050

Kirchmeier-Young, M.C., Zwiers, F.W., Gillett, N.P. & Cannon, A.J. (2017) Attributing extreme fire risk in Western Canada to human emissions. *Climatic Change*, 144(2), 365–379. https://doi.org/10.1007/s10584-017-2030-0

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H. *et al.* (2015) The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, 93(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Lange, S. (2018) Bias correction of surface downwelling longwave and shortwave radiation for the EWEMBI dataset. *Earth System Dynamics*, 9(2), 627–645. https://doi.org/10.5194/esd-9-627-2018

Lange, S. (2019a) *EartH2Observe, WFDEI and ERA-Interim Data Merged and Bias-Corrected for ISIMIP (EWEMBI)*. Potsdam, Germany: GFZ Data Services. https://doi.org/10.5880/pik.2019.004 Accessed 28 July 2021.

Lange, S. (2019b) Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0). *Geoscientific Model Development*, 12(7), 3055–3070. https://doi.org/10.5194/gmd-12-3055-2019

Lange, S. (2019c) *WFDE5 Over Land Merged with ERA5 over the Ocean (W5E5)*. Potsdam, Germany: GFZ Data Services 1.0. https://doi.org/10.5880/pik.2019.023 Accessed 28 July 2021.

Lehner, F., Wood, A.W., Vano, J.A., Lawrence, D.M., Clark, M.P. & Mankin, J.S. (2019) The potential to reduce uncertainty in regional runoff projections from climate models. *Nature Climate Change*, 9(12), 926–933. https://doi.org/10.1038/s41558-019-0639-x

Leonard, M., Westra, S., Phatak, A., Lambert, M., van den Hurk, B., McInnes, K. *et al.* (2014) A compound event framework for understanding extreme impacts. *Wiley Interdisciplinary Reviews: Climate Change*, 5(1), 113–128. https://doi.org/10.1002/wcc.252

Liang, X.U., Lettenmaier, D.P., Wood, E.F. & Burges, S.J. (1994) A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, 99(D7), 14415. https://doi.org/10.1029/94jd00483

Liljegren, J.C., Carhart, R.A., Lawday, P., Tschopp, S. & Sharp, R. (2008) Modeling the wet bulb globe temperature using standard meteorological measurements. *Journal of Occupational and Environmental Hygiene*, 5(10), 645–655. https://doi.org/10.1080/15459620802310770

Magnusson, J., Wever, N., Essery, R., Helbig, N., Winstral, A. & Jonas, T. (2015) Evaluating snow models with varying process representations for hydrological applications. *Water Resources Research*, 51(4), 2707–2723. https://doi.org/10.1002/2014wr016498

Masterton, J.M. & Richardson, F.A. (1979) *Humidex: A Method of Quantifying Human Discomfort due to Excessive Heat and Humidity*. Downsview, ON: Environment Canada, Atmospheric Environment Service. pp. 45.

McGinnis, S. & Mearns, L. (2021) Building a climate service for North America based on the NA-CORDEX data archive. *Climate Services*, 22, 100233. https://doi.org/10.1016/j.cliser.2021.100233

Mearns, L.O., Arritt, R., Biner, S., Bukovsky, M.S., McGinnis, S., Sain, S. *et al.* (2012) The North American Regional Climate Change Assessment Program: Overview of phase I results. *Bulletin of the American Meteorological Society*, 93(9), 1337–1362. https://doi.org/10.1175/bams-d-11-00223.1

Melton, J.R., Arora, V.K., Wisernig-Cojoc, E., Seiler, C., Fortier, M., Chan, E.D. *et al.* (2020) CLASSIC v1.0: The open-source community successor to the Canadian Land Surface Scheme (CLASS) and the Canadian Terrestrial Ecosystem Model (CTEM) – Part 1: Model framework and site-level performance. *Geoscientific Model Development*, 13(6), 2825–2850. https://doi.org/10.5194/gmd-13-2825-2020

Mengel, M., Treu, S., Lange, S. & Frieler, K. (2021) ATTRICI 1.1 counterfactual climate for impact attribution. *Geoscientific Model Development*, 14, 5269–5284. https://doi.org/10.5194/gmd-14-5269-2021

Meyer, J., Kohn, I., Stahl, K., Hakala, K., Seibert, J. & Cannon, A.J. (2019) Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments. *Hydrology and Earth System Sciences*, 23(3), 1339–1354. https://doi.org/10.5194/hess-23-1339-2019

Mizukami, N., Clark, M.P., Slater, A.G., Brekke, L.D., Elsner, M.M., Arnold, J.R. *et al.* (2014) Hydrologic implications of different large-scale meteorological model forcing datasets in mountainous regions. *Journal of Hydrometeorology*, 15(1), 474–488. https://doi.org/10.1175/jhm-d-13-036.1

Mizuta, R., Murata, A., Ishii, M., Shiogama, H., Hibino, K., Mori, N. *et al.* (2017) Over 5,000 years of ensemble future climate simulations by 60-km global and 20-km regional atmospheric models. *Bulletin of the American Meteorological Society*, 98(7), 1383–1398. https://doi.org/10.1175/bams-d-16-0099.1

Mudryk, L.R., Derksen, C., Howell, S., Laliberté, F., Thackeray, C., Sospedra-Alfonso, R. *et al.* (2018) Canadian snow and sea ice: Historical trends and projections. *The Cryosphere*, 12(4), 1157–1176. https://doi.org/10.5194/tc-12-1157-2018

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G. *et al.* (2021) ERA5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9), 4349–4383. https://doi.org/10.5194/essd-2021-82

Najafi, M.R., Zwiers, F.W. & Gillett, N.P. (2015) Attribution of Arctic temperature change to greenhouse-gas and aerosol influences. *Nature Climate Change*, 5(3), 246–249. https://doi.org/10.1038/nclimate2524

Najafi, M.R., Zwiers, F.W. & Gillett, N.P. (2017) Attribution of observed streamflow changes in key British Columbia drainage basins. *Geophysical Research Letters*, 44(21). 11012–11020. https://doi.org/10.1002/2017gl075016

Nikulin, G. & Legutke, S. (2016) *Data Reference Syntax (DRS) for Bias-Adjusted CORDEX Simulations*. Norrköping, Sweden: WCRP Coordinated Regional Downscaling Experiment (CORDEX). https://is-enes-data.github.io/CORDEX_adjust_drs.pdf Accessed 28 July 2021.

Osczevski, R. & Bluestein, M. (2005) The new wind chill equivalent temperature chart. *Bulletin of the American Meteorological Society*, 86(10), 1453–1458. https://doi.org/10.1175/BAMS-86-10-1453

Philip, S., Sparrow, S., Kew, S.F., van der Wiel, K., Wanders, N., Singh, R. *et al*. (2019) Attributing the 2017 Bangladesh floods from meteorological and hydrological perspectives. *Hydrology and Earth System Sciences*, 23(3), 1409–1429. https://doi.org/10.5194/hess-23-1409-2019

Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B. *et al*. (2007) Development of the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale. *Hydrology and Earth System Sciences*, 11(4), 1279–1294. https://doi.org/10.5194/hess-11-1279-2007

Raftery, A.E., Zimmer, A., Frierson, D.M.W., Startz, R. & Liu, P. (2017) Less than 2C warming by 2100 unlikely. *Nature Climate Change*, 7(9), 637–641. https://doi.org/10.1038/nclimate3352

Schwalm, C.R., Glendon, S. & Duffy, P.B. (2020) RCP8.5 tracks cumulative CO_2 emissions. *Proceedings of the National Academy of Sciences of the United States of America*, 117(33), 19656–19657. https://doi.org/10.1073/pnas.2007117117

Scinocca, J.F., Kharin, V.V., Jiao, Y., Qian, M.W., Lazare, M., Solheim, L. *et al*. (2016) Coordinated global and regional climate modeling. *Journal of Climate*, 29(1), 17–35. https://doi.org/10.1175/jcli-d-15-0161.1

Seneviratne, S.I., Donat, M.G., Pitman, A.J., Knutti, R. & Wilby, R.L. (2016) Allowable CO_2 emissions based on regional and impact-related climate targets. *Nature*, 529(7587), 477–483. https://doi.org/10.1038/nature16542

Shrestha, R.R., Bonsal, B.R., Bonnyman, J.M., Cannon, A.J. & Najafi, M.R. (2021) Heterogeneous snowpack response and snow drought occurrence across river basins of Northwestern North America under 1.0 to 4.0C global warming. *Climatic Change*, 164(3–4), 3–4. https://doi.org/10.1007/s10584-021-02968-7

Shrestha, R.R., Cannon, A.J., Schnorbus, M.A. & Alford, H. (2019) Climatic controls on future hydrologic changes in the subarctic river basin in Canada. *Journal of Hydrometeorology*, 20(9), 1757–1778. https://doi.org/10.1175/JHM-D-18-0262.1

Shrestha, R.R., Cannon, A.J., Schnorbus, M.A. & Zwiers, F.W. (2017) Projecting future nonstationary extreme streamflow for the Fraser River Canada. *Climatic Change*, 145(3–4), 289–303. https://doi.org/10.1007/s10584-017-2098-6

Sillmann, J., Kharin, V.V., Zhang, X., Zwiers, F.W. & Bronaugh, D. (2013) Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research: Atmospheres*, 118(4), 1716–1733. https://doi.org/10.1002/jgrd.50203

Singh, H. & Najafi, M.R. (2020) Evaluation of gridded climate datasets over Canada using univariate and bivariate approaches: Implications for hydrological modelling. *Journal of Hydrology*, 584(May), 124673. https://doi.org/10.1016/j.jhydrol.2020.124673

Singh, H., Najafi, M.R. & Cannon, A.J. (2021) Characterizing nonstationary compound extreme events in a changing climate based on large-ensemble climate simulations. *Climate Dynamics*, 56(5), 1389–1405. https://doi.org/10.1007/s00382-020-05538-2

Sørland, S.L., Fischer, A.M., Kotlarski, S., Künsch, H.R., Liniger, M.A., Rajczak, J. *et al*. (2020) CH2018 National Climate Scenarios for Switzerland: How to construct consistent multi-model projections from ensembles of opportunity. *Climate Services*, 20(December), 100196. https://doi.org/10.1016/j.cliser.2020.100196

Su, T., Chen, J., Cannon, A.J., Xie, P. & Guo, Q. (2020) Multisite bias correction of climate model outputs for hydro-meteorological impact studies: an application over a watershed in China. *Hydrological Processes*, 34, 2575–2598. https://doi.org/10.1002/hyp.13750.

Thornton, M.M., Shrestha, R., Wei, Y., Thornton, P.E., Kao, S. & Wilson, B.E. (2020) *Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 4*. Oak Ridge, TN: ORNL DAAC. https://doi.org/10.3334/ORNLDAAC/1840

Thrasher, B., Maurer, E.P., McKellar, C. & Duffy, P.B. (2012) Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrology and Earth System Sciences*, 16(9), 3309–3314. https://doi.org/10.5194/hess-16-3309-2012

Van Wagner, C.E. (1987) *Development and Structure of the Canadian Forest Fire Weather Index System*. Forest Technology Report 35. Ottawa, ON: Canadian Forestry Service. 37 pp. http://cfs.nrcan.gc.ca/pubwarehouse/pdfs/19927.pdf Accessed 28 July 2021

Vial, J., Dufresne, J.-L. & Bony, S. (2013) On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates. *Climate Dynamics*, 41(11–12), 3339–3362. https://doi.org/10.1007/s00382-013-1725-9

von Salzen, K., Scinocca, J.F., McFarlane, N.A., Li, J., Cole, J.N.S., Plummer, D. *et al*. (2013) The Canadian Fourth Generation Atmospheric Global Climate Model (CanAM4). Part I: Representation of physical processes. *Atmosphere-Ocean*, 51(1), 104–125. https://doi.org/10.1080/07055900.2012.755610

Walter, M.T., Brooks, E.S., McCool, D.K., King, L.G., Molnau, M. & Boll, J. (2005) Process based snowmelt modeling: Does it require more input data than temperature index modeling? *Journal of Hydrology*, 300, 65–75. https://doi.org/10.1016/j.jhydrol.2004.05.002

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O. & Schewe, J. (2013) The Inter-Sectoral Impact Model Intercomparison Project (ISIMIP): Project framework. *Proceedings of the National Academy of Sciences of the United States of America*, 111(9), 3228–3232. https://doi.org/10.1073/pnas.1312330110

Whan, K., Zscheischler, J., Jordan, A.I. & Ziegel, J.F. (2021) Novel multivariate quantile mapping methods for ensemble post-processing of medium-range forecasts. *Weather and Climate Extremes*, 32(June), 100310. https://doi.org/10.1016/j.wace.2021.100310

Whan, K. & Zwiers, F. (2015) Evaluation of extreme rainfall and temperature over North America in CanRCM4 and CRCM5.

Climate Dynamics, 46(11–12), 3821–3843. https://doi.org/10.1007/s00382-015-2807-7

Whan, K. & Zwiers, F. (2016) The impact of ENSO and the NAO on extreme winter precipitation in North America in observations and regional climate models. *Climate Dynamics*, 48(5–6), 1401–1411. https://doi.org/10.1007/s00382-016-3148-x

Wong, J.S., Razavi, S., Bonsal, B.R., Wheater, H.S. & Asong, Z.E. (2017) Inter-comparison of daily precipitation products for large-scale hydro-climatic applications over Canada. *Hydrology and Earth System Sciences*, 21(4), 2163–2185. https://doi.org/10.5194/hess-21-2163-2017

Zscheischler, J., Fischer, E.M. & Lange, S. (2019) The effect of univariate bias adjustment on multivariate hazard estimates. *Earth System Dynamics*, 10(1), 31–43. https://doi.org/10.5194/esd-10-31-2019

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R.M. *et al.* (2020) A typology of compound weather and climate events. *Nature Reviews Earth & Environment*, 1(7), 333–347. https://doi.org/10.1038/s43017-020-0060-z

Zscheischler, J., Westra, S., van den Hurk, B.J.J.M., Seneviratne, S.I., Ward, P.J., Pitman, A. *et al.* (2018) Future climate risk from compound events. *Nature Climate Change*, 8(6), 469–477. https://doi.org/10.1038/s41558-018-0156-3

How to cite this article: Cannon, A.J., Alford, H., Shrestha, R.R., Kirchmeier-Young, M.C. & Najafi, M.R. (2022) Canadian Large Ensembles Adjusted Dataset version 1 (CanLEADv1): Multivariate bias-corrected climate model outputs for terrestrial modelling and attribution studies in North America. *Geoscience Data Journal*, 9, 288–303. Available from: https://doi.org/10.1002/gdj3.142