Spatial datasets of CMIP6 climate change projections for Canada and the United States

Daniel W. McKenney, John H. Pedlar, Kevin Lawrence, Kaitlin DeBoer, & Tiziana Brescacin

2025

Pacific Climate Impacts Consortium (PCIC)

PCIC Publications

Original citation:

McKenney, D. W., Pedlar, J. H., Lawrence, K., Sobie, S. R., DeBoer, K., & Brescacin, T. (2025). Spatial datasets of CMIP6 climate change projections for Canada and the United States. *Data in Brief*, 58, 111246. https://doi.org/10.1016/j.dib.2024.111246

Data in Brief 58 (2025) 111246

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Spatial datasets of CMIP6 climate change projections for Canada and the United States

Daniel W. McKenney[a], John H. Pedlar[a,*], Kevin Lawrence[a], Stephen R. Sobie[b], Kaitlin DeBoer[a], Tiziana Brescacin[a]

[a] *Great Lakes Forestry Centre, Natural Resources Canada, 1219 Queen Street East, Sault Ste Marie, Ontario P6A 2E5, Canada*
[b] *Pacific Climate Impacts Consortium, University of Victoria, Victoria, British Columbia, Canada*

ARTICLE INFO

Article history:
Received 12 November 2024
Revised 10 December 2024
Accepted 17 December 2024
Available online 19 December 2024

Dataset link: Spatial datasets of CMIP6 climate change projections for Canada and the United States (Original data)

Keywords:
Climate change
Geospatial data
General circulation models
Shared socio-economic pathways
CMIP6
Downscaling
Canada
United States

ABSTRACT

Geospatial climate change projections are critical for assessing climate change impacts and adaptations across a wide range of disciplines. Here we present monthly-based grids of climate change projections at a 2-km resolution covering Canada and the United States. These data products are based on outputs from the 6th Coupled Model Intercomparison Project (CMIP6) and include projections for 13 General Circulation Models (GCMs), three Shared Socioeconomic Pathways (SSP1 2.6, SSP2 4.5, and SSP5 8.5), four 30-year time periods (2011–2040, 2021–2050, 2041–2070, and 2071–2100), and a suite of climate variables, including monthly maximum and minimum temperature, precipitation, climate moisture index, and various bioclimatic summaries. The products employ a delta downscaling method, which combines historical normal values at climate stations with broad-scale change projections (or deltas) from GCMs, followed by spatial interpolation using ANUSPLIN. Various quality control efforts, described herein, were undertaken to ensure that the final products provided reasonable estimates of future climate.

---

[*] Corresponding author.
E-mail address: john.pedlar@nrcan-rncan.gc.ca (J.H. Pedlar).

https://doi.org/10.1016/j.dib.2024.111246

Specifications Table

Subject	Earth and Plentary Sciences; Atmospheric Science
Specific subject area	Geospatial datasets of future climate projections for Canada and the United States at approximately 2-km resolution
Type of data	GeoTIFF spatial datasets
Data collection	Raw climate change projections were generated by General Circulation Model (GCM) teams around the world and made available at the Coupled Model Intercomparison Project v6 (CMIP6). Initial processing of these data were carried out by the Pacific Climate Impacts Consortium (PCIC), who provided the data as differences (or deltas) relative to the 1971–2000 period. Climate station data for the 1971–2000 normal period, which were required to downscale the GCM projections, were obtained from the North American monthly historical dataset (Northam version j).
Data source location	Canada and the United States
Data accessibility	Repository name: NRCan Climate Archives Data identification number: 10.23687/2921339c-f5a0-4407-a243-663c2cddf78d Direct URL to data: https://ftp.maps.canada.ca/pub/nrcan_rncan/Climate-archives_Archives-climatologiques/GCM/cmip6/

1. Value of the Data

- The grids of climate change projections presented here can be used in a variety of disciplines including agriculture, forestry, ecology, engineering, and urban planning.
- These products cover Canada and the United States at a 2-km resolution and a monthly time step for four future normal periods (2011–2040, 2021–2050, 2041–2070, and 2071–2100) – making them useful for many applications, while requiring modest levels of storage and memory.
- Many climate variables are provided, including monthly minimum temperature, maximum temperature, precipitation, and climate moisture index – as well as a suite of bioclimatic variables.
- The grids were downscaled using a delta method based on normal values at climate stations, which, by incorporating historical spatial and temporally varying lapse rates, provides an alternative product to downscaling using regional climate models.
- Quality control measures were undertaken to ensure the grids provide reasonable estimates of climate through time and space.

2. Background

This work was motivated by updated GCM projections that were recently made available as part of the CMIP6 experiment [1]. These projections incorporated recent GCM improvements and the latest round of emissions scenario narratives – termed shared socioeconomic pathways (SSPs) [2]. Though several downscaled spatial products from CMIP6 already exist for North America [3,4], we are unaware of any that cover the range of variables and time steps provided here. These products update previous work [5] and fill a niche as monthly, moderate resolution grids, that can be used in a wide variety of applications.

3. Data Description

Spatial data files are available in a single folder at the data storage location. Using the file EC-Earth3-Veg126cmi_2021-2050_01.tif as an example, the naming convention is as follows:

- "EC-Earth3-Veg" identifies the GCM

Table 1
The 13 general circulation models (GCM) used to develop the spatial climate data described here.

GCM	Country of Origin	Spatial Resolution (lat x long, degrees)
BCC-CSM2-MR	China	1.1° × 1.1°
CMCC-ESM2	Italy	0.9° x 1.25°
EC-Earth3-Veg	Europe	0.7° × 0.7°
FGOALS-g3	China	2° × 2.3°
INM-CM5-0	Russia	2° × 1.5°
IPSL-CM6A-LR	France	2.5° × 1.3°
MIROC-ES2L	Japan	2.8° × 2.8°
MPI-ESM1-2-HR	Germany	0.9° × 0.9°
MRI-ESM2-0	Japan	1.1° × 1.1°
NorESM2-LM	Norway	2.5° × 1.9°
TaiESM1	Taiwan	0.9° x 1.25°
UKESM1-0-LL	England	1.9° × 1.3°
CanESM5-1	Canada	2.8° × 2.8°

- "126" identifies the SSP
- "cmi" identifies the climate variable (climate moisture index in this case)
- "2021–2050" identifies the 30-year climate normal period
- "01" identifies the month of the year (January in this case)

More than 30 teams around the world have developed GCMs for projecting future global climate. Outputs from these models are made available as part of the CMIP experiments, the most recent being CMIP6. The various GCMs differ, to varying degrees, in how they model earth system responses to the radiative forcing caused by greenhouse gas emissions. A subset of 12 GCMs was selected for the current modelling exercise; this subset has been shown to be representative of the range of variation in future projections across North America [6]. A composite (or average) of these 12 models is included in our data products. Projections from the Canadian Earth System Model (CanESM5), which was not part of the original 12-model subset but was of interest in the Canadian context, are also included in our data package. The names of the GCMs used in this study, and related details, are listed in Table 1.

Shared socio-economic pathways characterize possible future development pathways for human society [2]. Each SSP is associated with several greenhouse gas emission pathways, which were used to drive the GCMs involved in the CMIP6 experiment. The scenarios included in our data products are SSP1-2.6, SSP2-4.5, and SSP5-8.5. Specifically, SSP1-2.6 simulates a world that focuses on sustainability, mitigation, and adaptation, resulting in low greenhouse gas emissions and a radiative forcing of 2.6 W/m^2 by 2100; SSP2-4.5 represents a moderate emissions future, wherein social policies do not change markedly from the past, though a transition towards greener energy sources limits radiative forcing to 4.5 W/m^2; and SSP5–8.5 simulates a future with ongoing fossil-fuelled development that pushes radiative forcing levels to 8.5 W/m^2.

Monthly average maximum and minimum temperature and total precipitation are widely used variables for climate change assessments. Minimum and maximum temperature variables were processed prior to receiving the data and were calculated as monthly averages of daily minimum and maximum temperature values respectively. Precipitation was calculated as the monthly sum of daily precipitation values. The climate moisture index (cmi) was calculated as the difference between monthly precipitation and potential evapotranspiration (the potential loss of water vapour from a landscape covered by vegetation). Monthly PET values were calculated using a simplified Penman-Monteith equation, which requires only mean monthly maximum and minimum temperature as detailed in [7]. Values of cmi typically range between −20 cm to 20 cm, where negative values are indicative of dry conditions. A suite of bioclimatic and growing season-related variables was also derived from the monthly temperature and precipitation models (see Table 2 for listing of all variables); see [8] for detailed variable definitions.

Table 2
Climate variables included in the geospatial datasets of future climate projections.

Climate Variable	Abbreviation	Units	Time Step
Minimum temperature	mint	°C	monthly
Maximum temperature	maxt	°C	monthly
Precipitation	pcp	mm	monthly
Climate Moisture Index	cmi	cm	monthly
annual mean temperature	bio_01	°C	annual
mean diurnal range	bio_02	°C	annual
isothermality	bio_03	°C	annual
temperature seasonality	bio_04	°C	annual
max temperature of warmest month	bio_05	°C	annual
min temperature of coldest month	bio_06	°C	annual
temperature annual range	bio_07	°C	annual
mean temperature of wettest quarter	bio_08	°C	annual
mean temperature of driest quarter	bio_09	°C	annual
mean temperature of warmest quarter	bio_10	°C	annual
mean temperature of coldest quarter	bio_11	°C	annual
annual precipitation	bio_12	mm	annual
precipitation of wettest month	bio_13	mm	annual
precipitation of driest month	bio_14	mm	annual
precipitation seasonality	bio_15	mm	annual
precipitation of wettest quarter	bio_16	mm	annual
precipitation of driest quarter s	bio_17	mm	annual
precipitation of warmest quarter	bio_18	mm	annual
precipitation of coldest quarter	bio_19	mm	annual
Growing season start	sg_01	day	annual
Growing season end	sg_01	day	annual
Growing season length	sg_02	days	annual
total precipitation period 1	sg_03	mm	annual
total precipitation period 2	sg_04	mm	annual
total precipitation period 3	sg_05	mm	annual
total precipitation period 4	sg_06	mm	annual
Growing degree days period 1	sg_07	°C	annual
Growing degree days period 2	sg_08	°C	annual
Growing degree days period 3	sg_09	°C	annual
Growing degree days period 4	sg_11	°C	annual
annual minimum temperature	sg_12	°C	annual
annual maximum temperature	sg_13	°C	annual
mean temperature period 3	sg_14	°C	annual
temperature range period 3	sg_15	°C	annual

A sample of our data products is provided in Fig. 1, which presents maps of maximum temperature in July for 4, 30-year periods. The grid for the 1971–2000 period was produced in a previous study [8], but is presented here for comparison with the future projections. The maps show typical patterns of spatial variation, with temperatures varying with respect to latitude, altitude, and proximity to large water bodies. Climate change impacts are also apparent, with increases of 5°C, or more, across much of the study area by the end of the current century.

We also illustrate variation between GCM outputs over time for a sample of climate variables and months under SSP2-4.5 (Fig. 2). The datapoints in Fig. 2 are averages across the spatial domain of the data for each GCM and time period. These plots illustrate the considerable variation across individual GCMs with respect to temperature and precipitation projections.

4. Experimental Design, Materials and Methods

4.1. Data sourcing

Datasets containing differences (or deltas) relative to the 1971–2000 period for each of the 13 selected GCMs were obtained from the Pacific Climate Impacts Consortium

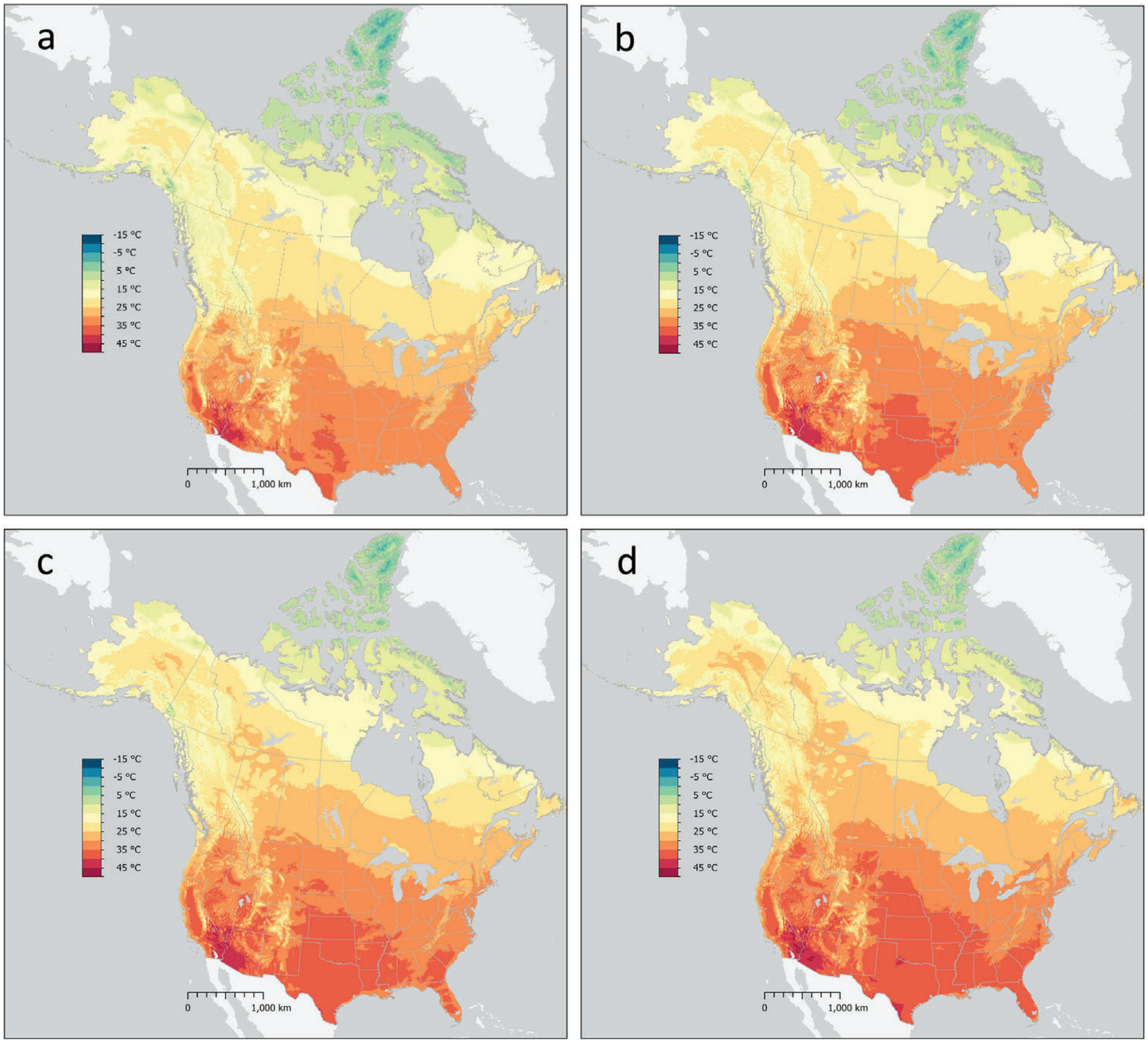

Fig. 1. Average July maximum temperature for a) 1971–2000, b) 2011–2040, c) 2041–2070, and d) 2071–2100. Future projections are based on a composite of the first 12 GCMs listed in Table 1 and the SSP2-4.5 emissions scenario.

(PCIC; https://www.pacificclimate.org/) [6]. The files were provided in NetCDF format and consisted of global estimates for the 3 primary variables of interest (monthly minimum and maximum temperature and precipitation) at a regular grid of locations defined by latitude and longitude variables in decimal degrees. Delta values for the temperature variables were calculated at each grid cell as differences (in °C):

$$\Delta\ T_{ijk} = T_{ijk} - T_{ink} \tag{1}$$

where T is average temperature, i is month, j is year, k is GCM, and n is normal period (i.e., 1971–2000). Alternatively, precipitation deltas were calculated as ratios:

$$\Delta\ P_{ijk} = P_{ijk}/P_{ink} \tag{2}$$

where P is total precipitation and i, j, k, and n are as defined above. Note that normal values were calculated by averaging historical simulations for each GCM at each grid cell over the 1971–2000 period.

Historical monthly temperature and precipitation data at climate stations across Canada and the United States were obtained from the North American monthly historical dataset (Northam version “j”; [9]) published by the National Oceanic and Atmospheric Administration’s (NOAA’s) National Centres for Environmental Information (NCEI).

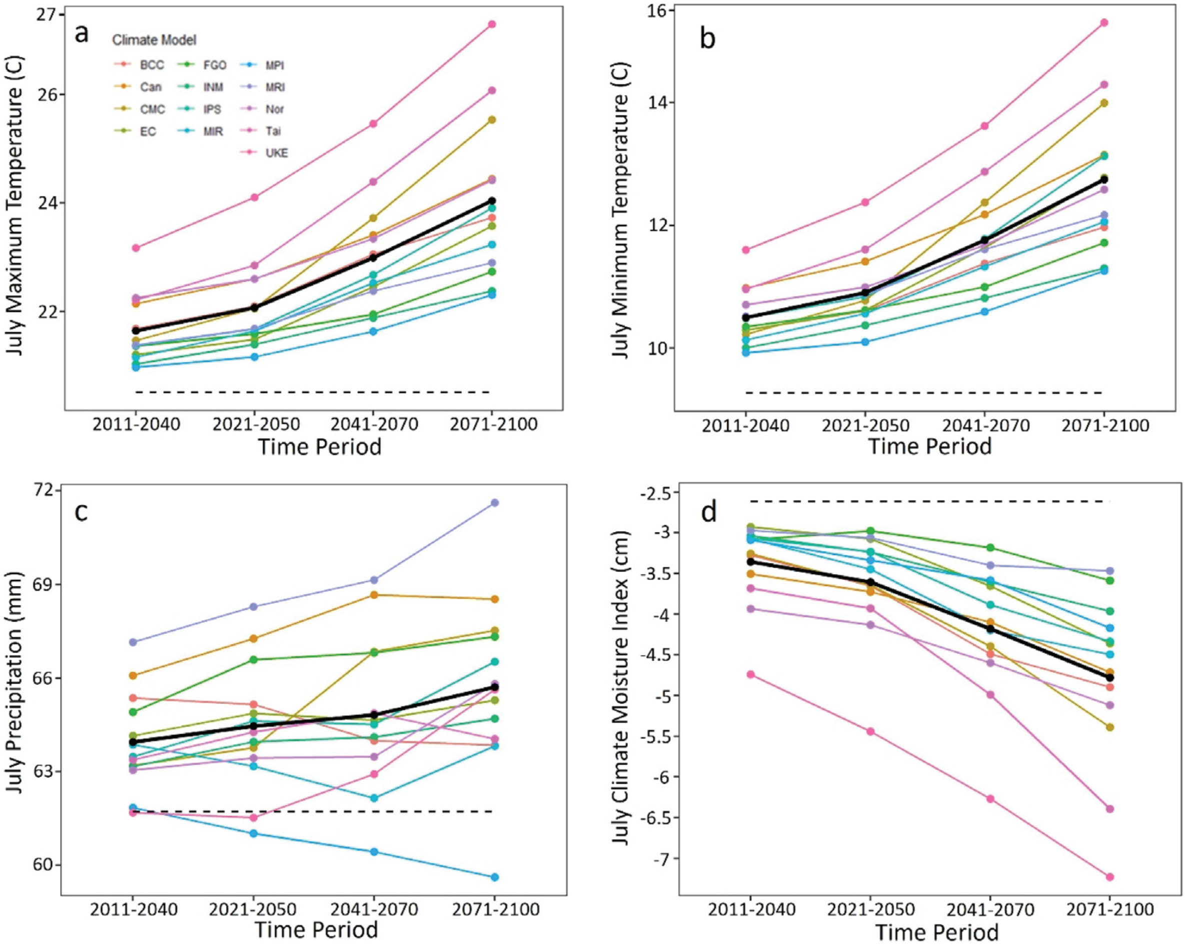

Fig. 2. Variation across 13 GCMs under the SSP2-4.5 emissions scenario and 4 time periods for a) July maximum temperature, b) July minimum temperature, c) July precipitation, and d) July climate moisture index. The average (or composite) value is shown as a solid black line and the baseline (1971–2000) value, obtained from historical climate stations, is shown as a dashed black line. Full GCM names are provided in Table 1.

4.2. Processing of GCM data

The spatial resolution of the raw GCM outputs varied from 0.7° to 2.8° (approximately 80 to 300 km) (Table 1). Though at a finer spatial resolution than previous releases, these outputs are still too coarse for most applications. Our approach for downscaling the GCM delta values involved the use of thin plate smoothing splines as formulated in the ANUSPLIN software [10]. These interpolations, which included both the raw delta values and the station network of climate change estimates, employed trivariate spline models with latitude, longitude, and scaled elevation as independent variables. Subsequent interrogation of the ANUSPLIN surface of delta values at climate station locations was undertaken using the SELNOT package in ANUSPLIN. Gridding of the final models was carried out by supplying a digital elevation model (DEM) at a 60 arcsec (approximately 2-km) resolution [11] within the ANUSPLIN workflow.

The procedure for downscaling the GCMs involved the following steps: 1) average the annual delta values over future normal period of interest; 2) create ANUSPLIN surface of these averaged delta values; 3) interrogate this surface at climate station locations across North America; 4) for each climate station, add the interrogated delta value to the 1971–2000 climate station normal value (or multiply the two values in the case of precipitation) to create a network of stations with future climate estimates; 5) create ANUSPLIN surface from network of future climate estimates; 6) grid ANUSPLIN surface at approximately 2-km resolution; and 7) repeat for each GCM, SSP, time period, climate variable, and month of interest. Composite grids, which are averages of

Table 3
Climate variable thresholds based on historical ranges. Any grids with values outside the ranges specified were inspected further.

Climate variable	Range
Maximum temperature (°C)	−42 to 52
Minimum temperature (°C)	−50 to 35
Total Monthly Precipitation (mm)	0–1000
Climate Moisture Index (cm)	−45 to 110

the first 12 GCMs in Table 1, were generated by averaging the 12 GCM estimates at each climate station in step 4 above; this network of averaged future estimates was then interpolated and gridded as outlined in steps 5 and 6 above.

Climate station averages for the 1971–2000 period were calculated from historical station data provided in the National Oceanic and Atmospheric Administration's (NOAA) Northam 'j' dataset [9]. Prior to the calculation, station-months with more than 11 missing days were first screened out of the analysis and then only stations with 10 or more years of values over the 1971–2000 period were included in the analysis. With these quality controls in place, normal values were calculated at a total of 8808 stations for temperature and 11,527 stations for precipitation across Canada and the United States.

4.3. Model assessment and quality control

Though not possible to test future climate projections for accuracy per se, we carried out a number of assessments to ensure that the gridded values appeared reasonable. The objective of these assessments was to identify grids of concern, which were then visually assessed for potential problems. In the event that errors were detected, efforts were made to identify and correct problematic processes and/or data, and the grids were re-generated. All testing was carried out in R [12], with Terra [13] as the primary package for raster operations.

As a first test, thresholds were set for each climate variable based on historical ranges, with adjustments for anticipated climate change (Table 3). Grids containing values outside threshold limits were submitted for further assessment.

As a second test, we compared each CMIP6 grid to a comparable CMIP5 grid developed previously using similar methodology [5]. These grids were overlaid, differences were calculated, and a suite of statistics were taken on the differences, including mean, minimum, maximum, and various percentiles. A Pearson Correlation Coefficient (r) was also calculated to assess the degree of correlation between the two grids. Histograms of the various difference statistics (particularly mean, min, and max) were used to quickly identify outliers, which indicated CMIP6 maps that differed greatly from their predecessors and thus required further visual inspection. Furthermore, CMIP6 grids with Pearson Correlation Coefficients < 0.95 were flagged for further inspection.

As a final means of outlier detection, the sets of statistics associated with the CMIP6 and difference grids were tested using Rosner methodology [14] as incorporated in the EnvStats package in R [15]. The Rosner test was used to detect outliers at the 95% significance level. Compared to other formal techniques that can only test one outlier at a time, the Rosner method detects multiple outliers at once. Importantly, the technique overcomes the concern of masking, where one outlier masks another since they are close in value. The function determines outliers based on a detection parameter, k, which sets the number of observations to test. Initially, k was set to 20 then increased iteratively to ensure all observations suspected as outliers were detected successfully. Statistics where the number of outliers exceeded 20 were further investigated through visual inspection of the histograms and/or the climate grids.

Limitations

The products provided here rely on historical climate station data as part of the process for downscaling the raw GCM delta values. Consequently, regions with limited numbers of climate stations – such as northern Canada – may have less reliable estimates than data-rich portions of the study area. Further, the delta downscaling approach employed in the current study implicitly assumes that existing climate relationships will remain constant in the future, which may limit the accuracy of future projections [16].

Ethics Statement

The authors have read the ethical requirements for publication in Data in Brief and confirm that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

Data Availability

Spatial datasets of CMIP6 climate change projections for Canada and the United States (Original data) (NRCAN Climate Archives)

CRediT Author Statement

Daniel W. McKenney: Conceptualization, Supervision, Project administration, Methodology, Writing – review & editing; **John H. Pedlar:** Methodology, Validation, Writing – original draft, Writing – review & editing; **Kevin Lawrence:** Methodology, Software, Writing – review & editing; **Stephen R. Sobie:** Data curation, Visualization, Writing – review & editing; **Kaitlin DeBoer:** Methodology, Data curation, Writing – review & editing; **Tiziana Brescacin:** Validation, Visualization, Writing – original draft, Writing – review & editing.

Acknowledgments

We thank David Price and Pia Papadopol for earlier support in developing the data and methodologies applied in the current work. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] V. Eyring, S. Bony, G.A. Meehl, C. Senio, B. Stevens, R.J. Stouffer, K.E. Taylor, Overview of the coupled model intercomparison project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev. 9 (5) (2016) 1937–1958, doi:10.5194/gmd-9-1937-2016.

[2] K. Riahi, D.P. Van Vuuren, E. Kriegler, J. Edmonds, B.C. O'neill, S. Fujimori, N. Bauer, K. Calvin, R. Dellink, O. Fricko, W. Lutz, A. Popp, J.C. Cuaresma, K.C. S, M. Leimbach, L. Jiang, T. Kram, S. Rao, K.Ebi J.Emmerling, T. Hasegawa, P. Havlik, F. Humpenoder, L.A. Da Silva, S. Smith, E. Stehfest, V. Bosetti, J. Eom, D. Gernaat, T. Masui, J. Rogelj, J. Strefler, L. Drouet, V. Krey, G. Luderer, M. Harmsen, K. Takahashi, L. Baumstark, J.C. Doelman, M. Kainuma, Z. Klimont,

G. Marangoni, H. Lotze-Campen, M. Obersteiner, A. Tabeau, M. Tavoni, The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview, Glob. Environ. Change 42 (2017) 153–168, doi:10.1016/j.gloenvcha.2016.05.009.

[3] S.R. Sobie, D. Ouali, C.L. Curry, F.W. Zwiers, Multivariate Canadian downscaled climate scenarios for CMIP6 (CanDCS-M6), Geosci. Data J. 11 (4) (2024) 806–824, doi:10.1002/gdj3.257.

[4] C.R. Mahony, T. Wang, A. Hamann, A.J. Cannon, A global climate model ensemble for downscaled monthly climate normals over North America, Int. J. Climatol. 42 (11) (2022) 5871–5891, doi:10.1002/joc.7566.

[5] D.T. Price, D.W. McKenney, L.A. Joyce, R.M. Siltanen, P. Papadopol, K. Lawrence, High-resolution interpolation of climate scenarios for Canada derived from General Circulation Model simulations, Nat. Resour. Can., Can. For. Serv., North. For. Cent. (2011) Rep. NOR-X-421.

[6] C.L. Curry, D. Ouali, S.R. Sobie, and F.W. Zwiers, Downscaled CMIP6 Climate Model Subset Selection, Pacific Climate Impacts Consortium, University of Victoria, Victoria, 2023.

[7] E.H. Hogg, Climate and the southern limit of the western Canadian boreal forest, Can. J. For. Res. 24 (9) (1994) 1835–1845, doi:10.1139/x94-237.

[8] D.W. McKenney, M.F. Hutchinson, P. Papadopol, K. Lawrence, J.H. Pedlar, K. Campbell, E. Milewska, R.F. Hopkinson, D. Price, T. Owen, Customized spatial climate models for North America, Bull. Am. Meteorol. Soc. 92 (2011) 1611, doi:10.1175/BAMS-D-10-3132.1.

[9] M.J. Menne, C.N. Williams, B.E. Gleason, J.J. Rennie, J.H. Lawrimore, The global historical climatology network monthly temperature dataset, version 4, J. Clim. 31 (24) (2018) 9835–9854, doi:10.1175/JCLI-D-18-0094.1.

[10] M. Hutchinson, T. Xu, ANUSPLIN Version 4.4 User Guide, The Australian National University, Fenner School of Environment and Society, Canberra, 2013.

[11] K.M. Lawrence, M.F. Hutchinson, D.W. McKenney, Multi-scale Digital Elevation Models for Canada, Natural Resources Canada, Great Lakes Forestry Centre, Sault Ste Marie, 2008.

[12] R Core TeamR: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, 2021.

[13] R.J. Hijmans, Terra: spatial data analysis, R package version 1 (2023) 7–23.

[14] B. Rosner, Percentage points for a generalized ESD many-outlier procedure, Technometrics 25 (1983) 165–172.

[15] S.P. Millard, EnvStats: An R Package for Environmental Statistics, Springer, New York, 2013.

[16] L.E. Hay, R.L. Wilby, G.H. Leavesley, A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States, JAWRA J. Am. Water Resourc. Assoc. 36 (2) (2000) 387–397, doi:10.1111/j.1752-1688.2000.tb04276.x.