



## Data Article

## Heating degree day spatial datasets for Canada



Heather MacDonald<sup>a,1,\*</sup>, John Pedlar<sup>a,1</sup>, Daniel W. McKenney<sup>a,1</sup>,  
Kevin Lawrence<sup>a,1</sup>, Kaitlin de Boer<sup>a,1</sup>, Michael F. Hutchinson<sup>b,1</sup>

<sup>a</sup> Great Lakes Forestry Centre, Canadian Forest Service, Natural Resources Canada, P6A 2E5 1219 Queen Street East, Sault Ste. Marie, Ontario, Canada

<sup>b</sup> Fenner School of Environment and Society, Australian National University, Australia

## ARTICLE INFO

## Article history:

Received 20 June 2023

Revised 10 July 2023

Accepted 21 July 2023

Available online 26 July 2023

Dataset link: [Heating Degree Days Canada 1951-1980, 1981-2010 \(Original data\)](#)

## Keywords:

Grids

Raster

Temperature

Spatial datasets

Heating degree days

HDD

Historical

Thin plate spline

Climate

ANUSPLIN

Canada

## ABSTRACT

Heating degree days (HDD) represent a concise measure of heating energy requirements used to inform decision making about the impact of climate change on heating energy demand. This data paper presents spatial datasets of heating degree days (HDD) for Canada for two thirty-year periods, 1951–1980 and 1981–2010, using daily temperature gauge observations over these time periods. Stations with fewer than nine missing days in a year and greater than nine years of data over each thirty-year period were included, resulting in 1339 and 1679 stations for the 1951–1980 and 1981–2010 periods respectively. Mean absolute error (MAE) of the spatial models ranged from 124.2 Celsius degree days (C-days) for the 1951–1980 model (2.4% of the surface mean) to 137.6 C-days for the 1981–2010 model (2.7%). This note presents maps illustrating cross validation errors at a set of representative stations. The grids are available at ~2 km resolutions.

Crown Copyright © 2023 Published by Elsevier Inc.  
This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

\* Corresponding author.

E-mail address: [heather.macdonald@nrcan-rncan.gc.ca](mailto:heather.macdonald@nrcan-rncan.gc.ca) (H. MacDonald).

<sup>1</sup> @socialscfgood

## Specifications Table

Subject	Earth and Planetary Sciences
Specific subject area	Thin plate spline datasets for heating degree days, Canada, spatial dataset, 1981–2010 & 1951–1980
Type of data	Geospatial grids
How data were acquired	Environment and Climate Change Canada (ECCC) provided daily minimum and maximum temperature values at meteorological stations across Canada (1950–2010).
Data format	Raw – delimited text/ascii Analysed – delimited text/ascii Final - geotiff
Description of data collection	Climate data in Canada are collected through a system of weather stations distributed unevenly across the country. Daily minimum and maximum temperature values from 1339 (1951–1980) and 1679 (1981–2010) weather stations were used to calculate HDD values for 1951–1980 and 1981–2010, which were then interpolated and mapped using ANUSPLIN via tri-variate thin-plate splines.
Data source location	Canada
Data accessibility	<a href="https://osf.io/xkpc7/">https://osf.io/xkpc7/</a>

## Value of the Data

- These datasets were developed in part to support updates to tax credits for northern and isolated areas in Canada for the Canadian Finance Department [1].
- Energy analyses rely on HDD to track changes in natural gas and other energy usage. Historical change in HDD is an important factor in energy consumption planning, particularly in northern areas.
- Users can use the dataset to obtain information about heating requirements for any location in Canada for two long-term periods, 1951–1980 and 1981–2010.
- This data description also provides a case study using published output from ANUSPLIN thin-plate spline program [2].

## 1. Objective

The ‘degree day’ method is used to calculate the difference between mean daily temperature and any given threshold – typically these differences are summed over a period of interest to provide a measure of heat or cold accumulation through time [3]. Heating degree days (HDD) sum the degree to which average daily temperatures are below the temperature of human comfort, defined as 65°F [4], or in Canada as 18 °C [3–6]. HDD have been analysed to estimate changes in energy usage [7], impacts of climate change [8–9], and historical trends [9] with respect to how often and how hard a furnace must work to keep a house warm.

The purpose of this brief report is to introduce HDD datasets for Canada for the 1951–1980 and 1981–2010 periods. These datasets were developed in part to support updates to tax benefits for northern and isolated areas in Canada which experience higher than average heating costs in Canada [1]. We describe these datasets and report on the quality and accuracy of the spatial datasets.

## 2. Data Description

### 2.1. Heating degree day (HDD) datasets

Canada-wide Heating Degree Day (HDD) gridded datasets were generated for two thirty-year periods, 1951–1980 and 1981–2020 (Fig. 1), using tri-variate thin-plate splines in ANUS-

PLIN [2] version 4.5 employing a 60' sec (approximately 2 km) Digital Elevation Model [10].

The datasets documented include:

1. Heating Degree Day Data Files containing Heating Degree Day values calculated for in situ temperature monitoring stations (see [11] detailing a rationale for a similar methodology). HDD, defined as the annual sum of the positive differences between the base temperature of 18 °C and daily temperature, was calculated using the average of maximum and minimum daily temperature according to the following formula:

$$\text{HDD} = \sum_i ((\theta_b - ((\theta_{\text{MAX}} + \theta_{\text{MIN}})/2))) \quad (1)$$

where  $i$  is the day of the year,  $\theta_{\text{MAX}}$  is the daily maximum temperature,  $\theta_{\text{MIN}}$  is the daily minimum temperature,  $\theta_b$  is the base temperature (18 °C), and  $\theta_b > (\theta_{\text{MAX}} + \theta_{\text{MIN}})/2$ .

Raw data file containing minimum and maximum temperatures by station:

<https://osf.io/x397p>

Average HDD calculated for the following 30-year periods:

1951–1980 (1339 stations): <https://osf.io/he3w8>

1981–2010 (1679 stations): <https://osf.io/x62vu>

The format used to read in these .dat files is provided at:

<https://osf.io/5m7d8>

COMBINED 1951–1980 and 1981–2010 Heating Degree Day Values (.xlsx format):

<https://osf.io/7u8wz>

1951–1980 and 1981–2010 HDD average values for stations with greater than 10 years of data and the count of number of years of observation data for 1951–1980 and 1981–2010.

2. Output from ANUSPLIN (Lis Files) – 1951–1980 and 1981–2010 “Lis” files contain Station coordinates (latitude, transformed longitude and transformed elevation), HDD value for the station, the fitted value (“Fitted\_estimate”), and the individual cross validated values (“CV\_estimate”) see [2] for a description of ANUSPLIN output).

Lis files:

1951–1980: <https://osf.io/29vk4>

1981–2010: <https://osf.io/5wyj6>

A genericized script to read in the “Lis files” is provided at:

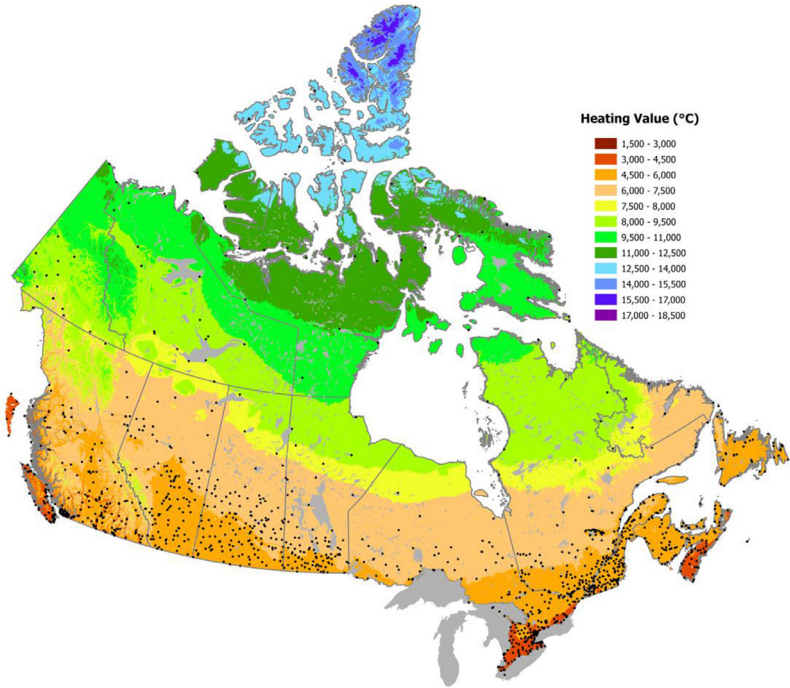
<https://osf.io/5m7d8>

3. Geotiff files – Canada-wide HDD surfaces

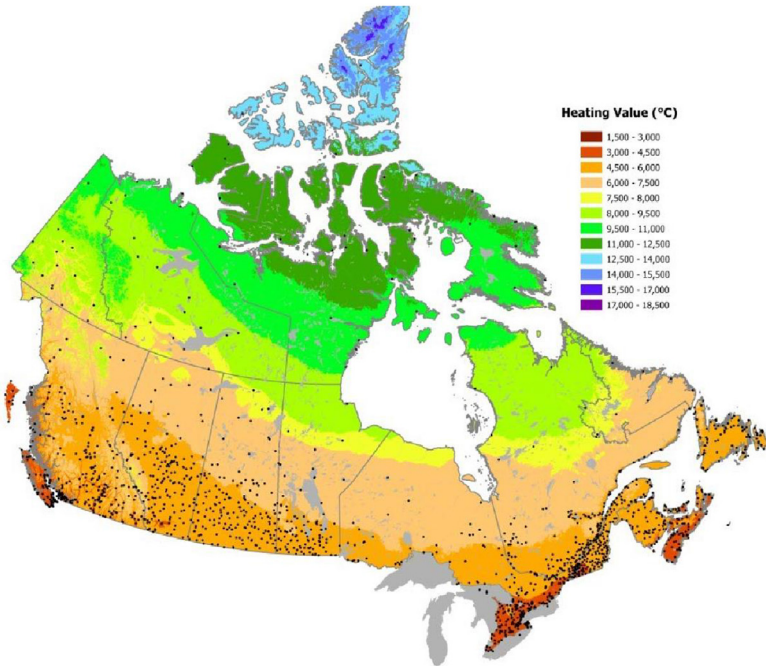
1951–1980: <https://osf.io/2zu5p>

1981–2010: <https://osf.io/sb5p3>

a)



b)



**Fig. 1.** Heating Degree Days for the 1951–1980 (a) and 1981–2010 (b) periods. Black dots show locations of meteorological stations used to develop the thin plate splines.

## 2.2. Predictive error of ANUSPLIN datasets

ANUSPLIN produces individual station cross-validation (CV) estimates (“CV\_Estimate”), which were compared to HDD calculated from station observations. The CV estimates are individually cross-validated values [2]. Mean error (ME) was calculated using the CV estimate minus calculated HDD. ME and Mean Absolute Error (MAE) are presented in C-days as well as a percentage of the surface mean.

ANUSPLIN CV estimates were biased on average by less than 1C-days for both periods (Table 1). Mean absolute error (MAE) of the ANUSPLIN models ranged from 124.2C-days for the 1951–1980 model to 135.3C-days for the 1981–2010 model. The average MAE for the 1981–2010 period represented 2.7% of the surface mean compared to 2.4% for the 1951–1980 period.

**Table 1**

HDD ME and MAE for 1951–1980 and 1981–2010 30-year periods in C-days and as a % of the surface mean.

Time Period	N	ME in C-days (% of Surface Mean)	MAE in C-days (% of Surface Mean)
1951–1980	1339	0.00 (0.0%)	124.2 (2.4%)
1981–2010	1679	−0.47 (0.0%)	137.6 (2.7%)

Plots of observed versus predicted values exhibited strong linear relationships with few outliers for both time periods (Fig. 2).

Predictive errors were plotted for 60 stations selected in previous Canadian studies to better reflect the range in latitude, longitude, and elevation across the country [12,14] as compared with the full set of stations, which are concentrated in southern Canada. Of these 60 stations, 56 stations met the criterion for inclusion in this analysis. Predictive errors at 56 selected stations (Fig. 3) were generally highest in mountainous and coastal regions. Higher errors in areas of complex terrain and coastal areas reflects known challenges with generating spatial models in these highly variable environments for sparse in-situ networks [12–14]. As a percentage, errors were greater in the 1981–2010 period compared to the 1951–80 period.

## 3. Experimental Design, Materials and Methods

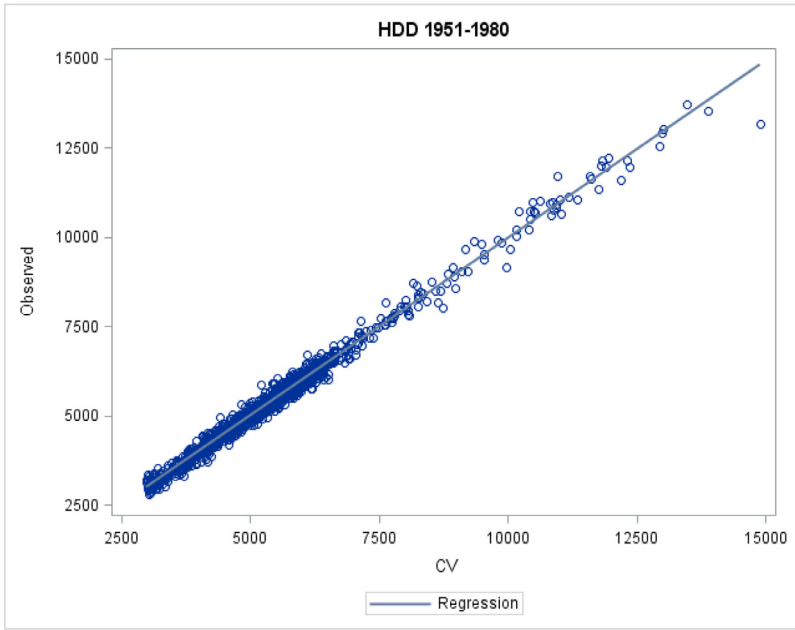
### 3.1. Data acquisition

Environment and Climate Change Canada (ECCC) provided daily minimum and maximum temperature values at meteorological stations across Canada from 1950 to 2010 [15].

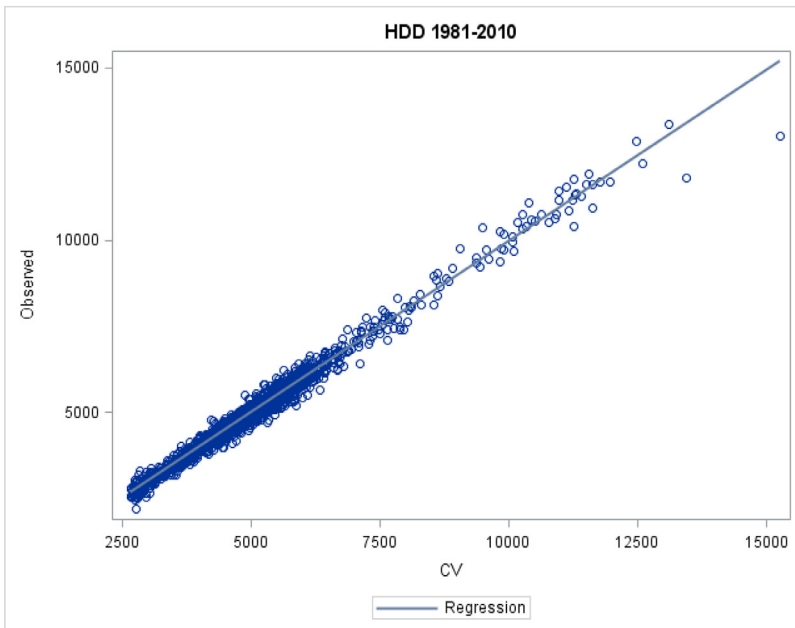
### 3.2. Data pre-processing

Plots were generated to examine the number of stations available for analysis based on cut-offs associated with the number of missing days in a year and the number of missing years in a normal period (Fig. 4). We selected stations with  $\leq 10$  missing days in a year and  $\geq 10$  years in a normal period for the spatial modelling. With these cut-offs, 1339 and 1679 stations were available for analysis in 1951–1980 and 1981–2010 respectively.

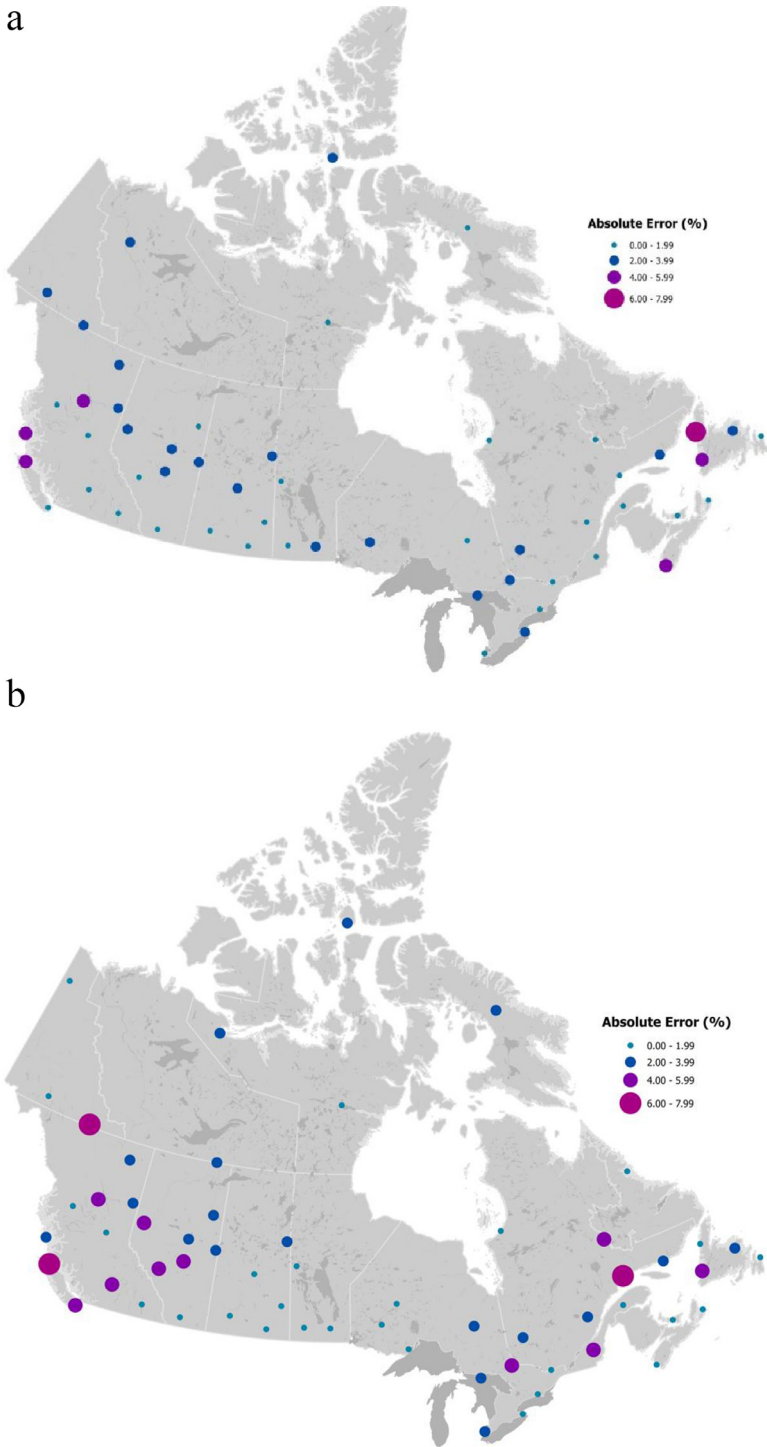
a



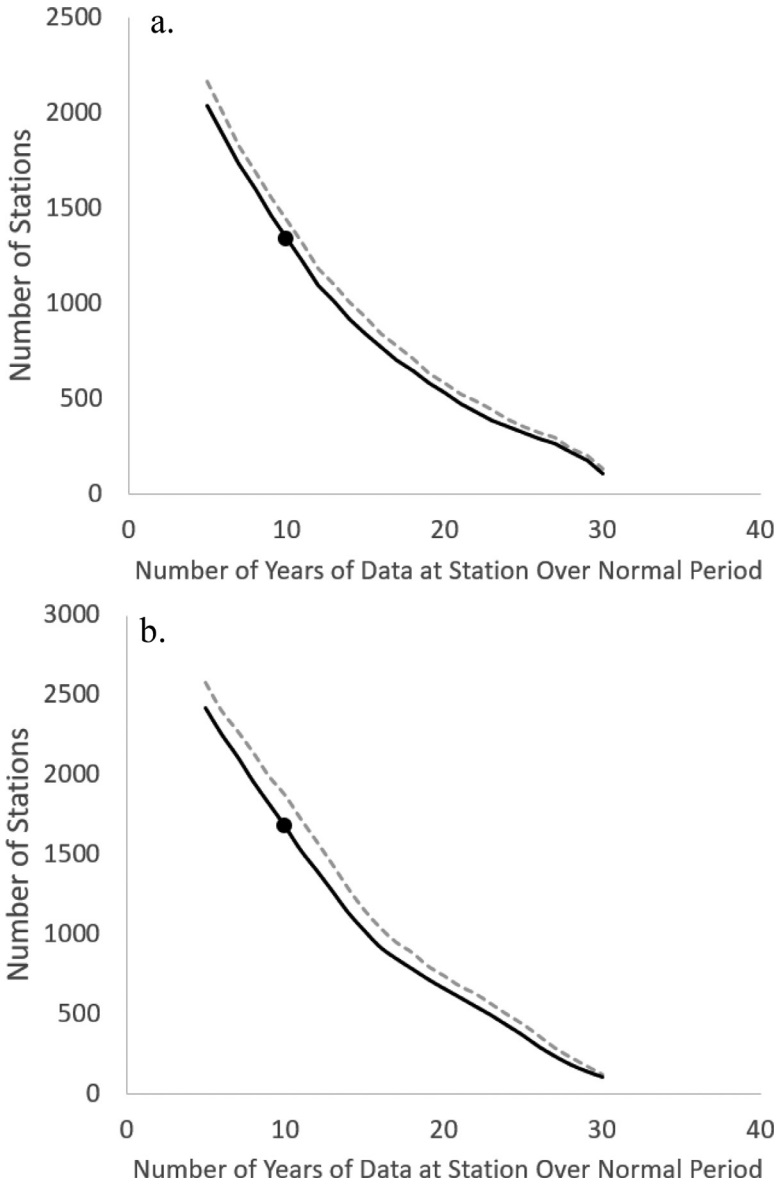
b



**Fig. 2.** Observed versus CV Estimates for the 1951–80 (a) and 1981–2010 (b) periods.



**Fig. 3.** Absolute prediction errors for 60 selected stations for 1951–1980 (a) and 1981–2010 (b) periods.



**Fig. 4.** Relationship between the number of years of station data and the number of stations included in the analysis for the a) 1951–80 and b) 1981–2010 normal periods. The solid and dashed lines were generated using only station-years with number of missing days  $\leq 10$  and  $\leq 20$  respectively. Cut-offs of 10 missing days per year and 10 years per normal period were used to define the station network employed in the current study (final station numbers shown by black dots).

### 3.3. Specifics of implementation

Spatial models were developed in ANUSPLIN [2] and resolved into map form using a 60' sec (approximately 2 km) DEM [10]. The ANUSPLIN grid was created using latitude, longitude (multiplied by 0.64279), and elevation (multiplied by 1000) as predictors. ANUSPLIN fits partial thin



plate smoothing splines constructed from a set of “knots” to noisy multivariate data. A portion of the available observations (in this case, 40%) are selected to limit the complexity of the fitted surface; however, all data points are used to calculate the fitted surface [2].

### 3.4. Experimental results

In addition to predictive error, the quality of the spatial datasets was evaluated using two diagnostic statistics output by ANUSPLIN:

- (a) The ratio of the “signal” ( $S$ ), which ranges between zero and the number of stations (or ‘knots’) selected by ANUSPLIN ( $nKTS$ ), to the number of knots ( $S:nKTS$ ). Ratios between 0.2 and 0.8 are considered acceptable [2,12]. HDD dataset ratios of 0.48 and 0.57 (Table 2) were non-problematic.
- (b) Root GCV (RtGCV). The GCV (Generalized Cross Validation) is calculated by removing each data point and summing the square of the difference of each omitted data point from a surface fitted to all remaining data points [16]. RtGCV, the square root of the GCV, essentially provides a spatially averaged estimate of standard error [14]. The RtGCV was 2.7% for 1951–1980 and 3.3% for 1981–2010 as a percentage of the surface mean (Table 2).

**Table 2**

HDD Signal to number of knots ( $S:nKTS$ ) ratio and root GCV for 1951–80 and 1981–2010 30-Year periods.

Time Period	Surface Mean (in C-days)	$S:nKTS$	RtGCV C-days (% of surface mean)
1951–1980	5400	0.57 (382:669)	147 (2.7%)
1981–2010	5115	0.48 (406:839)	167 (3.3%)

### 3.5. Limitations

Most stations were missing observations for at least some portion of the period considered for this study. With this data report, we published the number of years of data upon which the calculations are based to allow users to make decisions about the use of this dataset. Future work will consider the use of fully in-filled time series for a thirty-year period using estimates for missing HDD values. Canadian in situ stations were concentrated in southern latitudes. Notably much of northern Canada is monitored through a relatively sparse network. To address this feature of the datasets, ANUSPLIN predictions were evaluated for a set of 60 stations selected to better reflect the range in latitude, longitude, and elevation across the country [12].

### Ethics Statement

This work did not involve human subjects or experiments using animals.

### CRediT Author Statement

**Heather MacDonald:** Conceptualization, Methodology, Formal Analysis, Validation, Writing - Original draft preparation, Writing - review & editing. **John Pedlar:** Conceptualization, Methodology, Visualization, Formal Analysis, Validation, Writing -Original draft preparation, Writing - review & editing. **Daniel McKenney:** Conceptualization, Methodology, Writing -Original draft preparation, Writing - review & editing. **Kevin Lawrence:** Data curation, Investigation, Validation. **Kaitlin de Boer:** Data curation, Investigation, Visualization, Investigation, Writing - review &

editing. **Michael Hutchinson:** Software, Methodology, results validation and review, manuscript editing.

## 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.

## Data Availability

[Heating Degree Days Canada 1951-1980, 1981-2010 \(Original data\)](#) (Open Science Framework).

## Acknowledgments

Funding to support development of these datasets was provided to the Integrative Ecology and Economics Group at Great Lakes Forestry Centre (GLFC), Canadian Forest Service, Natural Resources Canada by Canada1Water (C1W) project, a collaboration between Natural Resources Canada, Agriculture and Agri-Food Canada, and Aquanty Inc. Funding is from the Canadian Safety and Security Program (CSSP), Defence Research and Development Canada (DRDC), and the Geological Survey of Canada Groundwater Geoscience and GEM-GeoNorth programs. Work completed in 2021 and 2022 was partly supported by Environment and Climate Change Canada funding for the “Disseminating high resolution historical spatial climate models for Canada” project. Funding to support this work was also provided by the Canadian Forest Service Forest Climate Change Program. Thanks also go to Pia Papadopol for reviews of earlier drafts of this manuscript.

## References

- [1] Government of Canada. Report on the Task Force of Tax Benefits for Northern and Isolated Areas. (1989).
- [2] M.F. Hutchinson, T. Xu, ANUSPLIN Version 4.4 User Guide, Australian National University, Fenner School of Environment and Society, Canberra, Canberra, Australia, 2013 Accessed from <http://fennerschool.anu.edu.au/files/anusplin44.pdf>.
- [3] Climate Atlas of Canada, version 2 (2019), Accessed from [https://climateatlas.ca/map/canada/hdd\\_2030\\_85#](https://climateatlas.ca/map/canada/hdd_2030_85#) on July 10, 2023.
- [4] Prairie Adaptation Research Collaborative. Sask Adapt. (2023). Accessed from <https://www.parc.ca/saskadapt/learn-more/degree-days.html> on July 10, 2023.
- [5] climatedata.ca. (2023). Accessed from <https://climatedata.ca/var-type/other/> July 10, 2023.
- [6] Government of British Columbia Determining ASHRAE 90.1-2010 Climate Zones, 2014 Accessed from [https://www2.gov.bc.ca/assets/gov/farming-natural-resources-and-industry/construction-industry/building-codes-and-standards/bulletins/b14-01\\_determining\\_ashrae\\_901-2010\\_climate\\_zones.pdf](https://www2.gov.bc.ca/assets/gov/farming-natural-resources-and-industry/construction-industry/building-codes-and-standards/bulletins/b14-01_determining_ashrae_901-2010_climate_zones.pdf) July 10, 2023.
- [7] H.C.S. Thom, The rational relationship between heating degree days and temperature, *Month. Weather Rev.* 82 (1954) 1–6.
- [8] M.E. Castañeda, F. Claus, Variability and trends of heating degree-days in Argentina, *Int. J. Climatol.* 33 (10) (2013) 2352–2361, doi:10.1002/joc.3583.
- [9] L.A. Vincent, X. Zhang, É. Mekis, H. Wan, E.J. Bush, Changes in Canada's climate: trends in indices based on daily temperature and precipitation data, *Atmosphere-Ocean* 56 (5) (2018) 332–349.
- [10] 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, Ontario, Frontli. Tech. Note 109 (2008) 4 <https://cfs.nrcan.gc.ca/publications?id=31499>.
- [11] A. D'Amico, G. Ciulla, D. Panno, S. Ferrari, Building energy demand assessment through heating degree days, *App. Energ.* 242 (2019) 1285–1306, doi:10.1016/j.apenergy.2019.03.167.
- [12] H. Macdonald, D.W. McKenney, X.L. Wang, J. Pedlar, P. Papadopol, K. Lawrence, Y. Feng, M.F. Hutchinson, Spatial models of adjusted precipitation for Canada at varying time scales, *J. Appl. Meteorol. Climatol.* 60 (3) (2021) 291–304, doi:10.1175/JAMC-D-20-0041.1.

- [13] M.F. Hutchinson, D.W. McKenney, K. Lawrence, J.H. Pedlar, R.F. Hopkinson, E. Milewska, P. Papadopol, Development and testing of Canada-wide interpolated spatial models of daily minimum–maximum temperature and precipitation for 1961–2003, *J. Appl. Meteor. Climatol.* 48 (2009) 725–741, doi:[10.1175/2008JAMC1979.1](https://doi.org/10.1175/2008JAMC1979.1).
- [14] H. MacDonald, D.W. McKenney, P. Papadopol, K. Lawrence, J. Pedlar, M.F. Hutchinson, North American historical monthly spatial climate dataset, 1901–2016, *Sci. Data* 7 (2020) (2020) 411–411, doi:[10.1038/s41597-020-00737-2](https://doi.org/10.1038/s41597-020-00737-2).
- [15] Environment and Climate Change Canada (ECCC)Climate Data Extraction Tool, 2023 <https://climate-change.canada.ca/climate-data/#/daily-climate-data>.
- [16] G. Wahba, *Spline Models For Observational Data*, Siam Publications, 1990 Vol. 59.