The HNA Blueprint

Climate Index Analysis

Methodology

May 2026

TABLE OF CONTENTS

1 OVERVIEW

The Housing Needs Assessment template includes select local climate estimates and their national percentile rankings across four categories: precipitation, heat, humidex, and cold related variables. The values are derived from projected climate indices rather than observed historical records, enabling the assessment to situate a community's current and near-future climate exposure relative to all other Canadian Census Subdivisions (CSDs).

Each index variable is processed through a shared five-step pipeline implemented in R across four separate workbooks, one per climate category. The pipeline loads rasters from source climate datasets, gap-fills values for coastal and island CSDs that fall outside the land mask, extracts area-averaged values to CSD boundaries, computes reference means at the Census Division and provincial level, and derives national percentile ranks.

The output of each workbook is a CSD-level CSV that feeds the HNA reporting pipeline. The values in the template are drawn from the 2025 and 2050 projection horizons under the SSP2-4.5 (intermediate emissions) scenario at the 50th percentile ensemble member.

2 SOURCE DATASETS

Two primary climate projection datasets are used, supplemented by a third for humidex and a fourth for wildfire. The table below summarises the dataset, file naming convention, and scenario variant associated with each workbook.

Category	Dataset	File pattern from download	Scenario / variant	Indices
Precipitation	CanDCS-U6 / BCCAQv2	*bccaqv2r_*_index_u6_ssp245_ens-enspctl50_…_2025-2050-yearly.nc*	SSP2-4.5, 50th percentile ensemble	r20mm, cdd, r1mm, sdii, r95days, r99days
Heat	CanDCS-U6 / BCCAQv2 (agro)	*bccaqv2r_*_index_u6-agro_ssp245_ens-enspctl50_…_2025-2050-yearly.nc*	SSP2-4.5, 50th percentile ensemble	Indices: tx30, tx35, wsdi
Cold	CanDCS-U6 / BCCAQv2 (agro)	*bccaqv2r_*_index_u6-agro_ssp245_ens-enspctl50_…_2025-2050-yearly.nc*	SSP2-4.5, 50th percentile ensemble	tnm15, tnm20, id
Humidity	MBCn / ERA5-Land	*HXmax*_MBCn_ERA5-Land_ensemble-percentiles_historical+ssp245_1950-2100.nc*	SSP2-4.5, p50 ensemble	hx30, hx35, hx40

Source: CMIP6 statistically downscaled climate scenarios (precipitation, heat, cold)[1] and CMIP6 Humidex Indices Bias Corrected and downscaled with MBCn against ERA5-Land (humidity)[2]

[1] Government of Canada. (2025, October). CMIP6 statistically downscaled climate datasets.
[2] Government of Canada. (2023, November). CMIP6 Humidex Indices Bias Corrected and downscaled with MBCn against ERA5-Land.

CanDCS-U6 / BCCAQv2
The Canadian Downscaled Climate Scenarios – Univariate Sixth Generation dataset (CanDCS-U6) provides statistically downscaled projections[3] of standard climate extremes indices at approximately 10 km resolution across Canada. Downscaling is performed using the BCCAQv2 method (Bias-Corrected Constructed Analogues with quantile mapping), applied to the CMIP6 multi-model ensemble.[4] Files are distributed as NetCDF time series spanning 2025–2050 on an annual time step, with a separate file per index and spatial domain. The SSP2-4.5 scenario (intermediate emissions pathway) at the 50th percentile ensemble member (ens-enspctl50) is used throughout.

MBCn / ERA5-Land humidex
Humidex indices are sourced from a separate ensemble-percentile dataset derived from ERA5-Land reanalysis using the Multivariate Bias-Corrected N-dimensional (MBCn) method. Three humidex threshold exceedance indices are available (HXmax30, HXmax35, HXmax40), each representing the number of days per year on which the daily maximum humidex exceeds the stated threshold. The files span 1950–2100 and include multiple ensemble percentiles; the p50 (median) is used. These files are structured as multi-variable SDS objects requiring selection by variable name rather than by time index.

3 INDEX VARIABLES

The table below maps each index code used in the R scripts to the corresponding row in Climate Analysis table in the Housing Needs Assessment Template.

Variable	Index code	Source dataset	Projection horizons
Precipitation			
Days with heavy rainfall	*r20mm*	CanDCS-U6 / BCCAQv2	2025, 2050
Typical rainfall on wet days	*sdii*	CanDCS-U6 / BCCAQv2	2025, 2050
Extreme precipitation days	*r95days*	CanDCS-U6 / BCCAQv2	2025, 2050
Total wet days	*r1mm*	CanDCS-U6 / BCCAQv2	2025, 2050
Total dry days	*cdd*	CanDCS-U6 / BCCAQv2	2025, 2050

[3] A Global Climate Model is like a very coarse grid laid over the entire planet and it simulates how temperature, rain, and other conditions change over time. Downscaling is the process of sharpening that picture by determining relationships between the coarse model output and detailed historical weather observations, then using those relationships to produce estimates at a much finer resolution. "Bias-corrected" part means they also fix known systematic errors in the model (e.g., if the model consistently overestimates rainy days, they adjust for that so the projections better match reality).

[4] A climate model requires hundreds of choices (e.g., about how to represent clouds, ocean circulation, ice sheets, and so on) and no single set of choices is definitively correct. An ensemble is a collection of model runs where those choices are varied, either across different research groups' models or by altering the same model's starting conditions. Each run produces a slightly different projection of the future. The 50th percentile member (ens-enspctl50) means that the chosen run sits at the middle of the ensemble's range, which is a common "middle-of-the-road" choice that avoids both the most optimistic and most pessimistic model runs.

Variable	Index code	Source dataset	Projection horizons
Maximum consecutive dry days	*cdd**	CanDCS-U6 / BCCAQv2	2025, 2050
Heat			
Days over 30°C (daily max)	*tx30*	CanDCS-U6 / BCCAQv2 (agro)	2025, 2050
Days over 35°C (daily max)	*tx35*	CanDCS-U6 / BCCAQv2 (agro)	2025, 2050
Warm spell duration index	*wsdi*	CanDCS-U6 / BCCAQv2	2025, 2050
Humidex			
Days humidex exceeds 30°C	*hx30*	MBCn / ERA5-Land	2025, 2050
Days humidex exceeds 35°C	*hx35*	MBCn / ERA5-Land	2025, 2050
Days humidex exceeds 40°C	*hx40*	MBCn / ERA5-Land	2025, 2050
Cold			
Days below -15°C (daily min)	*tnm15*	CanDCS-U6 / BCCAQv2 (agro)	2025, 2050
Days below -20°C (daily min)	*tnm20*	CanDCS-U6 / BCCAQv2 (agro)	2025, 2050
Icing days (max temp below 0°C)	*id*	CanDCS-U6 / BCCAQv2	2025, 2050

4 PROCESSING PIPELINE

All four R workbooks implemented the same seven-step pipeline. The steps below describe the shared logic, with workbook-specific differences noted in the relevant step.

#	Step	Detail
1	Load rasters	Each index is extracted from its source NetCDF file using terra::rast(). The CRS is set to EPSG:4326. For CanDCS-U6 / BCCAQv2 files (precipitation, heat, cold), the correct time slice is selected by matching the year embedded in the file's time dimension (load_tx()). For humidex files, the variable is selected by name from a multi-variable SDS object (load_hx()).
2	Gap-fill coastal NAs	Various coastal and island CSDs fall outside the land mask of the climate rasters, returning NA on direct extraction. Each raster is passed through a focal mean filter (terra::focal(), w = 19 × 19 cell window, na.policy = "only") that replaces NA cells with the mean of the surrounding communities in the 19 × 19 area. The gap-filled rasters are kept as a parallel set alongside the originals.
3	Load CSD boundaries	The 2022 Statistics Canada CSD boundary file (lcsd000a22a_e.gpkg) is read with sf::st_read(), repaired with st_make_valid(), and reprojected to match the raster CRS (EPSG:4326 for CanDCS-U6 and humidex; native CRS of the CWFIS normals for wildfire).

#	Step	Detail
4	Extract to CSDs	terra::extract() computes the mean pixel value within each CSD polygon (fun = mean, na.rm = TRUE, touches = TRUE). Two extractions are performed for CanDCS-U6 and humidex workbooks: one from the original raster and one from the gap-filled raster. For each variable, the gap-filled value is substituted only where the direct extraction returned NA. An extrapolated flag column records which CSDs required gap-filling.
5	Reference means	Area-weighted means are computed at two higher geographic levels using the CSD polygon area (km²) as the weight: Census Division (CD, first 4 digits of CSDUID) and Province (PR, first 2 digits). These reference means allow the report to benchmark a CSD's value against its regional and provincial context.
6	National percentiles	For each index variable, a national percentile rank is assigned to every CSD using R's rank() function divided by the count of non-NA CSDs, multiplied by 100, and rounded to one decimal place. This percentile is the figure reported in template's climate variable table. For example, a value of 88 means the community has more extreme conditions than 88% of Canadian CSDs.
7	Assemble and save	All extracted values, reference means, and percentiles are joined into a single CSD-level data frame and written to a CSV in the cleaned input data folder (precipitation_csd.csv, hotdays_csd.csv, colddays_csd.csv, humidex_csd.csv, wildfire_csd.csv). These files feed the HNA reporting pipeline.

5 COASTAL & ISLAND GAP-FILL

Climate rasters derived from land-surface models carry a land mask: grid cells classified as ocean receive NA values. CSDs that are entirely or predominantly coastal, insular, or estuarine may have their centroid or the majority of their polygon area over ocean cells, causing direct extraction to return NA even where the community clearly has a climate exposure.

To address this, each raster is passed through terra::focal() with a 19 × 19 cell window and a mean aggregation function applied only to NA cells (na.policy = "only"). The resulting gap-filled raster is a smoothed surface where former NA cells have been replaced by the area-weighted mean of their non-NA neighbours within the window. The gap-filled value is only substituted for a CSD where the direct extraction was NA; all CSDs with valid direct extractions retain their original value. Note that only 4 CSDs required extrapolation, being:

- Le Goulet, NB (1315030);
- Lamèque, NB (1315032);
- Sainte-Marie-Saint-Raphaël (1315033); and
- Igloolik, NVT (6204012).

An extrapolated flag column is added to each CSD-level output, set to TRUE where the gap-filled value was used. This allows downstream QA to identify and review communities whose climate values were inferred from neighbouring land cells rather than from grid cells directly underlying the CSD.

6 NATIONAL PERCENTILE RANKING

The national percentile reported in the template indicates where a community's value for a given index falls relative to all other Canadian CSDs. For instance, a percentile of 88 for days with 20mm+ precipitation means the community has more heavy-rain days than 88 percent of Canadian CSDs.

Calculation: For each index variable, all non-NA CSD values are ranked in ascending order using R's `rank()` function. The percentile is computed as:

```
percentile = round( rank(x) / count_non_NA(x) × 100, 1 )
```

This is a simple ordinal percentile. Ties receive the average of the ranks they span (R's default method), and NA values are excluded from the ranking and denominator (na.last = "keep"). The result is rounded to one decimal place.

Because the percentile is computed across all CSDs nationally (approximately 5,000 units), small communities in remote areas with fewer surrounding land cells may have gap-filled values that are less precise than those of larger or inland CSDs. The extrapolated flag identifies these cases.

7 REFERENCE MEANS (CD AND PROVINCIAL)

In addition to the national percentile, each workbook computes area-weighted reference means at the Census Division (CD) and provincial (PR) level. These are not currently displayed in the template, but are retained in the output CSVs for use in regional benchmarking, supplemental analysis, and QA.

Reference means are calculated using dplyr::summarise() with weighted.mean(), where the weight is the CSD polygon area in km² (derived from sf::st_area()). This area weighting prevents large, sparsely populated CSDs from dominating the CD or provincial average through sheer geographic extent, while still reflecting the spatial distribution of climate conditions.

8 OUTPUT FILES

Each workbook writes a single CSV to the cleaned input data folder. The key columns present in all CanDCS-U6 and humidex outputs are:

- **CSDUID** — 7-digit Census Subdivision identifier
- **CDUID**, **PRUID** — parent Census Division and Province identifiers
- **area_km2** — CSD polygon area used as weight in reference mean calculations
- **[index]_2025, [index]_2050** — extracted index values for each projection horizon
- **[index]_[year]_cd_mean, _pr_mean** — area-weighted CD and provincial reference means
- **[index]_[year]_pctile_national** — national percentile rank (0–100)
- **[theme]_extrapolated** — flag indicating gap-filled value was used

9 LIMITATIONS AND INTERPRETIVE NOTES

Scenario and ensemble selection
All CanDCS-U6 projections use the SSP2-4.5 scenario at the 50th percentile ensemble member. This represents an intermediate emissions trajectory and the median model response; it does not capture the range of possible futures. Communities particularly exposed to climate risk may wish to consider higher-emissions or higher-percentile scenarios for resilience planning, which are available in the source datasets but not currently processed in these workbooks.

Projection vs. observed climate
The values shown in the template are model projections anchored to the 2025 and 2050 calendar years, not observed historical averages. The 2025 values reflect modelled near-present conditions under the SSP2-4.5 scenario rather than a climatological normal period, and may differ from station-based observed values for the same community.

CSD-level averaging
Each CSD receives the mean pixel value across its polygon. For very large CSDs (particularly in northern Canada), this mean may smooth over considerable internal variation in climate exposure. The result is most reliable for small to medium CSDs where the polygon area is comparable to or smaller than the raster resolution.

Coastal gap-fill accuracy
Gap-filled values are informed by neighbouring land cells within a 19 × 19 window. For small island communities or narrow coastal strips, the nearest land cells may not accurately represent local climate conditions; particularly, where the surrounding ocean exerts a moderating influence on temperature and precipitation. Gap-filled values should be interpreted with additional caution for communities where the extrapolated flag is TRUE.

Humidex dataset difference
The humidex indices use a different source dataset and bias-correction method (MBCn / ERA5-Land) than the other temperature and precipitation indices (BCCAQv2 / CMIP6). While both use SSP2-4.5 and the p50 ensemble member, minor methodological differences in the downscaling and bias-correction approaches mean that the humidex values are not strictly comparable to the CanDCS-U6 indices in a cross-variable sense.