SimBuild 2024

Eleventh National Conference of IBPSA-USA
Denver, Colorado
May 21 – 23, 2024

Development of Regional-Scale Typical Meteorological Years for Canada

Manon Beaufort[1], Giulio Tonellato[1], and Michaël Kumert[1]
[1]Polytechnique Montréal, Montreal, Canada

Abstract

Typical Meteorological Years (TMY) are widely used in local-scale building simulation to reproduce realistic response to meteorological conditions using historical data. Current TMYs fail to reproduce building stock response to meteorological conditions on larger scale due to temporal incoherence of TMY from multiple stations. This study presents a method to determine a list of 12 “typical months” representative of all meteorological stations of Quebec province. This list of 12 provincial typical months can be used to form a provincial-purposed TMY for each station. This allows more realistic estimation of the peak demand in large scale simulations, without requiring the use of several actual meteorological years.

Nomenclature - Acronyms

CDF	Cumulative Distribution function
CWEC	Canadian Weather Year for Energy Calculation
CWEEDS	Canadian Weather Energy and Engineering Datasets
ECCC	Environment and Climate Change Canada
FSstat	Finkelstein-Schafer statistics
GHI	Global solar Irradiance on Horizontal surface
NRCan	Natural Resources Canada
NREL	National Renewable Energy Laboratory
TMM	Typical Meteorological Month
TMY	Typical Meteorological Year

Introduction

Typical Meteorological Year

Simulating energy demand and generation often requires meteorological data. One-year simulations can be based on a specific year to realistically represent peak consumption and production, but they do not represent the long-term trends.

A Typical Meteorological Year (TMY) concatenates meteorological data from months that are the most representative of a long-term dataset, allowing to obtain a realistic estimate of an “average year”. But because they are based on a concatenation of actual months, TMYs also allow to obtain a realistic estimate of peak demands.

National Resources Canada (NRCan) and Environment and Climate Change Canada (ECCC) provide a data base of meteorological data for 564 Canadian stations (ECC Canada 2020). For each station, a “Canadian Weather Energy and Engineering Dataset” (CWEEDS) compiles hourly meteorological data of the station from 1998 to 2017. A “Canadian Weather Year for Energy Calculation” (CWEC) is a type of TMY produced by NRCan and ECCC for building energy simulations. For each month, a Typical Meteorological Month (TMM) is selected among a pool of years, as the most statistically representative in terms of dry-bulb temperature, dew-point temperature, global solar irradiance on horizontal surface (GHI) and wind speed.

For instance, for the meteorological station of Amqui, Quebec, from all months of January from 1998 to 2017, January 2000 is the most statistically representative. 2000 is said to be the TMM of January for the Amqui station. The TMY of Amqui is obtained by concatenating data from the 12 TMMs of the station.

Thus, for each station, a CWEC can be produced based on the CWEEDS data, after determining the 12 TMM.

Large-scale TMY

Historically, building stock modeling has focused on annual energy use and greenhouse gas emissions. With increasing electrification of buildings, the focus is shifting towards predicting the building stock peak demand at the scale of electric utilities service areas. This require simulating a large number of buildings over a large geographical area with temporally coherent weather datafiles. As mentioned above, the TMY methodology allows to perform single year simulations to estimate those peak demands. However, TMY from individual stations can only be used for local-scale simulations for small geographical areas close to the stations. When covering a larger surface, one could want to use data from several stations covering the area. But using TMYs of different stations could use temporally incoherent data if they do not share the same TMMs. The simulation would then fail to reproduce a realistic peak load.

For instance, January 2012 is the typical month of January for the city-center McTavish station in Montréal, while it is January 1999 for the Montréal International Airport station, even if those two stations are only 13 km apart. Looking at the daily mean temperature of January for those two stations (Figure 1), we can see that they do not have their cold wave at the same time. In a simulation, a building simulated using the airport weather file would have a high daily peak demand on January 1st and 2nd, while a building simulated using the McTavish weather file would have a high daily peak demand on January 3rd and 4th. In real life conditions, the cold wave of those two stations would occur at the same time. This shows that simply using the TMY closest to each building is not an acceptable solution to simulate the dynamic demand of building stocks.

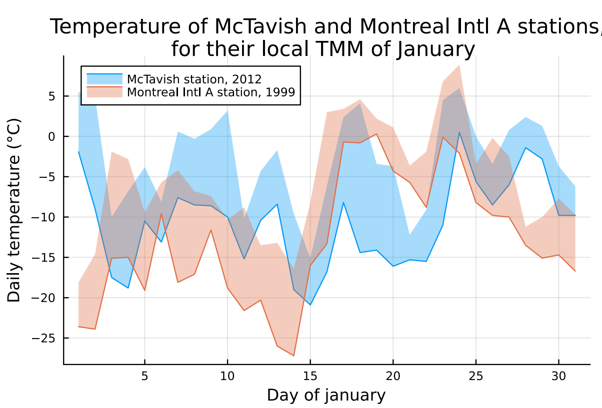

Figure 1: Daily temperature in January at McTavish station and Montreal Intl A station, for their local TMM

This study aims to produce a method that determines a list of 12 TMMs that would be representative of all stations of Quebec province at the same time. For each station, a new provincial-purposed TMY/CWEC will be produced by extracting data corresponding to the provincial TMMs of Quebec province. This will allow simulation at a provincial scale, maintaining station-scale granularity and coherent time period for realistic peak load modeling.

Background

One of the first and widely established methods to produce TMY was developed at Sandia Laboratory by Hall et al. (1978). This method is now called the Sandia method. It selects TMMs depending on daily mean, minimal and maximum values of 4 weather parameters: dry-bulb temperature, dew point temperature, wind speed and global radiation. Those parameters are given different weighting factor in the selection process.

This method was later used and adapted by researchers depending on the purpose of the TMY. Main modifications involve the choice of weather parameters and their weighting factors. For instance, Zang et al. (2012) included pressure and humidity as parameters in the selection of TMY for China, while Wilcox & Marion (2008) used direct radiation for United States' TMY selection. In Europe, TMYs are produced following the ISO 15927-4 procedure (ISO 2005). This method only uses the daily mean values of the 4 weather parameters previously mentioned. And each of those 4 parameters have the same weighting factor of 0.25.

Huang (2020) proposed a method to determine statewide TMMs for California meteorological stations. The method is based on the Sandia method to obtain TMYs, but the method is simplified to rank candidate months by their population-weighted Finkelstein-Schafer statistics (see Methodology below for details). In our work, we start from the Sandia method as adapted by Environment and Climate Change Canada and Natural Resources Canada to produce the Canadian Weather for Energy Calculations (CWEC) datasets. The TMY selection method of NRCan and ECCC is based on the Sandia method and its adaptation by Wilcox and Marion (2008) for NREL, and was last updated by Morris (2016). It includes different refining steps after an initial ranking by Finkelstein-Schafer statistics. We propose an adaptation of these steps to derive regional-scale TMYs and demonstrate the results obtained for the province of Quebec. We also document several steps used in deriving CWEC datasets which are only superficially – and ambiguously – described in their documentation (Morris 2016).

Methodology

This paper presents the 3 major steps in our methodology:

- The reproduction of the ECCC method that determines TMMs for individual Canadians stations.
- The reproduction of the ECCC algorithm that produces CWEC (TMY) from CWEEDS (long-term dataset)
- The adaptation of the TMM selection method to select a unique series of TMMs representative of all stations of Quebec province.

The first two steps were necessary because the documentation for the method used to produce TMYs is sometimes ambiguous, and we wanted to ensure that our provincial-level TMYs were based on the same rigorous foundations as the officially published “per-station” CWEC files.

Reproduction of the method to determine individual TMYs

The TMMs that were selected for each publicly available CWEC data file can be identified in the files themselves, and ECCC provides the long-term data sets (CWEEDS) and the resulting CWECs on their open-access platform (ECC Canada 2020).

We reproduced the ECCC algorithm that determines the TMMs and confirmed its performance by successfully selecting the 12 TMMs of each of the 81 meteorological stations for the Quebec province.

Some ambiguities in the CWEC documentation (Morris 2016) were resolved by examining the source code of the program used to generate the most recent published CWEC data files (Morris 2022).

The TMM selection algorithm is composed of the 4 following steps:

- Excluding years with missing data from the selection.
- Preselecting the five years with the smallest Finkelstein-Schafer statistics (FS statistics) ranked by ascending order.
- Reordering the 5 preselected years by closeness to the long-term values of GHI and temperature.
- Rejecting years using persistence criteria, i.e., years with extreme temperatures or GHI. The first year not rejected is selected as TMM.

Step 1: Excluding years with missing data

Years with missing data for a particular month are excluded from the TMM selection. This step is not necessary with the most recent long-term CWEEDS datasets published by ECCC because all missing data has been filled out as part of the data quality assurance procedure.

Step 2: Preselection of 5 years according to FS statistics

The Finkelstein-Schafer statistic (FS stat) evaluates the distance of a data subset to the complete dataset using their cumulative distribution function (CDF). The CDF of a dataset of n values is defined as follow:

$$CDF(x) = \begin{cases} 0 & for\ x < x_1, \\ \dfrac{i-0.5}{n} & for\ x_i \leq x < x_{i+1}, \\ 1 & for\ x_n < x. \end{cases} \qquad (1)$$

With n the number of values of the dataset, and x_1 to x_n those values in ascending order. Figure 2 shows CDFs of daily mean temperature in January at McTavish Station. Both long-term January CDF and January 2012 CDF are represented in this figure.

For the month m, the FS stat of the year y is the average distance between the CDF of this year y and the CDF of all years of the data set. The FS stat is defined as follow:

$$FS_x(m,y) = \frac{1}{n}\sum_{i=1}^{n}\left|CDF_m(x_i) - CDF_{m,y}(x_i)\right| \qquad (2)$$

With n the number of days in the month m, x the studied weather parameter, CDF_m the CDF of the month m on the long-term, $CDF_{m,y}$ the CDF of the month m for the year y of the dataset and $x_1$to x_n the values of the weather parameter on the month m of the year y.

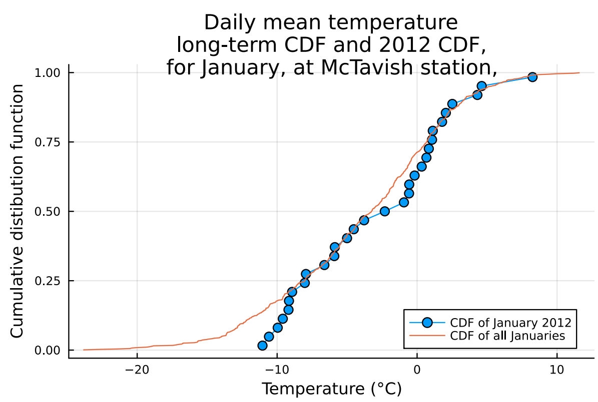

Figure 2: Daily mean temperature CDF at McTavish station

For a year y, 9 FS stat values are calculated from 9 different weather parameters listed below in table 1.

Those FS stats are weighted to produce a global FS stat for the year.

$$FS(m,y) = \frac{1}{n}\sum_{k=1}^{9}\omega_k \times FS_k(m,y) \qquad (3)$$

Where $FS(m, y)$ is the FS stat of the year y for the month m, k is the index of one of the 9 parameters, ω_k its weighting factor, and $FS_k(m, y)$ is the FS stat of the parameter k for year y for month m.

Table 1 Parameters and their weight for FS statistics, from CWEC documentation (Morris 2016)

Weather parameter		**Weight**
Daily maximum dry-bulb temperature,	T_{max}	5%
Daily minimum dry-bulb temperature,	T_{min}	5%
Daily mean dry-bulb temperature,	T_{mean}	30%
Daily maximum dew-point temperature,	Td_{max}	2.5%
Daily minimum dew-point temperature,	Td_{min}	2.5%
Daily mean dew-point temperature,	Td_{mean}	5%
Daily maximum wind speed,	WS_{max}	5%
Daily mean wind speed,	WS_{mean}	5%
Daily total of GHI,	GHI_{sum}	40%

Years of the dataset are then ordered by ascending FS stat, and the 5 years with the lower FS stats are preselected for the next steps.

Step 3: Reordering of the 5 preselected years

This step is described in the CWEC documentation (Morris 2016) by a short sentence: "The five candidate months are ranked with respect to the closeness of the mean and median of each month to that of the long-term data set". In practice, the ranking is based on daily mean temperature and daily total GHI, as described below (Morris 2022).

$$ERR_{Tmean}(m, y) = \left|mean\big(T_{mean}(m, y)\big) - mean\big(T_{mean}(m)\big)\right| + \left|med\big(T_{mean}(m, y)\big) - med\big(T_{mean}(m)\big)\right| \tag{4}$$

$$ERR_{GHIsum}(m, y) = \left|mean\big(GHI_{sum}(m, y)\big) - mean\big(GHI_{sum}(m)\big)\right| + \left|med\big(GHI_{sum}(m, y)\big) - med\big(GHI_{sum}(m)\big)\right| \tag{5}$$

Where $ERR_{Tmean}(m, y)$ is the closeness of the daily mean temperature median and mean of the month m of year y to the daily mean temperature median and mean of the month m of all years of the dataset. $T_{mean}(m, y)$ is the vector of daily mean temperatures of the month m of year y. $T_{mean}(m)$ is the vector of daily mean temperatures of all months m of the dataset.

Those two closeness components are normalized to form the final closeness value of each of the 5 preselected year:

$$closeness(m, y) = \frac{ERR_{Tmean}(m, y)}{\max\limits_{i=1,5}\big(ERR_{Tmean}(m, y_i)\big)} + \frac{ERR_{GHIsum}(m, y)}{\max\limits_{I=1,5}\big(ERR_{GHIsum}(m, y_i)\big)} \tag{6}$$

The closeness components are normalized using the maximal value of the components for the 5 preselected years y_1 to y_5.

The 5 preselected years are then reordered according to this closeness index, in ascending order.

Step 4: Rejection of years based on persistence criteria and final selection

This step is briefly described in the CWEC documentation (Morris 2016) but some details were obtained from the source code (Morris 2022). One by one, the 5 preselected years are tested using persistence criteria. If the year shows extreme behavior, it is rejected. The first year not rejected is selected as TMM.

For the 5 preselected years for the month m, the number and the length of runs above and below certain limits are counted:

- Runs of daily mean temperatures higher than the 67th percentile of long-term daily mean temperatures
- Runs of daily mean temperatures below the 33rd percentile of long-term daily mean temperatures
- Runs of daily total GHI below the 33rd percentile of long-term daily total GHI

For instance, for the McTavish station, the month of January 2012 counts 3 runs of 1 day, 2 runs of 2 days and 2 runs of 3 days above the 67th long-term daily mean temperature, as shown in Figure 3.

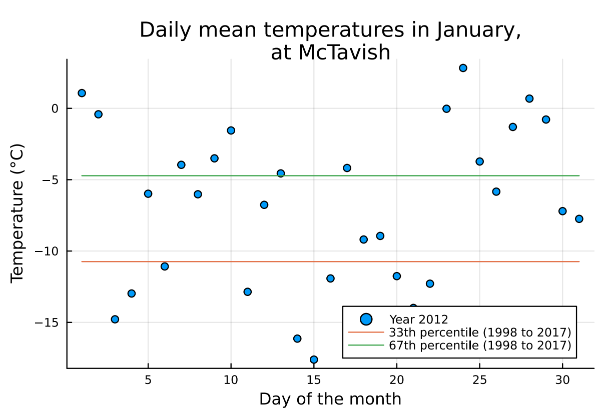

Figure 3: Daily mean temperature in January at McTavish station

The run counts are used to produce the two following types of ratios R_1 and R_2:

$$R_{1,j}\ (m,y) = \frac{n_{run,L\geq L_{y,m}}(m,y)}{n_{run,L\geq 1}(m)} \tag{7}$$

For j in [1,2,3], depending on the type of runs (above the 67th percentile of T_{mean}, below the 33rd percentile of T_{mean} or below the 33rd percentile of GHI_{sum}). With $L_{y,m}$ the maximum length of the runs of type j for the month m and the year y. $n_{run,L\geq L_{y,m}}(m,y)$ is the number of runs of type j of length $L_{y,m}$ or more of all the years of the dataset. $n_{run,L\geq 1}(m)$ is the total number of runs of type j for all years of the dataset.

$$R_{2,j}\ (m,y) = \frac{n_{year,N\geq N_{y,m}}(m,y)}{n_{year,N\geq 0}(m)} \tag{8}$$

With $N_{y,m}$ the number of runs of type j on month m and year y. $n_{year,N\geq N_{y,m}}(m,y)$ is the number of years with at least $N_{y,m}$ runs of type j. $n_{year,N\geq 0}(m)$ is the total number of years in the dataset.

The years with the smallest R_1 ratio are the ones with the longest runs, while years with small R_2 ratio are the ones with the most runs.

A year is rejected if it is one of the years with the most or the longest runs of extreme daily mean temperatures or daily total GHI. To be precise, the year is rejected if R_1 or R_2 are lower than 5% for high or low T_{mean} runs or if one of them is lower than 7.5% for low GHI_{sum} runs.

A year is also rejected if it does not have at least one run above or below those percentiles. This was presumably implemented to prevent selecting years that would be too "uneventful".

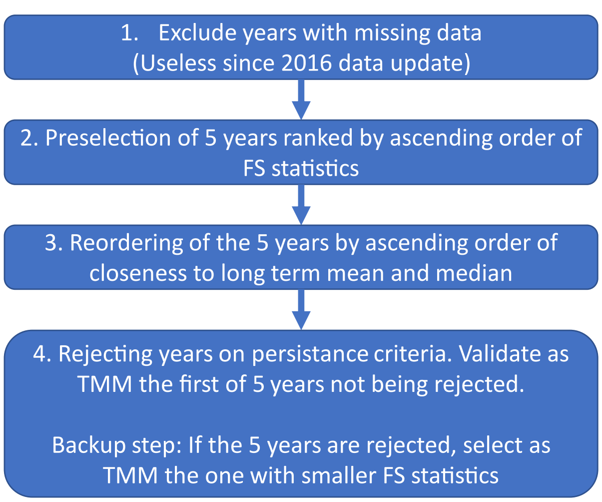

Figure 4: ECCC TMM selection method

Fallback step

If all 5 years are rejected, then a fallback step selects the year with the smallest FS stat of the dataset, as if the TMM was selected directly after step 2.

The fallback step is infrequent as it is only used for selecting 4 of the 972 TMMs of the 81 stations in Quebec.

The following figure summarizes the ECCC TMM selection method.

Reproduction of the CWEC

The production of CWEC from CWEEDS data and a TMM list consists in the concatenation of the data from the 12 TMMs. The pressure, dry-bulb temperature and dew-point temperature are then "smoothed" for the 6 first and last hour of each month to simulate a continuous evolution of those values through the year.

Those data are transformed as follows:

- Remove linear trend
- Apply Fourier transform
- Multiply by the window function
- Apply inverse Fourier transform
- Restore linear trend

The production of a provincial-purposed CWEC would follow the exact same protocol, except that the list of TMM would be representative of a provincial meteorological trend instead of a local trend.

Regional-scale method to determine a unique list of 12 "typical months" for the Quebec province

Once the TMM determination method of ECCC was reproduced, it was possible to adapt it to determine a unique provincial-scale list of TMMs. This study focuses on providing a provincial list of TMMs for Quebec province. The choice of a large-scale TMM depends on data from different meteorological stations. We adopted the same method as Huang (2020), and assigned weights to the weather stations based on the percentage of the Quebec population covered by this station. Thus, the selected months would be more representative of long-term meteorological conditions in area heavily populated than of under-populated areas. We could not find in (Huang 2020) the exact method used to assign a population to each weather station, so we adopted a method based on geographically referenced census data and Voronoi polygons.

Population data in Quebec

Canada publishes census data at various resolutions, This study used the 2021 census data for the 1282 "census subdivisions" of the Quebec Province (StatsCan 2022), shown in Figure 5.

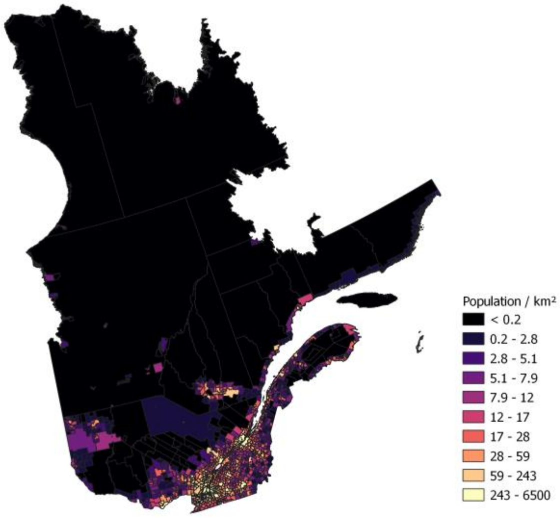

Figure 5: Census subdivisions in Quebec with their population density

CWEEDS stations for Quebec territory

CWEEDS are available for 564 meteorological stations in Canada, including 81 in Quebec. However, stations in neighbouring provinces can sometimes be closer to a given location than Quebec stations. So, we first selected 134 weather stations "in or around" Quebec.

Unfortunately, some of those stations went in operation later than 1997 and fail to represent the same 20 years period as the majority of the stations. Those stations are either close to other stations with full range datasets, or cover areas with very small population densities. We decided to exclude those stations from the study, to keep a larger pool of potential years in our selection process and to focus on areas with larger population densities connected to the main electricity grid. This resulted in a set of 94 stations.

To associate a census division to the nearest station, Voronoi polygons were used. Voronoi polygons were created in QGIS (QGIS.org 2022) to represent the proximity area covered by each station. The polygons are shown in Figure 6.

The weighting factor of each station was determined by associating its Voronoi polygon with the Quebec population covered by its area. To estimate that population, the population density from the census subdivisions was rasterized over a 1 km mesh to calculate the total population in each Voronoi polygon. 29 of the 94 stations had Voronoi polygons with zero population. Figure 7 shows the Voronoi polygons and associated population.

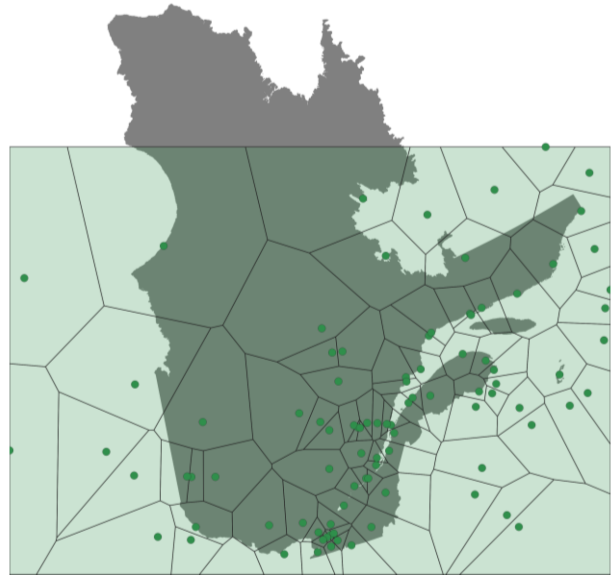

Figure 6: Voronoi polygons around the weather stations

Thus, we kept the 65 remaining stations, including 4 stations located outside of the province. The first 6 stations, located near Montreal, represent half of the population of Quebec. 6 other stations are required to represent up to 80% of the total population. The smallest station in terms of population is Ile aux Perroquets which only covers 0.0002% of the population.

Each of the 65 stations is given a weight corresponding to the proportion of the population of Quebec in its Voronoi polygon.

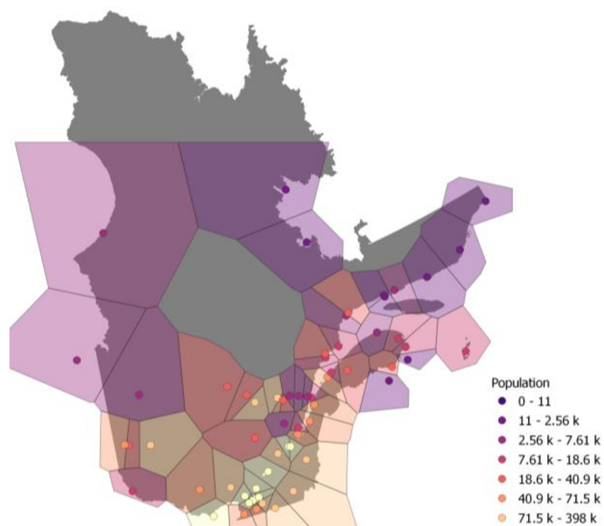

Figure 7: Stations and their Voronoi polygons colored by population

Adaptation of the ECCC method for regional-scale TMM

As mentioned previously, Huang (2020) relied on the ranking based on the FS statistics to select regional-scale TMMs – in other words, the applied methodology stopped at step 2 in the method described above to produce individual TMYs. We wanted to include in our method a reordering by closeness to long-term mean and median, and a rejecting step for persistence criteria, in order to match the ECCC method used for individual TMYs as close as possible. The adopted steps are described below.

Step 1: Excluding years with missing data

Years with missing data for at least one of the 65 stations are excluded from the selection. As explained above, this step is unnecessary since the published CWEEDS dataset is serially complete.

Step 2: Preselection of 5 years according to the FS statistics

For each year, a provincial FS statistic is calculated as the weighted sum of individual station's FS statistics, by population.

$$FS_{provincial}(m,y) = \sum_{z=1}^{n} \omega_z \times FS_z(m,y) \tag{9}$$

Where $FS_{provincial}(m,y)$ is the provincial FS statistic of the year y for the month m. With n the number of stations (65 in our case), ω_z the population weight of the station number z and $FS_z(m,y)$ the FS stat of the year y for the month m of the station number z.

The 5 years with the smallest provincial FS statistics are preselected for further steps.

Step 3: Reordering of the 5 preselected years

For each year, a provincial closeness index to long term mean and median of temperature and GHI is calculated according to the following formula.

$$closeness_{provincial}(m,y) = \sum_{z=1}^{n} \omega_z \times closeness_z(m,y) \tag{10}$$

Where $closeness_{provincial}(m,y)$ is the provincial closeness index of the year y for the month m. With n the number of stations (65 in our case), ω_z the population weight of the station number z and $closeness_z(m,y)$ the closeness index of the year y for the month m of the station number z.

The 5 preselected years are then reordered according to this provincial closeness index, in ascending order.

Step 4: Rejection of years on persistence criteria and final selection

While the first steps could intuitively be adapted to process several stations' data, the 4th step and the associated fallback step were more complex to adapt.

The first provincial rejecting method we tested was to reject a year from the provincial TMM selection if it was rejected by at least one station according to the individual station rejecting method. And if all 5 preselected years were rejected, the fallback step selected the year with the smallest provincial FS stat.

Unfortunately, all years were rejected by at least one station. Thus, this rejecting method resulted in rejecting all years, for each of the 12 TMM selections, and the fallback step was applied for each TMM. This rejecting method was judged too strict. For comparison, only 4 of the 780 TMMs were selected with the fallback step for the creation of the 65 official individual CWECs.

Different versions of rejecting method for provincial TMM

Two rejecting methods and two fallback selection methods were considered. For the rejecting method, we decided to reject years depending on the percentage of the population of Quebec covered by stations that individually rejected the year. In other words, a year is rejected if more than a certain percentage of the population lives in areas of stations rejecting the year. The 4 combinations of rejecting and fallback selection methods are listed in table 2.

Table 2 Different rejecting and fallback selection methods for step 4

Name	Description
A1	Reject year if it is rejected by a number of stations covering more than 1% of the population. If all 5 preselected years are rejected, select the year with the smallest provincial FS stat.
B1	Reject year if it is rejected by a number of stations covering more than 10% of the population. If all 5 preselected years are rejected, select the year with the smallest provincial FS stat.
A2	Reject year if it is rejected by a number of stations covering more than 1% of the population. If all 5 preselected years are rejected, select the year rejected by the smallest fraction of the population.
B2	Reject year if it is rejected by a number of stations covering more than 10% of the population. If all 5 preselected years are rejected, select the year rejected by the smallest fraction of the population.

Each method provides a list of 12 potential provincial TMMs. The choice of the method will depend on the percentage of the population located in areas of stations rejecting the TMMs, and the number of TMM selected by the fallback method. The results of those 4 methods are presented below in table 3 and 4.

Table 3 Results of the TMM selection method A1 and B1

Method	A1		B1	
Month	TMM	Rejecting % of pop	TMM	Rejecting % of pop
Jan	1999	0.3%	1999	0.3%
Feb	2006	97.9%	2009	7.6%
Mar	2011	72.8%	2009	1.4%
Apr	2015	55.6%	2013	8%
May	2014	44.8%	2014	44.8%
Jun	2017	3.1%	2017	3.1%
Jul	2016	49.4%	2002	4.8%
Aug	2010	3.1%	2010	3.1%
Sep	2005	1%	2005	1%
Oct	2000	16.1%	2000	16.1%
Nov	2003	69.8%	2010	8.6%
Dec	2009	46.6%	2005	9.4%

In green, the TMMs selected because the year was not rejected by more than 1 or 10% of the population, and in orange, the TMMs selected by the fallback method after the 5 preselected years have been rejected.

If we increase the percentage of population allowed to reject the year, then more TMMs are selected without going through the fallback selection. The selected months are also representative of a larger part of the population since they are rejected by a smaller proportion of the population.

It is important to note that the fallback strategy selecting the years with the smallest FS stat tends to select years with high rejection ratio from the population.

For instance, the A1 method selected February 2018 which has a 97.9% rejecting population ratio, and B1 selected May 2014 which has a 44.8% rejecting population ratio.

To avoid the selection of years rejected by a large part of the population, the fallback selection method was then modified to select the preselected year that is rejected by the less percentage of the population. This led to method A2 and B2.

Table 4 Results of the TMM selection method A2 and B2

Method	A2		B2	
Month	TMM	Rejecting % of pop	TMM	Rejecting % of pop
Jan	1999	0.3%	1999	0.3%
Feb	2009	7.6%	2009	7.6%
Mar	2009	1.4%	2009	1.4%
Apr	2013	8%	2013	8%
May	2003	21.7%	2003	21.7%
Jun	2017	3.1%	2017	3.1%
Jul	2002	4.8%	2002	4.8%
Aug	2010	3.1%	2010	3.1%
Sep	2005	1%	2005	1%
Oct	2000	16.1%	2000	16.1%
Nov	2010	8.6%	2010	8.6%
Dec	2005	9.4%	2005	9.4%

The modification of the fallback selection method improved the percentage of the population rejecting the TMMs. In our case, changing the ratio of allowed percentage of population rejecting the year does not impact the selected TMM. But it could have an impact with different data sets.

Since the fallback selection method was not frequently used for official TMY selection, we decided to set a limit percentage of rejecting population to 10%. Thus, the fallback selection method will be used less frequently than with a 1% population limit.

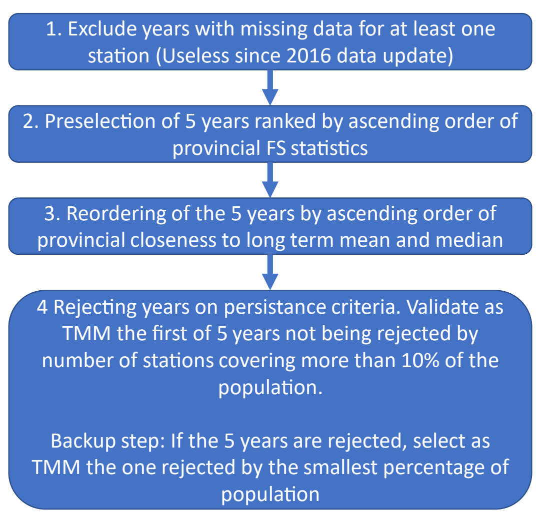

Figure 8: Provincial TMM selection method

Results

Finally, we chose to use the method B2 as final selection method. The provincial TMMs selection method is summarized in figure 8. The choice of the limit on the percentage of the population could be refined by testing this method on other Canadian provinces.

We select the following list of provincial TMMs in Table 5 for Quebec province.

Table 5 List of provincial TMMs for Quebec

Month	TMM	Month	TMM
Jan	1999	Jul	2002
Feb	2009	Aug	2010
Mar	2009	Sep	2005
Apr	2013	Oct	2000
May	2017	Nov	2010
Jun	2017	Dec	2005

These TMMs can be used to form provincial-purposed TMYs for each meteorological station of Quebec province, using their CWEEDS datafiles.

Figure 9 shows an interpolated raster map of the minimum temperature on January 13 according to the individual TMYs for each station (left) and according to the provincial-scale TMY (right). The figure shows that a typical “uniform” cold spell across the province results in temperatures ranging from -45 °C in the north of the province to -10 °C in locations with a marine climate. The peak demand simulated on that day would be much larger than the one obtained with the weather data from individual locations, even though the minimum temperature is similar in both cases – the temperature distribution in the left part of the figure is randomly affected by the individual months that were selected and cannot represent a “provincial-scale” cold spell.

Conclusion

This paper documents the method used to generate Typical Meteorological Years (TMYs) for individual weather stations according to the method used by ECCC and NRCan to develop the Canadian Weather for Energy Calculations (CWEC) datasets. We then propose a method to obtain provincial-scale TMYs suitable to simulate a building portfolio covering a large area and obtain a realistic estimate of their coincident peak demand. The method is inspired by previous work by Huang (2020) for California, and adds several refinements to the selection methodology instead of stopping at the initial ranking of candidate months by their Finkelstein-Schafer statistics. The proposed methodology is applied to the province of Quebec, and it successfully selects a series of 12 typical months representing the “best choices” for the largest share of the population and “good choices” for the rest of the population. The resulting regional-scale TMYs can be used to simulate coherent peak demands in building portfolios distributed across the province.

We intend to publish the codes selecting the 12 TMMs of a single station, selecting the 12 TMMs of a group of stations, and creating CWEC from CWEEDS.

This study only determined the provincial TMMs list for the Quebec province, but it can produce a similar list for every Canadian province, using their publicly available census and CWEEDS data.

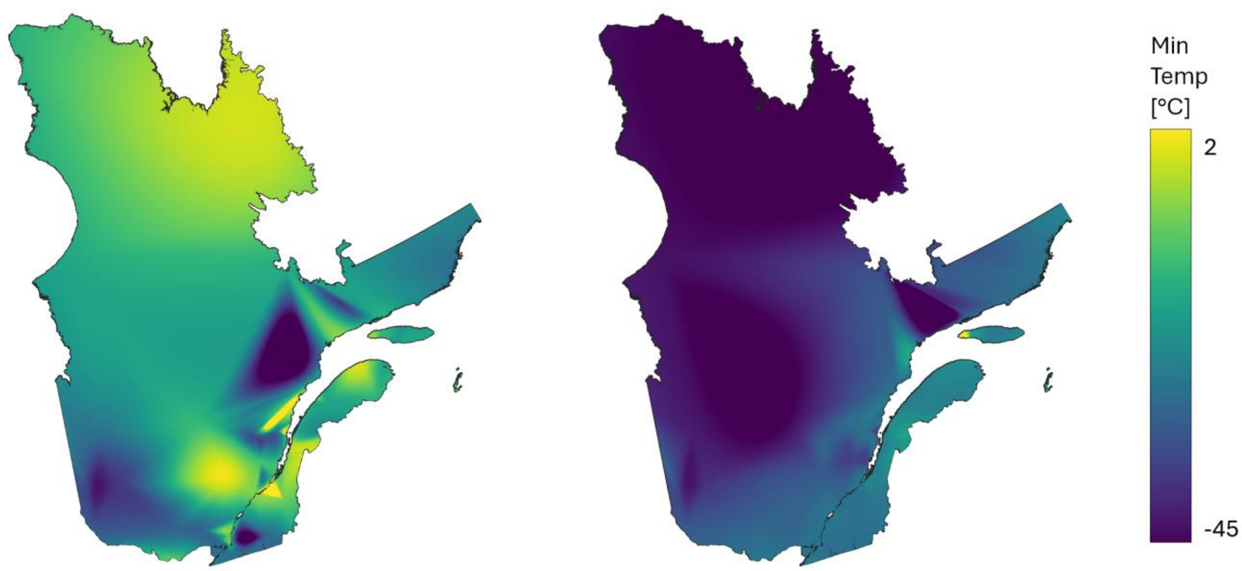

Figure 9: Minimum temperature on Jan. 13 according to the individual TMYs (left) or the provincial TMY (right).

The methodology could also be adapted to other countries, assuming that census and meteorological data, including dry-bulb and dew-point temperature, wind speed and GHI are available. The details of the methodology (for example weather parameters' weights) can also be adapted to the preferred variant of the Sandia method. Depending on the purpose of the large-scale TMY, it could be interesting to weigh the meteorological stations regarding to energy consumption instead of population, if the data are available.

Acknowledgment

The authors would like to thank Robert Morris for providing the original Fortran program used by ECCC to produce the Canadian TMYs (CWEC).

Nomenclature – Symbols

Table 6 List of variables

Name	Description
CDF	Cumulative distribution function
$closeness$	Normalized index evaluating the closeness of temperature and GHI mean and median for a particular year compared to the long-term values
ERR	Closeness of temperature or GHI mean and median for a particular year compared to the long-term values
FS	Finkelstein-Schafer statistic
j	Type of run for the persistence criteria
L	Length of runs
m	Month, from 1 to 12
N	Number of runs
R_1	Ratio evaluating the lengths of runs for a year. Close to 0 if the year has extremely long runs
R_2	Ratio evaluating the number of runs for a year. Close to 0 if the year has an extremely large number of runs
y	Year, from 1998 to 2017
z	Station number

References

ECC Canada. 2020. "Canadian Weather Energy and Engineering Datasets (CWEEDS) and Canadian Weather Year for Energy Calculation (CWEC) - 2020 Update." Toronto, ON, CAN: Environment and Climate Change Canada. https://climate.weather.gc.ca/prods_servs/engineering_e.html.

Hall, I., R. Prairie, H. Anderson, and E. Boes. 1978. *Generation of Typical Meteorological Years for 26 SOLMET Stations. Report SAND78-1601*. Albuquerque, NM: Sandia National Laboratories.

Huang, Joe. 2020. "Update of California Weather Files for Use in Utility Energy Efficiency Programs and Building Energy Standard Compliance Calculations." PGE0450. Oakland, CA, USA: Pacific Gas and Electric Company. http://weather.whiteboxtechnologies.com/pdf/PGE0450.pdf.

ISO. 2005. *ISO 15927-4:2005. Hygrothermal Performance of Buildings -- Calculation and Presentation of Climatic Data - Part 4: Hourly Data for Assessing the Annual Energy Use for Heating and Cooling*. Geneva, CHE: International Organization for Standardization.

Morris, Robert. 2016. *Final Report – Updating CWEEDS Weather Files*. Toronto, ON, CAN: Environment and Climate Change Canada. https://climate.weather.gc.ca/prods_servs/engineering_e.html.

———. 2022. "Fortran Program to Generate CWEC Files (Personal Communication)," 2022.

QGIS.org. 2022. "QGIS Geographic Information System." QGIS Association. http://www.qgis.org/.

StatsCan. 2022. "Population and Dwelling Counts: Canada and Census Subdivisions (Municipalities) - Table: 98-10-0002-01 (2021 Census)." Government of Canada, Statistics Canada. February 9, 2022. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=9810000201.

Wilcox, Steve, and W. Marion. 2008. *Users Manual for TMY3 Data Sets (NREL/TP-581-43156)*. Golden, CO, USA: National Renewable Energy Laboratory.

Zang, Haixiang, Qingshan Xu, Pengwei Du, and Katsuhiro Ichiyanagi. 2012. "A Modified Method to Generate Typical Meteorological Years from the Long-Term Weather Database." *International Journal of Photoenergy* 2012 (February): e538279. https://doi.org/10.1155/2012/538279.