NRC·CNRC CONSTRUCTION

Applicability of ASHRAE's damage function to predict moisture severity of climate for Canadian locations

Research Report No.: NRCC-CONST-56590E
Report Date: 16 May 2022
Author(s): Naman Bansal, Maurice Defo, and Michael A. Lacasse

CONSTRUCTION

National Research Council Canada
Conseil national de recherches Canada

Canada

Applicability of ASHRAE's damage function to predict moisture severity of climate for Canadian locations

Report No:	NRCC-CONST-56590E
Report Date:	16 May 2022
Author(s):	Naman Bansal, Maurice Defo, and Michael A. Lacasse

NRC CONSTRUCTION

(This page is intentionally left blank)

Table of Contents

List of Figures

List of Tables

Executive Summary

Hygrothermal simulation tools are commonly used to assess the moisture performance of building envelope components. Owing to the computational costs required to complete simulations over the long-term, one approach to reduce simulation time when undertaking hygrothermal design analysis is to select representative year(s) amongst sets of long-term climate data. To properly select these moisture reference year(s), a method is required to predict moisture performance and rank the climate years in terms of their moisture severity. To this end, several methods have been proposed in the literature, amongst which is the damage function method as reported in ASHRAE Project Report RP-1325. In this method, a stepwise regression model was developed to predict the damage function, as characterized by the RHT-index (integral of (Temperature - 0) (Relative Humidity – 70%)), in an OSB layer of a wood-framed wall as a function of several average yearly climate parameters for a North facing wall. The model was calibrated using climate and simulation data for eight cities in the USA and validated for three cities in the USA and one city in Canada (Winnipeg, MB). The method was found to be the most consistent and accurate amongst all ranking methods that have been evaluated. The objective of this paper was to: (1) evaluate the ASHRAE's method for several Canadian locations; (2) determine whether the original model can be recalibrated in the event it is shown to be deficient; and (3) explore the potential of improving the model using other approaches such as the Partial Least Squares Regression (PLSR), the Least Absolute Shrinkage and Selection Operator (LASSO) regression, and the combination of LASSO feature selection and Support Vector Regression (SVR). The results suggest that for some Canadian locations, as were investigated in this study, the use of the original model may not be appropriate for predicting the moisture performance and ranking of climate years in terms of their moisture severity. However, the model was shown to perform better after it was recalibrated for Canadian locations but without improvement in ranking performance. Furthermore, the damage function model can be improved by using either PLSR, LASSO or SVR in terms of prediction and ranking.

1 Introduction

Throughout its life, the building envelope is subjected to several stresses due to, *e.g.*, differences in outdoor and indoor temperatures, vapor pressure and air pressure. These stresses cause the movement of heat, air and moisture across the building envelope. If the wall assembly is not properly designed and maintained, it may suffer mould growth, which may pose health issues as well aesthetic concerns, decay of bio-based materials and components, frost damage to concrete and mortar, as well as brick and stone masonry units, and corrosion of metallic components. All these effects are associated with moisture accumulation within walls from either interstitial condensation, rain water leakage or rising damp. They can significantly reduce the service life of buildings by affecting, not only its structural integrity, but also the health and safety of building occupants. Thus, the building envelope must be designed and constructed in such a manner as to ensure its long-term serviceability. For this, the designer must be able to anticipate the effect of climatic loads on the moisture performance of envelope components and materials. This can be done through in-situ measurements, laboratory tests or hygrothermal simulations. Field measurements and laboratory tests are expensive and time-consuming, and limited to the specific wall and conditions of experience. Hence, the preferred method is to perform hygrothermal simulations. In fact, once the hygrothermal model has been developed and validated, it can be used in all situations to provide rapidly results from which to infer the durability of building envelope components and materials.

Numerical hygrothermal models permit simulation of coupled heat, air and moisture (HAM) transport in building envelope. There are a number of numerical simulation packages that are either wholly or partially dedicated to HAM simulations in building envelope (Hill, 2005; Delgado, 2010, 2013). Amongst all these models, a few are in the public domain or commercially available such as WUFI[1], DELPHIN[2] and COMSOL Multiphysics[3]; these are the most commonly used hygrothermal software platforms around the world by building practitioners, designers, academia or researchers. They provide a ready means of assessing the moisture performance of building envelope components and materials when the assembly is subjected to outdoor and indoor climatic loads. However, the year-to-year variation in climate data requires, in principle, that simulations be performed for longer terms, i.e., over 10 to 30 consecutive years. Thus, longer computing times are needed, especially when considering 2-D and 3-D simulations. This problem is exacerbated when uncertainties in simulation parameters and material properties are to be considered using a stochastic approach. Moreover, climate change needs now to be taken into account by designers as there is evidence around the globe, perhaps also evident in every day weather, that the climate is changing much more drastically than recorded in the past, and the expectation is that climate change will continue into the foreseeable future (IPCC, 2014). There are several scenarios of climate change and global climate models. As such, uncertainties related to the future climate also need to be taken into account during the design, which means more and more hygrothermal simulations to be performed.

Owing to the high computing costs of the long-term simulations, especially when uncertainties in simulation parameters, material properties and future climate are to be considered, one approach is to select representative year(s) among the climate data series, which should either give results similar to that which would be obtained using the entire climate data series or otherwise impose a severe stress on the building envelope to achieve the desired level of safety regarding the risk to the occurrence of moisture damage. These are called Moisture Reference Year (s) (MRYs). To properly select the MRYs, a method is required to estimate and rank the climate years in the climate series in terms of their moisture severity. To this end, several indices have been developed to assess the moisture severity of climate years. Among these indices are the: PI-factor (Hagentoft and Harderup, 1996), Moisture Index, MI, (Cornick *et al*., 2003), Index of moisture severity, Isev, (Salonvaara, 2010) developed in the framework of ASHRAE's research project RP-1325 (Salonvaara, 2011), and Climatic Index, CI, (Zhou *et al*., 2016).

In their study, Salonvaara *et al*. (2010, 2011) evaluated the moisture response of two wall systems: (i) A stucco-clad light-weight wall (LWW) consisting of (from exterior to interior): conventional stucco with an acrylic finish, 60

---

[1] https://wufi.de/en/software/wufi-pro/

[2] http://bauklimatik-dresden.de/delphin/index.php?aLa=en

[3] https://www.comsol.com/

min asphalt impregnated paper based water resistive barrier, oriented strand board (OSB) sheathing, 2×4 fiberglass insulation, kraft paper vapor retarder and drywall (gypsum board) with primer and latex paint; and (ii) A heavy-weight wall (HWW) consisting of (from exterior to interior): brick cladding, air cavity (non-vented), concrete masonry unit (CMU) block, R-13 fiberglass insulation with metal frame, kraft paper and drywall with one coat of latex primer and one coat paint layer. The climate data used comprised 30 consecutive years, from 1961 to 1990. The damage functions selected for ranking the years based on hygrothermal simulations were Time of Wetness (TOW), RHT-Index (Eq. (1)), Mold Growth Index, and Maximum Moisture Content, calculated for the critical component considered as OSB sheathing and outer wythe of the CMU block for LWW and HWW, respectively.

$$RHT = \sum_{i=1}^{n} (T_i - T_L)\,(RH_i - RH_L) \tag{1}$$

Where i is the current and hour; n is the total number of hours; T is the temperature (°); T_L is the critical temperature (°); RH is relative humidity (%); and RH_L is the critical relative humidity (%). The product $(T_i - T_L)(RH_i - RH_L)$ is computed in any given hour only if $T_i > T_L$ and $RH_i > RH_L$.

For computing the RHT index (Eq. (1)) from hygrothermal simulations results, the critical T and RH were set to 0ºC and 70%, respectively. This was called RHT70. The RHT70 values were used to rank the years in terms of their moisture severity and compared to other damage functions calculated using hygrothermal simulations results. The authors observed similar years ranking for both types of wall and for all other damage functions considered. Therefore, the moisture severity index (Eq. (2)) was developed with only the stucco-clad wood-framed wall and the RHT70 values derived from Eq. (1), using stepwise multiple regression analysis.

$$\begin{aligned} Isev = 108307 - 241SW - 1391C - 312326RH + 183308\,WDR + 15.2\,Pv + 27.3\,T^2 \\ + 261079RH^2 - 0.00972Pv^2 \end{aligned} \tag{2}$$

Where SW is the average yearly solar radiation normal to the north face of the building (W/m^2); C is the average yearly cloud index (Oktas); RH is yearly average relative humidity (-); WDR is the average yearly wind driving rain on the north orientation (mm), T is the average yearly temperature (ºC), and P_v is the average yearly vapour pressure (Pa).

The ranking of the years based on Isev (Eq. (2)) was compared to those obtained with other indices such as MI, PI-factor and CI and to those obtained using damage functions (Time of wetness, RHT70, Mould index, Maximum moisture content) derived from HAM simulations. The authors concluded that the ranking based on Isev was the most consistent and accurate of all analyzed methods in selecting the most severe years in terms of their hygrothermal performance in all locations considered in their study. They used the model to determine the design weather year (year having the third-highest moisture severity index (Isev) value in a 30-year series) in 100 US and 7 Canadian locations (Salonvaara *et al*., 2011). The Isev values were used to produce the contour plot of values for a map of North America. In a more recent study, Aggarwal *et al*. (2020) compared various moisture performance indices including the Isev, MI, and CI, for their accuracy in sorting the individual years in a series of climate data in terms of their potential for causing damage to building envelope components. The result showed that the Isev method performs better than either the MI or the CI index when compared to ranking using mould index calculated with hygrothermal simulations results.

Although the Isev method appears to be the most consistent and accurate method, it is not without drawbacks. The Isev equation was calibrated using stepwise multiple regression analysis and included several climatic and structural variables where some are correlated to each other, making the use of multiple linear regression inappropriate (James *et al.*, 2013). Moreover, it was developed using climatic data for eight locations across the USA (Seattle–Washington, New Orleans-Louisiana, Minneapolis-Minnesota, Chicago-Illinois, Atlanta-Georgia, Portland-Maine, Baltimore-Maryland, and San Francisco-California) and then validated using climate data of three US cities (Fairbanks-Alaska, Memphis-Tennessee, and Miami-Florida) and only one Canadian city (Winnipeg - Manitoba). On another side, the climate is evolving, and the recent data differ from those used for developing the model. In fact, in a preliminary study, it was found that the original model could not readily predict RHT70 in several Canadian cities with recent climate data. Hence, the objectives of this study were to:

- Recalibrate and re-evaluate the predictive and ranking performance of the Isev equation for Canadian locations.
- Explore the potential of improving the prediction and ranking performance using other methods such as the Partial Least Squares Regression (PLSR), the Least Absolute Shrinkage and Selection Operator (LASSO) regression, and the combined LASSO variable selection and Support Vector Regression (SVR).

2 Approach

The steps followed to realize the objectives of this study are:

1) Select 12 cities across Canada and group them into training and evaluation sets using principal components analysis (PCA).
2) Perform hygrothermal simulations to derive the response variable.
3) Evaluate the predictive and ranking accuracies of the original Isev equation for the Canadian cities.
4) Recalibrate the Isev equation using cities in the training set and validate using cities in the evaluation set.
5) Develop new models based on LASSO, PLSR and combined LASSO and SVR, and compare their performance to that of the Isev method.

These steps are described in the following sections.

2.1 Cities selected

Twelve Canadian cities were selected for this study. Their location and current climatic design data, as found in NBC (2015), are shown in Table 1. The selected cities, as shown in Figure 1, are located in the far north of Canada (Whitehorse, YT), as well as from the west to the east coast, and cover a range of values for moisture index (MI) and Heating Degree-Days. The definition of moisture index can be found in Cornick *et al*. (2003).

Table 1. Location and climatic design data of selected cities

City (Province)	Latitude (°)	Longitude (°)	Time zone	Climate zone[a]	HDD[b]	Moisture index	Rain (mm)
Whitehorse (YT)	60.7	-135.1	-8	7B	6580	0.5	170
Vancouver (BC)	49.3	-123.1	-8	4	2825	1.4	1325
Calgary (AB)	51.1	-114.1	-7	7A	5000	0.4	325
Saskatoon (SK)	52.1	-106.7	-6	7A	5700	0.4	265
Winnipeg (MB)	49.9	-97.1	-6	7A	5670	0.6	415
Toronto (ON)	43.7	-79.4	-5	5	3800	0.9	730
Ottawa (ON)	45.3	-75.4	-5	6	4500	0.8	750
Montreal (QC)	45.5	-73.6	-5	6	4200	0.9	830
Moncton (NB)	46.1	-64.8	-4	6	4680	1.0	850
Charlottetown (PE)	46.2	-63.1	-4	6	4460	1.1	900
Halifax (NS)	44.7	-63.6	-4	6	4000	1.5	1350
St-John's (NL)	47.6	-52.7	-4	6	4800	1.4	1200

[a]Climate zones 4, 5, 6, 7A and 7B correspond to zone with HDD ranging from 2000 to 2999, 3000 to 3999, 4000 to 4999, 5000 to 5999, and 6000 to 6999, respectively.

[b]HDD: Heating Degree-Days below 18 °C

YT: Yukon; BC: British Columbia; AB: Alberta; SK: Saskatchewan; MB: Manitoba; ON: Ontario; QC: Quebec; NB: New Brunswick; PE: Prince Edward Island; NS: Nova Scotia; NL: Newfoundland & Labrador

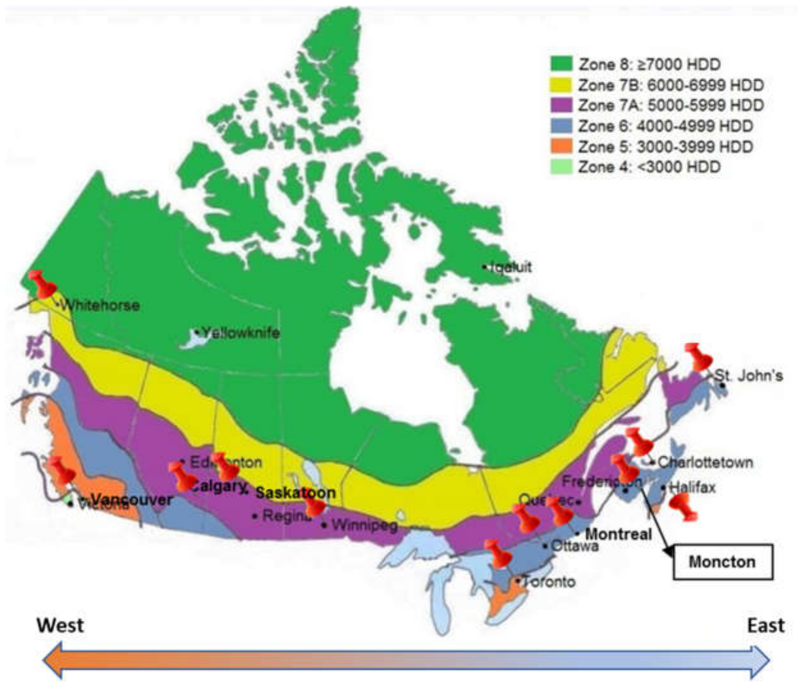


Figure 1. Geographical location of cities selected for this study.

2.2 Climate data

2.2.1 Source of climate data

Climate data used in this study were extracted from a large ensemble of modelled historical and future climate data developed by Gaur *et al.* (2019) for building energy and hygrothermal simulations. The datasets comprise, for each city, three 31-year long hourly time series corresponding to the historical baseline (1986-2016) and two future time periods, coincident with globally averaged future global warming of 2 ºC and 3.5 ºC. These two levels of global warming are expected to be reached in the future for a high greenhouse emission scenario (RCP8.5), respectively, over the time periods of 2034-2064 and 2062-2092. The climatic datasets were generated to capture the effects of the internal variability of the future climate in fifteen hourly realizations that are part of the 50-member datasets derived from the large ensemble of climates simulated by the Canadian Regional Climate Model - version 4 (CanRCM4). Each of the fifteen hourly realizations was initialized under a different set of initial conditions in the Canadian earth system global climate model (CanESM2). For the purpose of this study, only one realization of the ensemble modelled historical data was randomly selected in each city.

2.2.2 Variability of climate among cities

The differences in some of the climatic variables for the twelve cities considered are illustrated Figure 2. Each boxplot represents the 31 yearly averages or sums of the selected realization in each city. It can be observed that:

- Temperature (T) – Vancouver (west coast) is the warmest city, whereas Whitehorse is the coldest city.
- Relative humidity (RH) – cities on the east coast (Moncton, Charlottetown, Halifax and St. John's) exhibit higher average yearly RH values whereas Vancouver and Calgary have the lowest.
- Vapour pressure (P_v) – highest outdoor vapour pressure values are found in Vancouver, whereas the lowest values are found in Whitehorse.

- Cloudiness (C) – Cloud cover is relatively higher in Whitehorse, Vancouver and St. John's than in other cities.
- Wind direction (WD, Deg. from North) – predominant in the south-south-east direction in Whitehorse and Vancouver and between south-south-west and south-west for all other cities.
- Wind speed (V) – lower in Vancouver, moderate in Whitehorse, Calgary, Saskatoon, Winnipeg, Toronto, Ottawa, Montreal and relatively higher for cities on the east coast.
- Wind driving rain (WDR) (north orientation) – WDR values, calculated for the wall facing the north orientation, are relatively greater for the cities of Moncton and Charlottetown and smaller in the cities of Whitehorse and Vancouver.
- Global solar radiation or shortwave radiation (SW) (north orientation) – Solar radiation values normal to the north face of the building are higher in Toronto, Ottawa and Montreal, whereas lower values are found in Whitehorse.

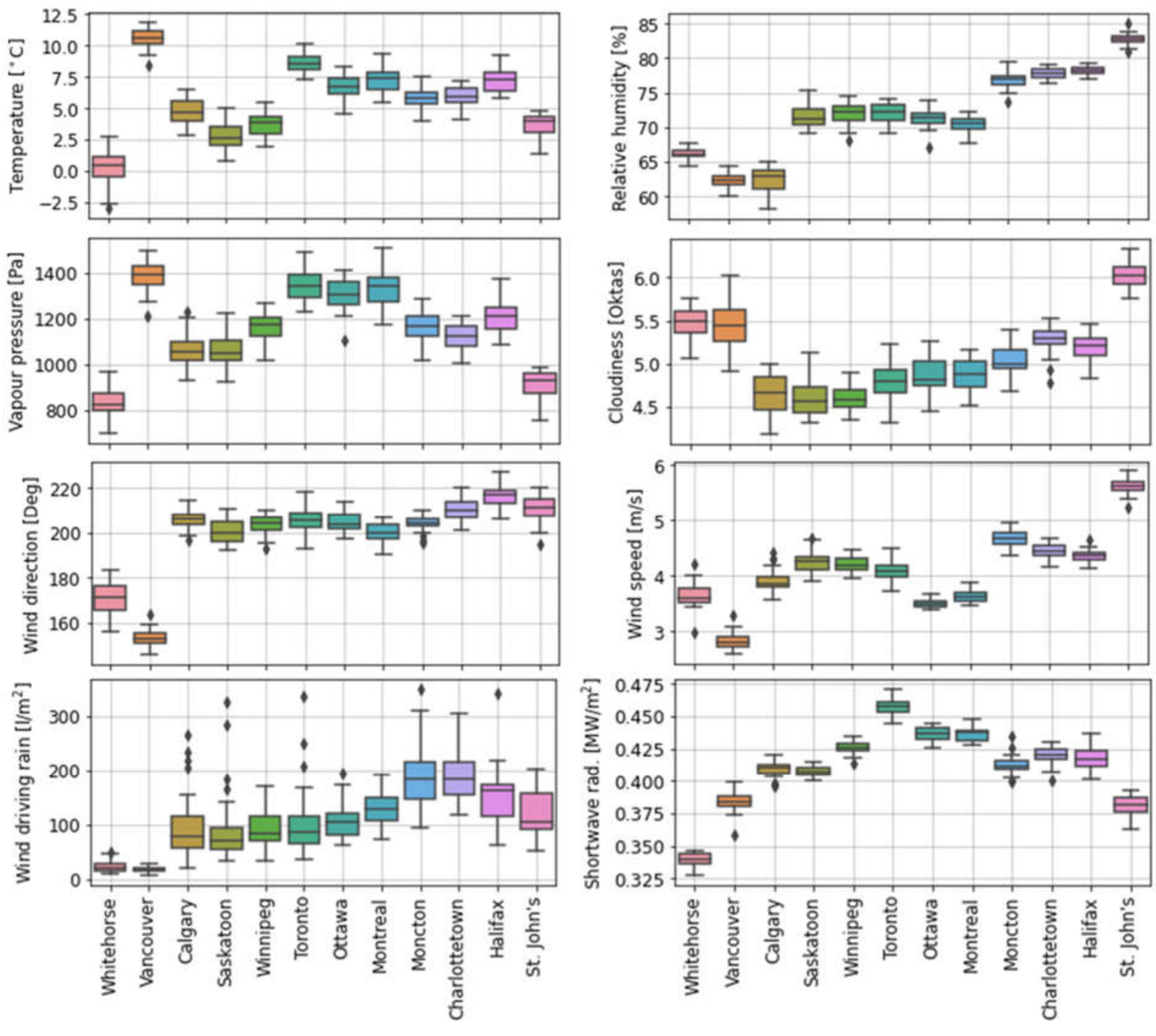


Figure 2. Boxplot of average the 31 yearly climatic variables in the cities selected for the study. Wind driving rain and shortwave radiation are annual sums and were calculated for the north orientation. All other variables are annual averages. The boxplot shows the: minimum (Q1 – 1.5*IQR), 25th percentile (Q1), median, 75th percentile (Q3), maximum (Q3 + 1.5*IQR) and the outliers (diamond). IQR is the interquartile range.

2.2.3 Correlation among climate variables

The scatter plots in Figure 3 show the nature of the relation among the variables, whereas the correlation matrix in Figure 4 shows the strength of the relationship among climate variables, for all the 12 cities considered. When considering all the cities together, some climate variables are highly positively correlated, like average vapour pressure and temperature (correlation coefficient (r) = 0.90), average relative humidity and wind speed (r = 0.79). Other combinations of climate variables show a moderately positive or negative correlation, such as sum shortwave radiation and average cloudiness (r = -0.60), while sum WDR and average relative humidity have a positive relationship (r = 0.57). Finally, some climate variable has a weak relationship between them, such as sum WDR and average temperature (r = 0.09).

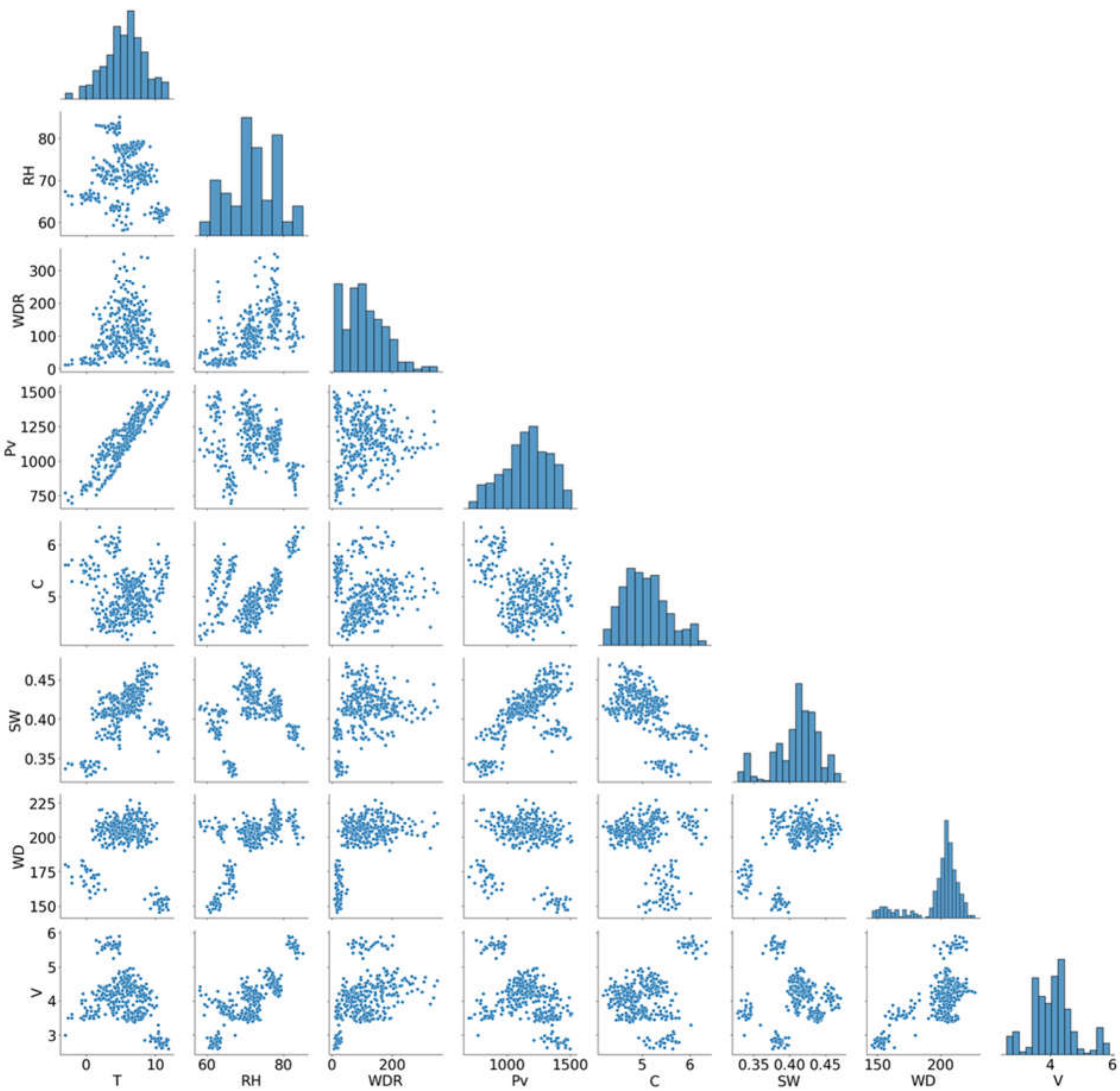


Figure 3. Scatter plot matrix for climate variables.

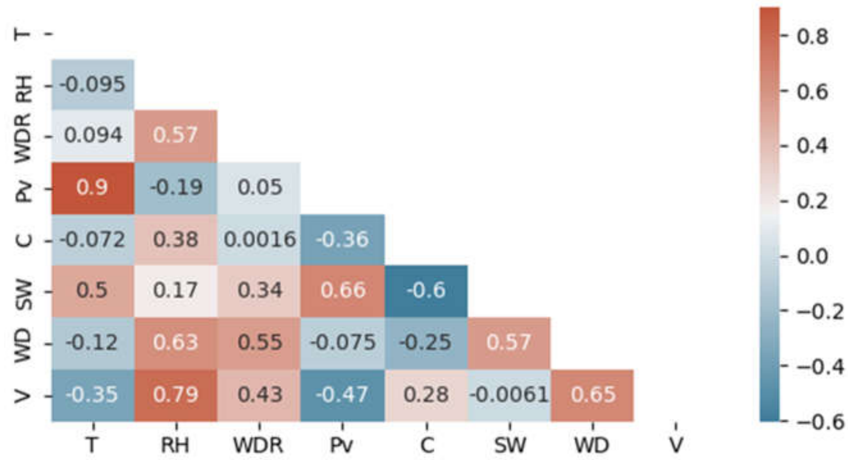


Figure 4. Coefficient of correlation among climate variables

2.3 Formation of training and test sets

To select a representative training subset of cities that provides uniform coverage over the data set and includes samples on the data set's boundary, principal components analysis (PCA) was performed using the climate data for all 12 cities. Section 2.2.3 shows the strength of correlation among climate variables. Section 2.3.1 presents an overview of the PCA method and Section 2.3.2 shows how the training and test cities were selected.

2.3.1 Principal components analysis

The idea supporting the PCA is that there are numerous features that are correlated among themselves and that can be combined linearly to form a new set of data with reduced features. As shown in Figure 4, there are many climate variables that are moderately or highly correlated. As such, these data are candidate for PCA analysis to extract the underlying structure. A description of the PCA can be found in James et al. (2013). It is an unsupervised dimensional reduction technique that is used to summarize a set of correlated explanatory variables into a smaller set of representative variables, referred to as principal components (PCs), which are orthogonal to each other. The PCs are a normalized linear combination of the original explanatory variables. Once the principal components are identified, the data are projected on a hyperplane defined by these principal components to produce a low dimensional view of the data and permit visualization of possible patterns in the data.

Given a data matrix with n observations and p explanatory variables, the first principal component or scores (projected observations) z_{i1} can be expressed as (Eq. (3)):

$$z_{i1} = \emptyset_{11}x_{i1} + \emptyset_{21}x_{i2} + \ldots.. + \emptyset_{P1}x_{iP} \tag{3}$$

Where $\emptyset_1 = (\emptyset_{11}, \ldots.. \emptyset_{p1})^T$ are referred to as the loading vectors or eigenvectors of the first principal component and $x_{i1}, x_{i2} \ldots \ldots x_{ip}$ are the explanatory variables. Normalization implies that $\sum_{j=1}^{p} \emptyset_{j1}^2 = 1$.

The first principal component accounts for the largest amount of variance in the explanatory variables, followed by the second principal component, which is orthogonal (uncorrelated) to the first one, and accounts for the largest amount of remaining variance. The determination of the PCs continues until the total number of principal components equals the number of explanatory variables, each succeeding principal component capturing the remaining variation without being correlated with the previous component. In practice, the first few PCs are sufficient as they contribute to a large percentage of the variance in the data.

Computing the first component involves looking for the linear combination of features value that has the largest sample variance. Mathematically, this can be framed as an optimization problem (Eq. (4)) which is solved via eigen decomposition. The computations are similar for other components with the added constraint of orthogonality with the previous component. The R package *Stats* (R Core Team, 2021) was used to perform the PCA.

$$\underset{\emptyset_{11},\ldots\ldots\emptyset_{p1}}{Maximize} \left\{ \frac{1}{n} \sum_{i=1}^{n} {z_{i1}}^2 \right\} \quad subject\ to \sum_{j=1}^{p} \emptyset_{j1}^2 = 1 \tag{4}$$

2.3.2 Formation of the training and test sets

First, individual yearly statistics, including average vapour pressure, temperature, wind speed, wind direction, relative humidity, cloudiness, the sum of shortwave radiation and the sum of WDR, were calculated. The annual sums of WDR and shortwave radiation were used rather than the annual average considered in the Isev model, given that they are usually reported as an annual sum.

The biplot in Figure 5 shows the placement of cities based on the linear combination of the 10 yearly climate characteristics considered. The plot was produced using the R package *factoextra* (Kassambara and Mundt, 2020). The first two principal components together explain about 73% of the variation in the data. Each city is composed of 31 data points representing the 31 individual yearly values from 1986 to 2016. From the biplot, it is evident that cities like Vancouver, Whitehorse and St. John's have unique climate conditions that make them appear as outliers compared to the remaining cities. For example, Vancouver receives the lowest WDR in the north orientation and has lower outdoor average relative humidity, while Whitehorse receives the lowest annual average global shortwave radiation. As well, St. John's has the highest annual average wind speed and relative humidity. These observations are in agreement with those in Figure 2.

On the other hand, cities in the middle are closer to each other in terms of their climate characteristics. For example, Montreal and Ottawa climates show an excellent overlap while being the closest to Toronto climate conditions. Similarly, Moncton, Charlottetown and Halifax have similar climate conditions on average, while there is a partial overlap between Saskatoon, Winnipeg, and Calgary climates.

Referring to the clustering shown in Figure 5, seven cities that included Whitehorse, Vancouver, St. John's, Winnipeg, Calgary, Ottawa, and Moncton were chosen to be part of the training set. These cities were explicitly chosen to be in the training set as they best represent the variation in the climate for all the 12 cities considered. The rest of the cities, *i.e.*, Charlottetown, Halifax, Montreal, Saskatoon, and Toronto, formed the test set.

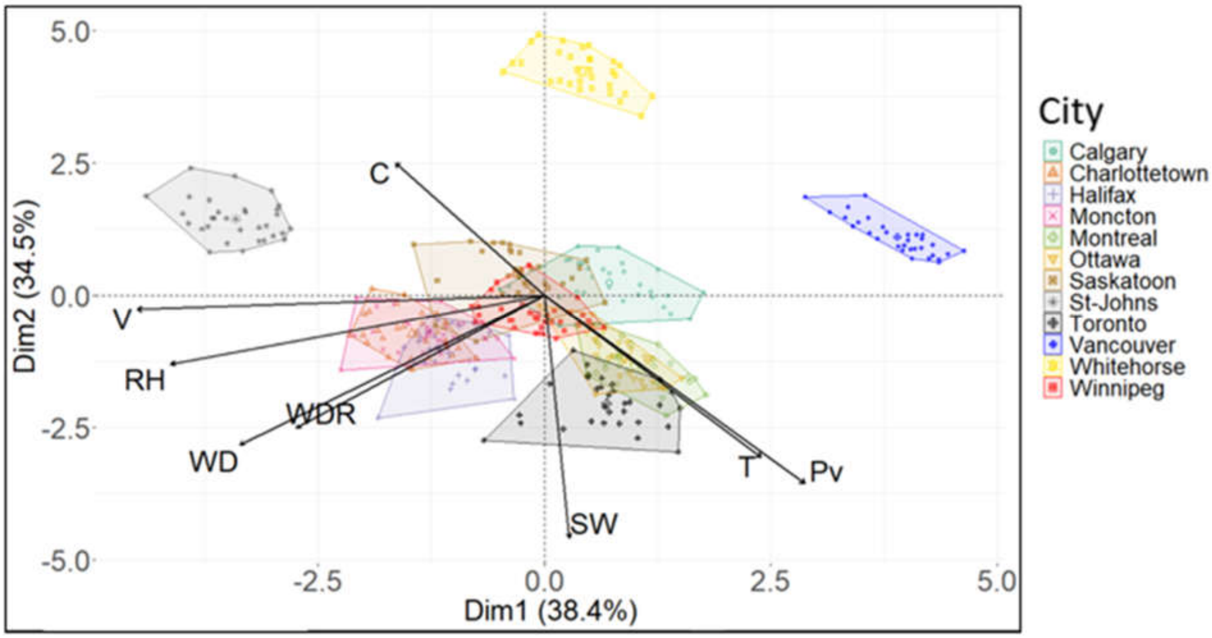


Figure 5. PCA biplot showing the clustering of cities by similarities based on the first two principal components.

2.4 Hygrothermal simulations

A wood-framed wall nearly similar to that used for developing the Isev model was used in this study (Figure 6). It was composed of (from exterior to interior):

- 19-mm regular Portland stucco with acrylic finish;
- 1.5-mm unvented air cavity;
- Weather Resistive Barrier (30 minute asphalt impregnated paper, 0.24 mm);
- 11-mm Oriented Strand Board (OSB);
- 140-mm mineral fibre insulation located in the stud cavity;
- 0.15-mm polyethylene vapour barrier;
- 12.7-mm interior grade gypsum panel with latex primer and one coat of latex paint.

The few differences between the walls used for developing the Isev model and that used in this study were the addition of a 1.5-mm unvented air cavity behind the stucco to account for the space created by the wire mesh and the use of a vapour barrier (current practice in Canada) rather than a vapour retarder.

Hygrothermal simulations were performed using DELPHIN v5.9 (Nicolai and Grunewald, 2002) on a vertical section of the wall passing through the middle of the stud cavity, far from and not including spruce wood studs. In this position in the wall, the heat and mass flow are almost unidirectional and can be represented by a one-dimensional configuration as shown in Figure 6. Simulations were performed for the 31-individual year of each selected realization to provide data for modelling.

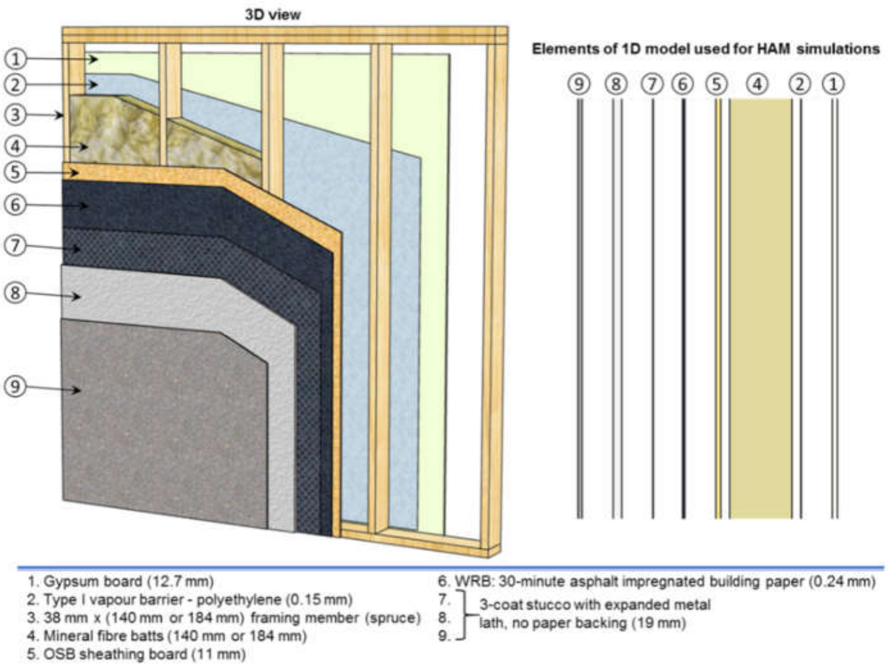


Figure 6. Stucco-clad wood-framed wall assembly used in this study.

2.4.1 Material properties

Material properties were retrieved the Kumaran *et al*. (2002). Some of the primary hygrothermal properties of materials used in the one-dimensional configuration are shown in Table 2.

Table 2. Thickness (e), dry density, thermal conductivity (λ) at standard temperature, porosity, equilibrium moisture content (%kg/kg), vapour permeance (ng/m^{2}sPa) and water absorption coefficient (A) of building materials used in the one-dimensional configuration.

	e	Density	λ	Porosity	Equilibrium moisture content			Vapor permeance			A
Component/material	(mm)	(kg/m^3)	(W/mK)	(m^3/m^3)	50%RH	80%RH	95%RH	10%RH	50%RH	90%RH	(kg/m^{2}s$^{0.5}$)
Cladding											
Regular Portland stucco	19	1960	0.407	0.24	3.55	5.27	7.63	30.6	94.7	160.5	0.01230
Weather resistive membrane											
30 minute building paper	0.15	464	0.248	0.01	-	-	-	4080.0	4080.0	4080.0	0.00031
Sheathing board											
OSB	11	600	0.094	0.96	6.70	11.60	21.50	23.3	111.8	370.9	0.00220
Insulation											
Mineral fibre	140	37	0.032	0.66	0.15	0.90	14.00	935.7	935.7	935.7	-
Vapour barrier											
Polyethylene	0.15	1256	0.159	0.25	-	-	-	-	-	-	-
Interior sheathing											
Gypsum + primer + latex	12.7	700	0.160	0.40	9.00	10.70	13.00	107.1	399.6	1629.9	0.00190

- : means do data

2.4.2 Wind driving rain calculation

The model from the ASHRAE Standard 160 (ASHRAE, 2016) is given by Eq. (5):

$$WDR = F_E.F_D.F_L.U_{10}.cos\theta.R_h \quad (5)$$

Where: WDR is the wind driving rain (kg/(m^2.h)) calculated on the north orientation; F_E is the rain exposure factor, depending on the building height, the exposure category (Table 3); F_D is the rain deposition factor accounting for the spatial distribution of the WDR on the façade; F_L is an empirical constant (= 0.2 kg·s/(m^3.mm)); U_{10} is the hourly mean wind velocity at 10 m; θ is the angle between the normal to the wall and the wind direction; and R_h is the rain intensity on the horizontal surface (mm/h).

Table 3. ASHRAE exposure factor, FE.

Building height (m)	Type of exposure category		
	Severe	Medium	Sheltered
≤ 10	1.4	1.0	0.7
> 10 and ≤ 20	1.4	1.2	1.0
> 20	1.5	1.5	1.5

There are three categories of exposure factor F_E (Table 3) that are a function of terrain type: severe, medium and sheltered. Severe exposure includes hilltops, coastal areas, and funnelled wind (*e.g*. wind tunnel effect caused by the proximity of two buildings). Sheltered exposure includes protection from nearby buildings or other permanent moderating features (e.g. trees). The rain deposition factor F_D is 0.35 for walls below a steep-slope roof, 0.5 for walls below a low-slope roof and 1.0 for walls subject to rain runoff.

The building considered in this study was a 3-storey (10-m height) residential building with low-slope roof located in a suburban area. As such, for calculating the WDR (Figure 2), the exposure and rain deposition factor were defined as 1.0 and 0.5, respectively.

2.4.3 Boundary conditions

2.4.3.1 Outdoor boundary conditions

The external boundary conditions are applied to the exterior surface of the cladding. The exterior conditions consist of the climate loads for a given location, which includes the following outdoor climate variables:

- Outdoor wind speed and direction
- Wind driving rain
- Outdoor temperature
- Outdoor relative humidity
- Solar radiations (direct and diffuse) normal to the wall
- Sky longwave emission or atmospheric counter radiation normal to the wall
- Ground longwave emission normal to the wall

The method of determining the WDR was described in Section 2.4.2.

Solar shortwave radiations (direct and diffuse) were provided directly to DELPHIN, with the information needed to compute the normal components, i.e., geographical location (latitude and longitude), time zone, and wall orientation. The shortwave absorption coefficient of the cladding was set to 0.35 for the stucco surface, assuming a white-colored surface (Henninger, 1984).

The sky and ground longwave emissions were explicitly calculated using Eqs. (6) and (7), respectively, and provided to DELPHIN, assuming longwave emissivity of 0.9 for the surrounding ground (ε_{ground}), and 1.0 for the sky (ε_{sky}):

$$q_{lw\_sky} = \sigma \varepsilon_{sky} T_{sky}^4 \tag{6}$$

$$q_{lw\_ground} = \sigma \varepsilon_{ground} T_{ground}^4 \quad (7)$$

Where: $q_{lw_{sky}}$is the atmospheric counter radiation (W/m^2); σ is the Boltzmann constant (W/m^{2}K^4); T_{sky} is the sky temperature (K); q_{lw_ground} is the ground counter radiation (W/m^2), and T_{ground} is the ground temperature (K). The sky temperature was calculated using equation (8) provided in DELPHIN manual:

$$T_{sky} = T_{air}\sqrt{\left(0.82 - 0.25\ 10^{-0.75006\ 10^{-3} p_v}\right)(1 + 0.2C)} \quad (8)$$

Where: T_{air} is the air temperature (K); p_v is the vapour pressure in the air (Pa) and C is the cloud covering factor.

The ground surface temperature and albedo were set to the air temperature and 0.2, respectively. The longwave emission coefficient of the wall surface was set to 0.9.

The outdoor convective heat transfer coefficient was calculated using Eq. (9) (ISO, 2009):

$$\alpha_c = 4 + 4.U \quad (9)$$

Where α_c is the outdoor convective heat transfer coefficient (W/m^2 K); and U is the wind speed corrected for the height of the building (m/s). The outdoor convective vapour transfer coefficient was calculated using the outdoor convective heat transfer coefficient and the Lewis number (Incropera and DeWitt, 1996).

2.4.3.2 Indoor boundary conditions

The indoor boundary conditions consist in defining a temperature and relative humidity condition. The indoor boundary conditions were selected as constants and set to 21^{o}C for temperature and 50% for relative humidity.

Referring to ISO 6946 Standard (ISO, 2009), the indoor convective heat transfer coefficient was set to 2.5 W/m^{2}K, whereas the indoor radiative heat transfer coefficient was set to 5.5 W/m^{2}K. The indoor vapour transfer coefficient was calculated using the convective indoor heat transfer coefficient and the Lewis number (Incropera and DeWitt, 1996). At normal pressure, the Lewis number, which is the ratio of thermal diffusion to mass diffusivity, is equal to 6.1 x 10^{-9}; given a convective heat transfer coefficient of 2.5, the indoor vapour transfer coefficient was determined to be 1.53 x 10^{-8} s/m.

2.4.4 Initial conditions

The initial T and RH for all components were set to 21 ^{o}C and 80%, respectively.

2.4.5 Simulation parameters

2.4.5.1 Meshing of the physical domain

For all the layers of the wall construction but the membranes (sheathing and vapour barrier), the size of the first and last element was set to 0.5 mm, then an expansion factor of 125% was used to generate the grids. For membranes, a constant mesh number of 3 was used.

2.4.5.2 Solver parameters

Table 4 shows the simulation parameters that were set in the DELPHIN solver. DELPHIN uses an implicit scheme for time stepping. The maximum time step was set to 1 h, corresponding to the time step of the climate data. Internally, the solver uses adaptive time steps, depending on the rate of convergence of the solution. The initial time step and the smallest time step permitted were set to 0.01 s and 10^{-5} s, respectively. The relative tolerance and the absolute tolerance (error level for which the iterations stop) were set to 10^{-6} and 10^{-7} and selected based on a compromise between the accuracy of the solution and computational time after preliminary evaluations.

Table 4. Parameters used for DELPHIN Solver

Parameter	Value
Relative tolerance	10^{-6}
Absolute tolerance for for moisture	10^{-7}
Initial time step (s)	10^{-2}
Maximum time step (h)	1
Smallest time step permitted	10^{-5}
Maximum method order	5

2.4.6 Calculation of the response variable

The RHT-index was used as the damage function. The OSB sheathing panel was used as the critical layer from which to extract results (*i.e*. RH and T) from hygrothermal simulations. These two variables were then used for calculating the cumulated RHT70 based on Eq. (1), with critical T and RH set to 0ºC and 70%, respectively.

2.5 Recalibration of the Isev equation

As mentioned in Section **Error! Reference source not found.**, the original Isev regression equation was developed using climate data (1961-1990) from 8 US cities and was only validated for one Canadian city, i.e., Winnipeg. The model was constructed by considering the yearly average of 8 climate elements and their second-order polynomial to predict the damage function RHT70 in the North Orientation. Then, using the Stepwise Regression method with criteria like R^2, adjusted R^2, and Mallows Cp statistics, the best model among the several possible regression models that best explained RHT70 was selected (Eq. (2)). A preliminary evaluation using more recent climate data (1986-2016) showed the model is not directly applicable to many Canadian cities (Section 3.2.1), thus the need to recalibrate the equation. The model was recalibrated using the training data as defined in Section 2.3.2 to capture the variation in the Canadian climate conditions. As a result, the regression coefficients in (Eq. (2)) were updated (Table 5), and the resulting regression model, shown in Eq. (10), was used to make predictions on the test set.

$$RHT70 = -144132.29 + 776.81SW - 122.51C + 247375.95RH + 248285.12WDR + 42.65\,Pv + 4.01T^2 - 157292.29RH^2 - 0.019Pv^2 \tag{10}$$

Table 5. New estimate of the coefficients of the Isev model

```
Coefficients:
                          Estimate     Std. Error t value
(Intercept)          -144132.293008   28234.764159  -5.105
AVG_Shortwave            776.818280     121.468239   6.395
AVG_Cloudiness          -122.512912     897.662951  -0.136
AVG_RH                247375.949307   67388.030564   3.671
AVG_WDR               248285.123102   28442.435832   8.729
AVG_Vapor_Pressure        42.657797      13.968358   3.054
AVG_Temp2                  4.009336      20.176563   0.199
AVG_RH2              -157292.289701   47094.505899  -3.340
AVG_Vapor_Pressure2       -0.019191       0.006609  -2.904

Residual standard error: 2342 on 208 degrees of freedom
Multiple R-squared:  0.852,   Adjusted R-squared:  0.8463
```

2.6 Development of new models

As shown in Figure 4, some of the climate variables are correlated to each other. This limits the use of multiple linear regression as one of the assumptions of multiple linear regression is the independence of feature variables. Therefore, new models that account for the collinearity in the feature variables were developed: PLSR, LASSO, and combined LASSO and SVR. Two new variables were also added: the annual average daily minimum (T_{min}) and maximum (T_{max}) temperature.

2.6.1 Relationship between the response variable RHT70 and the climate variables

Before developing the PLS, LASSO and combined LASSO and SVR models, the relationships between climatic variables and the response variable were first investigated using scatter plots (Figure 7). When considering data of all the 12 cities together, a linear trend can be seen with climate variables like the annual sum of shortwave, annual average temperature, vapour pressure, and wind direction against RHT70. However, the linear trend is not perfect, with cities like Vancouver and Whitehorse appearing as outliers, especially for temperature and vapour pressure. For the annual sum of WDR, the relationship is logarithm while the annual average relative humidity, wind speed and cloudiness appear to be polynomial of order 2 against RHT70, mainly due to the values in St. John's. To account for the logarithmic relationship between WDR and RHT70, the natural log of WDR was used as an input variable (Figure 8). For RH, cloudiness and wind speed, their square terms were added.

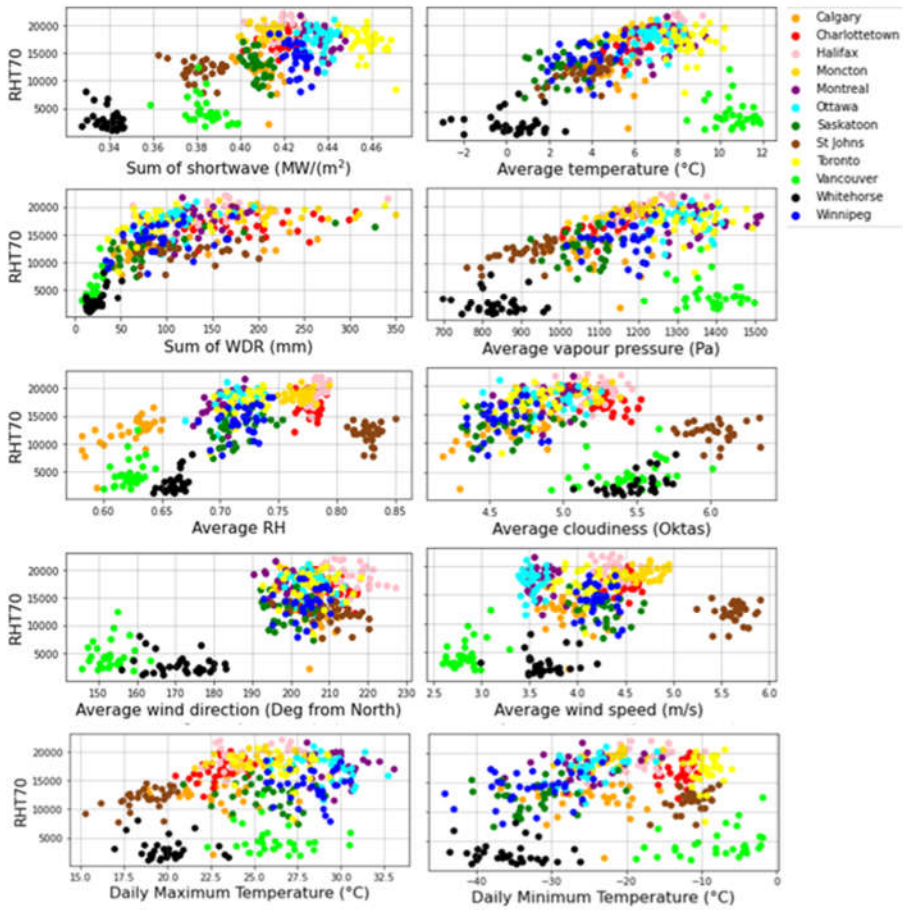


Figure 7. Relationship between annual value of climate variables and RHT70 for the 12 cities considered.

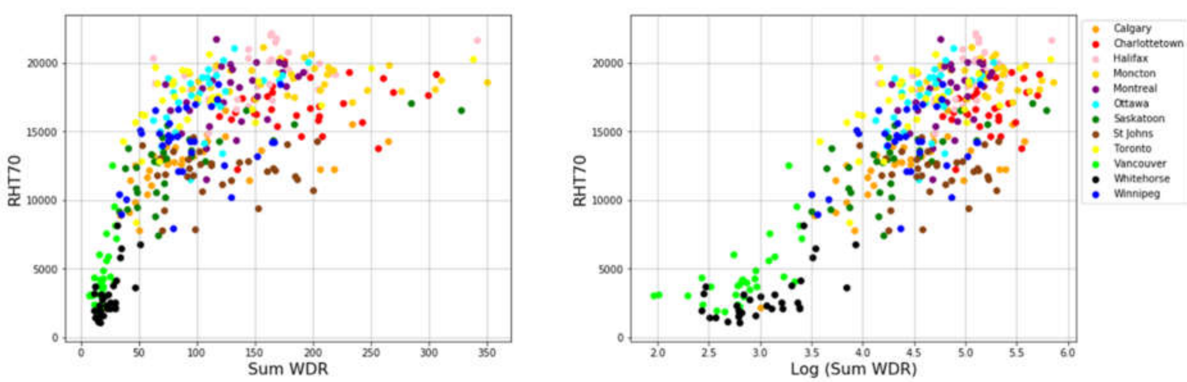


Figure 8. Relationship between the RHT70 and wind driving rain before and after logarithmic transformation.

2.6.2 Partial Least Squares Regression (PLSR)

2.6.2.1 Principle of PLSR method

Partial Least Squares Regression (PLSR) is a method used for developing predictive models when predictor variables are numerous, and some of the feature variables are collinear such that the use of multiple linear regression techniques is no longer appropriate (Sawastsky *et al*., 2015; James *et al*., 2013). It consists of identifying a linear regression model by projecting the predictor variables (X) and response variables (Y) into a new lower-dimensional space in order to control for collinearity among the variables. Fundamentally, it is a dimension reduction method that first identifies a new set of features referred to as latent variables or principal components (PCs) that are linear combinations of the original features and then fits a linear model using these new features. The new features are identified in a supervised way, i.e., it makes use of the response Y in order to identify new features that not only approximate the old features well, but also that are related to the response. The first component is a line in the X-space that well approximates the point-swarm and provides the best correlation with the y-vector. The second component is also a line in X-space and is orthogonal to the first component and finds the direction in X-space that improves the description of the X-data as much as possible, while providing a good correlation with the y-residuals remaining after the first component. This iterative procedure is repeated to identify all the PLS components (James *et al*., 2013).

The underlying model of PLSR is given in Eq. (11) (Sawatsky *et al.*, 2015):

$$X = TP^T + E$$
$$Y = UQ^T + F \tag{11}$$

Where:

X: n x m matrices of predictors where n is the number of observations and m is the number of features
Y: n x p matrice of responses where p is the number of response variables
T: projections of X, i.e., the X scores or component or factor
U: projections of Y, i.e., the Y scores
P: orthogonal loading matrix for the projected X scores
Q: orthogonal loading matrix for the projected Y scores
E: error terms for the predictor matrix
F: error terms for the response matrix

The goal of the PLSR is then to model Y scores U using X scores T. There are several algorithms that can be used to perform the analysis (Martins *et al*., 2010). In this study, the nonlinear iterative partial least squares (NIPALS) algorithm implemented in the R package *pls* (Mevik and Wehrens, 2007) was used. It starts with the singular value decomposition (SVD) of the cross-product matrix $S = X^TY$ which includes information about the variation in X and Y and the correlation between them. The matrix S can therefore be written as:

$$S = UDV^T \tag{12}$$

The column of U represents the left singular vector, w, while the column of V represents the right singular vector, q. These vectors, w and q, are used as weight vectors for X and Y, respectively, to obtain scores t and u as shown in equations (13) and **Error! Reference source not found.**:

$$t = Xw = Ew \tag{13}$$

$$u = Yq = Fq \tag{14}$$

Where E and F are initialized as X and Y, respectively. Additionally, X and Y loadings are computed by regressing against t (Eqs. (15) and (16):

$$p = E^Tt/t^Tt \tag{15}$$

$$q = F^Tt/t^Tt \tag{16}$$

Then, the information of the PLS factors (*i.e.*, the outer products, tp^t and tq^t) are subtracted from the current matrices, *E* and *F,* to obtain E_{n+1} and F_{n+1} needed to compute the next component (Eqs.(20) and (21):

$$E_{n+1} = E_n - tp^T \tag{17}$$

$$F_{n+1} = F_n - tq^T \tag{18}$$

The process is now repeated to find the next component by finding the SVD of the cross-product matrix $E_{n+1}^T F_{n+1}$. After every iteration, vectors w, t, p and q are saved as columns in matrices W, T, P and Q, respectively. Finally, the regression coefficients are calculated using the matrix formulation in Equation (22).

$$B = W(P^T W)^{-1} Q^T \tag{19}$$

2.6.2.2 Model based on PLSR

The package *pls* (Mevik and Wehrens, 2007, 2021) in R (R Core Team, 2021) was used to perform the analysis. Initially, all the original and structural variables were considered: T, T_{min}, T_{max}, RH, RH^2, SW, log(WDR), Pv, C, C^2, V, V^2, and WD. After centering (subtracting the mean) and scaling (dividing by the standard deviation) both the climate and response variables to allow for equal weight, a first model was calibrated. Then the Martens uncertainty test (Martens and Martens, 2000) implemented in the *pls* package was used to select the most important variables. This test is based on a combination of cross-validation, jack-knifing and significance testing. First, several sub models are created based on the samples not kept in the cross-validation segment. Then, for each segment M, the B coefficients are calculated. In addition, a model based on all the samples is created. For each regression coefficient obtained, the variance can be calculated by jack-knifing using Eq. (20) (Martens and Martens, 2000):

$$s^2 b = \left(\sum_{m=1}^{M} (b - b_m)^2\right)\left(\frac{(N-1)}{N}\right) \tag{20}$$

Where:

- N: number of samples
- M: number of cross-validation segments
- b: regression coefficient using all the N samples
- b_m: regression coefficient using samples not kept out in the cross-validation segment m.
- s^2b: uncertainty variance of the individual regression coefficient

Based on these uncertainty estimates of the model's parameters, a t-test is performed to check whether the regression coefficients were significantly different from zero. The *p*-value is calculated based on t-test [b, sqrt(s^2b), df = N]. If the *p*-value for a variable is below the threshold of 0.05, it is considered an important variable. Furthermore, the resulting regression coefficients can be presented with *a 95%* confidence interval that corresponds to 2 standard deviations. Formally the confidence interval is calculated as $b \pm sqrt(s^2 b) * t_{\frac{0.05}{2}, df=N}$.

Variables with confidence limits that do not cross the zero line are considered the most important variables. Figure 9 shows the regression coefficient with a 95% confidence interval corresponding to the 4^{th} principal component. The shaded bar represents significant variables with a p-value < 0.05 and a 95% confidence interval that never crosses 0. The variables cloudiness and daily maximum temperature were non-significant, so they were dropped from the analysis. Therefore, the yearly climate variables sum of SW, log(WDR), C^2, RH, RH^2, T, T_{min}, WD, V, V^2 and P_v were selected and included in the final model.

For the final model, the optimal number of principal components (PCs) was chosen based on the percentage of variance explained in the response variable, RHT70. As shown in Figure 10, the first 4 principal components explain most of the variance in the response. The residuals of the model were diagnosed and as shown in Figure 11, the plot of residuals versus predicted values for the training set (Figure 11a) shows nothing unusual. As well, most of the observations are close to the line of normality, as shown in the normal quantile plot (Figure 11b).

Therefore, the corresponding model with 4 PCs, expressed as a function of the original variables (Eq. (21)), was used to make predictions on the test set.

$$RHT70 = -41268.1758 + 0.031SSW + 3409.10\,log(SWDR) - 84.64C^2 + 16950.35RH + 10087.20RH^2 + 169.40T - 60.75T_{min} + 43.08WD - 525.59V - 105.14V^2 + 4.51P_v \quad (21)$$

Where:

SSW: annual sum of short wave radiation (W/m^2)
SWDR: annual sum of wind driving rain (L/m^2)

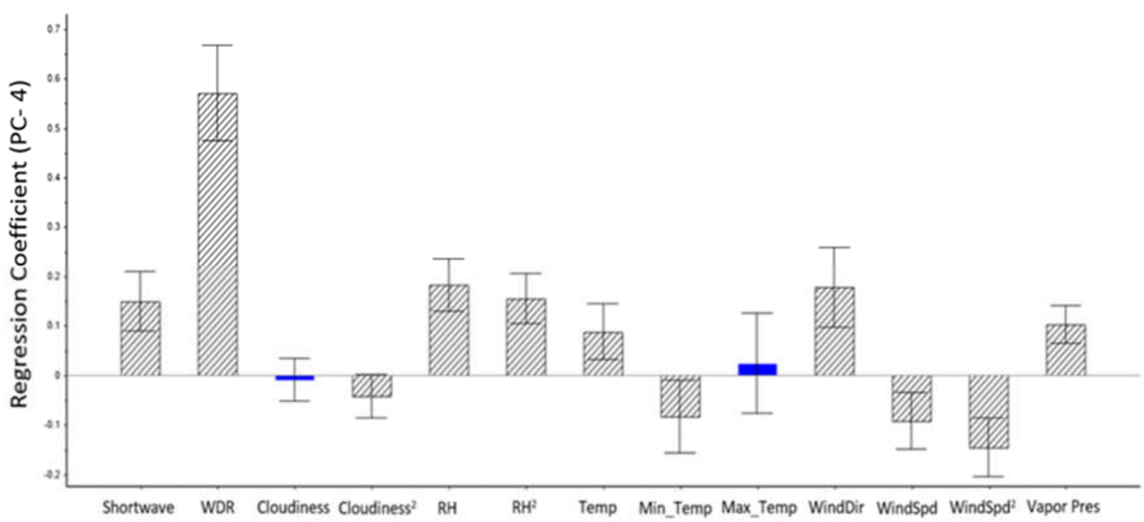


Figure 9. Regression coefficient b with uncertainty limit for RHT70 ($\pm\, 2\; std.dev.$).

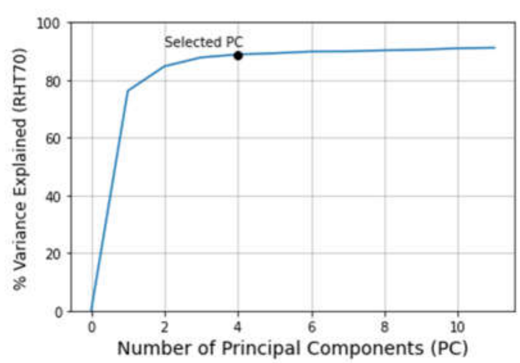
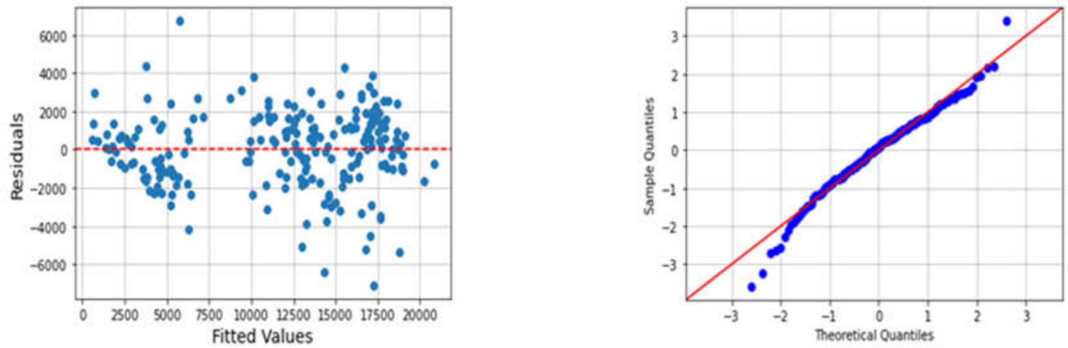


Figure 10. Percentage Variance in RHT70 explained by each additional principal component.



a) b)

Figure 11. PLS model diagnostic plots: a) Residuals vs. predicted values of RHT70 and b) Normal quantile plot of residuals for RHT70.

2.6.3 Least Absolute Shrinkage and Selection Operator (LASSO) regression

2.6.3.1 Theory

Least Absolute Shrinkage and Selection Operator (LASSO) belongs to a class of shrinkage methods that fit a model containing all the predictors but constraints or regularizes the coefficient estimates. In the case of LASSO, it is the L1 regularization that shrinks some of the coefficient estimates towards or exactly to zero, thereby performing feature selection, reducing complexity of the model and collinearity effects, and preventing over-fitting, which may result from simple linear regression. To understand the complete procedure, the reader can refer to Tibshirani (1996, 2011).

The cost function to minimize for ordinary linear least squares regression is given by Eq. (22).

$$\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{M}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j \times x_{ij}\right)^2 \tag{22}$$

Where:

M: number of observations
p: number of features
β_j: intercept
β_j: estimated coefficient for feature j

For LASSO regression, the cost function to minimize is given by Eq. (23):

$$\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{M}\left(y_i - \beta_0 \sum_{j=1}^{p}\beta_j \times x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j| \tag{23}$$

Where λ is the tuning parameter. The second term on the right hand side in Equation **Error! Reference source not found.** is the Lasso L1 penalty, which has the effect of shrinking some of the coefficients towards 0. Therefore, with the L1 regularisation, some features are completely discarded from the model and, as such, permits feature selection (James *et al*., 2013).

2.6.3.2 Model based on LASSO

LASSO was performed using the package *glmnet* (Friedman *et al*., 2010) in R (R Core Team, 2021) to select a subset of the most contributing climate variables among the 13 climate variables considered. The procedure used to select the most regularized model is the one described in Hastie *et al*. (2021). Leave-one-out cross-validation was performed over a grid of possible values to choose the regularization level, lambda (λ). Figure 12 shows the cross-validation curve (red dotted line) along with the upper and lower standard error curves (error bars) along the natural log of the λ sequence. Across the top is the number of variables included at each λ. Two values of lambda are indicated by the vertical dotted lines. The first line is the once closest to the y-axis. It is the one that indicates λ-value that results in the lowest MSE. MSE at this point was 3,483,411 and the standard error was 429,632. The second line is the λ corresponding to MSE above the minimum MSE + one error deviation (3,910,153). The λ at this second line (24.1) is the optimal λ according to Hastie *et al.* (2021) as it leads to a more parsimonious model with fewer included variables and a minimal loss in MSE. The 9 climate variables with nonzero coefficients corresponding to the chosen lambda were thus selected: log(sumWDR), C^2, RH, T, T_{min}, T_{ma}x, WD, V^2 and P_v. Other variables, *i.e*., the sum of SW, C, V, and RH^2, were thus discarded. The resulting model is given by Eq. (24), with the residuals shown in Figure 13.

$$\begin{aligned} RHT70 = -41047.33 \ + 3521.67\ Log(SWDR) - 148.77C^2 + \ 31626.30RH + \ 228.14T \\ - 44.60T_{min} \ - 114.58T_{max} + \ 84.41WD - 222.02V^2 + 7.04P_v \end{aligned} \tag{24}$$

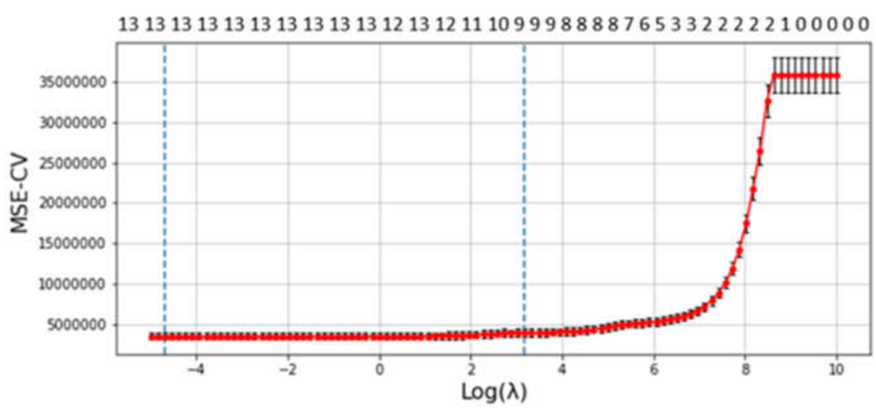
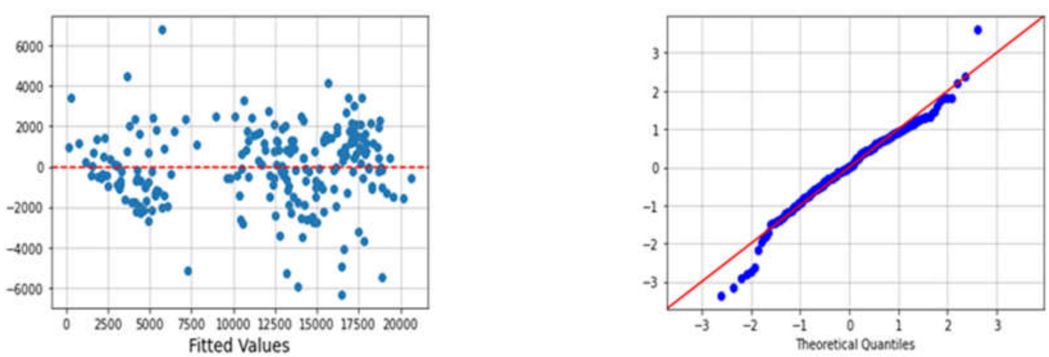


Figure 12. Mean square error of cross-validation (MSE-CV) as function of the regularization parameter lambda (λ).



a) b)

Figure 13. LASSO model diagnostic plots: a) Residuals vs. predicted values of RHT70 and b) Normal quantile plot of residuals for RHT70.

2.6.4 Support Vector Regression (SVR)

2.6.4.1 Principle of SVR

Support Vector Machines (SVM) is a supervised machine learning modeling technique introduced by Cortes & Vapnik (1995). In this section, an overview of the SVM that deal with modeling continuous response variable called Epsilon-based Support Vector Regressions (ε-SVRs) is given.

Support Vector Regression aims to find a function *f*(*x*) that can fit as many instances of the training data as possible with at most ε deviation from the response while also being as flat as possible. Assuming one is modeling a single output y as a function of n input variables x and is given a training dataset of length N, *i.e.*, $\{(x_1, y_1), (x_2, y_2) \dots \dots (x_N, y_N)\}$, where $x_i \in \Re^n$ and $y_i \in \Re$ for $i = 1, 2 \dots \mathrm{N}$. Then, SVR estimates the relationship between the explanatory variable and response using Eq. (25).

$$f(x) = \langle w, \phi(x) \rangle + b \tag{25}$$

where:

- $\langle \cdot, \cdot \rangle$: dot product;
- w : weight vector;
- *b*: bias term;
- $\phi(x)$: transformation from input space into feature space.

The objective of trying to find a flat function implies that the slope of the function *f*(*x*) is minimized, which yields a convex optimization problem (Eq. (26)).

$$\begin{gathered} Minimize \ \frac{||w||^2}{2} \\ \text{subject to } \begin{cases} y_i - < w, \phi(x_i) > -b \leq \varepsilon \\ < w, \phi(x_i) > +b - y_i \leq \varepsilon \end{cases} \end{gathered} \tag{26}$$

Where ε defines the margin of tolerance and $||w||$ represents the Euclidean norm of the weight vector. The above optimization problem is only feasible where *f*(*x*) exists and approximates all the training data with ε precision. Therefore, some deviations larger than ε are allowed by introducing the slack variables ξ_i and ξ_i^*. These slack variables measure how much the i-th training instance can violate the margin. This is called the ε-insensitive loss function, which is described by Eq. (27).

$$|\xi|_\varepsilon = \begin{cases} 0 & if \ |\xi| < \varepsilon \\ |\xi| - \varepsilon & otherwise \end{cases} \tag{27}$$

The ε-insensitive loss function implies that only those training instances that are greater than ε in magnitude are used to support or determine the function *f*(*x*). Adding more training instances within the allowed deviation ε does not have any effect on the predictions. Figure 14 depicts the ε-insensitive loss function graphically.

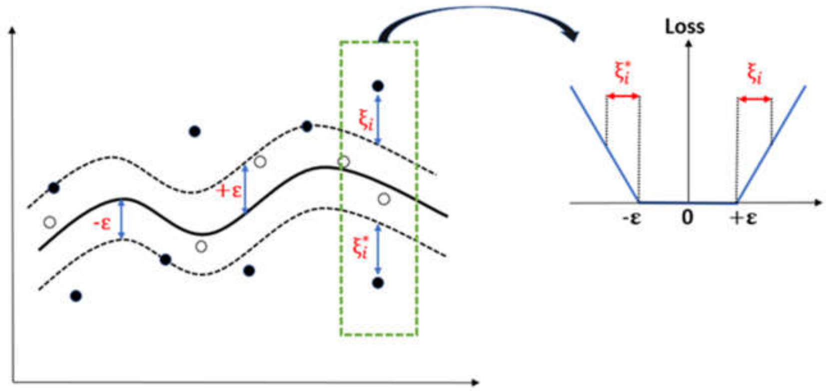


Figure 14. Loss setting for nonlinear SVR. The right side of the image shows soft ε-insensitive setting as described in Schölkopf *et a*l. (2002) for SVR. Any values between -ε and +ε are assigned a loss 0 and values outside the range is assigned a loss of $|\xi| - \varepsilon$.

Adding the objective of minimizing these deviations with Eq. (26) results in the following formulation (Eq. (28)):

$$\text{Minimize} \quad \frac{||w||^2}{2} + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$$

$$\text{subject to} \begin{cases} y_i - < w, \phi(x_i) > -b \leq \ \varepsilon \ + \ \xi_i \\ < w, \phi(x_i) > +b - y_i \ \leq \ \varepsilon \ + \ \xi_i^* \\ \xi_i, \xi_i^* \qquad \geq 0 \end{cases} \tag{28}$$

The value of the parameter ε controls the band's width and can affect the number of support vectors used to construct the regression function. The parameter C is a constant that determines the trade-off between the two conflicting objectives of trying to make the slack variables as small as possible to reduce margin violation and the flatness of function *f*(*x*). Both ε and C are hyperparameters that must be tuned according to the dataset used to train the SVR model to achieve maximum efficiency on the test dataset. A common approach to select the optimal values for these hyperparameters is to use a grid search where both ε and C are systematically varied, and the cross-validation error monitored. Further explanation of the optimization problem and the detailed mathematical solution can be found in Smola & Schölkopf (2004).

The Kernel function used in this paper is the Radial Basis Function (RBF) kernel. The hyperparameter γ in the function acts as a regularization that controls the spread of the function and must also be tuned during the training process. Given two instances of the input variables x_i, x_j, the RBF kernel evaluates nonlinearity between them, as described by Equation (29):

$$K(x_i, x_j) \ = \ e^{-\gamma||x_i - x_j||^2} \tag{29}$$

2.6.4.2 Model based on combined LASSO and Support Vector Regression

Only, the most important climate variables selected using LASSO, i.e., log(SWDR), C^2, RH, T, T_{min}, T_{max}, WD, V^2 and P_v, were used. It is thereafter simply called the SVR model. The training set was standardized to have zero mean and unit variance. The standardization ensures that all the variables are on the same scale, allowing equal weight in the model.

To optimize the hyper-parameters ε, C, and γ, a grid search was performed. The values of the parameters were systematically varied, and the cross-validation error was evaluated for all the possible combinations of hyperparameters values. Parameters combination, which yielded the lowest Mean Squared Error were then used to make predictions on the test set. A summary of the selected parameter combination for the best SVR model is shown in Table 6. For this case, the lowest Mean Squared Error of cross-validation (MSECV) was achieved using the RBF kernel with a gamma of 0.1, a cost of 1, and an epsilon of 0.1. Hence this parameter combination was used on the test set. The final SVR model used to make predictions on the test set is described in Appendix A.1. Figure 15 shows the diagnostic plots of the SVR model. At both ends of the quantile plot, there are a few points that depart from normality, which can be a sign that there is still room for improvement in the model.

Table 6: Summary of results from SVR parameter estimation with RBF kernel.

Gamma	Cost	Epsilon	MSECV
0.001	1	0.1	9678219
0.001	10	0.1	3964515
0.001	100	0.1	3052156
0.1	**1**	**0.1**	**2709400**
0.1	10	0.1	3117818
0.1	100	0.1	4257715
0.001	1	0.5	11774527
0.001	10	0.5	5382187
0.001	100	0.5	4087839
0.1	1	0.5	4683088
0.1	10	0.5	4844730
0.1	100	0.5	4880994

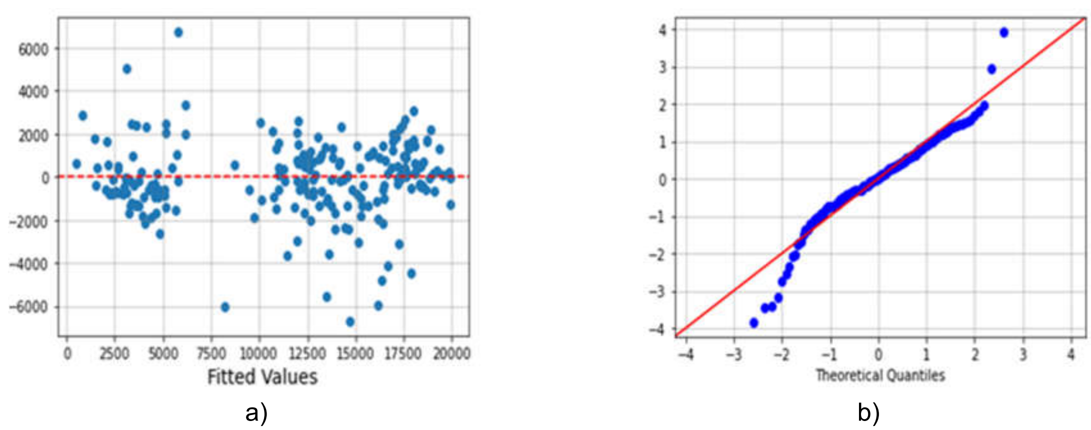


Figure 15. SVR model diagnostic plots: a) Residuals vs. predicted values of RHT70 and b) Normal quantile plot of residuals for RHT70.

2.7 Comparing the Isev model with new models

The five models, *i.e.*, original Isev regression equation, recalibrated Isev regression equation, and models based on PLSR, LASSO and SVR, were used to predict and rank the 31-year series of climate data in the test set in terms of RHT70. Three statistics were used to evaluate their performance: (i) the root mean square error of prediction (RMSEP), for their ability to predict the moisture severity (RHT70); (ii) the spearman rank correlation for their ability to rank all the 31 years (Eq. (30)) of a series, and (iii) the number of years correctly identified from the top 5 worst years identified from hygrothermal simulations, for their ability to rank the top five years and permit selection of the design weather year.

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \quad (30)$$

Where:

N: number of samples

D: difference between ranking pairs

3 Results and discussion

3.1 RHT70 from hygrothermal simulations

The boxplots in Figure 16 show the distribution of the 31 cumulative sums of RHT70 calculated using T and RH obtained from hygrothermal simulations on the outer side of OSB. These data are summarized in Table 7. There is a considerable variation of RHT70 values among the years in the same cities. The median value of RHT70 varies from less than 5000 in the dry city of Whitehorse (2338) and in the wet city of Vancouver (4010) to about 20 000 in the wet cities of Moncton (18745) and Halifax (19252). With the exception of Vancouver (MI = 1.4) and St. John's (MI = 1.4), RHT70 is relatively higher in cities with moderate to high MI (Toronto, Ottawa, Montreal, Moncton, Charlottetown and Halifax, with MI > 0.8) than in cities having lower MI (Whitehorse, Calgary, Saskatoon and Winnipeg, with MI < 0.7). As shown in Figure A1 in Appendix for the cities of Vancouver and St. John's, there is little WDR on the north face of the wall, which perhaps provides an explanation for the relatively lower values of RHT70 obtained in these wet cities.

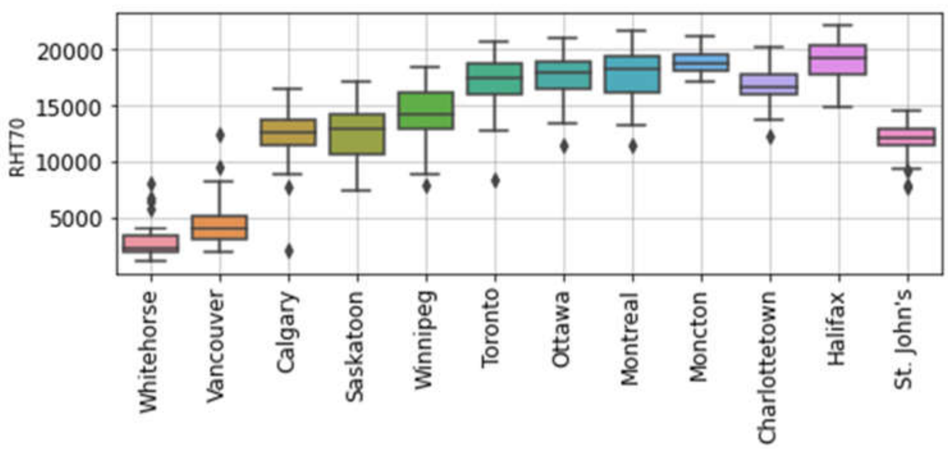


Figure 16. Boxplots of the 31 cumulative sums of RHT70 obtained from hygrothermal simulations in the 12 cities studied. The boxplot shows the: minimum (Q1 – 1.5*IQR), 25th percentile (Q1), median, 75th percentile (Q3), maximum (Q3 + 1.5*IQR) and the outliers (diamond). IQR is the interquartile range.

3.2 Predictive ability of Isev and new models

3.2.1 Original Isev model

The bar plots in Figure A 2 (Appendix A.3) show the yearly values of RHT70 obtained from hygrothermal simulations at the outer layer of OSB and predicted using the original Isev, for the 12 cities considered in this study. They are summarized in Figure 16 in the form of boxplots that show the distribution of the 31 cumulative sums of RHT70 values in both cases in each city. The difference between the minimum, median and maximum values of RHT70 obtained from hygrothermal simulations and that from calculation using original Isev equation (Eq. (2)) are also shown in Table 7.

In all the cities but St. John's and Whitehorse, RHT70 at the outer layer of OSB is generally underestimated. In Charlottetown, Vancouver and Whitehorse, the absolute difference between simulated and predicted RTH70 is less than 3100. In Montreal and Ottawa, this difference is more than 11000, while in the remaining cities, it lies between 5300 and 8000. There are also differences between the simulated and predicted minimum and maximum values of RHT70; minimum values are largely underestimated in Moncton, Montreal, and Ottawa and overestimated in St. John's with an absolute difference greater than 7000 while maximum values are largely underestimated in Calgary, Montreal, Ottawa, Toronto, and Vancouver and overestimated in St. John's, with an absolute difference greater than 7000, while maximum values are largely underestimated in Calgary, Montreal, Ottawa, Toronto, Vancouver and Winnipeg and overestimated in St. John's with an absolute difference greater than 6000.

Table 7 shows the mean bias and root mean square error of prediction (RMSEP) of RHT70 using the original Isev equation. The highest biases are found in Montreal, Ottawa and Toronto. RHT70 values are also poorly predicted in these cities with a RMSEP greater than 10000. In general, there is systematic bias when predicting RHT70 using the original Isev equation. This may be explained by the fact that the model was trained using only eight US cities where the RHT70 index varied from 1000 to 16000 compared to the Canadian cities where RHT70 ranges from 1000 to 21000. Another reason maybe the difference in the climate data set used. In fact, the climate data used to develop the original model ranged from 1961 to 1990, whereas those used in this study ranged from 1986 to 2016.

Table 7. Statistics of RHT70 obtained from hygrothermal simulations and estimated using the original Isev equation.

	Minimum			Median			Maximum				
City	**Sim[1]**	**Isev[2]**	**Diff[3]**	**Sim**	**Isev**	**Diff**	**Sim**	**Isev**	**Diff**	**Mean bias**	**RMSEP**
Calgary	2155	3041	-885	12600	4806	7794	16531	8810	7721	-6924	7281
Charlottetown	12223	11959	263	16664	14255	2409	20132	17224	2908	-2644	3328
Halifax	14838	11328	3510	19252	13823	5429	22140	17323	4817	-5287	5548
Moncton	17152	9518	7634	18745	13417	5328	21109	17639	3470	-5359	5778
Montreal	11507	3588	7919	18232	6284	11948	21708	8509	13198	-11526	11706
Ottawa	11534	3587	7947	17881	6740	11141	21059	8578	12481	-10976	11118
Saskatoon	7425	5809	1616	12988	7643	5346	17083	14113	2970	-4390	4734
St. John's	7798	15437	-7640	12133	18337	-6204	14561	21080	-6519	6379	6674
Toronto	8360	2713	5647	17443	6249	11194	20641	12525	8116	-10422	10669
Vancouver	1890	1892	-2	4010	2803	1207	12502	3672	8830	-1764	2818
Whitehorse	1099	4811	-3711	2338	5357	-3019	8133	5972	2161	2438	2871
Winnipeg	7967	4649	3318	14281	7647	6634	18471	11501	6970	-6591	7001

[1]Results from hygrothermal simulations; [2]Prediction using original Isev equation; [3]Difference bewteen Isev and simulations

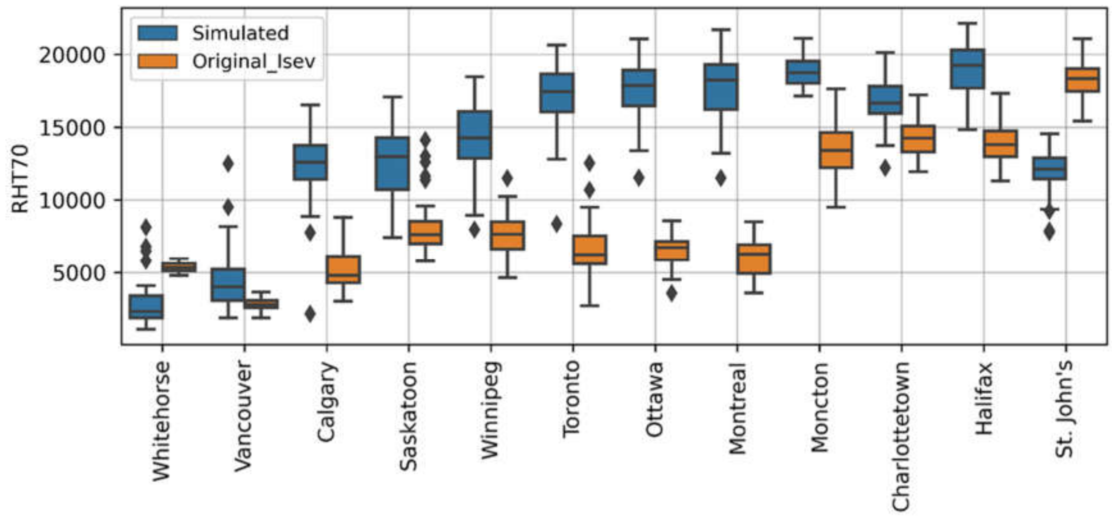


Figure 17. Boxplots of simulated (from hygrothermal simulations) and predicted RHT70 using the original Isev model in the 12 cities studied

3.2.2 Recalibrated Isev model

Given the poor prediction of RHT70 for Canadian locations, it was decided to recalibrate the model in order to verify if its predictive capacity could be improved (Eq. (10)). As explained in Section 2.3.2, the model was recalibrated using climate data of seven cities and then tested on five cities.

The bar plots in Figure A 3 (Appendix A.4) show the yearly values of RHT70 obtained from hygrothermal simulations at the outer layer of OSB and predicted using the recalibrated Isev equation for the five test cities considered. These results are summarized in Figure 18. Compared to the original Isev model results shown in Figure 18 for the same cities, there is an improvement in predicting the actual magnitude of RHT70 when the equation is recalibrated. The number of samples in the test set was limited to 5 and the distribution of RMSEP values was asymmetric. As such, it was not possible to test statistically the significance of the difference between the original and recalibrated Isev predictions by using, for example, the Wilcoxon signed-rank test. However, the boxplots in Figure 19 and quartile values reported in Table 8 show that there is a clear improvement of the predictions using the recalibrated Isev model. As a result, cities such as Montreal and Toronto are now well predicted.

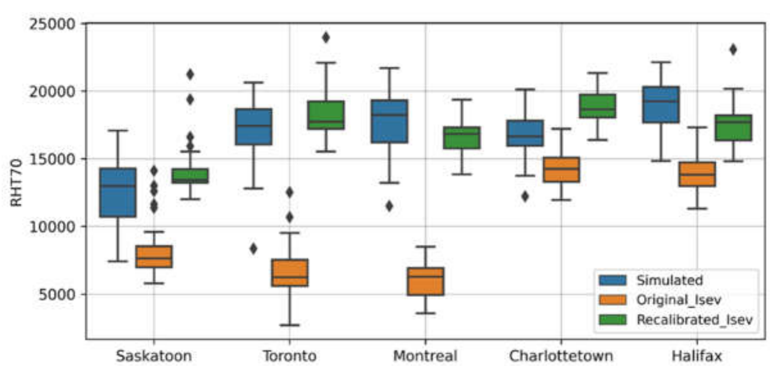


Figure 18. Boxplots of simulated (from hygrothermal simulations) and predicted RHT70 using the original and recalibrated Isev models in the five test cities.

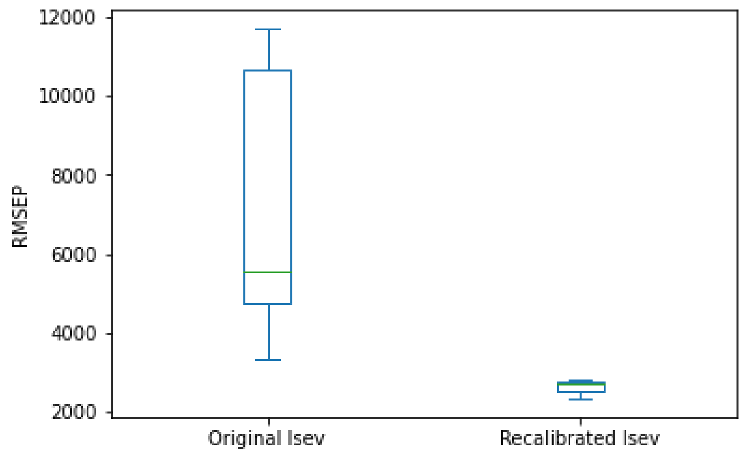


Figure 19. Boxplots of root mean square errors of prediction for the original and recalibrated Isev models in the five test cities.

Table 8. Comparison of RMSEP between the original and recalibrated Isev models in the five test cities.

City	Original Isev equation	Recalibrated Isev equation
Saskatoon	4734	2331
Toronto	10668	2772
Montreal	11706	2497
Charlottetown	3328	2816
Halifax	5548	2686
Mean	*7197*	*2620*
Minimum	*3328*	*2331*
25%	*4734*	*2497*
Median	*5548*	*2686*
75%	*10668*	*2772*
Maximum	*11706*	*2816*

3.2.3 Models based on PLS, LASSO and SVR

The original Isev equation was calibrated using stepwise multiple regression analysis, but the final model included several climatic variables that are correlated to some extent. The collinearity among some model variables can increase the variance of the regression coefficient estimates and may make the model unstable, *i.e*. sensitive to minor changes in the features. Owing to these limitations, three modelling strategies capable of handling collinearity at diverse degrees in the climate variables were evaluated: Partial Least Squares Regression (PLSR) and Least Absolute Shrinkage and Selection Operator (LASSO) and a combination of LASSO and Support Vector Regression (SVR). The PLSR model is described in Eq. (21), the LASSO model in Eq. (24) and the SVR model is described in Appendix A.1. These models were trained using data from 7 cities and were evaluated in 5 cities.

The yearly values of RHT70 obtained from hygrothermal simulations at the outer layer of OSB and predicted using the PLSR, LASSO and SVR for the five test cities considered are shown in Figure A 4, Figure A 5, and Figure A 6, respectively, in the Appendix. All these results are summarized in Figure 20 and Table 9. As with the recalibrated Isev model, the PLSR, LASSO and SVR models capture the year-to-year variation of RHT70, with no excessive and systematic biases as observed with the original Isev equation. However, it can be noticed that:

- The minimum values of RHT70 are overestimated in all the five test cities for all three models.
- In Saskatoon, Montreal and Halifax, the median values of RHT70 obtained with the three models are close to those from hygrothermal simulations. In Toronto, median values obtained with PLSR,

and LASSO are lower than those obtained from hygrothermal simulations. In Charlottetown, the median values of RHT70 obtained with the three models are higher than those obtained from hygrothermal simulations.

- Compared to RHT70 from hygrothermal simulations, the maximum values obtained with PLSR and LASSO are higher in Saskatoon and Toronto, lower in Montreal and almost similar in Halifax whereas in Charlottetown, PLSR and LASSO give similar and higher maximum value, respectively. On the other side, SVR gives maximum RHT70 values similar to those from hygrothermal simulations in Saskatoon and Toronto but gives lower values in Montreal, Charlottetown and Halifax.

The RMSEP values and their quartile values for the three new models and those for the original and recalibrated Isev models are shown in Table 10. They are illustrated in Figure 21. While no statistical test was possible, it is clear that the performance of the new models are relatively good in all the five test cities. Although the SVR model seems to have the best predictive ability with an improvement in all the five test cities, there is no clear evidence, based on statistics in Table 10 and boxplots in Figure 21, to conclude that its predictions are significantly better than those for the PLSR and LASSO models.

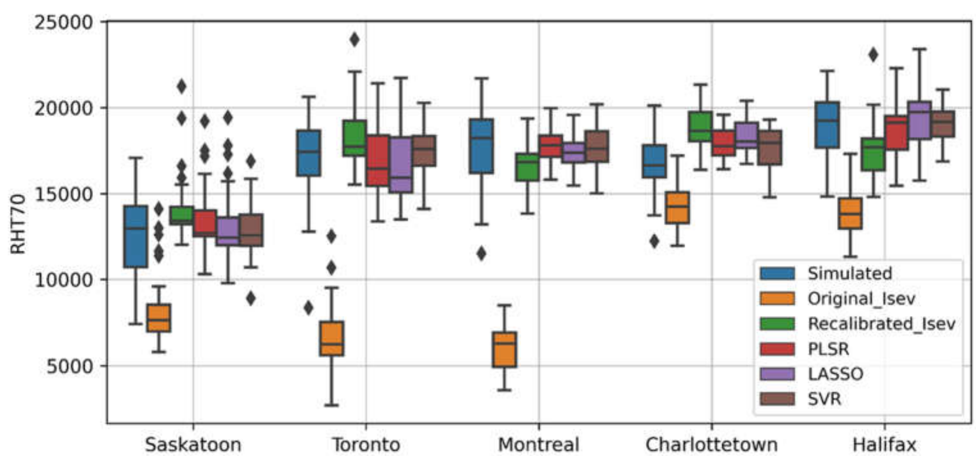


Figure 20. Boxplots of simulated (from hygrothermal simulations) and predicted RHT70 using the original and recalibrated Isev, PLSR, LASSO and SVR models in the five test cities.

Table 9. Statistics of RHT70 obtained from hygrothermal simulations and estimated using PLSR, LASSO and SVR models.

	Minimum				Median				Maximum			
City	Sim[1]	PLSR	LASSO	SVR	Sim	PLSR	LASSO	SVR	Sim	PLSR	LASSO	SVR
Saskatoon	7425	10318	9801	8909	12988	12715	12430	12581	17083	19220	19430	16898
Toronto	8360	13396	13520	14129	17443	16461	15946	17597	20641	21422	21734	20283
Montreal	11507	15823	15481	15029	18232	17817	17388	17628	21708	19969	19575	20195
Charlottetown	12223	16431	16730	14806	16664	17766	18044	17968	20132	19595	20396	19312
Halifax	14838	15477	15770	16885	19252	19136	19745	19181	22140	22289	23398	21049

[1]Results obtained from hygrothermal simulations

Table 10. Comparison of the RMSEP between the original Isev, recalibrated Isev, PLSR, LASSO and SVR models in the five cities of the test set.

City	Original Isev	Recalibrated Isev	PLSR	LASSO	SVR
Charlottetown	3328	2816	1963	2233	1496
Halifax	5548	2686	1790	1777	1494
Montreal	11706	2497	2131	2099	1926
Saskatoon	4734	2331	1800	1714	1591
Toronto	10668	2772	2170	2221	2035
Mean	*7197*	*2620*	*1971*	*2009*	*1708*
Minimum	*3328*	*2331*	*1790*	*1714*	*1494*
25%	*4734*	*2497*	*1800*	*1777*	*1496*
Median	*5548*	*2686*	*1963*	*2099*	*1591*
75%	*10668*	*2772*	*2131*	*2221*	*1926*
Maximum	*11706*	*2816*	*2170*	*2233*	*2035*

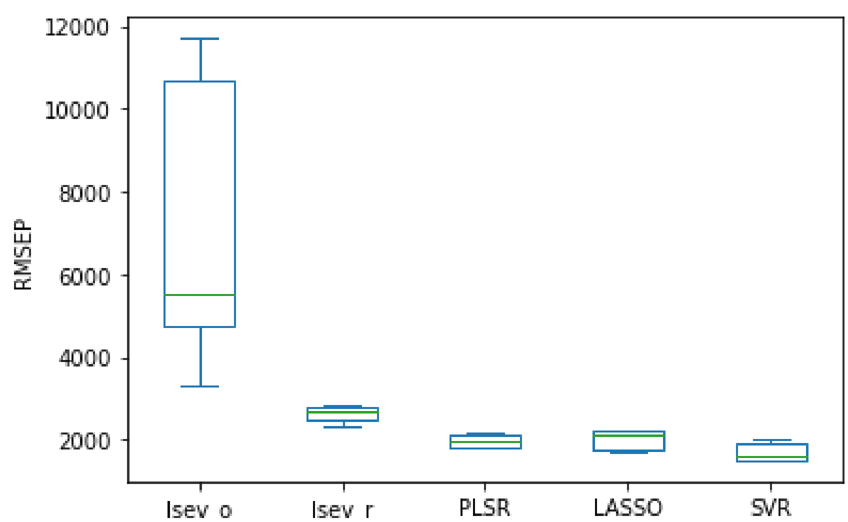


Figure 21. Boxplots of root mean square errors of prediction for the original Isev (Isev_o), recalibrated Isev (Isev_r), PLSR, LASSO and SVR models in the five test cities.

3.3 Comparison of the ranking performance of the models

One objective of the ASHRAE's Research Project RP-1325 was to develop a model capable of ranking the climate years in terms of their moisture severity for the purpose of selecting the design weather year. In this study, the rankings of the climate years obtained using the original and recalibrated Isev, PLSR, LASSO and SVR models were compared to that obtained using hygrothermal simulations, using the Spearman's rank correlation (ρ) obtained for all the 31 years of the series (Table 11 and Figure 22), and the number of similar years classified in the top five severe years (Table 12 and Table 13, Figure 23). It should be noted that the last ranking performance metric does not take into account the position of the years amongst the top five.

The Spearman's rank correlation for the original Isev model ranges from 0.10 (very poor, in Charlottetown) to 0.71 (strong, in Saskatoon), with a median value of 0.51 for the five test cities. For the recalibrated Isev model, there is a decrease in the median value of ρ (0.32) and in all the cities except for Charlottetown, despite the improvement in the prediction of RHT70. The PLSR and LASSO models, with a median values of 0.50 and 0.53, slightly improve ρ in all the cities except for Montreal, in comparison with the recalibrated Isev. The SVR model shows the best-ranking performance with improvement in all cities and a median value of 0.66, almost twice the value obtained with recalibrated Isev. It is however difficult to conclude that the new models improved significantly the ranking of the years, compared to the original Isev model but there might be a significant difference between the recalibrated Isev and the new models, given the spread of the data.

When focusing on the selection of the top five more severe years (Table 12 and Table 13, Figure 23), the original Isev model detected, on average, 2.4 years amongst the ones depicted from hygrothermal simulations results whereas the recalibrated Isev model detected 1.8 years, less than the mean of 2.6 years detected by the PLSR, LASSO and SVR models. But here too, it is difficult to conclude that there is any significant difference among the five models.

The improvement of the predictive capacity of the recalibrated Isev model did not translate into a good ranking of the years. In all the five test cities but Charlottetown, the spearman correlation decreases. Regarding the detection of the top five severe years, in cities like Charlottetown and Montreal, there is a decrease in the number of years correctly ranked, whereas, in Halifax, Saskatoon and Toronto, the number of years correctly ranked in the top five remains the same for both original and recalibrated Isev models. This lack of improvement in the ranking using the recalibrated model may be due to the collinearity among some variables of the models, which makes it unstable given the uncertainty in the coefficient estimates (James et *al.*, 2013).

Table 11. Comparison of the Spearman's Rank Correlation between the recalibrated Isev, PLSR and SVR model in the five test cities.

City	Original Isev	Recalibrated Isev	PLSR	LASSO	SVR
Charlottetown	0.10	0.22	0.41	0.38	0.72
Halifax	0.51	0.15	0.60	0.53	0.61
Montreal	0.62	0.38	0.37	0.41	0.66
Saskatoon	0.71	0.65	0.72	0.76	0.72
Toronto	0.39	0.32	0.50	0.53	0.47
Mean	*0.47*	*0.34*	*0.52*	*0.52*	*0.64*
Minimum	*0.10*	*0.15*	*0.37*	*0.52*	*0.64*
25%	*0.39*	*0.22*	*0.41*	*0.41*	*0.61*
Median	*0.51*	*0.32*	*0.50*	*0.53*	*0.66*
75%	*0.62*	*0.38*	*0.60*	*0.53*	*0.72*
Maximum	*0.71*	*0.65*	*0.72*	*0.76*	*0.72*

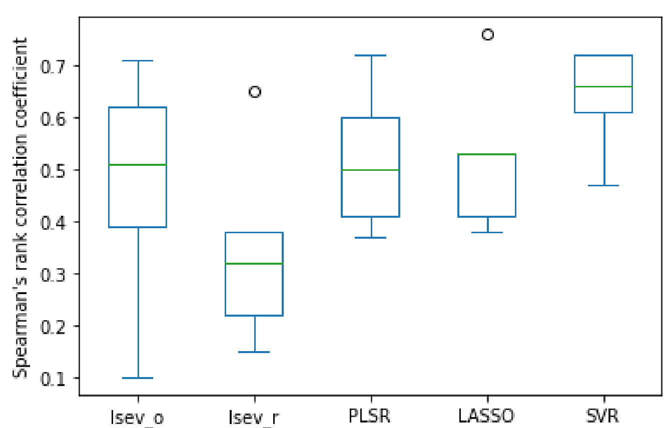


Figure 22. Boxplots of Spearman's rank correlation coefficients for the original Isev (Isev_o), recalibrated Isev (Isev_r), PLSR, LASSO and SVR models for all the 31 years in the five test cities.

Table 12. Top five severe years depicted by simulations, original and recalibrated Isev, PLSR, LASSO and SVR methods. The years highlighted are those that are determined from both simulations and the given method.

Method	Charlottetown	Halifax	Montreal	Saskatoon	Toronto
Simulations	2007	2014	2012	2014	2016
	2008	2013	2016	2006	2005
	2009	2011	2003	2010	2015
	2010	2005	2004	2002	1988
	2006	2009	2008	2001	2006
Original Isev	**2010**	**2005**	1987	**2010**	**2005**
	2000	1993	**2004**	**2014**	**1988**
	1999	2004	2002	1988	2010
	2006	1986	**2003**	**2002**	1992
	2016	2007	**2016**	**2006**	2007
Recalibrated Isev	**2010**	**2005**	1987	**2010**	**2005**
	2016	2007	**2004**	**2014**	**1988**
	1990	1986	1998	**2002**	2010
	2000	1998	2002	**2006**	1992
	1999	1994	1991	1988	2012
PLSR	**2009**	**2005**	2010	**2010**	**2005**
	2016	2007	**2004**	**2014**	2010
	2014	**2011**	**2003**	**2002**	**1988**
	2010	2010	2002	**2006**	2012
	2000	**2013**	2014	1988	2004
LASSO	**2009**	**2005**	2010	**2010**	**2005**
	2014	2007	2002	**2014**	2010
	2000	**2011**	**2003**	**2002**	**1988**
	2016	1994	**2004**	**2006**	**2016**
	2010	2006	2014	1988	2012
SVR	2013	2010	2002	**2014**	**2005**
	2016	**2014**	**2003**	**2010**	2010
	2008	**2011**	1991	2009	**1988**
	2011	2007	**2016**	2015	2004
	2007	**2009**	**2012**	**2002**	1992

Table 13. Comparison of the ability to select the top 5 worst years between recalibrated Isev, PLSR and SVR models in the five test cities.

City	Original Isev	Recalibrated Isev	PLSR	LASSO	SVR
Charlottetown	2	1	2	2	2
Halifax	1	1	3	2	3
Montreal	3	1	2	2	3
Saskatoon	4	4	4	4	3
Toronto	2	2	2	3	2
Mean	*2.4*	*1.8*	*2.6*	*2.6*	*2.6*
Minimum	*1.0*	*1.0*	*2.0*	*2.0*	*2.0*
25%	*2.0*	*1.0*	*2.0*	*2.0*	*2.0*
Median	*2.0*	*1.0*	*2.0*	*2.0*	*3.0*
75%	*3.0*	*2.0*	*3.0*	*3.0*	*3.0*
Maximum	*4.0*	*4.0*	*4.0*	*4.0*	*3.0*

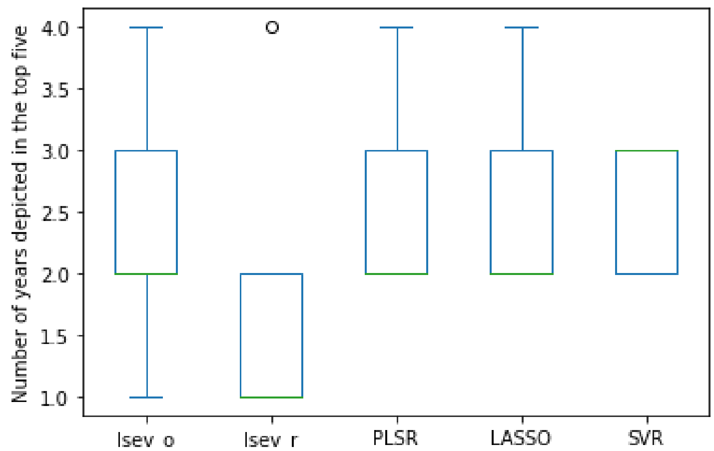


Figure 23. Boxplots of Spearman's rank correlation coefficients for the original Isev (Isev_o), recalibrated Isev (Isev_r), PLSR, LASSO and SVR models for all the 31 years in the five test cities.

The above results suggest that if the only objective is the relative ranking of all the years in a series in terms of their moisture severity, either PLSR, LASSO or SVR could be a good choice as they have highest Spearman correlation, in comparison with the recalibrated Isev model. However, this option is valid only when there is a need to construct moisture reference years that comprise a combination of severe and least severe years. For design purpose, only one year among the most severe years is generally considered (ASHRAE, 2016). In this case, given that the PLSR, LASSO and SVR models depict the same number of years in the top five, any of them could be used. One could also suggest the use of the original Isev equation for depicting the top five severe years given there is no clear difference with the new models, but the results would be unpredictable, due to the instability of the model coefficients.

Overall, it is clear, based on the results of this study, that the performance of the Isev model can be improved. PLSR, LASSO and SVR algorithms were evaluated, but there are still many other algorithms developed especially to tackle this type of problem that could be evaluated: Elastic Net (Jas *et al*., 2020), Generalized Additive Model (Wood and Wood, 2015), and Machine learning algorithms such as Random Forest (Breiman *et al*., 2001), Generalized Boosted Machine (Ridgeway *et al*., 2013) and Artificial Neural Network (Ripley *et al*. 2016).

The results obtained in this study are based on the moisture performance of a North facing wall, as was the case in the original work (Salonvaara *et al*., 2011). In a more recent study, Aggarwal et al. (2021) found that, for several cities in Canada, the critical orientation to assess the moisture performance of building envelope is the one receiving the highest amount of WDR. As such, further development should consider, for each city, the orientation that leads to the worst moisture response.

4 Conclusion

The objective of this study was to verify the applicability of the ASHRAE's moisture severity index (Isev) for Canadian locations in predicting and ranking a 31-year series of climate data in terms of moisture severity (RHT70) and explore the potential to improve the model by recalibrating the equations and by using the Partial Least Squares Regression (PLSR), Least Absolute Shrinkage and Selection Operator (LASSO) regression or Support Vector Regression. First, DELPHIN was used to perform one-dimensional hygrothermal simulations for individual years in a 31-year long series of climate data in 12 cities across Canada: Whitehorse (YT), Vancouver

(BC), Calgary (AB), Saskatoon (SK), Winnipeg (MB), Toronto (ON), Ottawa (ON), Montreal (QC), Moncton (NB), Charlottetown (PE), Halifax (NS) and St. John’s (NL). Then, principal component analysis was performed using the yearly climate statistics to identify the variability in the climate data and construct a representative data set for training and validation. Seven representative cities that included Whitehorse, Vancouver, St. John's, Winnipeg, Calgary, Ottawa, and Moncton were chosen to be part of the training set, and the remaining 5 cities, i.e., Charlottetown, Halifax, Montreal, Saskatoon, and Toronto formed the testing set on which the five models, i.e., original Isev, recalibrated Isev, PLSR, LASSO and SVR, were evaluated. Finally, the results obtained were validated with the actual simulation results from DELPHIN using the root mean squared error of prediction (RMSEP), the Spearman’s rank correlation for the 31-year series and the number of years selected out of the top five worst years.

The results suggest that the original Isev model is unsuitable for predicting the RHT70 index for Canadian cities with more recent climate data, with a median value of RMSEP over the five test cities of 5548. However, when trained on the Canadian cities, the recalibrated Isev model performed better in predicting the RHT70 index (median value of RMSEP of 2686) but failed to improve the ranking of the years. There was a decrease of RMSEP of RHT70 when using either PLSR, LASSO or SVR, with median values over the five test cities of 1963, 2099, and 1591, respectively. The median Spearman’s rank correlation coefficient obtained with the original Isev model over the five test cities was 0.51, compared to 0.32, 0.50, 0.53 and 0.66 for recalibrated Isev, PLSR, LASSO, and SVR, respectively. The original Isev model detected, on average, 2.4 years amongst the ones depicted from hygrothermal simulations results whereas the recalibrated Isev model detected 1.8 years, less than the 2.6 years detected by the PLS, LASSO and SVR models. The regression-based Isev model could therefore be improved by using methods that either accounts for the collinearity present in the climate variables such as PLSR, performs feature selection such as LASSO regression or combines LASSO feature selection with SVR. Other regularization and machine learning algorithms that deal with situations like this, with many and possible correlated predictor variables, will be evaluated in future works.

5 Acknowledgments

The authors would like to express their gratitude to Dr. Abhishek Gaur, for his assistance in providing the climate data. This work is a part of the Climate Resilient Built Environment Program supported by Infrastructure Canada and the National Research Council of Canada

6 References

Aggarwal, C., Defo, M., Ge, H., & Lacasse, M. A. (2021). Identifying the critical orientation in assessing moisture risks of wood-frame walls in hygrothermal simulation for Canadian cities. *Paper accepted for upcoming 8th International Buildings Physics Conference (IBPC).* Denmark.

ASHRAE. (2016). *ASHRAE Standard 160-2016. Criteria for Moisture-Control Design Analysis in Buildings.* Atlanta: ASHRAE.

Breiman, L. (2001). Random forest. *Machine Learning, 45*(1), 5-32.

Cornick, S., Djebbar, R., & Dalgliesh, W. A. (2003). Selecting moisture reference years using a Moisture Index approach. *Building and Environment, 38*(12), 1367–1379.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 273-297.

Delgado, J. M., Barreira, E., Ramos, N. M., & de Freitas, V. P. (2013). *Hygrothermal numerical simulation tools applied to building physics.* Berlin: Springer Science & Business Media.

Delgado, J. M., Ramos, N. M., Barreira, E., & de Freitas, V. P. (2010). A critical review of hygrothermal models used in porous building materials. *Journal of Porous Media, 13*(3), 221-234.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software, 33*(1), 1-22.

Gaur, A., Lacasse, M., & Armstrong, M. (2019). Climate data to undertake hygrothermal and whole building simulations under projected climate change influences for 11 Canadian cities. *Data, 4*(2), 72. doi:10.3390/data4020072

Hastie, T., Qian, J., & Tay, K. (2021, March 15). An introduction to glmnet. Retrieved from https://glmnet.stanford.edu/articles/glmnet.html

Henninger, J. H. (1984). *Solar absorbance and thermal emittance of some common space craft thermal-control coatings. NASA RP-1121.* Washington, DC: NASA.

Hill, D. (2005). Review of Hygrothermal Models for Building Envelope Retrofit Analysis. *Research Highlights*, pp. 1-4.

Incropera, F. P., & DeWitt, D. P. (1996). *Fundamentals of heat andmass transfer* (4th ed.). New York: John Wiley and Sons.

IPCC. (2014). *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change.* Geneva: IPCC.

ISO. (2009). *ISO 15927-3 Hygrothermal performance of buildings - Calculation and presentation of climatic data.* Switzerland: International Organization for Standardization.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R* (1st ed.). New York, NY: Springer.

Kassambara, A., & Mundt, F. (2020, April 1). Package factoextra: Extract and Visualize the results of multivariate data analyses. Vienna, Austria: R Foundation for Statistic Computing. Retrieved 2022, from https://cloud.r-project.org/web/packages/factoextra/factoextra.pdf

Kumaran, M. K., Lackey, J. C., Normandin, N., Tariku, F., & van reenen, D. (2002). *A thermal and moisture transport property database for common building and insulation materials. Final report from ASHRAE Research Project 1018-RP.* ASHRAE.

Martens, H., & Martens, M. (2000). Modified Jack-knife Estimation of Parameter Uncertainty in Bilinear Modelling by Partial Least Squares Regression (PLSR). *Food Quality and Preference, 11*, 5-16.

Martins, J. P., Teofilo, R. F., & Ferreira, M. M. (2010). Computational performance and cross-validation error precision of five PLS algorithms using designed and real data sets. *Journal of Chemometrics, 24*(2010), 320-332.

Mevik, B. H., & Wehrens, R. (2007). The pls package: principal component and partial least squares regression in R. *Journal of Statistical Software, 18*(2), 1-23.

Mevik, B.-H., & Wehrens, R. (n.d.). Introduction to pls Package. Retrieved April 1, 2022, from https://cran.r-project.org/web/packages/pls/vignettes/pls-manual.pdf

NBC. (2015). *National Building Code of Canada.* Ottawa: National Research Council of Canada.

Nicolai, A., & Grunewald, J. (2014). *DELPHIN 5 User Manual and Program Reference.* Dresden, Germany.: Institute for Building Climatology/University of Technology Dresden (TUD).

R Core Team. (2021). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Ridgeway, G., Southworth, M. H., & RUnit, S. (2013). Package ‘gbm’. *Viitattu, 10*(2013), 40.

Ripley, B., & Venables, W. (2016). *Package ‘nnet’: Feed-forward neural networks and multinomial log-linear models.* CRAN R Core Team, 11.

Salonvaara, M. (2011). *Environmental weather loads dor hygrothermal analysis and design of buildings. ASHRAE RP-1325.* ASHRAE.

Salonvaara, M., Sedlbauer, K., Holm, A., & Pazera, M. (2010). Effect of selected weather year for hygrothermal analyses. *Thermal performance of the exterior envelopes of whole buildings XI.* Atlanta, Georgia: ASHRAE.

Sawatsky, M. L., Clyde, M., & Meek, F. (2015). Partial Least Squares Regression in the Social Sciences. *The Quantitative Methods for Psychology, 11*(2), 52-62.

Schölkopf, B., Smola, A., Bach, F., & al., e. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond.* MIT Press.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing, 14*(3), 199-222.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *58*(1), pp. 267-288.

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a restropective. *Journal of the Royal Statistical Society Series B, 73*, pp. 273-282.

Wood, S., & Wood, M. (2015). *Package 'mgcv. R package version 1-7.*

Zhou, X., Derome, D., & Carmeliet, J. (2016). Robust moisture reference year methodology for hygrothermal simulations. *Building and Environment, 110*, 23-35.

7 Appendix

A.1 SVR model

The final SVR equation used to make predictions on the test set is described below:

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(x_i, x) + b$$

$$K(x_i, x) = e^{-\gamma ||x_i - x||^2}$$

Where:

- $\gamma = 0.1$
- $b = -0.40$
- x is a 2D array of size 130 x 9 (Table A2)
- x_i are the new test data
- $\alpha_i - \alpha_i^*$ is a 2D array of size 1 x 130

$\alpha_i - \alpha_i^* =$
```
array([[ 1.  ,  1.  , -1.  ,  1.  , -1.  , -1.  , -1.  ,  0.62,  0.06,
        -0.41, -1.  ,  1.  , -1.  ,  1.  , -0.19,  1.  , -1.  ,  0.06,
         0.02,  1.  ,  0.24, -1.  , -1.  , -0.37,  1.  , -1.  ,  0.21,
         1.  ,  1.  , -1.  , -1.  ,  1.  , -1.  , -1.  ,  0.74,  0.26,
         1.  , -1.  , -1.  ,  1.  , -1.  , -0.19,  1.  ,  1.  ,  0.3 ,
         0.75,  1.  ,  0.86, -1.  ,  1.  , -1.  , -0.03,  0.87,  1.  ,
         0.64, -1.  ,  1.  , -1.  ,  1.  , -0.96, -1.  , -0.17, -1.  ,
         0.84, -1.  ,  1.  ,  1.  ,  1.  , -1.  , -1.  ,  1.  , -0.82,
         1.  , -1.  , -0.98,  1.  , -1.  ,  1.  ,  1.  ,  1.  ,  0.51,
        -1.  ,  1.  , -0.41, -1.  , -1.  , -0.59, -0.56, -1.  ,  1.  ,
        -1.  , -0.28,  1.  , -0.86,  1.  , -1.  , -1.  ,  1.  ,  0.84,
         1.  , -1.  , -1.  ,  1.  , -0.53, -0.1 , -1.  , -0.34,  1.  ,
         1.  ,  1.  ,  1.  , -0.37, -1.  , -0.02, -1.  , -1.  , -1.  ,
         0.38, -1.  , -1.  ,  1.  ,  1.  ,  1.  ,  1.  , -1.  ,  1.  ,
        -1.  , -1.  ,  1.  ,  1.  ]])
```

Sample calculation for one random test sample.

The sample considered is the year 1986 in Halifax. The steps to estimate the RHT70 for that year are:

1. Calculate the yearly climate parameters for that year.

Table A 1. Yearly climate values for the sample.

Parameter	Yearly value
log of the sum of wind driving rain (log(SWDR), L/m^2	5.1
Avg square cloudiness (C^2), Ocktas	27.96
Avg relative humidity (RH)	0.77
Avg temperature (T), ºC	5.53
Avg daily minimum temperature (T_{min}), ºC	-14.18
Avg daily maximum temperature (T_{max}), ºC	22.34
Avg wind direction (WD), Deg from North	207.11
Avg square wind speed (V^2), m^2/s^2	19.79
Avg vapour pressure (P_v), Pa	1079.2

Avg: Annual average

Table A 2. Values of array x of size 130 x 9

```
array([[-0.37, -1.24, -1.57, -0.37, -0.02, -0.79,  0.88, -0.61, -0.41],
       [ 0.13, -0.58, -0.96, -0.47, -0.87, -0.73,  0.42, -0.2 , -0.78],
       [-0.49, -0.54, -0.97, -0.4 , -0.63, -0.14,  0.55, -0.53, -0.48],
       [ 0.15, -0.86, -0.97, -0.5 , -0.19, -1.33,  0.35, -0.38, -0.8 ],
       [ 0.87, -0.93, -1.38, -0.21,  0.14,  0.01,  0.89, -0.11, -0.25],
       [ 0.3 , -0.66, -0.79, -0.13,  0.09, -0.39,  0.71, -0.42, -0.32],
       [ 1.25, -0.74, -1.04, -0.06,  0.02, -0.46,  0.49, -0.33, -0.42],
       [ 0.23, -0.74, -1.03, -0.32, -0.9 , -0.38,  0.84, -0.31, -0.42],
       [ 0.21, -0.77, -0.94, -0.35, -0.87, -0.05,  0.26, -0.44, -0.23],
       [-0.37, -1.45, -1.47,  0.03, -0.4 , -0.24,  0.69, -0.34, -0.16],
       [-1.29, -1.54, -1.55,  0.21, -0.03, -0.41,  0.54, -0.27,  0.19],
       [ 0.69, -0.59, -0.87,  0.3 , -0.19, -0.07,  0.15, -0.52,  0.2 ],
       [-0.29, -1.72, -1.7 ,  0.29, -0.77,  0.78,  0.68,  0.01,  0.46],
       [-0.49, -1.29, -1.73,  0.14, -0.57,  0.83,  0.76,  0.31,  0.6 ],
       [ 1.39, -0.75, -1.01,  0.3 , -0.49, -0.22,  0.67, -0.2 ,  0.16],
       [-0.06, -1.26, -1.28,  0.22, -0.32, -0.04,  0.64, -0.21,  0.04],
       [ 0.9 ,  0.05,  0.91,  0.14, -0.18,  0.04,  0.11,  0.73,  0.04],
       [ 1.69, -0.14,  0.62,  0.16,  0.47,  1.08,  0.47,  0.67,  0.06],
       [ 1.57, -0.34,  0.9 , -0.27, -0.24, -0.53,  0.76,  0.64, -0.27],
       [ 1.28,  0.  ,  0.92,  0.17,  0.03, -0.4 ,  0.3 ,  0.84,  0.12],
       [ 1.05, -0.1 ,  0.76,  0.41,  0.16,  0.75,  0.51,  1.01,  0.61],
       [ 0.47, -0.12,  0.91,  0.43,  0.22,  0.56,  0.78,  0.3 ,  0.46],
       [ 1.25, -0.69,  0.73,  0.09, -0.4 , -0.14,  0.51,  0.87,  0.23],
       [ 0.57, -0.64,  0.89,  0.6 ,  0.98,  0.4 ,  0.61,  0.45,  0.55],
       [ 0.63, -0.92,  0.44,  0.28, -0.26,  0.78,  0.62,  0.93,  0.44],
       [ 1.26, -0.34,  0.71,  0.3 ,  0.  ,  0.26,  0.62,  0.39,  0.33],
       [ 1.21,  0.29,  1.24,  0.79,  0.72,  0.1 ,  0.17,  0.74,  0.88],
       [ 0.95, -0.06,  0.91,  0.54,  0.2 ,  0.34,  0.54,  0.74,  0.64],
       [ 1.17, -0.4 ,  1.04,  0.25,  0.23,  0.32,  0.24,  0.49,  0.35],
       [ 0.39, -0.78, -0.1 ,  0.21, -0.27,  1.17,  0.62, -0.73,  0.77],
       [ 0.23, -0.7 ,  0.  ,  0.26, -0.01,  1.3 ,  0.8 , -0.78,  0.53],
       [ 0.56, -0.8 ,  0.15,  0.58, -0.6 ,  1.01,  0.49, -0.76,  0.72],
       [ 0.52, -1.14, -0.48,  0.51, -0.64,  1.63,  0.95, -0.64,  1.4 ],
       [ 0.05, -0.72, -0.06,  0.18, -0.53,  1.15,  0.49, -0.52,  0.76],
       [ 0.46, -0.  ,  0.25, -0.16, -0.32,  0.41,  0.97, -0.7 , -0.06],
       [ 0.55, -0.05,  0.47,  0.26, -0.11,  0.78,  0.38, -0.73,  0.61],
       [ 0.5 , -0.93, -0.07,  0.49, -0.57,  1.06,  0.67, -0.67,  0.81],
       [ 0.4 , -0.34,  0.21,  0.48, -0.35,  0.66,  0.74, -0.66,  0.59],
       [ 0.67, -1.18,  0.13,  0.8 , -0.35,  1.6 ,  0.36, -0.73,  1.42],
       [ 0.14, -1.29,  0.06,  0.57, -0.12,  1.29,  0.87, -0.74,  1.04],
       [ 0.66, -0.51,  0.18,  0.32, -0.32,  1.5 ,  0.44, -0.57,  0.77],
       [ 0.7 , -0.69,  0.21,  0.55, -0.26,  1.29,  0.72, -0.78,  1.07],
       [ 0.68, -0.36,  0.13,  0.78,  0.65,  1.14,  0.45, -0.51,  0.99],
       [ 0.39, -0.81,  0.03,  0.63,  0.22,  0.87,  0.58, -0.8 ,  0.9 ],
       [-0.07, -0.34,  0.04,  0.83, -0.25,  2.1 ,  0.87, -0.64,  1.3 ],
       [ 1.06, -0.42,  0.07,  0.44, -0.35,  1.4 ,  0.19, -0.52,  0.92],
       [ 0.76, -1.09,  0.02,  0.64, -0.  ,  1.03,  0.53, -0.72,  1.14],
       [ 1.19, -0.15,  0.38,  0.98, -0.09,  1.77,  0.56, -0.68,  1.49],
       [ 0.07,  1.83,  1.77, -1.1 ,  0.56, -1.78,  0.7 ,  1.65, -1.64],
       [ 0.85,  1.63,  1.72, -0.61,  0.98, -1.48,  0.75,  2.12, -1.23],
       [ 1.21,  1.67,  1.75, -0.58,  1.21, -1.62,  0.57,  2.  , -1.24],
       [ 0.09,  1.84,  1.77, -1.16,  0.89, -2.25,  0.88,  2.23, -1.84],
       [ 0.67,  1.95,  1.75, -0.9 ,  0.83, -2.02,  1.27,  1.81, -1.54],
       [ 0.48,  1.36,  1.66, -0.32,  1.05, -1.42,  0.71,  1.95, -0.92],
       [ 1.23,  1.81,  1.52, -0.1 ,  0.95, -1.04,  0.87,  2.43, -0.68],
       [ 0.92,  2.46,  1.75, -1.  ,  0.96, -1.61,  1.27,  2.16, -1.65],
       [ 0.79,  1.67,  1.69, -0.73,  1.05, -1.6 ,  0.85,  2.03, -1.3 ],
       [ 0.58,  1.97,  1.82, -0.15,  1.22, -1.11,  0.3 ,  1.84, -0.72],
       [ 0.04,  1.71,  1.7 , -0.25,  1.11, -1.03,  1.02,  2.07, -0.84],
       [-0.01,  1.19,  1.58, -0.3 ,  1.17, -1.22,  0.95,  1.9 , -0.92],
       [ 0.38,  1.48,  1.87, -0.12,  1.25, -1.07,  0.66,  1.68, -0.73],
       [ 1.01,  1.56,  1.53, -0.29,  1.16, -1.39,  0.99,  2.05, -0.94],
       [ 0.5 ,  1.95,  1.68, -0.34,  1.26, -1.2 ,  0.8 ,  2.02, -0.96],
       [ 1.09,  2.23,  1.87, -0.1 ,  1.34, -0.98,  0.08,  1.98, -0.73],
       [ 0.44,  1.58,  1.66, -0.21,  1.15, -1.12,  1.07,  1.9 , -0.8 ],
       [ 0.42,  2.44,  2.04, -0.07,  1.45, -1.37,  0.3 ,  1.64, -0.78],
       [-0.24,  1.43,  1.67, -0.08,  1.36, -1.25,  0.81,  2.07, -0.67],
       [-0.9 , -0.05, -1.18,  1.78,  1.7 ,  0.28, -1.87, -1.26,  1.4 ],
       [-1.51, -0.21, -1.35,  1.77,  0.65,  0.25, -2.26, -1.36,  1.42],
       [-0.92,  0.7 , -1.13,  1.53,  0.88,  0.29, -1.98, -1.21,  1.06],
       [-0.87,  0.9 , -0.96,  1.41,  1.53, -0.12, -1.67, -1.2 ,  0.8 ],
       [-1.76, -0.48, -1.47,  1.76,  0.38,  0.67, -2.06, -1.2 ,  1.48],
       [-1.19,  0.97, -1.05,  1.7 ,  1.68,  0.01, -1.9 , -1.1 ,  1.22],
       [-1.67,  0.7 , -1.14,  1.36,  1.09,  0.84, -1.62, -1.15,  0.93],
       [-1.4 ,  0.43, -1.21,  1.09,  0.88, -0.5 , -1.81, -1.25,  0.5 ],
       [-1.44, -0.28, -1.33,  1.67,  0.49,  0.74, -1.97, -1.27,  1.35],
       [-1.55,  0.38, -1.14,  1.91,  1.78,  0.46, -2.04, -1.39,  1.46],
       [-1.58, -0.05, -1.39,  1.93,  1.41,  1.56, -2.18, -1.42,  1.78],
       [-1.34,  0.5 , -1.15,  1.67,  1.17,  0.28, -1.92, -1.22,  1.18],
       [-0.86,  1.03, -1.17,  1.35,  0.66,  0.67, -2.13, -1.28,  0.92],
       [-2.37,  0.37, -1.11,  2.08,  1.74,  0.45, -1.83, -1.36,  1.78],
       [-1.33,  0.47, -1.18,  1.97,  1.78,  0.53, -1.81, -1.31,  1.57],
       [-0.99,  0.93, -1.06,  1.72,  2.01,  0.29, -1.82, -1.05,  1.26],
       [-1.91, -0.19, -1.24,  1.91,  1.69,  0.11, -2.  , -1.36,  1.5 ],
       [-1.47,  0.22, -1.15,  1.87,  1.54,  1.04, -1.91, -1.32,  1.5 ],
       [-1.21,  1.72, -0.85,  1.72,  2.01,  0.3 , -1.62, -0.88,  1.23],
       [-1.53,  0.73, -1.03,  2.16,  1.95,  1.55, -2.17, -1.24,  1.86],
       [-1.53,  0.55, -0.73, -1.67, -0.63, -1.08, -0.68, -0.39, -1.56],
       [-0.89,  0.31, -0.66, -1.24, -1.31, -1.06, -1.1 , -0.59, -1.2 ],
       [-1.88,  0.85, -0.58, -2.46, -1.65, -1.23, -0.7 , -0.68, -2.05],
       [-1.64,  0.21, -0.87, -2.27, -1.72, -1.34, -1.26, -0.57, -1.9 ],
       [-1.51,  0.44, -0.62, -1.47, -1.16, -1.18, -1.51, -0.7 , -1.49],
       [-0.74,  0.62, -0.55, -1.73, -0.87, -1.05, -1.35, -0.73, -1.62],
       [-1.49,  0.82, -0.55, -1.61, -1.67, -1.07, -0.86, -0.54, -1.68],
       [-0.7 ,  0.9 , -0.54, -1.65, -1.04, -1.67, -0.79, -0.15, -1.65],
       [-0.88,  0.69, -0.51, -1.45, -1.33, -0.99, -1.77, -0.62, -1.38],
       [-1.06,  0.19, -0.85, -1.82, -1.1 , -1.32, -0.65, -0.25, -1.67],
       [-0.83,  1.18, -0.41, -1.44, -1.17, -1.5 , -1.55, -0.68, -1.51],
       [-0.86,  0.84, -0.56, -1.86, -1.55, -0.91, -1.03, -0.69, -1.34],
       [-1.47,  0.02, -0.74, -1.16, -1.35, -1.05, -0.98, -0.57, -1.24],
       [-0.87, -0.21, -0.78, -1.03, -0.98, -0.24, -1.28, -0.41, -0.77],
       [-1.17,  0.86, -0.51, -1.29, -1.4 , -0.89, -0.99, -0.62, -1.16],
       [-0.28,  0.82, -0.53, -1.06, -1.96, -0.68, -1.48, -0.3 , -1.02],
       [-0.38,  0.63, -0.67, -1.19, -0.41, -0.33, -1.62, -0.49, -1.21],
       [-1.83,  0.65, -0.62, -1.05, -1.21, -1.3 , -0.96, -0.65, -1.27],
       [-1.14,  1.04, -0.63, -1.77, -1.57, -1.83, -0.5 ,  0.07, -1.56],
       [-1.52,  0.88, -0.49, -1.28, -1.08, -0.96, -1.38, -0.7 , -1.19],
       [-1.91,  0.39, -0.66, -1.06, -0.79, -0.89, -1.28, -0.48, -1.09],
       [ 0.15, -0.62,  0.36, -0.89, -0.52,  0.36,  0.53,  0.03, -0.33],
       [-0.25, -0.9 ,  0.21, -0.33, -1.35,  1.47, -0.03,  0.11,  0.38],
       [ 0.58, -1.27,  0.32, -0.83, -1.43,  0.99,  0.74, -0.08,  0.08],
       [ 1.  , -1.22,  0.37, -1.  , -1.11,  0.8 ,  0.66, -0.08, -0.06],
       [ 0.3 , -1.12,  0.34, -1.02, -0.31,  0.71,  0.79,  0.35, -0.33],
       [ 0.43, -1.18,  0.23, -0.61, -0.94,  0.55,  0.56, -0.14, -0.15],
       [ 0.2 , -1.24, -0.04, -0.32, -1.34,  1.27,  0.39,  0.32,  0.44],
       [ 0.26, -0.93,  0.33, -0.37, -1.23,  1.18,  0.25, -0.  ,  0.39],
       [ 0.4 , -1.17,  0.15, -0.67, -0.66,  1.08,  0.56,  0.02,  0.23],
       [ 0.19, -1.47, -0.09, -0.21, -0.69,  0.88,  0.24,  0.32,  0.48],
       [-0.69, -1.34, -0.19, -0.53, -0.89,  1.2 ,  0.5 ,  0.37,  0.29],
       [-0.56, -1.28,  0.04, -0.08, -0.42,  0.81,  0.6 ,  0.07,  0.53],
       [-0.27, -1.28, -0.06, -0.24, -1.39,  1.57,  0.5 ,  0.05,  0.57],
       [-0.04, -1.08,  0.02, -0.13, -0.45,  0.36,  0.78,  0.23,  0.39],
       [ 0.64, -0.9 ,  0.37, -0.45, -1.49,  1.36,  0.75, -0.03,  0.21],
       [ 0.44, -0.67,  0.42, -0.52, -0.94,  0.68,  0.68,  0.04,  0.05],
       [ 0.74, -0.61,  0.39, -0.37, -1.42,  1.17,  0.09, -0.2 ,  0.49],
       [ 0.26, -1.2 ,  0.23, -0.38, -1.16,  1.04,  0.67,  0.07,  0.35],
       [ 0.68, -0.86,  0.27, -0.33, -1.21,  1.11,  0.46,  0.01,  0.55],
       [ 0.14, -1.07,  0.45,  0.03, -0.54,  1.36,  0.37, -0.06,  0.36],
       [ 0.64, -0.78,  0.51,  0.14, -0.62,  1.42,  0.47, -0.03,  0.65],
       [ 0.03, -1.04,  0.21, -0.68, -0.62,  0.42,  0.62,  0.17,  0.01]])
```

2. Standardize the yearly value of climate variables using the mean and standard deviation calculated from the training set (Table A3). The resulting transformed values are shown in Table A4.

Table A 3. Mean and Standard Deviation used to standardize the training data.

Yearly climate parameter	Mean	Standard Deviation
Log(SWDR)	4.19	0.92
C^2	26.82	5.44
RH	0.70	0.07
T	5.05	3.12
Tmin	-22.72	10.39
Tmax	24.28	4.01
WD	193.35	21.11
V^2	17.22	7.25
P_v	1115.35	194.64

Table A 4. Transformed climate values used for SVR prediction.

Yearly climate parameter	Original scale	Transformed scale
Log(SWDR)	5.10	1.00
C^2	27.96	0.21
RH	0.77	0.91
T	5.53	0.15
Tmin	-14.18	0.82
Tmax	22.34	-0.48
WD	207.11	0.65
V^2	19.79	0.35
P_v	1079.20	-0.18

3. Calculate $\mathrm{K}(\mathrm{x_i}, \mathrm{x})$, where $\mathrm{x_i}$ is the standardized climate values of the test sample, and x is the 2D array given above. The resulting new feature space $\mathrm{K}(\mathrm{x_i}, \mathrm{x})$ is a 2D array of size 1 X 130.

$\mathrm{K}(\mathrm{x_i}, \mathrm{x}) =$
```
array([[0.29, 0.4 , 0.37, 0.42, 0.45, 0.58, 0.54, 0.4 , 0.4 , 0.29, 0.2 ,
        0.53, 0.18, 0.2 , 0.49, 0.36, 0.83, 0.71, 0.82, 0.89, 0.72, 0.79,
        0.74, 0.78, 0.6 , 0.82, 0.79, 0.81, 0.83, 0.43, 0.45, 0.43, 0.24,
        0.41, 0.65, 0.61, 0.4 , 0.57, 0.28, 0.34, 0.44, 0.42, 0.53, 0.49,
        0.25, 0.43, 0.42, 0.33, 0.32, 0.43, 0.42, 0.22, 0.31, 0.53, 0.45,
        0.24, 0.41, 0.48, 0.46, 0.54, 0.58, 0.5 , 0.44, 0.42, 0.53, 0.37,
        0.46, 0.09, 0.05, 0.11, 0.15, 0.05, 0.1 , 0.09, 0.13, 0.07, 0.06,
        0.03, 0.09, 0.1 , 0.03, 0.07, 0.1 , 0.05, 0.06, 0.09, 0.03, 0.14,
        0.17, 0.05, 0.05, 0.1 , 0.14, 0.09, 0.16, 0.12, 0.13, 0.12, 0.12,
        0.13, 0.2 , 0.15, 0.15, 0.22, 0.11, 0.11, 0.12, 0.13, 0.58, 0.27,
        0.32, 0.39, 0.51, 0.46, 0.3 , 0.36, 0.43, 0.39, 0.27, 0.39, 0.23,
        0.53, 0.33, 0.52, 0.36, 0.37, 0.39, 0.43, 0.45, 0.51]])
```

4. Next perform the dot product between $\alpha_i - \alpha_i^*$ and the transpose of $K(x_i, x)$. Then add the intercept term to obtain the scaled prediction for RHT70. In this sample calculation this is equal to 0.93

5. Finally inverse transform the scaled prediction by multiplying by 5972.75 (standard deviation) and adding the mean (11730.06). In this example this is equal to $(0.93 * 5972.75) + 11730.06 = 17284.72$. This value is the prediction of RHT70 in the original scale for the year 1986 in Halifax.

A.2 Wind driving rain rose

The formula used to calculate the wind driving rain for the purpose of making the wind driving rain rose plots is:

$$WDR = 0.22 \times V \times R^{0.88} \times Cos(\theta)$$

Where:

V: wind speed (m/s)
R: horizontal rain (mm/h)
θ: angle between the normal to the wall and the wind direction

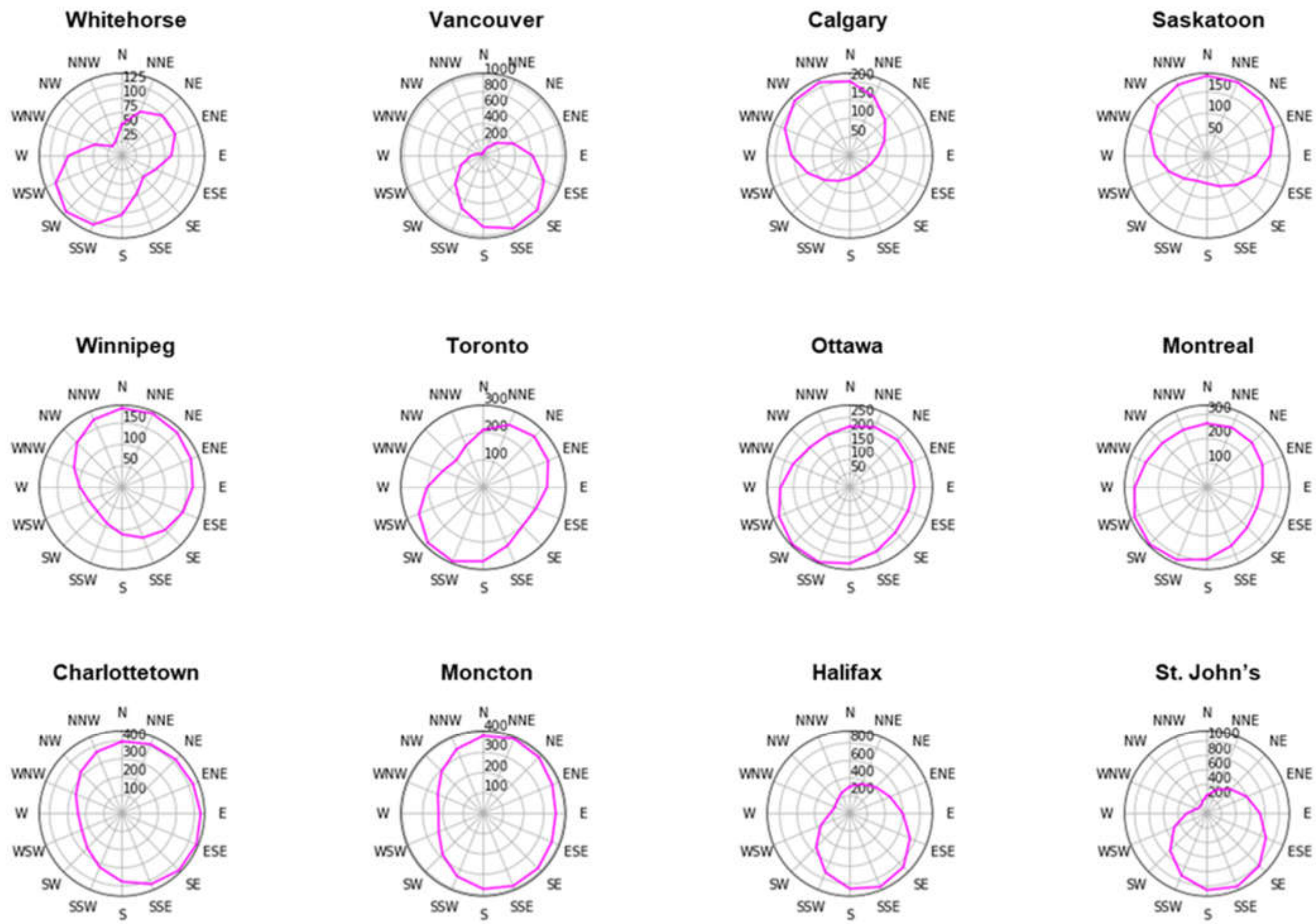


Figure A 1. Distribution of wind driving rain for the selected run in each city. Values are average of yearly summ over the 31 years of the run (L/m^2).

A.3 Comparison of yearly values of RHT70 obtained from hygrothermal simulation and original Isev equation

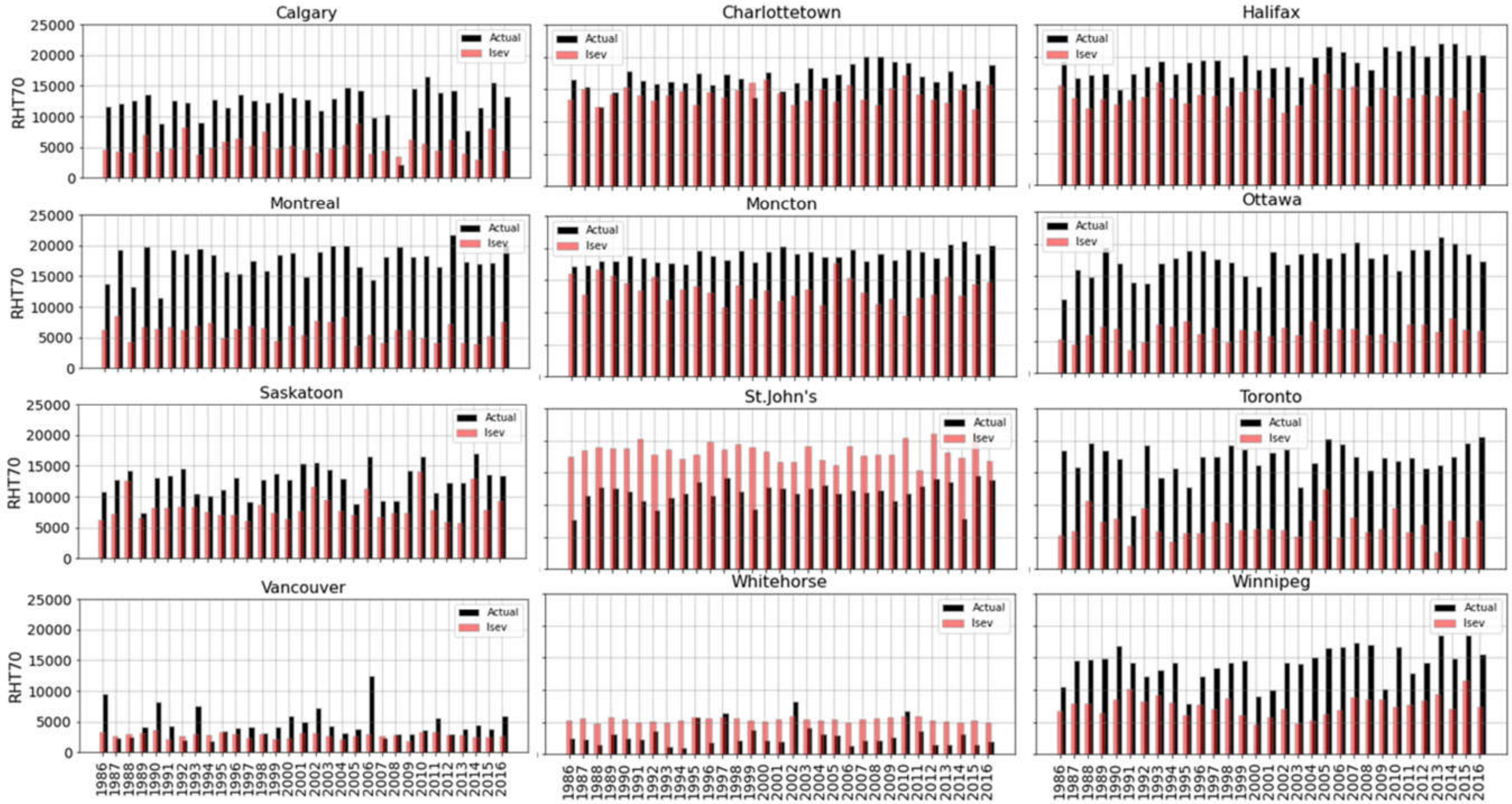


Figure A 2. Simulated (Actual) and predicted (Isev) RHT70 using the original Isev equation in the 12 Canadian cities considered.

A.4 Comparison of yearly values of RHT70 obtained from hygrothermal simulations and recalibrated Isev equation

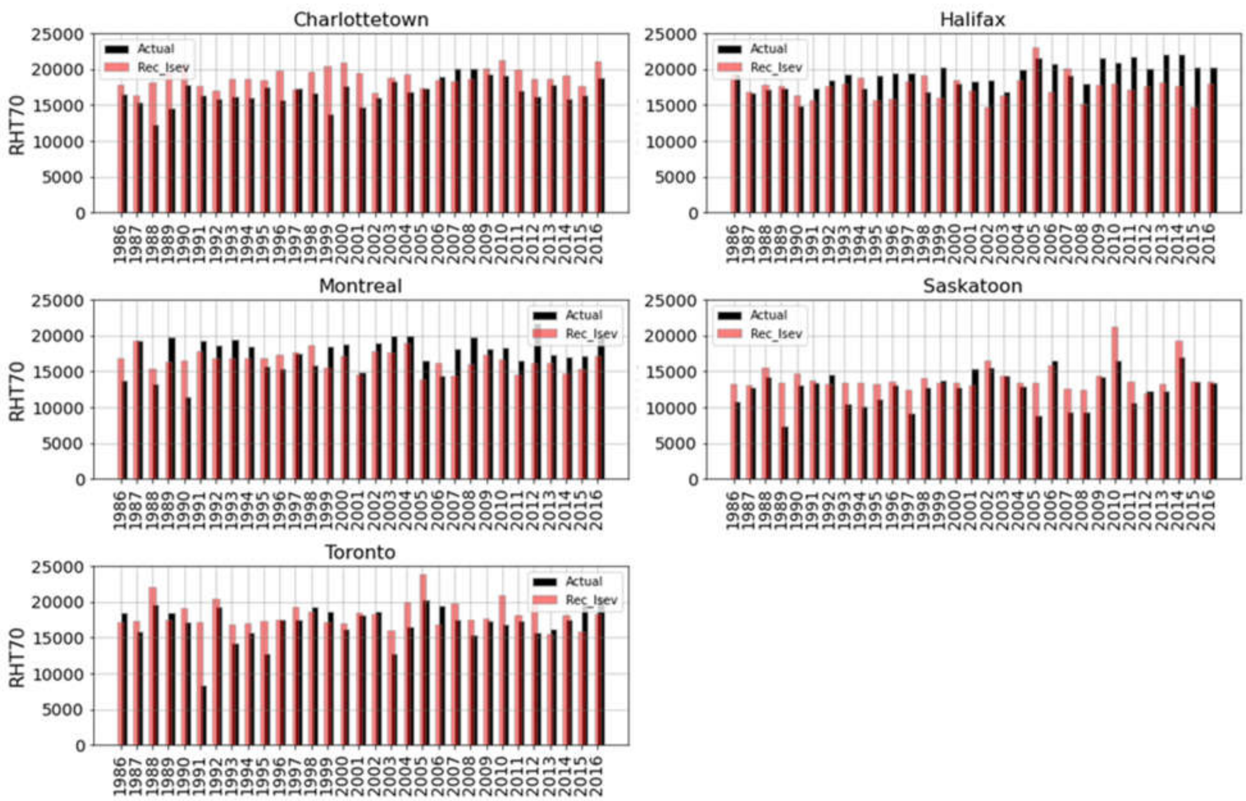


Figure A 3. Comparison of RHT70 calculated from hygrothermal simulation results (Actual) and that obtained using the recalibrated Isev equation (Rec_Isev) in the test cities.

A.5 Comparison of yearly values of RHT70 obtained from hygrothermal simulations and PLSR model

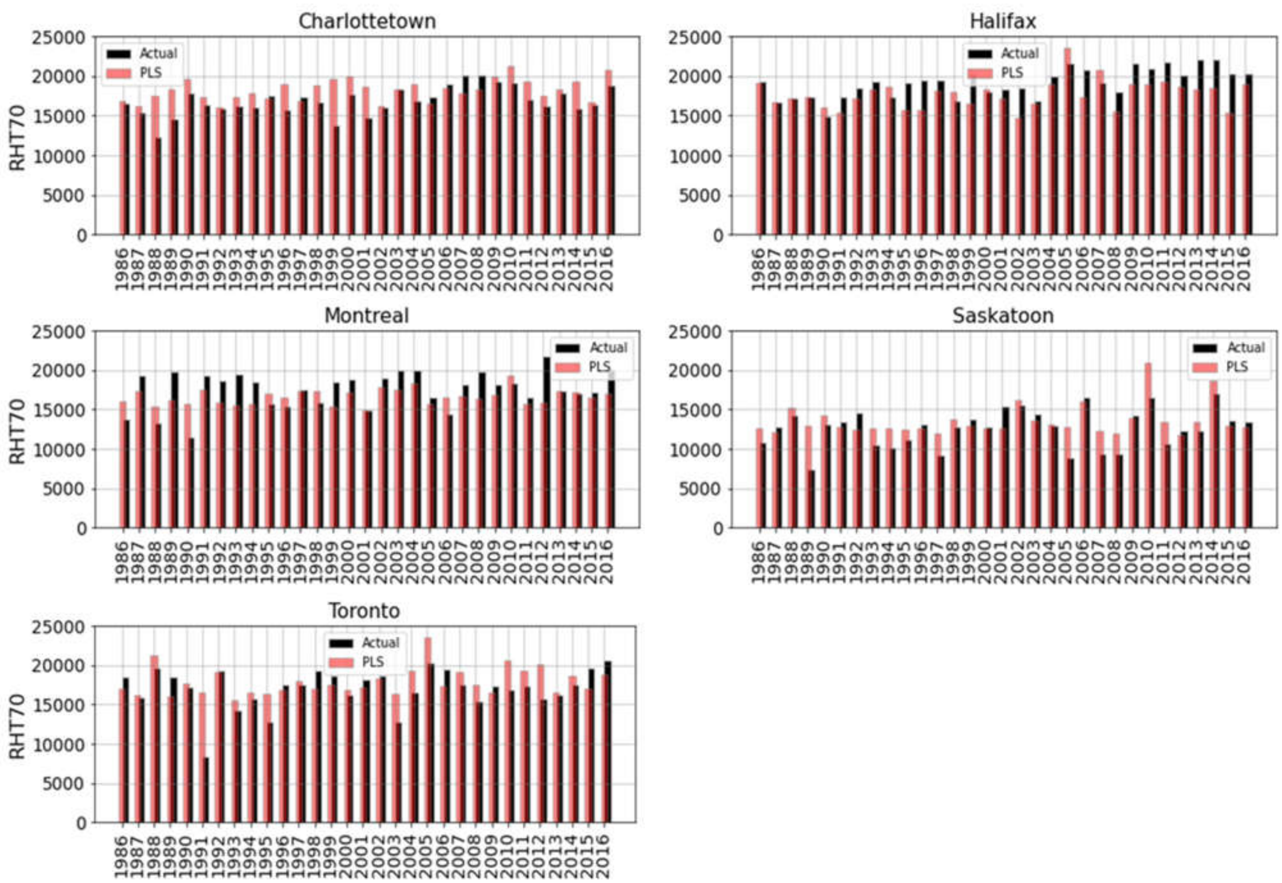


Figure A 4. Comparison of RHT70 calculated from hygrothermal simulation results (Actual) and that obtained using the partial lest squares regression (PLSR) in the test cities.

A.6 Comparison of yearly values of RHT70 obtained from hygrothermal simulations and LASSO model

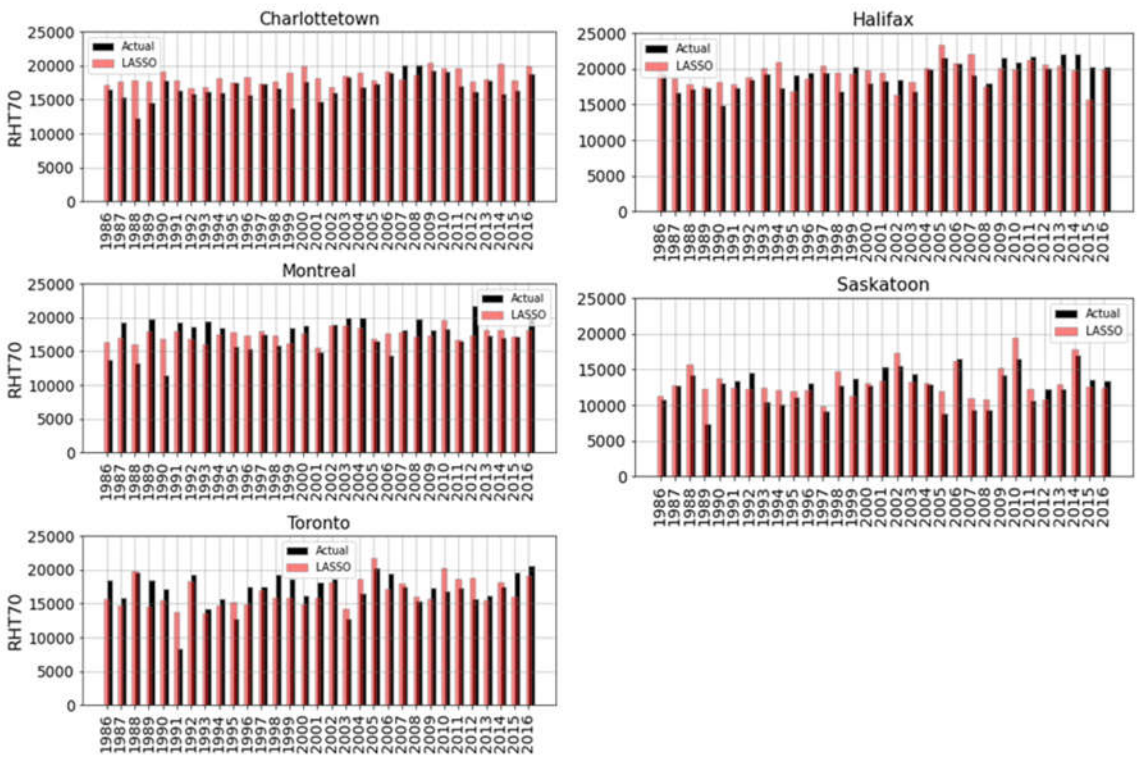


Figure A 5. Comparison of RHT70 calculated from hygrothermal simulation results (Actual) and that obtained using the LASSO regression (LASSO) in the test cities.

A.7 Comparison of yearly values of RHT70 obtained from hygrothermal simulations and SVR model

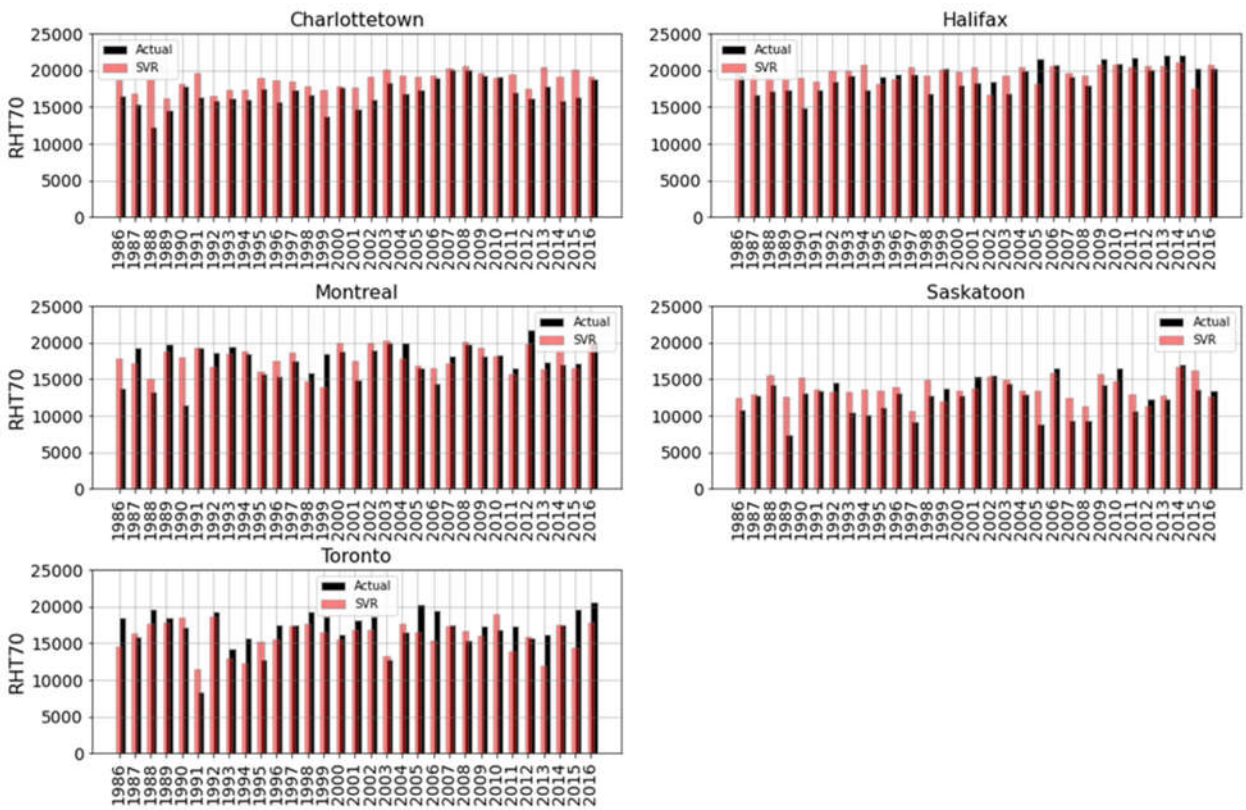


Figure A 6. Comparison of RHT70 calculated from hygrothermal simulation results (Actual) and that obtained using the support vector regression (SVR) in the test cities.