



This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.

 **creative commons**  
C O M M O N S D E E D

**Attribution-NonCommercial-NoDerivs 2.5**

**You are free:**

- to copy, distribute, display, and perform the work

**Under the following conditions:**

 **Attribution.** You must attribute the work in the manner specified by the author or licensor.

 **Noncommercial.** You may not use this work for commercial purposes.

 **No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

**Your fair use and other rights are in no way affected by the above.**

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:  
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

# Relationship between annual mean temperature and degree-days

Monjur Mourshed<sup>a,\*</sup>

<sup>a</sup>*Building Energy Research Group (BERG), School of Civil and Building Engineering, Loughborough University, Loughborough, Leicestershire, LE11 3TU, United Kingdom*

---

## Abstract

Degree-days are a versatile climatic indicator and used for many applications in the design and operation of energy efficient buildings – from the estimation of energy consumption and carbon emissions due to space heating and cooling to the energy and environmental monitoring of buildings. This research is aimed at developing an equation for calculating degree-days from low-resolution temperature data by exploring the relationship between degree-days and annual mean temperature of 5511 locations around the world, using multiple non-linear regression. Results suggest a very strong relationship between annual mean temperature and degree-days. Incorporating standard deviation (SD) of monthly mean temperature and latitude increases the accuracy of prediction ( $R^2 > .99$ ), demonstrating the strength of the location-agnostic relationship in predicting degree-days from two temperature parameters: annual mean and SD of monthly mean. Research findings can be used to calculate degree-days of locations, for which daily temperature data may not be available. The equation can also be used to calculate degree-days from low-resolution global circulation model (GCM) projections of increasing temperature, for investigating the impact of climate change on building heating and cooling energy demand at global scale without the need to create synthetic weather series through morphing or downscaling.

*Keywords:* Annual mean temperature, Global relationship, Heating degree-days, Cooling degree-days

---

## Symbols and units

$T$	annual mean outdoor air temperature (°C)
$T_d$	daily mean outdoor air temperature (°C)
$T_m$	monthly mean outdoor air temperature (°C)
$T_i$	outdoor air temperature at the $i$ -th hour of the day (°C)
$T_{max}$	daily maximum outdoor air temperature (°C)
$T_{min}$	daily minimum outdoor air temperature (°C)
$S_d$	standard deviation of daily mean temperature in a month(°C)
$S_m$	standard deviation of monthly mean temperature (°C)
$T_b$	base temperature (°C)
$DD$	degree-days (°C-day)
$HDD$	annual heating degree-days (°C-day)
$HDD_d$	daily heating degree-days (°C-day)
$HDD_m$	monthly heating degree-days (°C-day)
$CDD$	annual cooling degree-days (°C-day)
$CDD_d$	daily cooling degree-days (°C-day)
$CDD_m$	monthly cooling degree-days (°C-day)

---

\*Corresponding author

Email address: [m.m.mourshed@lboro.ac.uk](mailto:m.m.mourshed@lboro.ac.uk) (Monjur Mourshed)

URL: <http://monjur.mourshed.org/> (Monjur Mourshed), +44 (0)1509 228792 (Monjur Mourshed), +44 (0)1509 223981 (Monjur Mourshed)

## 1. Introduction

Degree-days are an important climatic design indicator that captures the extremity and duration of ambient temperature [1]. They are essentially the summation of temperature differences between the ambient or outdoor air temperature and a reference temperature, which is also known as the base or balance point temperature. The base temperature,  $T_b$ , is referred to as the outdoor air temperature at which the heating or cooling systems do not need to run in order to maintain comfort conditions [1]. In other words, for the specified value of the indoor air temperature; i.e., set point temperature, the total heat loss from the space is equal to the heat gain from sun, occupants, lights, equipment, etc. [2]. When outdoor air temperature is below the base temperature, the heating system needs to provide heat. As heat loss from a building is directly proportional to the differences between indoor and outdoor air temperature, the energy consumption of a heated building over a period of time relates to the sum of these temperature differences over this period. On the other hand, cooling systems need to operate if outdoor air temperature is above the base temperature. The summed temperature differences or the cooling degree-days also have a relationship with cooling energy consumptions.

Degree-days and their uses depend on two distinct but essentially unrelated issues: the way degree-days are calculated and the way they are applied to building energy [1]. Calculation techniques for degree-days vary depending on the temporal resolution of temperature data. Hourly temperature (outdoor air) data produces better estimates of degree-days than daily or monthly methods. On the other hand, the application of degree-days varies depending on whether it is for heating or cooling season. Thermal response of the building; i.e., the overall heat transfer coefficient is an important factor for the application of heating degree-days as it influences the base temperature, specific to the building. Another factor affecting the use of heating degree-days is whether the building is continuously or intermittently heated. For the application of cooling degree-days, the type of cooling systems affects its use. This paper is concerned with the calculation of degree-days from a low resolution temperature dataset and, therefore, the reader is directed to the associated references for detailed discussions on: general overview [1, 2]; the use and identification of base temperature [3]; estimates of chiller energy consumption using degree-days [4]; use in energy management [5] and uncertainties in energy estimates using degree-days [6].

Apart from their use in estimating energy demand and associated carbon emissions, degree-days are widely used as a climatic indicator for the assessment of the impact of climate change, in particular the increasing temperature. Degree-days are particularly suited for the analysis of extremity and duration of increased outdoor temperature [7–9] due to the simplicity of its concept and its extensibility in investigating the inherent uncertainties in climate projection from global circulation models (GCM). The uncertainties in GCM projection primarily originate from the use of various emissions scenarios<sup>1</sup>, which result in a range of temperature projections instead of a single figure [10]. Other forms of uncertainties originate from the ocean-atmosphere coupling in the underlying model; i.e., the way atmospheric interactions are modeled in a particular GCM. Multi-model ensembles are, therefore, used to quantify uncertainties and to generate projections of future climate [11] for wider applications such as risk and vulnerability assessments.

On the other hand, GCM projections of surface air temperature are typically reported as anomalies at monthly scale. Only a few models, 8 out of 23 in the IPCC fourth assessment report (FAR), give air temperature outputs at daily scale and only for a few selected marker scenarios. These outputs are not readily suitable for building applications. GCM outputs need to be temporally downscaled to finer resolutions;

---

<sup>1</sup>Climate change projections depend on future human activities: economic, environmental and technological. Future climate projections are, therefore, based on several scenarios, each making different assumptions for future technological and economic development, affecting the concentration of greenhouse gas in the atmosphere. There are six families of scenarios adopted in the assessment reports by the Intergovernmental Panel on Climate Change (IPCC) [10]. The six illustrative marker scenarios are B1, A1T, B2, A1B, A2, and A1FI, representing an approximate CO<sub>2</sub>-eq concentrations in 2100 of about 600, 700, 800, 850, 1250 and 1550 ppm, respectively. The use of different scenarios results in uncertainties in climate projection; e.g., the best estimate of temperature change at 2090-2099 is 1.8°C and 4.0°C for B1 and A1FI scenarios respectively.

e.g., hourly, if detailed-based building simulation tools are to be used. Such downscaling techniques introduce uncertainties in the synthetic future weather series, mainly because of the assumption that the variability and distribution of weather pattern is similar in present-day and future climates [12], which may not necessarily be the case. The alternative is to develop a technique that uses coarse resolution (e.g., monthly or annual) GCM outputs of surface air temperature to determine changes in degree-days. This would eliminate uncertainties associated with downscaling and enable the reconciliation of multiple marker scenarios for improved reliability of projections.

Considering the importance of calculating degree-days from a low temporal resolution temperature series, this paper explores the relationship between annual mean temperature and degree-days of 5511 locations around the world. The aim is to develop an equation to predict heating and cooling degree-days from annual mean temperature and site specific geographical parameters such as latitude. The rest of the paper is organized as follows. Commonly used methods for calculating degree-days are discussed next, followed by a discussion on the methods adopted in this research and sources of data. The results and directions of future research are deliberated, with an emphasis on the accuracy of the developed equation. The paper ends with concluding remarks.

## 2. Calculation of degree-days

Depending on the availability of outdoor air temperature data, different methods are used for calculating heating and cooling degree-days. The hourly or *ideal* method produces the most accurate estimate. However, the hourly method is not suitable for all applications due to the unavailability of hourly temperature series for many locations. Therefore, several attempts have been made in the past to develop methods for calculating degree-days from reduced datasets. Of note are the works by Thom [13], Erbs et al. [14], Hitchin [15] and Schoenau and Kehrig [16]. Of available techniques using reduced datasets, the following are commonly used and, therefore, are discussed in this article: the American Society of Heating, Refrigerating and Air-Conditioning (ASHRAE) daily mean temperature method [2]; UK Meteorological Office (UKMO) daily maximum and minimum temperature method [1] and Schoenau and Kehrig's monthly mean temperature method [16].

### 2.1. Hourly method

Hourly temperature data for a location are used in this method to sum differences between the base temperature and hourly temperature measurements; i.e., degree-hours. The cumulative degree-hours of a day is divided by 24 to get the mean degree-hours or degree-days. Daily heating degree-days,  $HDD_d$  and daily cooling degree-days,  $CDD_d$  are given by Equations 1 and 2 respectively.

$$HDD_d = \left( \sum_{i=1}^{24} (T_b - T_i)^+ \right) / 24 \quad (1)$$

$$CDD_d = \left( \sum_{i=1}^{24} (T_i - T_b)^+ \right) / 24 \quad (2)$$

where  $T_b$  is base temperature and  $T_i$  is outdoor air temperature at the  $i$ -th hour of the day. The subscript plus symbol (+) means that only positive differences between  $T_b$  and  $T_i$  are taken into account.

Monthly degree-days,  $DD_m$  is calculated by summing up the daily degree-days,  $DD_d$  in a month:

$$DD_m = \sum_{j=1}^P (DD_{d,j}) \quad (3)$$

where  $P$  is number of days in a month and  $DD_{d,j}$  is daily degree-days on the  $j$ -th day of the month.

Annual degree days,  $DD_a$  is calculated by summing up monthly degree-days,  $DD_m$  :

$$DD_a = \sum_{k=1}^{12} (DD_{m,k}) \quad (4)$$

where  $DD_{m,k}$  is monthly degree-days of the  $k$ -th month of the year.

## 2.2. ASHRAE formula

According to the ASHRAE daily mean temperature method, the daily degree-days is the difference between the daily mean temperature,  $T_d$  and base temperature,  $T_b$  [2]. The calculations for daily heating degree-days,  $HDD_d$  and daily cooling degree-days,  $CDD_d$  are given by Equations 5 and 6 respectively.

$$HDD_d = (T_b - T_d)^+ \quad (5)$$

$$CDD_d = (T_d - T_b)^+ \quad (6)$$

$T_d$  is calculated from daily maximum and minimum temperatures,  $T_{max}$  and  $T_{min}$  respectively:

$$T_d = (T_{max} + T_{min})/2 \quad (7)$$

## 2.3. UKMO equations

The UKMO method uses a single daily reading of a maximum and minimum thermometer. There are four possible relationships between the base temperature and diurnal temperature variation, resulting in four different scenarios. Depending on these four scenarios, daily heating degree-days,  $HDD_d$ , is calculated from the base temperature,  $T_b$ , and daily maximum and minimum temperatures,  $T_{max}$  and  $T_{min}$  respectively, using Equation 8.

$$HDD_d = \begin{cases} T_b - 0.5(T_{max} + T_{min}), & T_{max} \leq T_b; \\ 0.5(T_b - T_{min}) - 0.25(T_{max} - T_b), & T_{min} < T_b; \text{ and } (T_{max} - T_b) < (T_b - T_{min}); \\ 0.25(T_b - T_{min}), & T_{max} > T_b; \text{ and } (T_{max} - T_b) > (T_b - T_{min}); \\ 0, & T_{min} \geq T_b. \end{cases} \quad (8)$$

Daily cooling degree-days,  $CDD_d$ , can be calculated using the same parameters:

$$CDD_d = \begin{cases} 0.5(T_{max} + T_{min}) - T_b, & T_{min} \geq T_b; \\ 0.5(T_{max} - T_b) - 0.25(T_b - T_{min}), & T_{max} > T_b; \text{ and } (T_{max} - T_b) > (T_b - T_{min}); \\ 0.25(T_{max} - T_b), & T_{min} < T_b; \text{ and } (T_{max} - T_b) < (T_b - T_{min}); \\ 0, & T_{max} \leq T_b. \end{cases} \quad (9)$$

The coefficients of 0.5 and 0.25 in Equations 8 and 9 are originally determined by trial and error. The accuracy of calculations using these coefficients has been found to vary from location to location and there have been suggestions in favor of location specific values. A study on the effect of coefficients on the accuracy of these equations was conducted by Day and Karayiannis [17].

## 2.4. Schoenau-Kehrig monthly mean temperature method

The method was developed by Schoenau and Kehrig [16] and is one of the widely adopted method. The formula for calculating monthly heating degree-days,  $HDD_m$  to base  $T_b$  is given by:

$$HDD_m = NS_d [Z_b F(Z_b) + f(Z_b)] \quad (10)$$

where  $N$  is the number of days in the month,  $Z_b$  is the difference between base temperature,  $T_b$  and monthly mean temperature,  $T_m$ , normalized by the standard deviation (SD) of the daily mean temperature,  $S_d$ , which can be found from the following expression:

$$Z_b = \frac{T_b - T_m}{S_d} \quad (11)$$

Function  $f$  is the Gaussian probability density function with mean 0 and SD 1, and function  $F$  is the equivalent cumulative normal probability function:

$$f(Z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{Z^2}{2}\right) \quad (12)$$

$$F(Z) = \int_{-\infty}^Z f(z) dz \quad (13)$$

Monthly cooling degree-days,  $CDD_m$  are calculated using Equation 14, which is similar to the method for calculating heating degree-days, described in Equation 10.

$$CDD_m = N S_d [Z_b F(Z_b) + f(Z_b)] \quad (14)$$

However,  $Z_b$  for cooling degree-days has the following from:

$$Z_b = \frac{T_m - T_b}{S_d} \quad (15)$$

### 2.5. Summary of calculation procedures

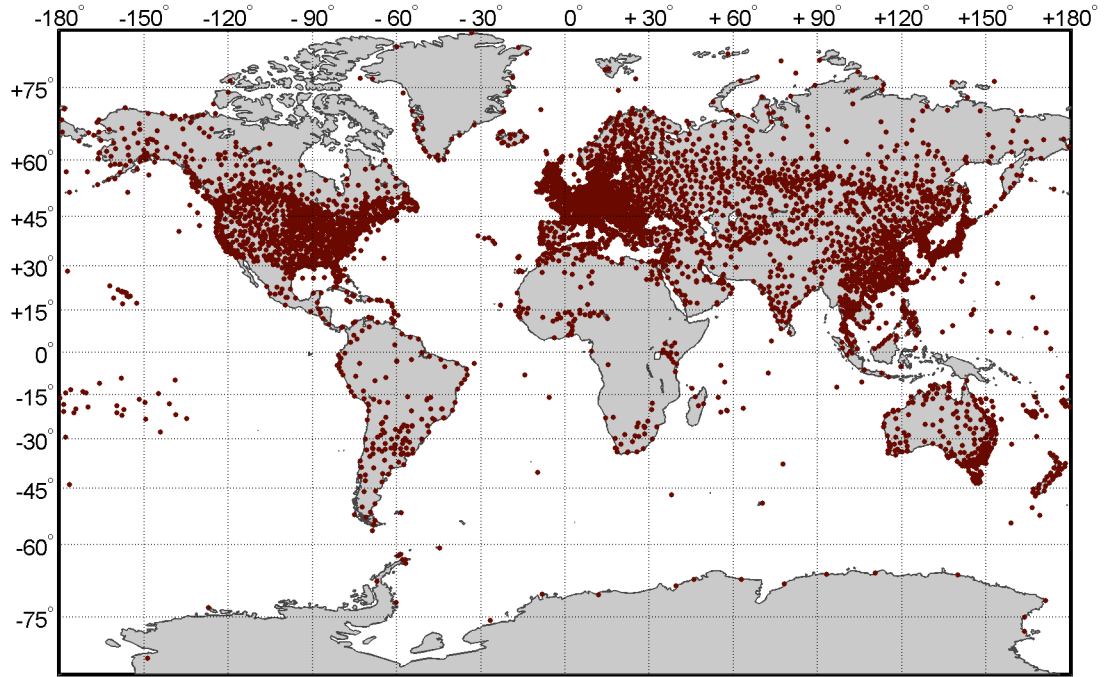
All three methods for calculating degree-days from a reduced dataset involve daily readings of temperature in some form. Both ASHRAE and UKMO equations require daily maximum and minimum temperatures. On the other hand, Shoenu-Kehrig method requires daily mean temperature to calculate monthly standard deviation. For a number of meteorological stations worldwide, in particular the regions that are underrepresented in global climatological database, daily mean temperature data are not readily available. The situation worsens for historical climate data because of the prevalence of manual observations and recordings of data. Although, daily mean can be calculated from daily maximum and minimum temperatures (using Equation 7), there is an underlying assumption that diurnal distribution of temperature is normal. There is, therefore, a need for further development of a calculation procedure from annual mean temperature.

## 3. Methods

### 3.1. Data source

Degree-days and temperature data from ASHRAE are used in this study [2]. The dataset is part of the ‘2009 ASHRAE Handbook – Fundamentals’ and often referred to as the “HOF’09” dataset. The dataset reportedly contains climatic design conditions for 5564 meteorological stations, worldwide. Closer inspection reveals that some of the stations such as Canadian stations from Nova Scotia, Northwest Territories, etc. are reported twice. Duplicate stations are included only once in this study. In total, 5511 unique stations are considered, locations of which are presented as points in Figure 1. There are 1085 stations from the United States, 427 from Canada<sup>2</sup> and 3999 from the rest of world. The concentration of stations is higher in North America. The reporting stations in the HOF’09 dataset are widely distributed and representative of major climatic zones [18]; hence, suitable for a global analysis of the relationship between annual mean temperature and degree-days. Further details on the dataset and underlying methodology are discussed in [2, 19, 20].

<sup>2</sup>The number of Canadian stations in HOF’09 dataset was 480, of which 53 were duplicates.



**Figure 1:** Location map of 5511 reporting stations.

ASHRAE publishes heating and cooling degree-days for two base temperatures: 18.3°C (65°F) and 10°C (50°F) respectively, as they are the most commonly used bases for building energy applications. Degree-days to other base temperatures can be calculated using the Schoenau-Kehrig procedure, described in Equations 10 to 15. In this research,  $HDD_{18.3}$  and  $CDD_{10}$  are used for exploring the relationship between annual mean temperature and degree-days.

### 3.2. Modeling

Multiple regression has been selected as the method for building predictive models to investigate the relationship between annual degree-days and climate parameters, mainly annual mean temperature. This research employed the stepwise procedure, which is a popular and highly effective method for building regression models [21]. According to Draper and Smith [22], stepwise regression is one of the best approaches for variable selection. In stepwise regression, independent variables are successively added or removed based on partial  $F$ -statistic. It begins with no explanatory variable in the model and sequentially adds a variable according to entry/removal criteria. At each step, a variable whose partial  $F$ -statistic yields the smallest  $p$ -value is added to the model. The procedure stops when the addition of any of the remaining variables yields a partial  $p$ -value greater than a specific maximum value. The difference between stepwise and forward selection procedures is that the stepwise method also permits the elimination of variables that become statistically non-significant.

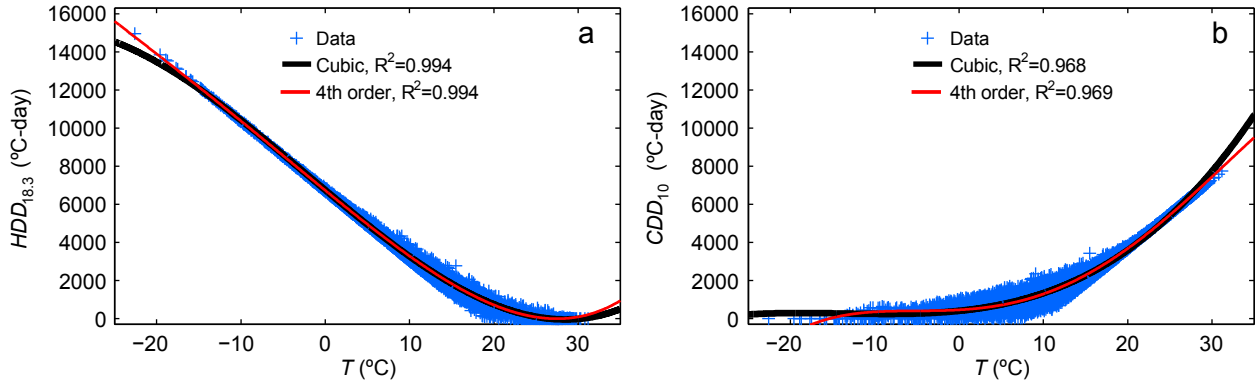
In this study, separate models are constructed using annual heating degree-days,  $HDD$  (°C-day) and cooling degree-days  $CDD$  (°C-day) as dependent variables. Annual mean temperature,  $T$  (°C), standard deviation of monthly mean temperatures  $S_m$  (°C), elevation,  $E$  (m) and latitude,  $L$  (°) of the location are chosen as independent variables. The choice of variables is mostly based on the availability of data. Annual mean temperature, latitude and elevation are available from a variety of sources. It is relatively easy to calculate  $S_m$  of a particular location from monthly mean temperatures using Equation 16.

$$S_m = \left( \frac{\sum (T_{m,i} - \bar{T}_m)^2}{n - 1} \right)^{0.5} \quad (16)$$

where,  $T_{m,i}$  is mean temperature of the  $i$ -th month,  $\bar{T}_m$  is the mean of all monthly mean temperatures of the year and  $n$  (=12) is number of months in a year.

Initial explorations of the data indicated that there is a non-linear association between  $T$  and degree-days, as shown in Figure 2. 4<sup>th</sup> order polynomials best describe the relationships, with  $R^2$  estimates of 0.994 and 0.969 for  $HDD_{18.3}$  and  $CDD_{10}$  respectively. Non-linear regression is, therefore, considered to be the most appropriate. A simple polynomial form of the model, as shown in Equation 17, is used. This is primarily because that  $R^2$  estimates for one independent variable ( $T$ ) is quite high (>0.95), indicating the importance of  $T$  in explaining the variance. Moreover, by defining different orders of  $T$  as independent variables, it is possible to convert the non-linear relationship to a simpler linear model, the solution and interpretation of which is relatively straightforward.

$$Y_i = b_0 + b_1 X_i + b_2 X_i^2 + \dots + b_n X_i^n + \varepsilon_i \quad (17)$$



**Figure 2:** Cubic and 4<sup>th</sup> order curve-fit of  $T$  vs. degree-days for 5511 locations. (a)  $HDD_{18.3}$  and (b)  $CDD_{10}$ .

Statistical analyses are carried out using IBM PASW Statistics 18.0 for Mac OS X [23]. The samples are tested for normal distribution.  $T$  is normally distributed while the remaining variables are not. The entry or removal criteria for stepwise regression is the probability of  $F$ . The probability of  $F$  to enter is  $\leq .050$  and the probability of  $F$  to remove is  $\geq .100$ .

## 4. Results and discussion

### 4.1. Heating degree-days

Results from the stepwise regression for heating degree-days at 18.3°C base temperature,  $HDD_{18.3}$  are given in Tables 1 and 2. Table 1 summarizes six models with change statistics such as  $R^2$ ,  $F$ , degrees of freedom and significance. In all cases,  $R^2$  and adjusted  $R^2$  values are same and therefore only  $R^2$  values are used for interpretation of results. Of note is the order of the models.  $T$  appears as an important predictor (standardized coefficient,  $\beta = -.975$ ) in model #1 with  $R^2 = .952$ . The next model (#2) with  $T^3$  improves the  $R^2$  statistic. The change in  $R^2$  is .042, resulting in the overall  $R^2 = .993$ . The next improvement comes in model 3, when  $S_m$  is added, with an overall  $R^2 = .997$ . For the next three models (#4-6), changes in  $F$ -statistic is progressively smaller. However, model 5 ( $R^2 = .997$ ) incorporates lower and higher orders of  $T$  and reduces standard error to 123.419. Although, there is no change in  $R^2$  from the previous one, model 5 agrees with the 4<sup>th</sup> order fit in scatter plot (Figure 2a). The addition of latitude



in the regression model reduces the standard error, but the change is quite small. There is no change in  $R^2$  from model #3 upwards.

$\beta$  and  $t$ -statistics of explanatory variables demonstrate that first order  $T$  has the highest degree of importance in the model, apart from the constant term,  $b_0$  (see Table 2).  $T$  is negatively associated in all models, indicating that an increase in  $T$  will result in a corresponding decrease in  $HDD_{18.3}$ , which agrees with conventional wisdom. All predictor variables are significant at .000, indicating their importance in explaining the variance.

**Table 1:** Summary of stepwise regression for  $HDD_{18.3}$ .

Model	Predictors	$R$	$R^2$	Std. error	Change Statistics				
					$R^2$	$F$	df1	df2	Sig.
1	$b_0, \bar{T}$	.975	.952	532.411	.952	108170.063	1	5509	.000
2	$b_0, \bar{T}, \bar{T}^3$	.997	.993	195.290	.042	35437.687	1	5508	.000
3	$b_0, \bar{T}, \bar{T}^3, S_m$	.998	.997	141.845	.003	4933.577	1	5507	.000
4	$b_0, \bar{T}, \bar{T}^3, S_m, \bar{T}^4$	.999	.997	128.307	.001	1224.456	1	5506	.000
5	$b_0, \bar{T}, \bar{T}^3, S_m, \bar{T}^4, \bar{T}^2$	.999	.997	123.419	.000	445.777	1	5505	.000
6	$b_0, \bar{T}, \bar{T}^3, S_m, \bar{T}^4, \bar{T}^2, L$	.999	.997	122.510	.000	82.962	1	5504	.000

#### 4.2. Cooling degree-days

Different models of stepwise regression for cooling degree-days at 10°C base temperature ( $CDD_{10}$ ) are summarized in Table 3. The key difference with  $HDD_{18.3}$  models is the presence of 2<sup>nd</sup> order  $T$  in the first and subsequent models for  $CDD_{10}$ , compared to 1<sup>st</sup> order  $T$  in different models for  $HDD_{18.3}$ .  $T^2$  has the highest degree of importance in predicting  $CDD_{10}$ . However, the introduction of  $T^3$  in model #4 improves the  $R^2$  to a large extent. The scatter plot in Figure 2b and the corresponding cubic curve fit illustrates the importance of  $T^3$ . The shape of curves for both the cubic and 4<sup>th</sup> order fits are almost same, except at the tail-ends for extremes of  $T$ . There are gradual improvements in  $R$  and  $R^2$  statistics for models #1-5. The initial  $R^2$  in model #1 is .952, increasing to .994 in model #6 that incorporates  $T$ ,  $T^2$ ,  $T^3$ ,  $T^4$ ,  $L$ ,  $S_m$  and  $b_0$ . The addition of  $T^4$  in model #6 does little to improve the  $F$ -statistic (=22.236). There appears to be no change in  $R^2$  values up to three decimal points. Change in standard error is also small (=267). However the significance level is .000, which indicates that the effect of  $T^4$  should not be ignored.

Regression coefficients for  $CDD_{10}$  models are given in Table 4. Standardized coefficient,  $\beta$  for  $T^2$  is the highest among the independent variables in models #1-6. For models #1-3,  $\beta$  for  $T^2$  is between .785 and .976. In the rest of the models (#4-6),  $\beta$  for  $T^2$  ranges between .468 and .492, and is more than the remaining predictor variables but its influence is smaller. Standard deviation of monthly mean temperature,  $S_m$  has greater influence in  $CDD_{10}$  models, compared to the  $HDD_{18.3}$  models.  $\beta$  for  $S_m$  in  $CDD_{10}$  models varies between .227 and .234 whereas in  $HDD_{18.3}$  models the range is .079–.087. The greater influence of  $S_m$  may be due to the greater variance in mid-ranges of  $T$ .  $T$ ,  $T^2$ ,  $T^3$  and  $S_m$  are positively associated with  $CDD_{10}$ , indicating that with an increase in their values, there is a corresponding increase in  $CDD_{10}$ . The constant term,  $b_0$ , as well as  $T^4$  and  $L$  are negatively associated with  $CDD_{10}$ .

**Table 2:** Regression coefficients for  $HDD_{18.3}$  models.

Model	Predictors	Unstandardized coefficients		Standardized coefficients, $\beta$	$t$	Sig.
		$b$	Std. error			
1	$b_0$	6361.474	11.730		542.336	.000
	$\bar{T}$	-274.492	.835	-.975	-328.892	.000
2	$b_0$	6783.879	4.852		1398.025	.000
	$\bar{T}$	-366.805	.578	-1.304	-634.514	.000
	$\bar{T}^3$	.158	.001	.387	188.249	.000
3	$b_0$	6211.852	8.874		700.013	.000
	$\bar{T}$	-350.743	.478	-1.246	-733.589	.000
	$\bar{T}^3$	.158	.001	.386	258.451	.000
	$S_m$	51.711	.736	.079	70.239	.000
4	$b_0$	6104.809	8.590		710.681	.000
	$\bar{T}$	-332.197	.684	-1.181	-485.615	.000
	$\bar{T}^3$	.040	.003	.097	11.665	.000
	$S_m$	56.685	.681	.087	83.245	.000
	$\bar{T}^4$	.004	.000	.240	34.992	.000
5	$b_0$	6082.231	8.332		730.007	.000
	$\bar{T}$	-337.004	.696	-1.198	-483.999	.000
	$\bar{T}^3$	.025	.003	.062	7.500	.000
	$S_m$	56.602	.655	.086	86.413	.000
	$\bar{T}^4$	.003	.000	.221	33.078	.000
	$\bar{T}^2$	.801	.038	.071	21.113	.000
6	$b_0$	6059.131	8.651		700.436	.000
	$\bar{T}$	-337.420	.693	-1.199	-487.127	.000
	$\bar{T}^3$	.026	.003	.064	7.815	.000
	$S_m$	54.913	.676	.084	81.218	.000
	$\bar{T}^4$	.003	.000	.214	32.126	.000
	$\bar{T}^2$	.896	.039	.080	22.929	.000
	$L$	.756	.083	.008	9.108	.000

#### 4.3. Equations for heating and cooling degree-days

Given the regression coefficients for  $HDD_{18.3}$  and  $CDD_{10}$  in Tables 2 and 4 respectively, the relationship between degree-days and the independent variables in model #6 can be described as Equations 18 and 19.

$$HDD_{18.3} = 6059.131 - 337.420T + .896T^2 + .026T^3 + .003T^4 + 54.913S_m + .756L \quad (18)$$

$$CDD_{10} = -718.905 + 74.871T + 3.823T^2 + .109T^3 - .001T^4 + 109.959S_m - 1.585L \quad (19)$$

where,  $HDD_{18.3}$  is annual heating degree-days at 18.3°C (°C-day),  $CDD_{10}$  is annual cooling degree-days at 10°C (°C-day),  $T$  is annual mean temperature (°C),  $S_m$  is standard deviation of monthly mean temperatures (°C) and  $L$  is the latitude of the location in decimal degree (°) where north latitudes are positive and south latitudes are negative.

Errors in Equations 18 and 19 tend to occur at lower tail ends of  $HDD_{18.3}$  and  $CDD_{10}$  respectively. This is primarily because of points of inflection in 4<sup>th</sup> order curves in Figure 2. Closer inspection of the original data reveals that degree-days around the inflection points are very low – either zero or very close to zero. The inflections result in over estimates of  $HDD_{18.3}$  and under estimates of  $CDD_{10}$  at very high and very low temperatures respectively. Simple corrections are made to Equations 18 and 19 around the points of inflection ( $\approx 25^\circ\text{C}$  for  $HDD_{18.3}$  and  $-15^\circ\text{C}$  for  $CDD_{10}$ ) to overcome this issue. Modified relation-

**Table 3:** Summary of stepwise regression for  $CDD_{10}$ .

Model	Predictors	$R$	$R^2$	Std. error	Change Statistics				
					$R^2$	$F$	df1	df2	Sig.
1	$b_0, \bar{T}^2$	.976	.952	383.880	.952	108406.754	1	5509	.000
2	$b_0, \bar{T}^2, \bar{T}$	.981	.963	336.491	.011	1661.966	1	5508	.000
3	$b_0, \bar{T}^2, \bar{T}, S_m$	.994	.989	183.777	.026	12958.410	1	5507	.000
4	$b_0, \bar{T}^2, \bar{T}, S_m, \bar{T}^3$	.997	.993	142.562	.004	3645.409	1	5506	.000
5	$b_0, \bar{T}^2, \bar{T}, S_m, \bar{T}^3, L$	.997	.994	138.778	.000	305.413	1	5505	.000
6	$b_0, \bar{T}^2, \bar{T}, S_m, \bar{T}^3, L, \bar{T}^4$	.997	.994	138.511	.000	22.236	1	5504	.000

ships are given in Equations 20 and 21.

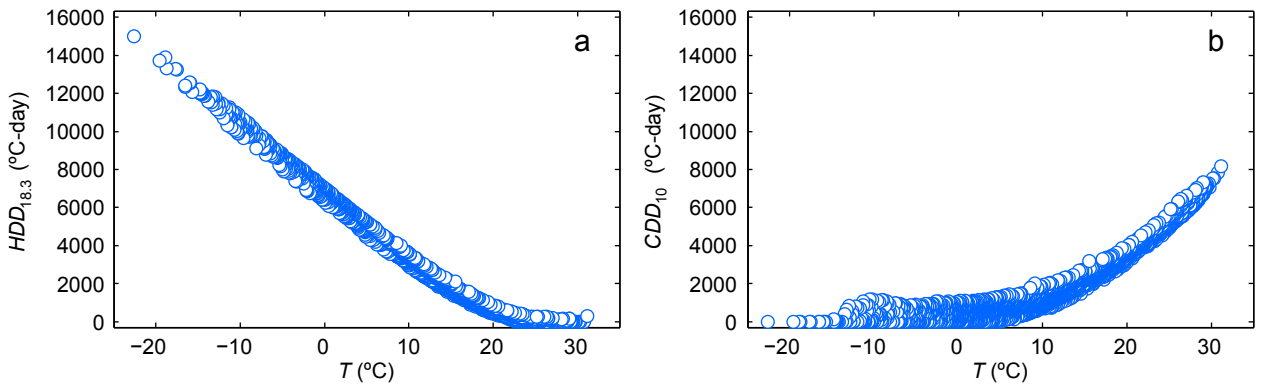
$$HDD_{18.3} = \begin{cases} 0, & T \geq 25; \text{ and } S_m > 4; \\ (6059.131 - 337.420T + .896T^2 + .026T^3 + .003T^4 + 54.913S_m + .756L)^+, & T < 25. \end{cases} \quad (20)$$

$$CDD_{10} = \begin{cases} 0, & T \leq -15; \\ (-718.905 + 74.871T + 3.823T^2 + .109T^3 - .001T^4 + 109.959S_m - 1.585L)^+, & T > -15. \end{cases} \quad (21)$$

Dividing the equation into temperature ranges and taking only positive values of degree-days improves  $R^2$  statistic but only marginally.  $R^2$  increases were .0005 and .0007 for  $HDD_{18.3}$  and  $CDD_{10}$  respectively. Few locations are affected by the conditional relationship on temperature and standard deviation; 11 for ( $T \geq -15^\circ\text{C}$ ) and 67 for ( $T \geq 25$ ; and  $S_m > 4$ ).

#### 4.4. Validation and accuracy of derived relationships

Calculated degree-days for all 5511 locations using Equations 20 and 21 are illustrated in Figure 3. Note the similarities between scatter plots in Figures 2 and 3, in particular in the shape and characteristics – illustrating the effectiveness and accuracy of the derived relationships. Standard deviation of error for Equations 20 and 21 are 120 and 135  $^\circ\text{C-day}$  respectively, which represents .328 and .370  $^\circ\text{C-day per day}$  for  $HDD_{18.3}$  and  $CDD_{10}$  respectively. The estimates can be said to be highly accurate and well within the range of typical measurement errors.



**Figure 3:** Heating and cooling degree-days calculated using the derived relationships. (a)  $HDD_{18.3}$  and (b)  $CDD_{10}$ .

**Table 4:** Regression coefficients for  $CDD_{10}$  models.

Model	Predictors	Unstandardized coefficients		Standardized coefficients, $\beta$	$t$	Sig.
		$b$	Std. error			
1	$b_0$	461.552	7.014		65.808	.000
	$\bar{T}^2$	7.898	.024	.976	329.252	.000
2	$b_0$	273.925	7.680		35.669	.000
	$\bar{T}^2$	6.356	.043	.785	146.854	.000
	$\bar{T}$	44.262	1.086	.218	40.767	.000
3	$b_0$	-924.599	11.333		-81.582	.000
	$\bar{T}^2$	6.358	.024	.785	268.976	.000
	$\bar{T}$	77.408	.661	.381	117.176	.000
	$S_m$	108.576	.954	.230	113.835	.000
4	$b_0$	-787.818	9.079		-86.775	.000
	$\bar{T}^2$	3.983	.043	.492	91.754	.000
	$\bar{T}$	77.975	.513	.384	152.133	.000
	$S_m$	107.426	.740	.227	145.142	.000
	$\bar{T}^3$	.088	.001	.297	60.377	.000
5	$b_0$	-731.968	9.398		-77.886	.000
	$\bar{T}^2$	3.789	.044	.468	86.731	.000
	$\bar{T}$	77.720	.499	.383	155.705	.000
	$S_m$	110.782	.746	.234	148.571	.000
	$\bar{T}^3$	.093	.001	.313	64.220	.000
6	$L$	-1.633	.093	-.024	-17.476	.000
	$b_0$	-718.905	9.780		-73.505	.000
	$\bar{T}^2$	3.823	.044	.472	86.506	.000
	$\bar{T}$	74.871	.783	.369	95.603	.000
	$S_m$	109.959	.764	.233	143.846	.000
	$\bar{T}^3$	.109	.004	.369	28.895	.000
	$L$	-1.585	.094	-.023	-16.889	.000
	$\bar{T}^4$	-.001	.000	-.049	-4.715	.000

The highest  $R^2$  statistic for  $CDD_{10}$  models is .994; in contrast the highest  $R^2$  statistic for  $HDD_{18.3}$  models is .997. The resulting regression fits, therefore, describe the variance quite well; i.e., 99.7% and 99.4% for  $HDD_{18.3}$  and  $CDD_{10}$  respectively. Final models (#6) in both regression experiments have strong predictive power, when the distribution and coverage of the data are considered. 4<sup>th</sup> order non-linear relationship appears to be more appropriate than the 3<sup>rd</sup> order or lower. 4<sup>th</sup> order curves match the shape of the data at tail ends (see Figure 2), at lower and higher temperatures for  $HDD_{18.3}$  and  $CDD_{10}$  respectively – resulting in reduced errors for higher degree-days on both occasions. What sets apart this research from previous studies is that only three predictor variables ( $T$ ,  $S_m$  and,  $L$ ) are required to determine annual degree-days and the identified relationship is based on a wider coverage of locations. For example, Schoenau-Kehrig’s paper was based on four locations only, compared to 5511 in this research. If locations were regionalized the  $R^2$  estimates could have been or very close to unity. However, the purpose of this research was to find a globally applicable relationship, not regionalized equations.

## 5. Conclusion

Degree-days is an important climatic design indicator and is widely used in the design and operation of energy-efficient buildings. The calculation of degree-days has always been a topic of focus by researchers. Attempts have been made in the past to calculate degree-days from datasets of reduced temporal resolution, the lowest being the daily mean temperature. This research went further and inves-

tigated the relationship between annual mean temperature and annual heating and cooling degree-days using multiple non-linear regression. Developed equations are able to predict annual degree-days from one reading of annual mean temperature, standard deviation of monthly mean temperature and signed decimal latitude of the location. The predictions are highly accurate with standard errors of .328 and .370 °C-day per day for heating and cooling degree-days respectively.

The relationship investigated in this research and the developed equations will open up new avenues of research in climate and built environment. To begin with, it will now be possible to calculate historical annual degree-days from reduced temperature observations for many locations in the world, for which only monthly mean temperature data are available. Such historical analyses will enable us to understand the impact of observed changes in climate. Other immediate applications will be the investigations into the impact of the projected climate change. Long term monthly projections of temperature from global circulation models (GCM) can now be transformed into degree-days to investigate the severity and duration of temperature that have a bearing on heating and cooling energy consumption in buildings.

Other directions for future research can include the extension of this work to include the relationship between annual mean temperature and degree-days at variable base temperatures. The method can also be assessed against other reduced dataset methods such as UKMO, ASHRAE and Schoenau-Kehrig for accuracy, ease of use and potential for application in geographical regions, underrepresented in global climatological databases.

## Acknowledgements

This research was funded by the British Council through the International Strategic Partnerships in Research and Education (INSPIRE) program. Grant reference: 10/SP150.

## References

- [1] CIBSE, Degree-days: Theory and application, TM41, Chartered Institution of Building Services Engineers, London, UK, 2006.
- [2] ASHRAE, ASHRAE Handbook: Fundamentals, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, 2009.
- [3] A. R. Day, T. G. Karayiannis, A new degree-day model for estimating energy demand in buildings, *Building Services Engineering Research and Technology* 20 (4) (1999) 173–178.
- [4] A. R. Day, An improved use of cooling degree-days for analysing chiller energy consumption in buildings, *Building Services Engineering Research and Technology* 26 (2) (2005) 115–127.
- [5] Carbon Trust, Degree days for energy management – A practical introduction, Carbon Trust, London, UK, 2007.
- [6] A. R. Day, T. G. Karayiannis, Identification of the uncertainties in degree-day-based energy estimates, *Building Services Engineering Research and Technology* 20 (4) (1999) 165–172.
- [7] M. Christenson, H. Manz, D. Gyalistras, Climate warming impact on degree-days and building energy demand in Switzerland, *Energy Conversion and Management* 47 (6) (2006) 671–686.
- [8] CIBSE, Climate change and the indoor environment: impacts and adaptation, TM36, Chartered Institution of Building Services Engineers, London, UK, 2005.
- [9] M. Mourshed, The impact of the projected changes in temperature on heating and cooling requirements in buildings in Dhaka, Bangladesh, *Applied Energy* 88 (11) (2011) 3737–3746.
- [10] IPCC, Summary for policymakers, in: *Climate Change 2007: Synthesis Report: Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, New York, NY, 2007, pp. 1–22.
- [11] C. Tebaldi, R. Knutti, The use of the multi-model ensemble in probabilistic climate projections, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 365 (1857) (2007) 2053–2075.
- [12] S. E. Belcher, J. N. Hacker, D. S. Powell, Constructing design weather data for future climates, *Building Services Engineering Research and Technology* 26 (1) (2005) 49–61.
- [13] H. C. S. Thom, Normal degree days above any base by the universal truncation coefficient, *Monthly Weather Review* 94 (7) (1966) 461–465.
- [14] D. Erbs, W. Beckman, S. Klein, Estimation of degree-days and ambient temperature bin data from monthly-average temperatures, *ASHRAE Journal* 25 (6) (1983) 60–65.
- [15] E. R. Hitchin, Estimating monthly degree-days, *Building Services Engineering Research and Technology* 4 (4) (1983) 159–162.
- [16] G. J. Schoenau, R. A. Kehrig, Method for calculating degree-days to any base temperature, *Energy and Buildings* 14 (4) (1990) 299–302.
- [17] A. R. Day, T. G. Karayiannis, Degree-days: Comparison of calculation methods, *Building Services Engineering Research and Technology* 19 (1) (1998) 7–13.

- [18] ASHRAE, Energy Standard for Buildings Except Low-Rise Residential Buildings, American Society of Heating Refrigeration and Air-Conditioning Engineers, Atlanta, GA, 2007.
- [19] D. Thevenard, R. Humphries, The calculation of climatic design conditions in the 2005 ASHRAE Handbook – Fundamentals, ASHRAE Transactions 111 (1) (2005) 457–466.
- [20] D. Thevenard, Updating the ASHRAE climatic data for design and standards (RP-1453), American Society of Heating Refrigeration and Air-Conditioning Engineers, Atlanta, GA, 2009.
- [21] J. Davidson, D. Savic, G. Walters, Symbolic and numerical regression: Experiments and applications, in: R. John, R. Birkenhead (Eds.), Developments in soft computing, Physica-Verlag, Heidelberg, Germany, 2001, pp. 175–182.
- [22] N. Draper, H. Smith, Applied regression analysis, John Wiley and Sons, New York, NY, 1998.
- [23] IBM-SPSS, Predictive Analytics Software (PASW) Statistics v 18.0 [computer software], International Business Machines (IBM) SPSS, Somers, NY, 2010.