A climate zone approach to global solar radiation modelling using artificial neural networks

Liu Yang 1, Xujie Huo 1, Danny H W Li 2, Joseph C Lam 2
1 School of Architecture, Xi’an University of Architecture and Technology, Shaanxi 710055, China
2 Building Energy Research Group, Department of Architecture and Civil Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR, China
E-mail: hxujie2002@gmail.com

Abstract. Information on solar availability is crucial in the study of both passive building designs and active solar energy systems. There are still many locations within different climate zones worldwide that do not have solar radiation measurements. Correlation between solar radiation and the more commonly measured meteorological variables such as temperature and sunshine hours is useful for locations with no measured solar radiation data. This is also useful for locations with measured solar radiation data, in that any missing data due to equipment breakdown or malfunction can be modelled. Global solar radiation (GSR) was modelled for 96 cities in different climate zones across China using artificial neural networks (ANNs). The novelty of this study is the climate zone approach, by which locations with similar climates were modelled together. Climate classification was based on both the traditional thermal climates and the solar climates. Two sets of models were developed based on measured diurnal temperatures and sunshine hours. Model performance in terms of the predictive power of ANN solar radiation models was evaluated through the Nash-Sutcliffe efficiency coefficient (NSEC). Error analysis of the predicted solar radiation as compared with the measured data was also conducted for each of the 96 cities.

Keywords: Energy use; Solar radiation; Solar and thermal climate zones; Artificial neural networks; temperatures and sunshine hours

1. Introduction
Recently there is a growing awareness of the likely rise in energy consumption in the built environment due to climate change [1] and an increase in the desire for better thermal comfort through space conditioning [2]. These have resulted in more attention being paid to the development and application of renewable technologies when formulating energy and environmental policies. For instance, the Energy Performance of Buildings Directive requires all new buildings to be “nearly zero energy buildings” by the end of 2020 in EU countries, in which energy-efficient building designs and solar energy applications (in terms of photovoltaic (PV) and solar thermal) tend to play an important role in the development of zero energy buildings. Information on solar availability is essential in the assessment of the potential of any solar applications [3]; whether it is related to passive solar building designs, PV/building integrated photovoltaic (BIPV), etc. Because of the costs and regular maintenance (including calibration) of the equipment, long and sustained records of solar radiation data are less readily available compared with other meteorological variables such as temperatures, cloud cover and sunshine hours. As a result, a number of empirical models have been developed to estimate the global solar radiation using the more readily available measured weather data.
One of the most widely used climatic parameters to estimate global solar radiation (GSR) is the possible bright sunshine. Compared with solar radiation measurements, sunshine data tends to be simpler and more readily available from measuring stations in many locations worldwide. A number of researchers had employed regression techniques to correlate global solar radiation with the corresponding possible bright sunshine [4-7] These studies were largely based on linear and/or non-linear regression techniques (e.g. the two-parameter Angstrom-Prescott regression model [8, 9]). Subsequent to these earlier works, there had been two new developments. First, an alternative modelling approach based on artificial neural networks (ANNs) was used in estimating solar radiation for locations with different latitudes and climates [10, 11]. In general, the results indicated an improvement in solar radiation modelling accuracy compared with the conventional regression techniques. Second, temperature data (especially daily minimum and daily maximum) in general are much more readily available than both solar radiation and sunshine hours, particularly in developing countries/regions. Recently, it had been shown that daily global solar radiation correlated reasonable well with daily diurnal temperature for some locations [12]. This work was based on the much earlier work on solar radiation modelling for agricultural and forest meteorology applications. The general approach was to express the daily global solar radiation on a horizontal surface as a function of the corresponding diurnal temperature and several empirical coefficients. These empirical coefficients were then calibrated or determined for a particular location from measured global solar radiation data using regression technique and/or the ordinary least square method by comparing the observed and estimated data.

In the present work, we have developed solar radiation models for different climate zones across China using ANN technique. Both the much more readily available meteorological data (i.e. daily minimum and maximum temperatures) and sunshine duration were considered. Most of the previous modelling works were individual-location-based (i.e. one model for each location using solar radiation data measured at that location). Our work is climate-zone-based. In other words, measured solar radiation data from all the locations with the same climate zones were considered in the ANN modelling. The aim was to develop one solar radiation estimation model for each of the major climate zones, so that any location with no solar radiation measurement but lies within the same climate zone could use the same model to assess its solar potential. This is an extension of our earlier work on regression and ANN solar radiation models for major climate zones in China [10]. Model performances of the temperature-based and sunshine-based ANN models were compared to get a better understanding of the differences in modelling accuracy between these two types of solar radiation models. Comparative studies were conducted for each major climate zone as well as individual cities.

2. **Climate classifications and major climate zones**

In China, for the thermal design of buildings, the national design guidelines stipulate five major climate types - severe cold (SC), cold(C), hot summer and cold winter (HSCW), mild(M), and hot summer and warm winter (HSWW) [13]. Figure 1 shows an overall layout of the nine thermal climate zones and sub-zones. For assessing solar potentials among different locations, a clearness index (K) was used to describe the prevailing solar climates in China (details can be found in Ref. [14]). Five major solar climates were identified with annual mean monthly-average-daily clearness index ranging from 0.3 in the Sichuan Basin to 0.65 in the west and north-west. Both Zone II and Zone III have two regions – Basin/Plateau and Plain/Plateau, respectively. Figure 2 shows the seven solar climate zones identified.

3. **Data gathering and quality control**

In this study, daily mean, minimum and maximum temperatures, relative humidity, GSR and sunshine duration measured at 96 stations in different climate zones across China were gathered and analysed. The period of records ranged from 5 to 55 years covering the period between 1958 and 2012. The measured GSR and sunshine duration data were checked for errors and inconsistencies. This quality control was to eliminate spurious data and inaccurate measurements [15].
4. Artificial neural networks

A total of six geographical and climatic variables were used as the input parameters for the input nodes of the input layer. Both the diurnal temperature model and the sunshine duration model have four geographical variables (i.e. the day-number, latitude, longitude and altitude). The climatic variables for diurnal temperature model were relative daily mean humidity and daily diurnal temperature. Daily mean temperature and normalized daily sunshine duration (i.e. measured sunshine duration ÷ daily maximum possible sunshine duration) were adopted for the sunshine duration model.

There was one single node at the output layer with the estimated daily GSR as the output. The transfer function adopted for the neurons was radial basis function.

In this study, 70% and 30% of the data were used for training and testing, respectively. ANNs were trained and two corresponding GSR prediction models (one using diurnal temperature and the other sunshine duration) were developed for each of the nine thermal climate zones and sub-zones. Likewise, two ANN GSR prediction models (one using diurnal temperature and the other sunshine hours) were developed for each of the seven solar zones and sub-zones.

To get a general idea about the predictive power of the ANN models developed, the Nash-Sutcliffe efficiency coefficient (NSEC) was determined. A summary for the thermal and solar model climate zones is shown in Tables 1, respectively. It can be seen, in Table 1 (thermal zone models), that NSEC for the ANN models based on temperature varies from 0.7 in SC-II climates to 0.81 in SC-I with a zone average of 0.76. NSEC based on sunshine duration tends to be larger, ranging from 0.83 also in SC-II to 0.94 in SC-I with a zone average of 0.89. Similarly, in the solar climate zones, NSEC for the ANN models based on diurnal temperature varies from 0.7 in solar zone IIIB to 0.82 in solar zone V with a zone average of 0.76. Again, NSEC based on sunshine duration tends to be larger, ranging from 0.83 in Zone IIIB to 0.92 in Zone IIA with a zone average of 0.89. It suggests that sunshine-based ANN models in general would have better performance than temperature-based. It is interesting to note that the overall mean NSEC (i.e. average NSEC of the climate zones) is the same for the thermal and solar climates, 0.76 for diurnal temperature and 0.89 for sunshine duration. This indicates that there is no significant difference in model performance, in terms of predictive power, whether the ANN solar radiation models are developed using thermal or solar climate classification.

5. Error analysis of the ANN models for individual cities

To have a better understanding of how well the ANN models could predict GSR, performance of the thermal climate zone ANN models based on diurnal temperature was evaluated using measured temperature and solar radiation for the three-year period 2010-2012, which were excluded in the ANN training and testing. To quantify the performance of the ANN models and ascertain whether there was any underlying trend in the different climates, mean bias error (MBE) and root-mean-square error (RMSE) were determined for the 96 cities.

A summary of the MBE and RMSE of the temperature-based ANN models for the 96 cities grouped into the nine thermal climate zones and sub-zones is shown in Table 2. In terms of the actual GSR, Sanya (HSWW) has the largest underestimation of -5.03 MJ/m² and Hami (SC-I) the largest overestimation of 1.94 MJ/m². Most of the MBEs (about two-third) are within the ±5% range and most of the RMSEs (about three-quarter) are less than 25%. In general, cities with warmer climates (i.e. HSCW and HSWW) in the south tended to have larger MBEs and RMSEs. Similarly, an error analysis
of the sunshine-based ANN models was conducted for the 96 cities within the nine thermal zones using sunshine duration, and a summary is shown in Table 3. Similar to the temperature-based ANN models, most of the MBEs (again about two-third) are within the ±5% range. RMSEs, however, are much smaller; nearly all of them (about 97%) are less than 25%.

### Table 1. Nash-Sutcliffe efficiency coefficient (NSEC) for the ANN thermal and solar climate zone models.

| Thermal zone | NSEC | Solar zone | Sunlight hour | NSEC | Diurnal temperature | Sunshine hour |
|--------------|------|------------|---------------|------|---------------------|---------------|
| SC-I         | 0.81 | I          | 0.76          | 0.89 |
| SC-II        | 0.70 | II         | 0.79          | 0.92 |
| SC-III       | 0.78 | IIA        | 0.76          | 0.90 |
| C-I          | 0.80 | IIB        | 0.75          | 0.90 |
| C-II         | 0.71 | IV         | 0.76          | 0.89 |
| C-III        | 0.75 | V          | 0.82          | 0.88 |
| HSCW         | 0.79 | SC-I       | 0.70          | 0.83 |
| Mld          | 0.72 | SC-II      | 0.75          | 0.76 |
| HSWWW        | 0.75 | SC-III     | 0.76          | 0.70 |

Likewise, error analysis results grouped into the seven solar zones for temperature-based and sunshine-based ANN models are shown in Tables 4 and 5, respectively. It can be seen from Table 4 that MBE varies from -30.3% underestimation in Sanya (solar zone IV) to 20.1% overestimation in Gushi (solar zone IIIA). RMSE varies from 12.9% in Lhasa (solar zone I) to 42.9% in Gushi (solar zone IIIA). Table 5 shows that MBE for sunshine-based ANN models ranges from -20.9% in Changning (solar zone IV) to 17.8% in Hami (solar zone I), and RMSE ranges from 7.9% in both Golmud (solar zone I) and Yinchun (solar zone IIB) to 28.7% in Gushi (solar zone IIIA). Again, similar to thermal zone results, about two-third of the MBEs are within the ±5% range for both the temperature-based and sunshine-based models, and about three-quarter of the RMSEs are less than 25% for the temperature-based models and nearly all (97%) for the sunshine-based models.

### 6. Comparisons of results and discussion

#### 6.1. Individual cities

From the error analysis results shown in Tables 4 and 5 for temperature-based models, there is very little difference in both the MBEs and the RMSEs for all the 96 cities. Similar finding has been observed for sunshine-based models shown in Table 6. It suggests that grouping of cities into thermal or solar climate zones makes very little difference in terms of model performance for individual cities. There are, however, differences in the error analysis results between sunshine-based and temperature-based models. As for the MBE, there is no specific trend indicating whether temperature-based or sunshine-based model would perform better. Different cities perform differently. For example, in Altay (SC-I) temperature-based has slightly less underestimation (-6.6%, see Table 2) than sunshine-based (-6.9%, see Table 3), whereas in Tacheng (SC-I) temperature-based has a much larger overestimation (2.7%) compared with sunshine-based (1.5%). In terms of long-term performance, therefore, sunshine-based models do not tend to out-perform temperature-based. This finding does not tend to support the generally held view that sunshine hours is a better meteorological variable than temperature in solar radiation modelling. As for the RMSE, the situation is different. Sunshine-based models tend to have a much smaller RMSEs (mostly 30-50% smaller) than temperature-based. In terms of short-term performance, sunshine-based models would be more appropriate than temperature-based. For both the temperature-based and sunshine-based models, MBEs in most cities (92%) are within ±10% (Tables 4 and 5). This suggests that the ANN models developed should be able to predict long-term (e.g. annually) solar availability to within ±10%. For short-term prediction, however, the variations would be larger. For temperature-based (Table 2), MBE varies from 12.9% in Lhasa (C-II) to 42.5% in Gushi (HSCW). This implies that predicted daily GSR could deviate from the measured data by up to 12.9% in Lhasa and 42.5% in Gushi. Only 16 cities have RMSEs less than 20%. For sunshine-based models, the expected deviations are much smaller. Table 3 shows that RMSE ranges from 7.8% in Yinchuan (C-III) to 28.7% in Gushi (HSCW). Most of the cities (92 out of 96) have RMSEs smaller than 20%.

#### 6.2. Zone averages

Performances of the temperature-based and sunshine-based ANN models were compared into terms of the zone averages. Table 6 shows a summary of the comparison for the nine thermal climate zones and
Table 2. Error analysis of individual cities within the 9 thermal climate zones (temperature-based models).

| Thermal zone | MBE (%) | RSME (%) | Solar zone | MBE (%) | RSME (%) |
|--------------|---------|----------|------------|---------|----------|
|              | Temperature | Sunshine | Temperature | Sunshine | Temperature | Sunshine |
| SC-I         | 2.0     | -1.3     | 23.4       | 15.4    | 0.8       | 1.5       |
| SC-II        | -1.3    | -2.2     | 18.7       | 12.0    | 0.1       | 2.5       |
| SC-III       | -3.4    | -2.2     | 22.3       | 11.2    | -0.4      | -0.2      |
| C-I          | 1.4     | 3.8      | 20.8       | 11.0    | 0.9       | -1.8      |
| C-II         | -1.2    | -0.1     | 18.2       | 11.2    | -3.3      | -0.6      |
| C-III        | 1.3     | -0.3     | 23.3       | 12.3    | 1.6       | -2.8      |
| HSCW         | 4.8     | -1.8     | 28.0       | 13.4    | 0.2       | -5.2      |
| HSWW         | -2.6    | -4.9     | 25.3       | 11.7    |           |           |
| Mild         | -1.5    | -2.4     | 20.7       | 14.0    |           |           |

Table 3. Error analysis of individual cities within the 9 thermal climate zones (sunshine-based models).

Table 4. Error analysis of individual cities within the 7 solar climate zones (temperature-based models).

Table 5. Error analysis of individual cities within the 7 solar climate zones (sunshine-based models).

Table 6. Comparison of error analysis between temperature-based and sunshine-based models (thermal and solar climate zone averages).

| Thermal zone | MBE (%) | RSME (%) | Solar zone | MBE (%) | RSME (%) |
|--------------|---------|----------|------------|---------|----------|
|              | Temperature | Sunshine | Temperature | Sunshine | Temperature | Sunshine |
| SC-I         | 2.0     | -1.3     | 23.4       | 15.4    | 0.8       | 1.5       |
| SC-II        | -1.3    | -2.2     | 18.7       | 12.0    | 0.1       | 2.5       |
| SC-III       | -3.4    | -2.2     | 22.3       | 11.2    | -0.4      | -0.2      |
| C-I          | 1.4     | 3.8      | 20.8       | 11.0    | 0.9       | -1.8      |
| C-II         | -1.2    | -0.1     | 18.2       | 11.2    | -3.3      | -0.6      |
| C-III        | 1.3     | -0.3     | 23.3       | 12.3    | 1.6       | -2.8      |
| HSCW         | 4.8     | -1.8     | 28.0       | 13.4    | 0.2       | -5.2      |
| HSWW         | -2.6    | -4.9     | 25.3       | 11.7    |           |           |
| Mild         | -1.5    | -2.4     | 20.7       | 14.0    |           |           |

As for the MBEs, there is no distinct pattern or trend indicating whether temperature-based models or sunshine-based models would perform better. Similar to individual cities, sunshine-based models perform better than temperature-based. For temperature-based models, MBE ranges from 18.2% in thermal zone C-II to 28% in HSCW, whereas the range for sunshine-based models is from 11% in C-I to 15.4% in SC-I. Similar comparison was conducted for the solar climate zone results, and a summary of the zone-averaged MBEs and RMSEs is shown in Table 6. As for the RMSEs, sunshine-based models have smaller errors (between 30-50% smaller) than temperature-based. Again, it is interesting to note that there is very little difference between the overall average RMSEs for the two climate classifications. This is in with the observation based on the overall mean NSEC reported earlier.

7. Conclusions
In terms of predictive power and error analysis MBEs and RMSEs, performance of the ANN solar models developed differ very little whether the ANNs are grouped under the more traditional thermal climate classification or the more solar-related solar climates.

The ranges of NSEC are 0.7-0.81 for temperature-based ANN models (thermal climates), 0.83-0.94 for sunshine-based ANN models (thermal climates), 0.7-0.82 for temperature-based ANN models (solar climates), and 0.83-0.92 for sunshine-based ANN models (solar climates). In terms of predictive power, sunshine-based ANN models have larger NSECs and hence, in general, would perform better than temperature-based models.

Comparisons of the MBEs for each of the 96 cities between temperature-based and sunshine-based ANN models show that there is no significant difference in the long-term performance. Both the temperature-based and sunshine-based models for most cities (92%) could predict annual GSR to within ±10% of the corresponding measured GSR data. The RMSE results, however, show a different picture. Sunshine-based models tend to indicate better short-term performance. Only 16 cities (about 15%) have RMSEs less than 20%. For sunshine-based models, the expected deviations are much smaller. Most of the cities (92 out of 96) have RMSEs smaller than 20%.

Acknowledgements
The research described in this paper was also supported by National Key Research and Development Program of China during the 13th Five-year Plan Period (NO. 2018YFC070004). The authors would like to thank their colleague Dr. Eric W.M. Lee for his advice on the artificial neural networks.

References
[1] Wan K K W, Li D H W, Pan W and Lam J C 2012 Impact of climate change on building energy use in different climate zones and mitigation and adaptation implications Applied Energy 97 274-82
[2] Yang L, Yan H and Lam J C 2014 Thermal comfort and building energy consumption implications – A review Applied Energy 115 164-73
[3] Muneer T and Kambezidis H 1997 Solar radiation and daylight models for the energy efficient design of buildings (Oxford: Architectural Press)
[4] Gueymard C 1993 Analysis of monthly average solar radiation and bright sunshine for different thresholds at Cape Canaveral, Florida Solar Energy 51 139-45
[5] Lam J C, Wan K K W, Lau C C S and Yang L 2008 Climatic influences on solar modelling in China renewable Energy 33 1591-604
[6] Lam J C and Li D H W 1996 Regression Analysis of Solar Radiation and Sunshine Duration Architectural Science Review 39 15-23
[7] Şen Z 2007 Simple nonlinear solar irradiation estimation model Renewable Energy 32 342-50
[8] Prescott J A 1940 Evaporation from a water surface in relation to solar radiation Trans Royal Soc South Australia 61
[9] Angstrom A 1924 Solar and terrestrial radiation Quart J Royal Meteorol Soc 24
[10] Wan K K W, Tang H L, Yang L and Lam J C 2008 An analysis of thermal and solar zone radiation models using an Angstrom–Prescott equation and artificial neural networks Energy 33 1115-27
[11] Lam J C, Wan K K W and Yang L 2008 Solar radiation modelling using ANNs for different climates in China Energy Conversion and Management 49 1080-90
[12] Almorox J, Hontoria C and Benito M 2011 Models for obtaining daily global solar radiation with measured air temperature data in Madrid (Spain) Applied Energy 88 1703-9
[13] Thermal Design Code for Civil Building (GB 50176-93) (Beijing: China Architecture and Building Press)
[14] Lau C C S, Lam J C and Yang L 2007 Climate classification and passive solar design implications in China Energy Conversion and Management 48 2006-15
[15] Muneer T and Fairooz F 2016 Quality control of solar radiation and sunshine measurements – lessons learnt from processing worldwide databases Building Services Engineering Research and Technology 23 151-66