Energy 82 (2015) 589–600

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Changes in the relationship between solar radiation and sunshine duration in large cities of China

Jiandong Liu [a, b, *], Hans Linderholm [b], Deliang Chen [c], Xiuji Zhou [a], Gerald N. Flerchinger [c], Qiang Yu [d], Jun Du [e], Dingrong Wu [a], Yanbo Shen [f], Zhenbin Yang [f]

[a] Center for Ago-meteorology, Chinese Academy of Meteorological Sciences, China Meteorological Administration, Beijing, China
[b] Regional Climate Group, Department of Earth Sciences, University of Gothenburg, Gothenburg, Sweden
[c] USDA-ARS, Northwest Watershed Research Center, Boise, ID, USA
[d] Plant Functional and Climate Change Cluster, University of Technology, Sydney, Australia
[e] Tibet Climate Center, Tibet Autonomous Meteorological Administration, Lhasa, China
[f] Centre for Solar and Wind Energy Research, China Meteorological Administration, Beijing, China

ARTICLE INFO

Article history:
Received 23 July 2014
Received in revised form
20 January 2015
Accepted 21 January 2015
Available online 18 February 2015

Keywords:
Solar radiation
Sunshine duration
Angstrom model
Linear relationship
Calibration

ABSTRACT

Based on the linear relationship between solar radiation and sunshine duration, the Angstrom model is widely used to estimate solar radiation from routinely observed meteorological variables for energy exploitation. However, the relationship may have changed in quickly developing regions in the recent decades under global “dimming” and “brightening” context, with increasing aerosols due to industrial pollutions. Solar radiation stations under different climate conditions in six large cities in China are selected to test this hypothesis. Analysis of the related meteorological items shows that Guiyang has the lowest solar radiation with the average annual value of 10.5 $MJm^{-2}d^{-1}$, while Lhasa on the Tibetan Plateau has the highest of 20.1 $MJm^{-2}d^{-1}$. Both radiation and sunshine hours decreased from 1961 to 2010, but at different rates. A moving linear regression method is used to investigate the changes in the relationship between radiation and sunshine duration, the results indicate an abrupt change in the correlation coefficients in 1980–1990s, which can be attributed to the aerosol load resulting from air pollution caused by the industrial development in 1980s under China's Open Door Policy. The sky condition has been changing from clean to dirty, thus the relationship between solar radiation and duration changes in the 1980's and has recovered in the recent decades. This finding implies that it might not necessarily be right to use long data sets for model calibration. Further investigation confirms that the Angstrom model performs the best with higher *NSE* (nash-sutcliffe efficiency) of 0.914 and lower *MAPE* (mean absolute percentage error) and *RMSE* (root mean square error) values of 13.7 w/m^2 and 23.9 w/m^2 respectively, when calibrated with a 10-year data set. In contrast, the model performs worst when it is calibrated with a 40-year data set, with *NSE*, *MAPE* and *RMSE* values of 0.891, 15.1 w/m^2 and 25.3 w/m^2, respectively. Based on the findings of this research, a 10-year data set is recommended as the national standard for model calibration in rapidly developing regions of China. Further analogous investigations are needed in other industrial regions to make an international standard for Angstrom model calibration.

1. Introduction

Compared to the energy from fossil fuels, solar energy is believed to be even more important as clean energy not contributing to global warming [1]. Accurate identification of the temporal and spatial distribution of solar radiation is the essential premise to solar energy exploitation. However, solar radiation is not a routinely observed meteorological item as temperature or rainfall due to the high cost of instruments and maintenance. Up to now, solar radiation is only observed at very few particular stations in the developing nations, which means that the estimation of solar radiation is of vital importance for scientists in the domain of energy, meteorology, and agronomy, among others e.g. Refs. [2–5].

* Corresponding author. Center for Agro-meteorology, Chinese Academy of Meteorological Sciences, No. 46 Zhongguancun South Street, Beijing 100081, China. Tel.: +86 10 68408957; fax: +86 10 62175931.
E-mail address: liujd2001@263.net (J. Liu).

http://dx.doi.org/10.1016/j.energy.2015.01.068
0360-5442/

Robust numerical models have already been established by meteorologists, but the requirement of too many input parameters makes this kind of model suitable only for theoretical analysis rather than practical applications [6,7]. In contrast, simple empirical models are still the main tool for solar radiation estimation [8]. Based on the relationship between solar radiation and readily available meteorological parameters, such as sunshine duration, temperature, rainfall etc., different kind of simple empirical models have been developed [9–12]. Comparison of different kinds of empirical models identified that the sunshine-based model performed better than the temperature- and rainfall-based ones [4,13–15], even under the unique alpine conditions [16], and the sunshine-based Angstrom model is likely the most popular empirical model for solar radiation estimations [13,17–21]. Since the coefficients of the Angstrom model are site-specific, the first step for the model application is to calibrate the model with the collected radiation and sunshine data set. However, the length of the data set used for the calibration seems casual. Different lengths of the data set have been used for model calibration, varying from 1 year [4], 5 years [13], 10 years [20] to 40 year or so [22]. There seems an intuitive tendency that the calibration should be made with data set as long as possible, if only the related data set can be obtained [23]. Of course, in the view of statistics, longer data sets for model calibration always means more reliable parameters [24], under assumption that the relationship between solar radiation and sunshine duration is stationary.

However, there is increasing evidence that the amount of solar radiation incident at the earth's surface is not stable over the years but undergoes significant decadal variations [25], which is referred to as "global dimming" and "brightening" [26]. As for the quickly developing regions like China, Qian et al. [27] suggested that the increasing aerosol loading from emissions of pollutants is responsible for the observed reduced global radiation. It was identified that both radiation and sunshine duration have decreased in the recent decades, partly or entirely due to changes in aerosol concentrations under global "dimming" and "brightening" context [25–28]. If the changing trend of radiation and sunshine duration were not synchronized, we can cautiously envisage that the relationship between solar radiation and sunshine duration may have changed. But up to now, no research to our knowledge has been conducted to test this hypothesis.

With the Open Door Policy in the 1980s, China developed quickly with significant environmental issues, especially air pollution due to increasing aerosols [27]. In this research, solar radiation and sunshine duration data were collected at radiation stations in six large cities of China under different climate conditions. The objectives of the study are: (1) to test the hypothesis that the relationship between radiation and sunshine duration has changed in the large cities of China, under "global dimming and brightening" context; (2) to address a corresponding strategy for solar radiation estimation, if the hypothesis can be validated; and (3) to give the scientific basis for a national standard of solar radiation estimation.

2. Data collection and methodology

2.1. Database

For this research, access to the database of the NMIC (national meteorological information center) was given by the CMA (China meteorological administration). Study of the influence of different time scales on the Angstrom calibration indicated that calibration made on daily basis has important advantages on monthly scale with stabilized coefficients and more accurate prediction [22], therefore daily solar radiation and the related meteorological parameters, including sunshine duration, cloud coverage and rainy days, were collected from the NMIC/CMA database. The data for radiation and sunshine duration were screened according to the rules described by Persand et al. [29]: (1) daily observations were excluded from the database if either sunshine duration or solar radiation was missing; (2) daily observations were excluded if the measured radiation/extra-terrestrial radiation or the actual sunshine duration/potential sunshine duration was greater than 1; and (3) the data for an entire month were excluded if more than 10 days of radiation or sunshine duration were missing for that month. Additionally, stations were selected for this research based on two more strict criteria: (1) the selected stations should measure 4 parameters, i.e. radiation, sunshine duration, cloud cover and rainy days simultaneously; and (2) stations were excluded if the corresponding visibility was not measured. In the end, six radiation stations were selected with data from 1961 to 2010. The distribution of the selected stations is shown in Fig. 1, and more detailed information can be seen in Table 1 [30].

2.2. Methodology

Based on the linear relationship between solar radiation and sunshine duration, the Angstrom model was first suggested in 1924 [17] and modified by Prescott in 1940 [18] as follows:

$$\frac{R_a}{R_e} = a + b\frac{S}{S_0} \tag{1}$$

in which R_a is the actual solar radiation, R_e the daily extra terrestrial solar radiation, S the actual sunshine duration, and S_0 the theoretical sunshine duration. R_e can be calculated according to equations described by Allen et al. [19], and the relationship between R_a/R_e and S/S_0 can be fitted by linear regression [24], through which coefficients of the model and correlation coefficient r between solar radiation and sunshine duration can be obtained.

The *NSE* (nash-sutcliffe efficiency), the *MAPE* (mean absolute percentage error), and the *RMSE* (root mean squared error) [13,14,16,20,22,23], were used as criteria in evaluating the model performance in this study; these statistics can be expressed as follows:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - S_i)^2}{\sum_{i=1}^{n}\left(O_i - \overline{O}\right)^2} \tag{2}$$

$$MAPE = \frac{\sum_{i=1}^{n}\frac{|O_i - S_i|}{O_i} \times 100}{n} \tag{3}$$

$$RMSE = \left[\frac{1}{n}\sum_{i=1}^{n}(O_i - S_i)^2\right]^{\frac{1}{2}} \tag{4}$$

in which O_i is the observed value, S_i the simulated value, $\overline{O}$ the average value of the observed radiation, and n the number of data pairs. %*MAPE* and %*RMSE*, which are the ratios of *MAPE* and *RMSE* to the average value of the observed radiation $\overline{O}$ respectively, are also used for model evaluation. The model is believed to be performed better with a higher value of *NSE* and lower values of *MAPE*, *RMSE*, %*MAPE* and %*RMSE*.

The t-test is used to identify the significant differences between parameters before and after the 1980s. The value of t is calculated as [16,23]:

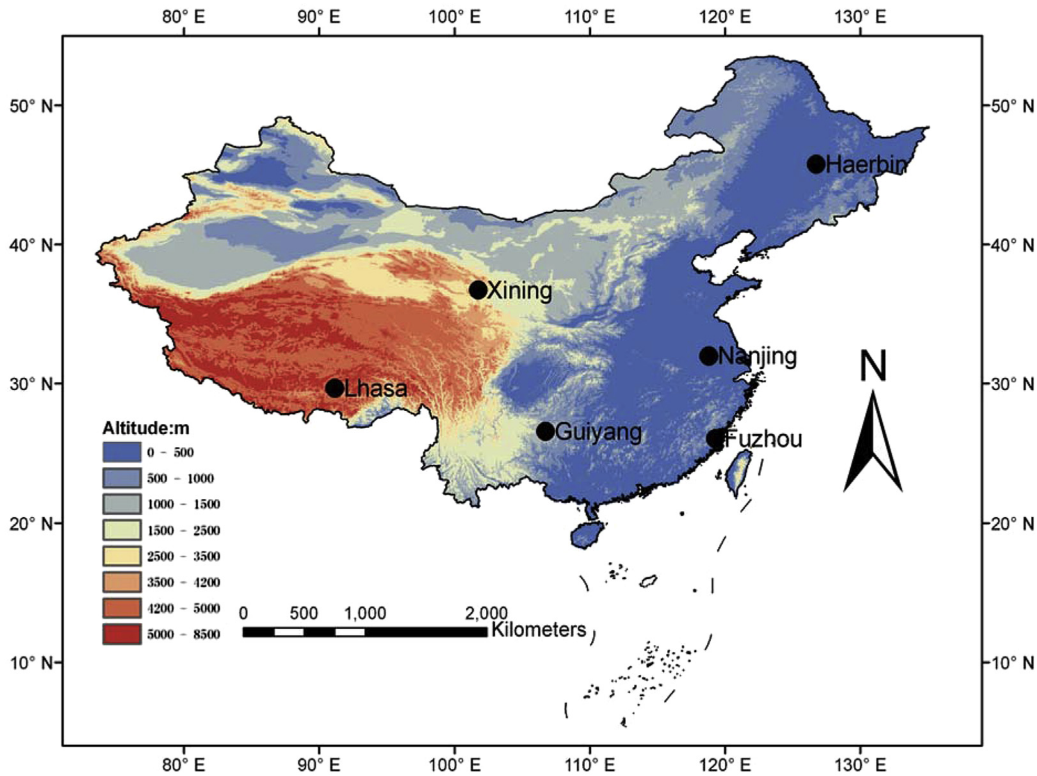

Fig. 1. Distribution of the solar radiation observation stations used in this study.

$$t = \sqrt{\frac{(n-1) \times MBE^2}{(RMSE^2 - MBE^2)}} \quad (5)$$

where *MBE* is the mean bias error, which can be calculated as [16,23]:

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(S_i - O_i) \quad (6)$$

When the calculated $|t| \geq t_{0.05}$ (critical value), the two groups of the data differ significantly, and vice versa. Trends of changes in radiation, sunshine duration, cloud coverage and rainy days were analyzed by the linear regression method with *F*-test [24].

3. Results

3.1. Statistical analysis of solar radiation and related meteorological items

Long-term variations of solar radiation and sunshine duration are shown in Fig. 2. Lhasa on the Tibetan plateau has the highest average annual solar radiation of 20.1 $MJm^{-2}d^{-1}$, which can be attributed to its high altitude (3650.1 m a.s.l.). In contrast, Guiyang has the lowest solar radiation with the average annual value of 10.1 $MJm^{-2}d^{-1}$, nearly half of that in Lhasa, which can be attributed to its high cloud coverage and excessive rainy days. The highest sunshine duration is also found in Lhasa, with the average annual value of 8.2 hd^{-1}, much higher than that in Guiyang, with the corresponding value of 3.2 hd^{-1}.

Superficially speaking, radiation and sunshine duration changes steadily in the recent decades (Fig. 2), and the ratio of radiation to the sunshine duration appears stable with little variation. However, trend analysis (Table 2a) indicates obvious changes in solar radiation in the recent decades. Solar radiation at all six stations has a decreasing trend in 1961–2010. Lhasa and Xining have the higher decline of 0.037 $MJm^{-2}d^{-1}$ and 0.036 $MJm^{-2}d^{-1}$, respectively. Nanjing, one of the most quickly developing regions in the industrial Yangtze River Delta, also has a high decreasing trend of 0.033 $MJm^{-2}d^{-1}$. Similar to the variation features of the solar radiation, the sunshine duration at all six stations also decreased in the period of 1961–2010. Nevertheless, the decreasing rates between solar radiation and sunshine duration are quite different. Reduced radiation and sunshine duration can be caused by the increasing cloud cover and rainy days. However, the trend analysis of cloud cover indicates that cloud cover in four out of the six stations has decreased in 1961–2010. Only Fuzhou and Xining have increasing trend in clouds, but the decreasing trends are not significant with *F*-test at >25% level. All of the six stations have decreasing trend in rainy days, and four out of the six stations, including Fuzhou, Nanjing, Guiyang and Harebin decreased significantly (*F*-test at the significant level of 1%). This indicates that the decreased radiation and sunshine duration cannot be attributed to the variation of clouds and rainy days. This result is in good agreement with previous reports [31,32].

Table 1
Detailed information of the solar radiation stations in this research work.

Station	Latitude/N°	Longitude/E°	Altitude/m asl	Climate conditions
Fuzhou	26.08	119.28	85.4	South subtropical
Nanjing	32.00	118.80	7.5	North subtropical
Guiyang	26.58	106.73	1223.2	South subtropical
Haerbin	45.75	126.77	143.0	Cold temperate
Xining	36.72	101.75	2296.2	Sub-alpine plateau
Lhasa	29.67	91.33	3650.1	Arid plateau

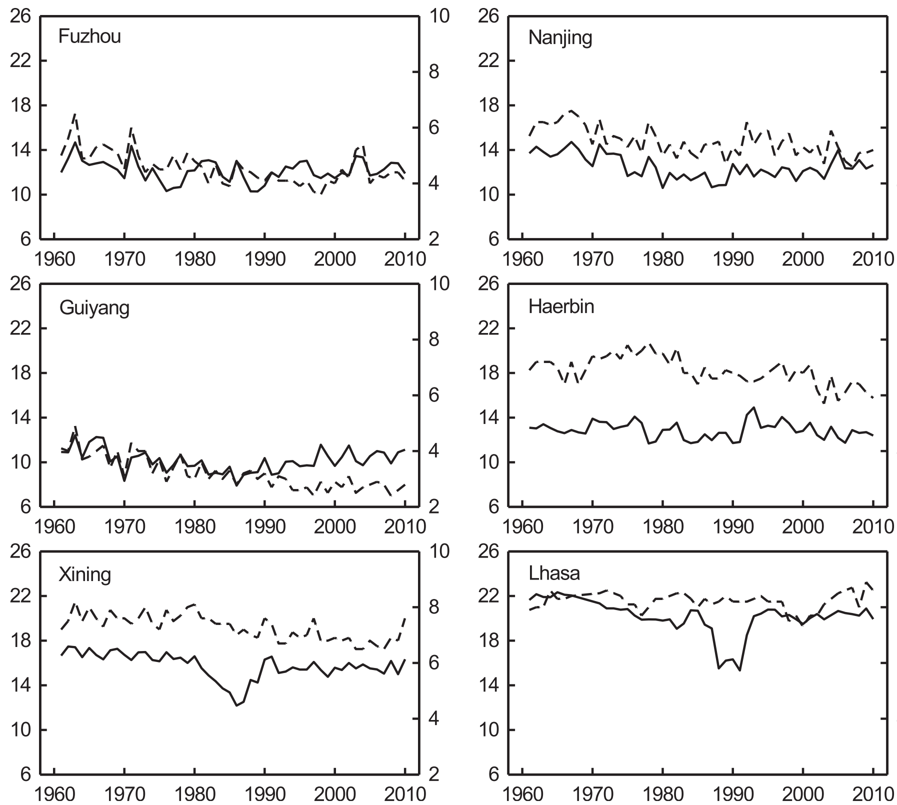

Fig. 2. Variation of annual solar radiation and sunshine hours (The horizontal axis denotes time, the left vertical axis solar radiation [MJ $m^{-2}d^{-1}$]), and the right vertical axis sunshine duration [h]). The solid line expresses solar radiation, and the dash line sunshine duration.

When the period from 1961 to 2010 is divided into two parts, i.e. from 1961 to 1990 and from 1991 to 2010, the trend analysis indicates a more obvious "dimming" (Table 2b) and an obscure "brightening" period (Table 2c). This result is in well agreement with the previous researches [25,26,33], and the relationship between radiation and sunshine duration might have been changing during the period from dimming to brightening.

Table 2
Trend of the solar radiation and the related meteorological parameters in the recent decades.

	Radiation ($MJm^{-2}d^{-1}$)		Sunshine duration(h)		Cloud coverage (1/10)		Rainy day (d)	
	Trend	F-test	Trend	F-test	Trend	F-test	Trend	F-test
a. 1961–2010								
Fuzhou	−0.009	>25%	−0.025	1%	0.004	>25%	−0.435	1%
Nanjing	−0.033	1%	−0.022	1%	−0.005	25%	−0.285	1%
Guiyang	−0.010	>25%	−0.033	1%	−0.005	5%	−0.633	1%
Haerbin	−0.008	>25%	−0.022	1%	−0.002	>25%	−0.309	1%
Xining	−0.036	1%	−0.023	1%	0.007	5%	−0.133	10%
Lhasa	−0.037	5%	−0.001	>25%	−0.013	1%	−0.092	>25%
b. 1961–1990								
Fuzhou	−0.068	1%	−0.047	1%	0.014	10%	0.095	>25%
Nanjing	−0.115	1%	−0.044	1%	0.002	>25%	−0.179	>25%
Guiyang	−0.090	1%	−0.042	1%	0.003	>25%	−0.842	1%
Haerbin	−0.038	1%	−0.007	>25%	−0.002	1%	−0.376	10%
Xining	−0.128	1%	−0.014	5%	0.013	10%	−0.129	>25%
Lhasa	−0.169	1%	0.001	>25%	−0.008	25%	−0.438	25%
c. 1991–2010								
Fuzhou	0.016	>25%	0.022	25%	−0.005	>25%	−0.837	25%
Nanjing	0.044	10%	−0.035	5%	0.016	25%	0.019	>25%
Guiyang	0.076	1%	−0.005	>25%	−0.015	10%	−0.729	25%
Haerbin	−0.068	5%	−0.039	5%	0.027	5%	−0.727	5%
Xining	0.003	>25%	−0.010	>25%	0.033	1%	−0.376	>25%
Lhasa	0.094	5%	0.023	25%	−0.011	25%	0.966	10%

− Denotes decreasing trend.

3.2. Determination of the suitable data length for fitting Angstrom model

To explore the change of relationship between solar radiation and sunshine duration, the moving linear regression method is used to calculate the coefficients a, b and the correlation coefficient r at different periods. The moving linear regression method is actually the linear regression used for fitting the coefficients at different periods as $X - Y, X+1 - Y+1, X+2 - Y+2, \ldots, X+W - Y+W$. The first step for this method is to seek the suitable data length, i.e. $Y - X + 1$.

Fig. 3 shows the influence of the calibration data length on the fitted coefficients. The coefficients vary obviously when the Angstrom model is calibrated with 1-year data pairs, but become stable enough under 10-year data pairs (Fig. 3). When the calibration data length becomes longer, the coefficients exhibit little variation. This implies two meanings: (1) 10-year is long enough for model calibration, and (2) the coefficients are changing in the recent decades, but too long data set will distort the real changing patterns, e.g. the coefficients fitted by the period 1961–1990 might obscured the real changes in the relationship between radiation and sunshine durations in the 1961–1970, 1971–1980 and 1981–1990. So, 10-year data length might be suitable for fitting coefficients.

The fitted coefficients can also be influenced by the number of the missing data. If the coefficients are significantly affected by the number of missing data, then the changes in the coefficients at different periods might be caused by the missing data rather than "dimming" and "brightening" effects. To clarify this situation,

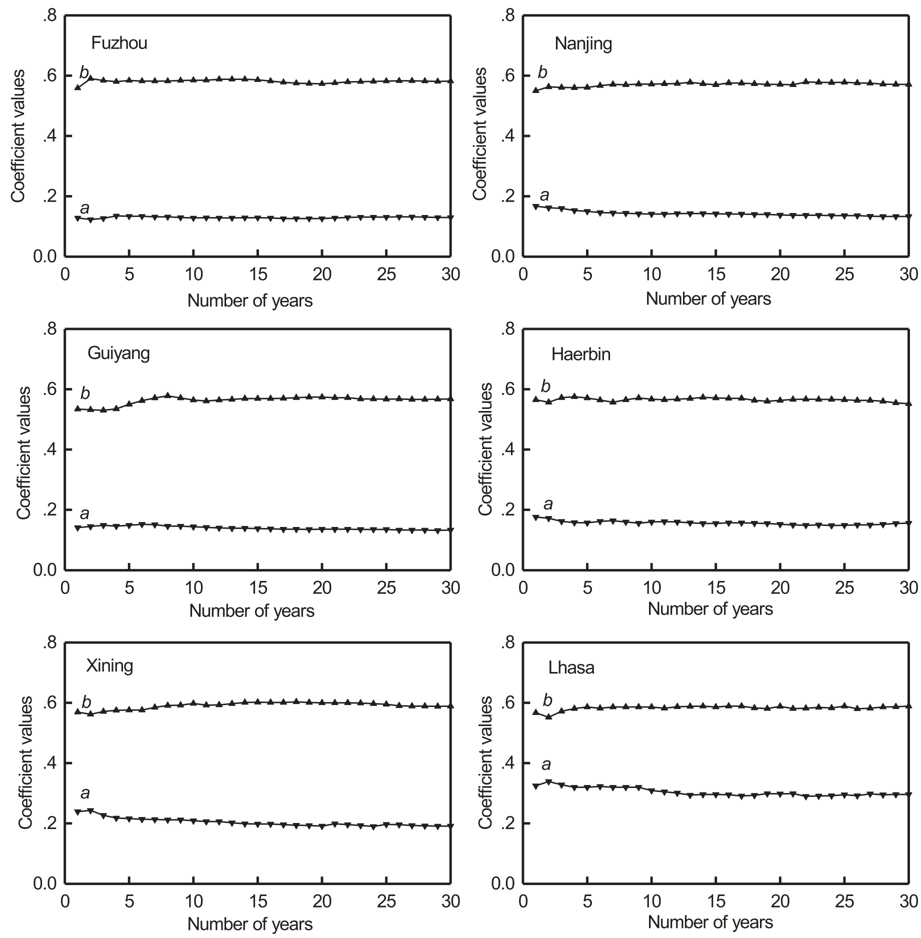

Fig. 3. Influence of the calibration data length on the stability of the fitted coefficients.

sensitivity of the coefficients fitted by data pairs in 1975–1984, 1985–1994 and 1995–2004 in Haerbin to the number of missing data is made. First, the data set P1975 (PX denotes the 10-year period from X to X + 9, hereinafter) is fitted, and the coefficients a, b and r are obtained. Then, 1 data pair is excluded randomly from the data set, and the corresponding coefficients are fitted. This process is repeated for 100 times, so we can identify the sensitivity of the coefficients to 1 missing data. Likewise, the sensitivity of the coefficients to different number of missing data is identified. It is shown vividly in Fig. 4 that the trend in the coefficients (P1975, P1985 and P1995) can not be distorted by even as many as 100 missing data pairs, if 10-year data is used as data length for the moving linear regression method. According to Figs. 3 and 4, 10-year data length is used as the suitable data length for moving linear regression method in this study.

3.3. Variation of the coefficient b of the Angstrom model

The coefficient b in the Angstrom model reflects the transmission characteristics under the average sky conditions, mainly determined by the water content and aerosols [23]. The coefficient b varies from 0.45 to 0.63 for all stations from P1961 to P2001 (Fig. 5). Fuzhou has the highest mean coefficient b of 0.57, averaged from P1961 to P2010. Though the solar radiation is the highest in Lhasa, the averaged coefficient b in P1961–P2010 shows no obvious difference to those for the other stations, with a value of 0.53. Distribution of the coefficients of Angstrom model in China has been calculated by Chen et al. [13]; the coefficient b in Lhasa was reported as 0.54, which can be regarded as a very ordinary value among the 48 solar radiation stations in China. High solar radiation on the Tibetan Plateau may have resulted in the higher value of coefficient a apart from the coefficient b, which was identified by Ref. [13] and further confirmed by our research in Gaize on the western part of the Tibetan Pleatau [16].

The coefficient b has obvious decreasing trend for most of the stations, including Fuzhou, Nanjing, Harebin, Xining and Lhasa in the recent decades. For all of these five stations, the quickly decreasing period can be found around 1980s or so (Fig. 5). The coefficient b increases a little in Fuzhou and Lhasa in the recent years. Guiyang is an exception with the coefficient b increasing steadily after P1981. This might be attributed to a decrease in cloud cover and rainy days and an abrupt increase in light fogs in Guiyang, which resulted in the higher radiation and the lower sunshine duration.

3.4. Variation of the coefficient (a + b) of the Angstrom model

The coefficient $(a + b)$ in the Angstrom model reflects the transmission characteristics under the clear sky conditions. The coefficient $(a + b)$ varies from 0.66 to 0.92 for all stations from P1961 to P2001 (Fig. 6). Nanjing has the lowest mean coefficient $(a + b)$ of 0.69, averaged from P1961 to P2001. Unlike the coefficient b as mentioned above, the coefficient $(a + b)$ are much higher on the Tibetan Plateau than those in any other regions, with the highest average value of 0.80 in Lhasa and the second highest value of 0.75 in Xining. The high values of the coefficient $(a + b)$ over the

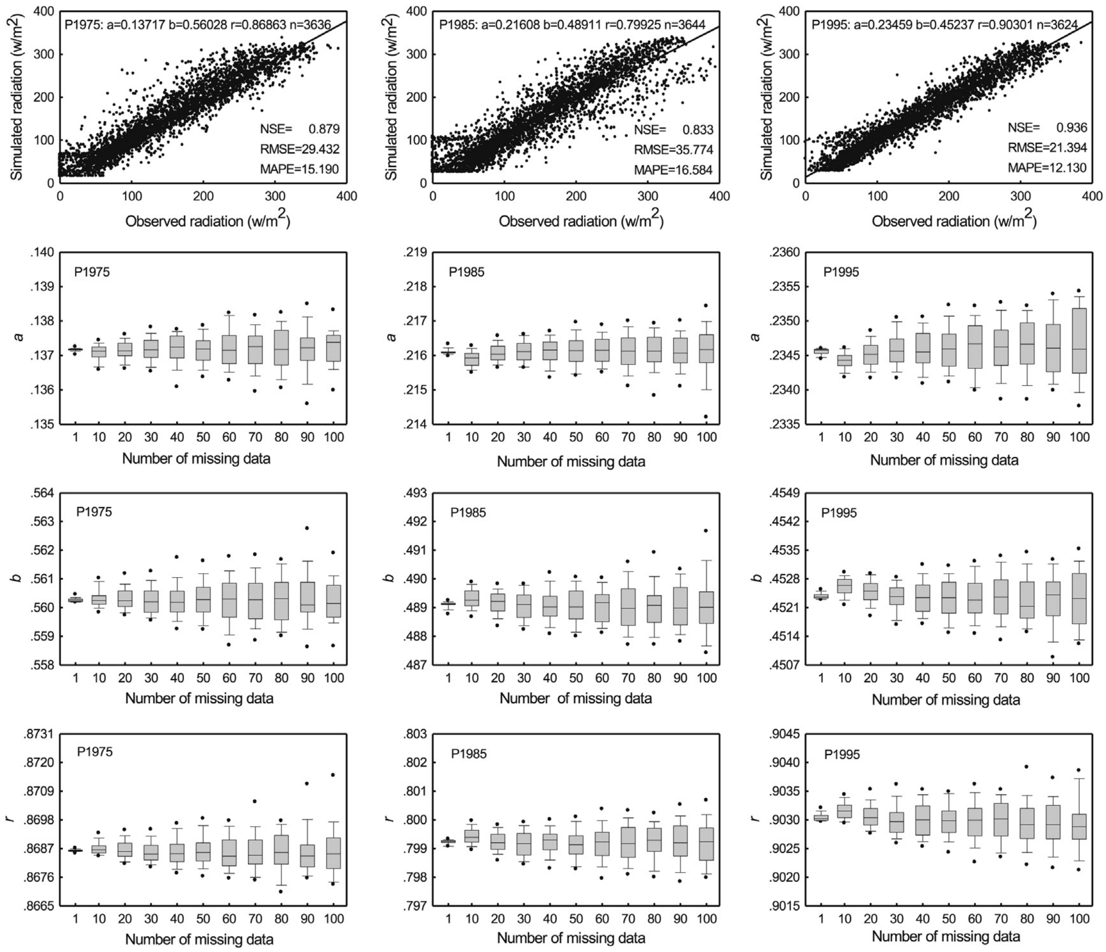

Fig. 4. Sensitivity of the fitted coefficients to the number of missing data under 10-year calibration conditions. Each box plot showing statistics of the coefficients calibrated with 100 data sets randomly generated. Top dot marks the maximum value, vertical line from top to bottom marks 95th, 75th, 50th, 25th and 5th percentile, and the bottom dot marks the maximum value calculated from 100 set of data.

Tibetan Plateau can be attributed to the high altitude and less vapor effects [16].

Generally speaking, the coefficient $(a+b)$ has no obvious trends in Fuzhou and Haerbin. However, obvious change in the coefficient $(a+b)$ can be found in Nanjing, Guiyang, Xining and Lhasa. Changes in the coefficient $(a+b)$ among these cities have the similar trends. First, the coefficient $(a+b)$ decreases from P1961 to P1981, then the coefficient $(a+b)$ increases from P1981 to P2001 (Fig. 6).

3.5. Changes in the correlation coefficients r

The Angstrom model is based on the linear relationship between solar radiation and sunshine duration. Variation in the correlation coefficient r in P1961–P2001 is shown in Fig. 7. The correlation coefficient r changes greatly in the recent decades (Fig. 7). Except for Lhasa, no r lower than 0.795 can be found for the other five stations, indicating the robustness of the Angstrom model for radiation estimation. The lowest r occurs in Lhasa in P1982–P1989, with the average value of 0.713. This may have resulted from the abnormal decrease in solar radiation in Lhasa around 1990, which could markedly disturb the relationship between radiation and sunshine duration in that period. Note that the lowest radiation began in 1988 (Fig. 2) while the lowest r began in P1982 (Fig. 7). This can be explained by the definition of P1982, which referred to the period from 1982 to 1991. In other words, the coefficient r in P1982 was obtained by linear regression with the sub data set 1981–1991. All of the four abnormal years from 1988 to 1991 were mixed with 6 ordinary years for calibration, leading to the very low r in P1982. Likewise, r becomes higher in P1992 corresponding to the period from 1992 to 2001, without any abnormal data mixed in the data set for calibration.

Obvious “troughs” can be found in the variation curve of r in Fig. 7. The “troughs” of r are especially obvious for most stations such as Fuzhou, Nanjing, Haerbin, Xining and Lhasa. Though shapes of the “troughs” differ among the six stations, all of them occur around P1980 and end no later than P1995. In general, r has high values before P1980, then decreases abruptly around P1980–P1995, but returns to even higher values afterwards. The troughs of variation in r indicate that the relationships between solar radiation and sunshine duration have been changed and recovered at these six stations in the period from 1960 to 2010, which might be attributed to the “dimming” and “brightening” effects discussed below.

4. Discussion

Previous research has focused on the trend analysis of solar radiation and sunshine duration by the linear regression method, [32–35]. However, the simple linear regression cannot reveal the characteristics of the relationship between radiation and sunshine duration, which is the essential scientific basis for solar radiation estimation with the Angstrom model. In this research, the moving

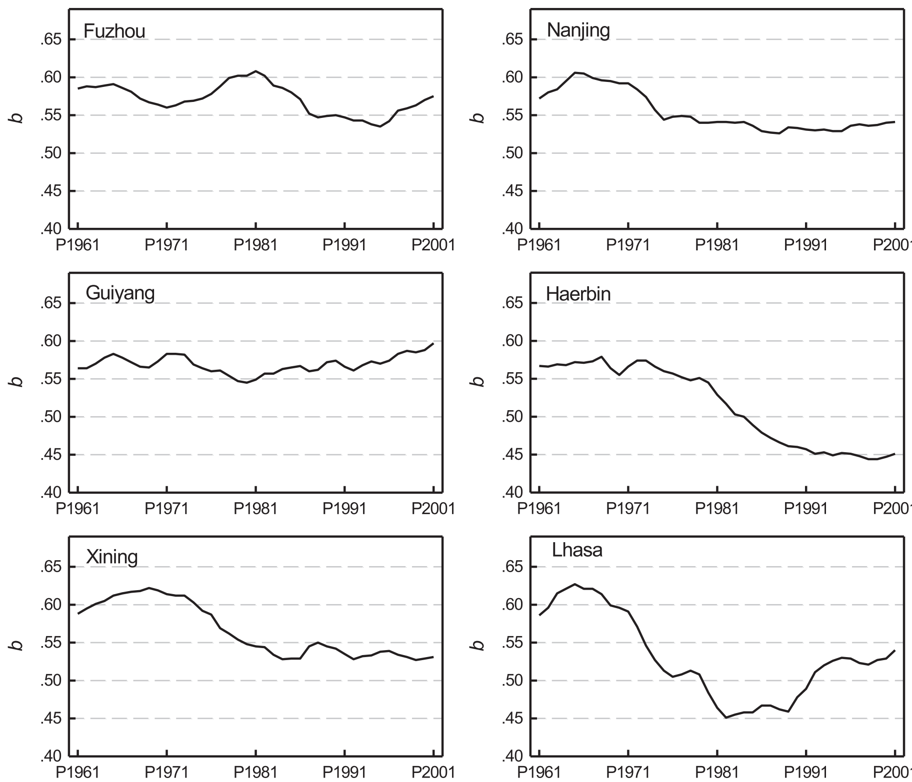

Fig. 5. Variation in the coefficient b of the Angstrom model in the period 1961–2010.

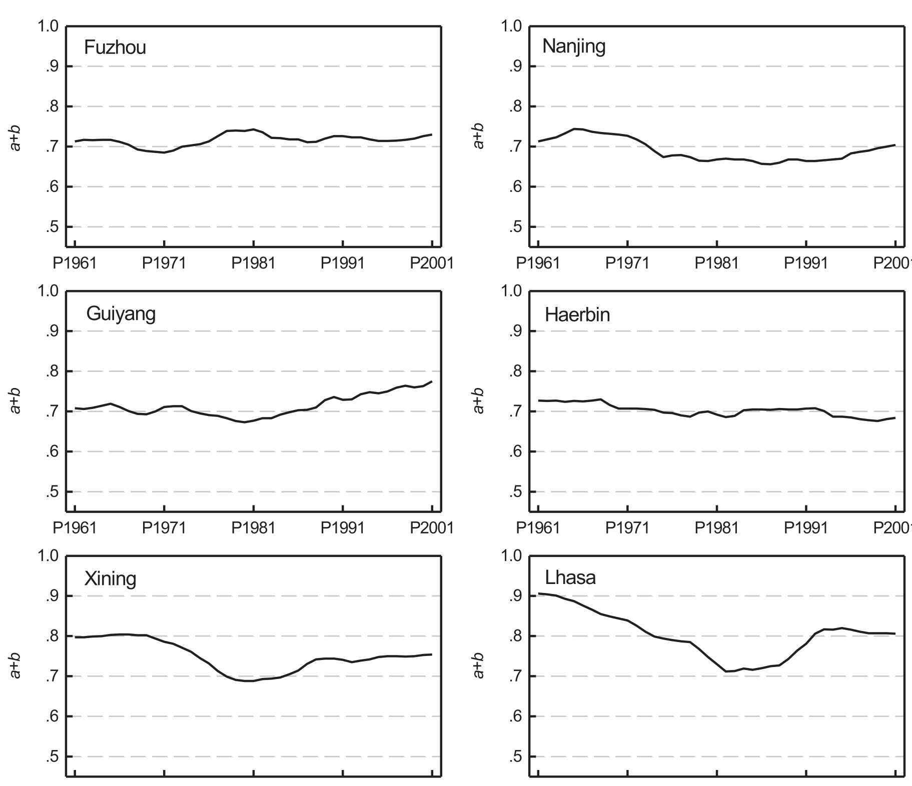

Fig. 6. Variation in the coefficient $(a + b)$ of the Angstrom model in the period 1961–2010.

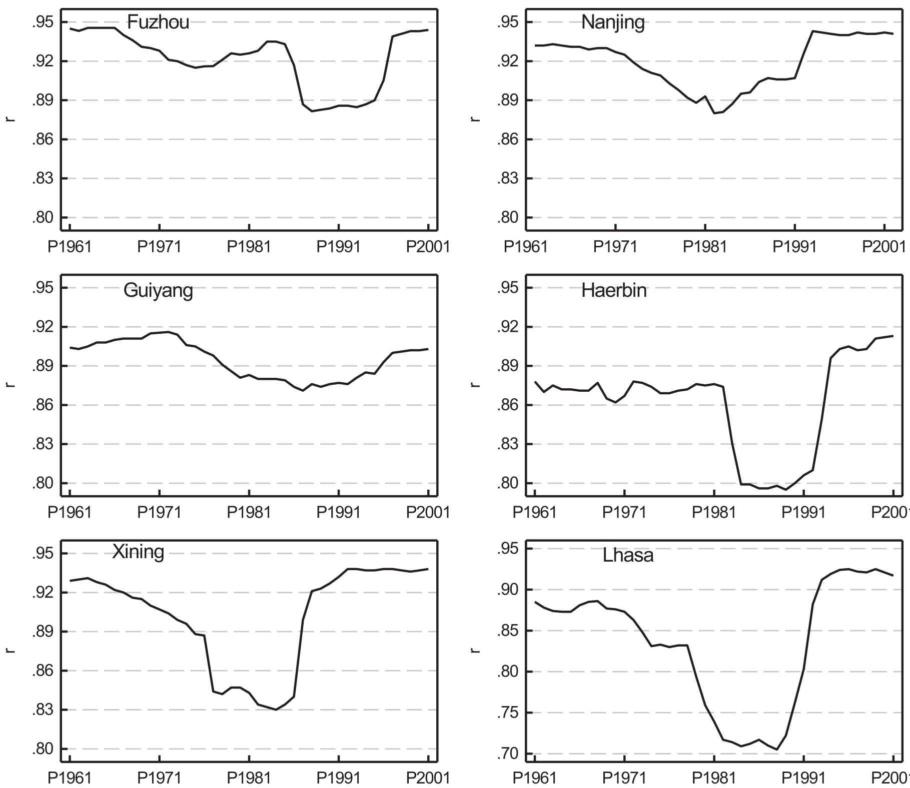

Fig. 7. Variation in the correlation coefficient r of the Angstrom model in the period 1961–2010.

linear regression method was used and successfully depicted the variation processes of the relationship (Figs. 5–7).

The obvious variation of r is caused by the unsynchronized trends in radiation and sunshine duration. Analysis of the meteorological parameters indicated an obvious "dimming" and an obscure "brightening" exist in large cites of China in 1961–1990 and 1991–2010 (Table 2), respectively. This agrees well with lots of previous reports [25,26]. The decreasing trend in solar radiation and sunshine duration cannot be attributed to the changes in cloud cover and rainy days, which are also decreasing rather than increasing in the same period (Table 2) [31–33,36]. So, aerosols are speculated to be the main factor giving rise to the "dimming" and "brightening" in large cites of China, which resulted in the obvious variation of r due to their asynchrony. Visibility can be used as an important index indicating the aerosol concentration in the atmosphere, through the Aerosol Optical Depth (AOD) [36,37]. Visibility data at the six stations in the period 1961–2010 were provided by the NMIC and the Tibetan Autonomous Meteorological Administration. However, the data were recorded very coarsely before 1980, so only the accurate observed visibility data from 1980 to 2010 have been used for trend analysis. The visibility data were averaged by the moving average method at 10-year interval, to correspond to the correlation coefficient r and $(a + b)$ obtained over the same period (Fig. 8). Vivid decreasing trend in visibility can be seen in Fuzhou, Guiyang, Haerbin and Nanjing from P1980 to P1995. An obvious deceasing trend can also be found in Xinjing in P1985–P1995. Though coarse data were used to analyze the long-term variation of visibility in a previous research [36], results from present study and [36] agree very well. Since the decreasing

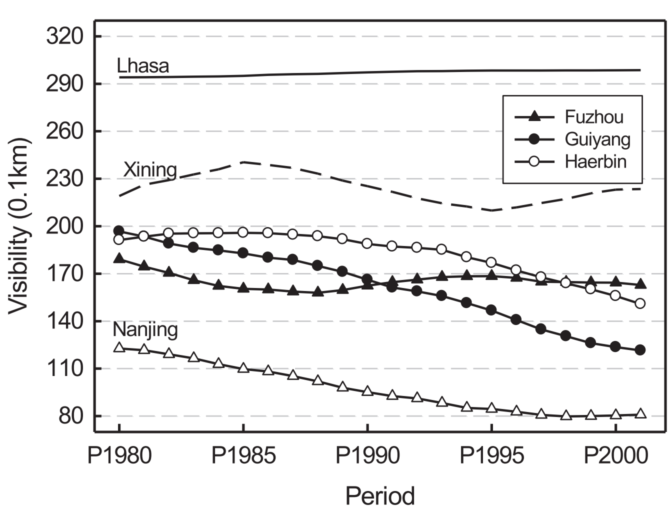

Fig. 8. Variability in decadal averages of visibility in the period 1980–2010.

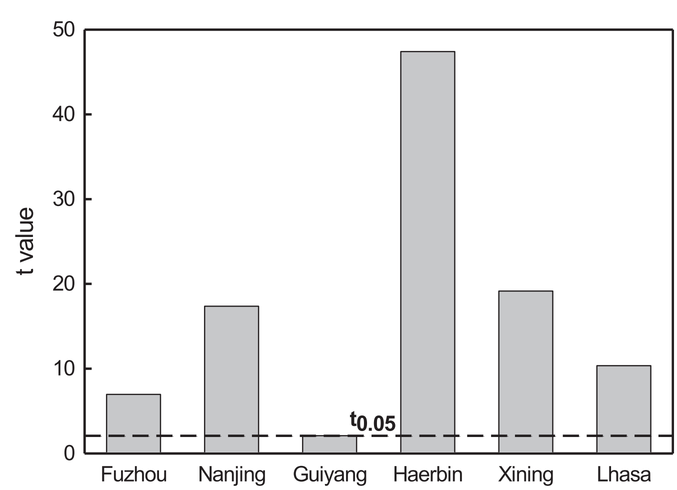

Fig. 9. The t-statistic of the coefficients b before and after 1978.

trend in visibility suggests the increasing concentration of aerosol pollution [36], the decreasing trends in solar radiation and sunshine duration, which correspond to the obvious decreasing visibility, can be attributed to the increasing aerosol pollutions. This conclusion is also in agreement with the findings of Qian et al. [27]. The trend in visibility has changed after P1995 in Xining, Fuzhou, and Nannjing. Increase in the visibility corresponds to the increase in $(a + b)$ in these large cities, indicating "brightening" in large cities of China in the recent years might resulted from less pollutants. This was identified by the measurements of the air pollution index from 2000 to 2011 [33]. Air pollution has resulted in the "dimming" and "brightening" in large cities of China [33]. Here we cautiously make the further conclusion that relationship between radiation and sunshine duration has also changed, due to their different decreasing rates under "dimming" and "brightening" context.

The coefficient b in the Angstrom model represents the transmittance of the atmosphere under average sky conditions [23], and the increasing aerosol concentration might result in significant change in b. The values of the coefficient b were divided into two groups before and after 1978 (the year of Open Door Policy in China) for the t-test statistic analysis for each station. Significant changes in b were identified for most of the six stations (Fig. 9). In other words, the sky conditions are significantly different before and after 1978, which implies that most stations have changed from "clean" to "dirty" sky conditions in the 1980s and 1990s. For a certain station, the relationship between radiation and sunshine duration is stable under clean sky conditions, which results in a

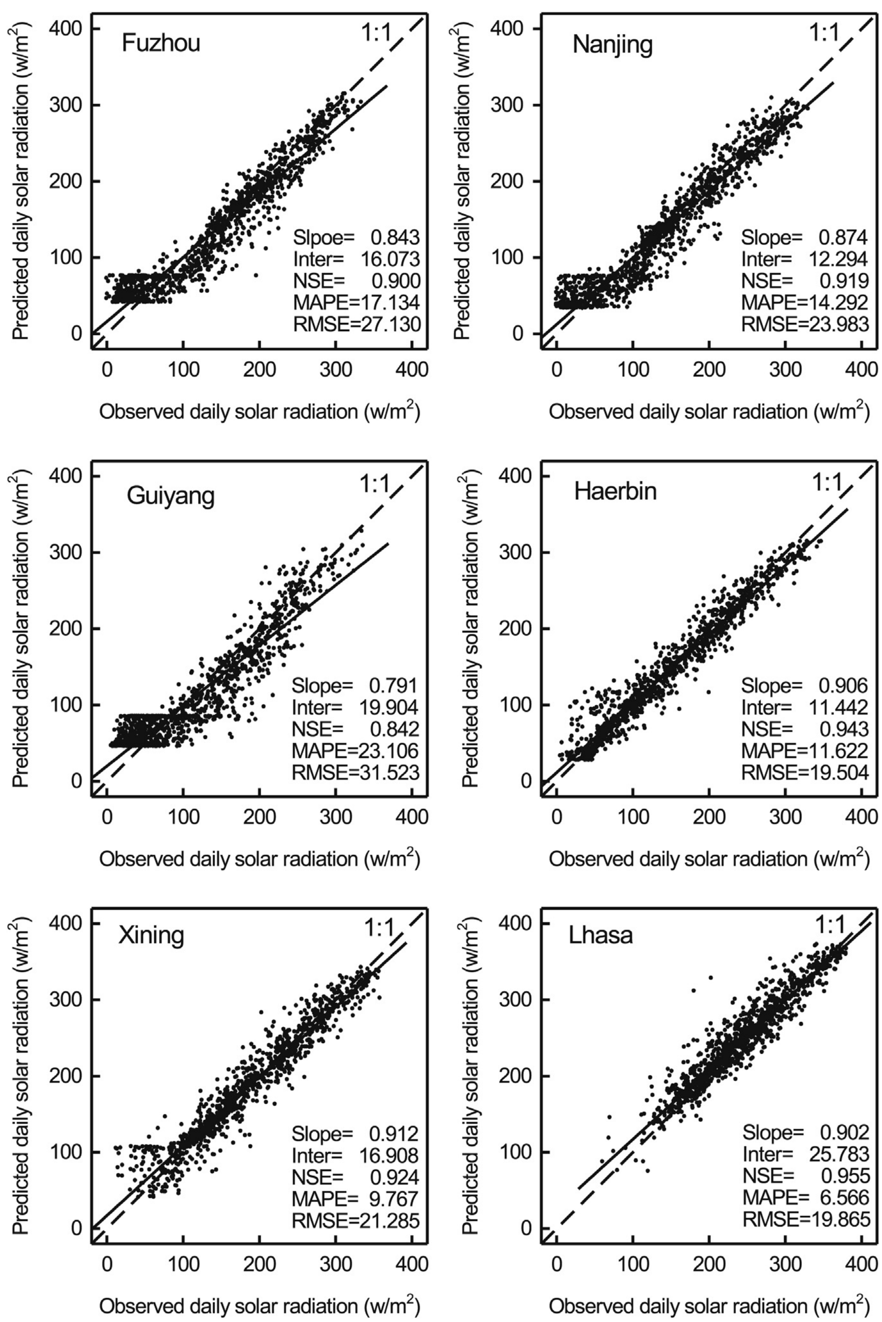

Fig. 10. Model validation with 10-year data length for calibration.

higher value of r and higher b values in the Angstrom model. The coefficient b would become smaller due to higher aerosol pollution under dirty atmospheric conditions, at the same time the correlation coefficient should also be higher, since the data are collected likewise from the stable population (all under dirty atmospheric conditions). However, r would decrease in the transitional period from clean to dirty atmospheric conditions, as the data set used for regression are actually from two different populations, i.e. the first part of the data set corresponds to clean-sky conditions, while the latter from dirty atmospheric ones. In view of statistics [24], the data set from two different populations will disturb the linear relationship and lead to a lower r. This is the main reason of the obvious "troughs" in Fig. 7. Herewith, we can conclude that the relationship between radiation and sunshine duration have changed over some large cities of China under "dimming" and "brightening" context in the recent decades, mostly due to increasing aerosols caused by rapid industrial development after China's Open Door Policy of 1978. The shapes of the variation of r (including start and end of the trough) at different stations should not be expected to be the same, due to uneven development and unequal degrees of pollution in China [36,38].

So, what does this mean for solar radiation estimation? As mentioned above, the relationship between solar radiation and sunshine duration has obviously changed. This implies that the tendency to calibrate Angstrom model with the longest data set might be wrong e.g. Refs. [13,14,22,23], since this is based on the assumption of a stationary linear relationship between the two parameters. It may be more reasonable to calibrate the Angstrom model with shorter data sets collected in recent years, rather than with longer data set including samples from different sky conditions. To test this speculation, the data set in 1961–2010 was divided into different sub data sets for further investigation. The sub data sets of 1968–2007, 1978–2007, 1988–2007 and 1998–2007 were used for model calibration. The sub data set of 2008–2010 was left for model validation. As expected (see Fig. 10 and Table 3), the Angstrom model performs best with the higher *NSE* value of 0.914 and lower *MAPE*, *RMSE*, *%MAPE* and *%RMSE* values of 13.7 w/m^2, 23.9 w/m^2, 9.4% and 15.7%, respectively, when calibrated with a 10-year data set. When calibrated with a 40-year data set, the model performs worst with a lower *NSE* value of 0.891 and higher *MAPE*, *RMSE*, *%MAPE* and *%RMSE* values of 15.1 w/m^2, 25.3 w/m^2, 10.4% and 16.8%, respectively. Note that higher bias can be found in Fuzhou, Nanjing and Guiyang when the value of radiation is low, which is very similar to the previous reports by Wu and Liu et al. [14,23].

Accurate estimation of the current distribution of solar radiation in China becomes a significant project for the Research Center for Wind and Solar Energy in CMA [38,39], under the National Ministry of Science and Technology. The Angstrom model was selected to be used as the main method for estimating solar radiation from sunshine duration [38,39]. However, the length of the data set for calibration (the first step of the model application) has currently caused serious disagreement. Some scientists believe that 30–40 year data sets should be used as standard for model calibration, as linear relationship between radiation and sunshine duration might have kept stable in the recent decades; others think that a calibration with a shorter period of 5–10 years would be more reasonable, considering air pollution caused by the rapid industrial development in the recent years. Without scientific basis, however, either data length advised for calibration is subject to speculation. In contrast, this research theoretically identified and explained the change of relationship between solar radiation and sunshine duration in China, and showed vividly that the Angstrom model performs best when calibrated with a 10-year data set. In view of

Table 3
Effect of the calibration data length on simulation accuracy.

	Calibration								Validation					
	a	*b*	*NSE*	*MAPE*	*RMSE*	*%MAPE*	*%RMSE*	*n*	*NSE*	*MAPE*	*RMSE*	*%MAPE*	*%RMSE*	*n*
a. 40 Year: 1968–2007 data set for calibration; 2008–2010 data set for validation														
Fuzhou	0.149	0.568	0.849	18.644	33.890	13.420	24.395	14,175	0.893	18.064	28.034	12.463	19.341	1094
Nanjing	0.137	0.554	0.890	16.515	28.314	11.602	19.891	14,534	0.906	16.304	25.775	11.101	17.550	1096
Guiyang	0.147	0.562	0.819	23.477	34.203	20.538	29.921	14,488	0.777	27.268	37.526	22.155	30.489	1095
Haerbin	0.196	0.503	0.817	14.621	30.082	9.759	20.078	14,515	0.945	12.348	19.183	8.486	13.183	1096
Xining	0.185	0.562	0.877	12.765	29.774	7.021	16.377	14,587	0.923	10.322	21.719	5.634	11.855	1095
Lhasa	0.269	0.518	0.714	11.168	31.750	4.785	13.602	13,271	0.899	6.482	19.546	2.716	8.190	1094
Average			0.828	16.198	31.336	11.188	20.711	14,261	0.891	15.131	25.297	10.426	16.768	1095
b. 30 Year: 1978–2007 data set for calibration; 2008–2010 data set for validation														
Fuzhou	0.154	0.569	0.839	18.404	34.789	13.190	24.934	10,893	0.901	17.564	27.061	12.118	18.670	1094
Nanjing	0.138	0.536	0.904	15.572	25.446	11.250	18.384	10,885	0.908	15.640	25.476	10.649	17.346	1096
Guiyang	0.152	0.564	0.814	23.752	34.526	20.852	30.311	10,836	0.790	26.915	36.341	21.868	29.526	1095
Haerbin	0.209	0.480	0.876	14.534	30.014	9.818	20.275	10,909	0.945	12.148	19.151	8.348	13.161	1096
Xining	0.189	0.538	0.875	12.603	29.294	7.151	16.621	10,934	0.915	10.735	22.760	5.859	12.423	1095
Lhasa	0.276	0.495	0.747	11.429	31.628	5.029	13.918	10,715	0.890	6.830	20.344	2.862	8.525	1094
Average			0.843	16.049	30.950	11.215	20.740	10,862	0.891	14.972	25.189	10.284	16.609	1095
c. 20 Year: 1988–2007 data set for calibration; 2008–2010 data set for validation														
Fuzhou	0.161	0.553	0.828	17.501	34.919	12.610	25.161	7242	0.899	17.096	27.269	11.795	18.813	1094
Nanjing	0.144	0.530	0.907	15.245	24.871	10.896	17.776	7237	0.900	16.300	26.554	11.098	18.080	1096
Guiyang	0.163	0.574	0.822	23.789	34.233	20.356	29.293	7186	0.819	24.983	33.783	20.298	27.448	1095
Haerbin	0.236	0.456	0.889	14.014	28.495	9.290	18.890	7268	0.946	11.473	19.075	7.885	13.109	1096
Xining	0.206	0.539	0.925	10.408	22.310	5.805	12.444	7282	0.926	10.114	21.275	5.520	11.612	1095
Lhasa	0.279	0.485	0.735	12.111	32.781	5.189	14.044	7094	0.884	7.001	20.905	2.934	8.760	1094
Average			0.851	15.511	29.602	10.691	19.601	7218	0.896	14.494	24.810	9.922	16.304	1095
d. 10 Year: 1998–2007 data set for calibration; 2008–2010 data set for validation														
Fuzhou	0.158	0.559	0.913	15.374	24.738	10.962	17.638	3638	0.900	17.134	27.130	11.821	18.717	1094
Nanjing	0.154	0.536	0.893	15.772	26.806	11.076	18.825	3650	0.919	14.292	23.983	9.731	16.330	1096
Guiyang	0.177	0.587	0.866	22.412	29.203	18.216	23.736	3626	0.842	23.106	31.523	18.773	25.612	1095
Haerbin	0.234	0.444	0.940	12.110	19.969	8.215	13.546	3623	0.943	11.622	19.504	7.987	13.404	1096
Xining	0.218	0.531	0.936	9.440	20.346	5.281	11.382	3642	0.924	9.767	21.285	5.331	11.618	1095
Lhasa	0.286	0.521	0.900	6.338	19.298	2.715	8.268	3470	0.955	6.566	19.865	2.751	8.324	1094
Average			0.908	13.574	23.393	9.411	15.566	3608	0.914	13.748	23.882	9.399	15.667	1095

Note: Average value only given to the evaluation criteria, while the coefficients of *a* and *b* are not averaged in the table as they are meaningless for comparison.

climatic statistics, a shorter data length for model calibration might result in unstable parameters, leading to unreliable prediction outputs. However, results from many previous researches on the Angstrom model have indicted that stable and reliable parameters could be obtained with data length fewer than 10 years for calibration [e.g. 4, 13, 20]. Liu et al. [22] have explored the effect of calibration data length on the stabilization of the parameters in the Angstrom model, finding that for most of the study locations the stable parameters could be obtained with data length longer than 5 years. Calibration with 20-year data length did not make the parameters more stable, but led to abrupt changes in both parameters *a* and *b* of Minqin station [22]. Comparison of the parameters *a* and *b* calibrated with different data lengths (Table 3) also indicates clearly that the parameters calibrated with a 10-year data set are similar to those calibrated with 20- or 40-years in each study station, and no obvious abrupt change occurred in either of the parameters *a* and *b*. So, 10 year data set can be considered to be long enough to retrieve the stable and reliable parameters for model calibration, with which the model performs best. Therefore, period of 10 years is recommended as the national standard for Angstrom calibration in China.

Though the main objectives of the study have been achieved, we must acknowledge that there are still some uncertainties unresolved in the research work. First, radiation in Lhasa on the Tibetan Plateau also has decreased in the recent years, certifying the obvious changes in the relationship between solar radiation and sunshine duration. However, up to now, no explanation can be given on this topic. An increasing rather than a decreasing trend in visibility can be found in Fig. 8, which is in agreement with the results given by Yang et al. [40]. A further investigation should be made to identify the cause of changes in solar radiation, sunshine duration and their relationship in this unique alpine region. Secondly, 10 years were used for the time interval for the moving regression in this study, as the shorter length of the data set, e.g. 2 or 3 years, contains fewer samples for linear regression, which may lead to unreliable fitting parameters (see Fig. 3) [22]. On the other hand, longer lengths of the data set will obscure the features of variation in the parameters. Different time intervals of 5, 8, 12, 15 years were also used to duplicate the moving linear regression method, and resulted findings were very similar to Fig. 5–7 (not shown). Even so, further research is still needed to investigate the influence of the time intervals on the statistical results made by the moving linear regression method.

In addition, the "troughs" in *r* time series was interpreted as the effects of transition from "clean" to "dirty" sky conditions, under "dimming" and "brightening" context. However, all such abrupt variations also might be connected with long term oscillations of the climate of earth, such as ENSO, NAO, PNA, ElNino etc. Therefore, further investigations on the effect of long term oscillations of the climate of earth on the variations in *r* are essentially important. Finally, we cautiously envisage that the relationship between solar radiation and sunshine duration might have also changed in the other regions under global "dimming" and "brightening" context, especially the quickly developing regions such as India and Brazil [34,41]; this means that an analogous research should be important and necessary in other parts of the world.

5. Summary and conclusions

Solar radiations under different climate conditions in six large cities in China were selected to detect the changes in the relationship between solar radiation and sunshine duration in recent decades. Analysis of solar radiation and its related meteorological variables indicated that Lhasa has the highest solar radiation with average annual value of 20.1 $MJm^{-2}d^{-1}$, while Guiyang the lowest of 10.5 $MJm^{-2}d^{-1}$. Both solar radiation and sunshine duration have decreased in the recent decades, but at different rates. An obvious "trough" can be found in the variation of the temporal correlation coefficient between solar radiation and sunshine duration between 1980 and 1995. However, the decrease in solar radiation and sunshine duration, together with the obvious changes in the correlation coefficient, cannot be attributed to changes in cloud cover and rainy days, as both of them also decreased during the same period. Further investigation of the changes in the visibility has identified that the atmospheric conditions have changed over the large cities of China, due to "global dimming" effect [42] caused by the increasing aerosol concentrations. The relationships between solar radiation and sunshine duration have changed in the 1980's and have recovered during the recent decades over the large cities of China, due to the changed atmospheric conditions caused by the rapidly increasing industrial pollutions after China's Open Door policy in the 1980s.

The findings mentioned above imply that the Angstrom model should be calibrated with a data set collected in the recent years, instead of the longest data set possible. Influence of the data set length on the model accuracy was investigated, and the Angstrom model performed best when calibrated with a 10-year data set, while it performed worst with a 40-year calibration data set. Thus, a 10-year data set is recommended as the national standard for Angstrom model calibration in large cities of China.

There are still some uncertainties unresolved in the research work. The relationship has changed in Lhasa, on the Tibetan Plateau. However, the reason is not yet clear. In addition, further investigations on the effect of long term oscillations of the climate of earth on the variations in *r* are essentially important, and further investigations are encouraged to explore whether the relationship has also altered in other fast developing industrial regions of the world.

Acknowledgment

This research is jointly supported by the National Science Foundation of China (No. 41105079) and Special Key Project on "Estimation and Prediction of the Solar Radiation in China" (No. GYHY201306048). The authors sincerely thank the anonymous referees for their strict but instructive comments on the manuscript. Special thanks are also extended to the weathermen at Lhasa station on the Tibetan Plateau under altitude hypoxia conditions.

References

[1] Dincer I. Renewable energy and sustainable development: a crucial review. Renew Sust Energy Rev 2000;4:157–75.
[2] Chen R, Kang E, Ji X, Yang J, Wang J. An hourly solar radiation model under actual weather and terrain conditions: a case study in Heihe river basin. Energy 2007;32:1148–57.
[3] Shazly SME. Estimation of hourly and daily global solar radiation at clear days using an approach based on modified version of gaussian distribution. Adv Atmos Sci 1996;13:349–58.
[4] Podesta GP, Nunez L, Villanueva CA, Skansi MA. Estimating daily solar radiation in the argentine pampas. Agric For Meteorol 2004;123:41–53.
[5] Brock TD. Calculating solar radiation for ecological studies. Ecol Model 1981;14:1–19.
[6] Morcette JJ, Mozdzynski G, Leutbecher M. A reduced radiation grid for the ECMWF integrated forecasting system. Mon Wea Rev 2008;136:4760–72. http://dx.doi.org/10.1175/2008MWR2590.1.
[7] Manners J, Thelen C, Petch J, Hill P, Edwards JM. Two fast radiative transfer methods to improve the temporal sampling of clouds in numerical weather prediction and climate models. Quart J R Meteorol Soc 2009;135:457–68.
[8] Linacre E. Climate data and resources. a reference and guide. NY: Routledge; 1992. p. 366.
[9] Ogelman H, Ecevit A, Tasdemiroglu E. A new method for estimating solar radiation from bright sunshine data. Sol Energy 1984;33:619–25.
[10] Bahel V, Bakhsh H, Srinivasan R. A correlation for estimation of global solar radiation. Energy 1987;12:131–5.

[11] Psiloglou BE, Kambezidis HD. Performance of the meteorological radiation model during the solar eclipse of 29 march 2006. Atmos Chem Phys 2007;7:6047–59.
[12] Liu DL, Scott BJ. Estimation of solar radiation in Australia from rainfall and temperature observations. Agric For Meteorol 2001;106:41–59.
[13] Chen R, Kang E, Yang J, Lu S, Zhao W. Validation of five global radiation models with measured daily data in China. Energy Consers Manage 2004;45: 1759–69.
[14] Wu GF, Liu YL, Wang TJ. Methods and strategy for modeling daily global radiation with measured meteorological data-a case study in Nanchang station, China. Energy Convers Manage 2007;48:2447–52.
[15] Trnka M, Zalud Z, Eitzinger J, Dubrovsky M. Global solar radiation in central European lowlands estimated by various empirical formulae. Agric For Meteorol 2005;131:54–76.
[16] Liu JD, Liu JM, Linderholm HW, Chen DL, Yu Q, Wu DR, et al. Observation and calculation of the solar radiation on the Tibetan Plateau. Energy Convers Manage 2012;57:23–32.
[17] Angstrom A. Solar and terrestrial radiation. Quart J R Meteorol Soc 1924;50: 121–6.
[18] Prescott JA. Evaporation from a water surface in relation to solar radiation. Trans Roy Soc Aust 1940;64:114–25.
[19] Allen RG, Pereira LS, Raes D, Smith M. Crop evapotranspiration-guidelines for computing crop water requirements. FAO irrigation and drainage paper 56. Rome: Food and Agriculture Organization of the United Nations; 1998.
[20] Li H, Ma W, Lian Y, Wang X, Zhao L. Global solar radiation estimation with sunshine duration in Tibet, China. Renew Energy 2011;36:3141–5.
[21] Kambezidis HD, Adamopoulos AD, Zevgolis D. Determination of Angstrom and Schuepp's parameters from ground-based spectral measurements of beam irradiance in the ultraviolet and visible spectrum in Athens, Greece. Pure Appl Geophys 2001;158:821–38.
[22] Liu XY, Mei XR, Li YZ, Zhang YQ, Wang QS, Jensen JR, et al. Calibration of the Angstrom-Prescott coefficients (a, b) under different time scales and their impacts in estimating global solar radiation in the Yellow River basin. Agric For Meteorol 2009;149:697–710.
[23] Liu XY, Mei XR, Li YZ, Porter JR, Wang QS, Zhang YQ. Choice of Angstrom-Prescott coefficients: are time-dependent ones better than fixed ones in modeling global solar irradiance? Energy Convers Manage 2010;51:2565–74.
[24] Hans VS, Francos WZ. Statistical analysis in climate research. Cambridge: Cambridge University Press; 1999.
[25] Wild M. Global dimming and brightening: a review. J Geophys Res 2009;114: D00D16. http://dx.doi.org/10.1029/2008JD011470.
[26] Wild M. Enlightening global dimming and brightening. Bull Amer Meteor Soc 2012;93:27–37.
[27] Qian Y, Wang W, Leung LR, Kaiser DP. Variability of solar radiation under cloud-free skies in China: the role of aerosols. Geophys Res Lett 2007;34: L12804. http://dx.doi.org/10.1029/2006GL028800.
[28] Gueymard CA. Importance of atmospheric turbidity and associated uncertainties in solar radiation and luminous efficacy modeling. Energy 2005;30: 1603–21.
[29] Persand N, Lesolle D, Ouattara M. Coefficients of the Angstrom-Prescott equation for estimating global irradiance from hours of bright sunshine in Botswana and Niger. Agric For Meteorol 1997;88:27–35.
[30] Domros M, Peng GB. The climate of China. Berlin: Springer-Verlag; 1988.
[31] Kaiser DP. Decreasing cloudiness over China: an updated analysis examinging additional variables. Geophys Res Lett 2000;27:2193–6.
[32] Qian Y, Kaiser DP, Leung LR, Xu M. More frequent cloud-free sky and less surface solar radiation in China from 1955 to 2000. Geophys Res Lett 2006;33:L01812.
[33] Wang YW, Wang YH, Han SM, Wang QX, Zhang JH. Sunshine dimming and brightening in Chinese cities (1955–2011) was driven by air pollution rather than clouds. Clim Res 2013;56:11–20.
[34] Soni VK, Pandithurai G, Pai DS. Evaluation of long-term changes of solar radiation in India. Int J Climatol 2012;32:540–51.
[35] Kaiser DP, Qian Y. Decreasing trends in sunshine duration over China for 1954-1998: Indication of increased haze pollution? Geophys Res Lett 2002;29:2042. http://dx.doi.org/10.1029/2002GL016057.
[36] Wu J, Fu C, Zhang L, Tang J. Trends of visibility on sunny days in China in the recent 50 years. Atmos Environ 2012;55:339–46.
[37] Kambezidis HD, Kaskaoutis DG. Aerosol climatology over four AERONET sites: an overview. Atmos Environ 2008;42:1892–906.
[38] Zhu ZR. Atlas of solar energy distribution in China. Beijing: Meteorological Press; 2006 [in Chinese].
[39] Wang BZ. Calculation of the solar energy resources in China. Beijing: Meteorological Press; 2008 [in Chinese].
[40] Yang K, Ding B, Qin J, Tang W, Lu N, Lin C. Can aerosol loading explain the solar dimming over the Tibetan Plateau? Geophys Res Lett 2012;39:L20710. http://dx.doi.org/10.1029/2012GL053733.
[41] Santos CMD, Souza JLD, Junior RAF, Tiba C, Melo RO, Lyra GB, et al. On modeling global solar irradiation using air temperature for Alagoas state, Northeastern Brazil. Energy 2014;71:388–98.
[42] Kambezidis HD, Kaskaoutis DG, Kharol SK, Moorthy KK, Satheesh SK, Kalapureddy MCR, et al. Muti-decadal variation of the net downward shortwave radiation over south Asia: the solar dimming effect. Atmos Environ 2012;50:360–72.