Solar radiation analysis and regression coefficients for the Vhembe Region, Limpopo Province, South Africa

Sophie T Mulaudzi
Department of Physics, University of Venda

Vaithianathaswami Sankaran
Department of Physics, University of Venda

Meena D Lysko
Optronic Sensor Systems, Defence Peace Safety Security, Council for Scientific and Industrial Research

Abstract
Given the limited observed and reliable data for solar irradiance in rural parts in South Africa, a correlation equation of the Angström-Prescott linear type has been used to estimate the regression coefficients in the Vhembe District, Limpopo Province, South Africa. Five stations were selected for the study, with the greatest distance between stations less than 180 km. Monthly regression coefficients were derived for each station based on an observation dataset of sunshine duration hours and global horizontal irradiance. The correlation coefficients appear to be above 0.9. The representative Angström-Prescott model for the Vhembe Region was found by collating the data for each station and then averaging the respective correlation coefficients. This paper presents the generated regression coefficients for each station and for the Vhembe Region.

Keywords: South Africa, global horizontal irradiance, Angstrom Prescott, linear regression, solar radiation

1. Introduction

Limpopo Province is situated in the north-eastern part of South Africa and it incorporates about 10% of South Africa's total land. Over 75% of households within this province rely on alternative fuel for cooking and heating (El-Sebaii *et al.*, 2005) of which less than 1% is reliant on solar systems; even though the province has solar energy potential to support all of the province's energy demand. It is identified that the quantitative knowledge of solar surface irradiance would better help in initiatives such as the leveraging of financial support for solar energy projects, for designing of efficient solar energy systems and for the planning of cost effective maintenance solutions.

An attempt is now made for a quantitative knowledge of the solar resource. The limited in-situ observations for solar irradiance data across the province makes it necessary to consider model predictions. Modelling provides the means for generating global solar irradiation predictions at different sites where the measured data is unavailable. This paper reports on the regression coefficients of the Angström-Prescott linear regression model to predict average daily global solar radiation for one of the five regions in the Limpopo Province, namely the Vhembe Region.

2. Methodology
2.1 Regression model
The total solar irradiance (H_T) on a horizontal surface at ground is the sum of the sun's direct beam () ($B_{Horizontal}$) and sky diffuse ($D_{Horizontal}$) solar irradiance. That is:

$$H_T = D_{Horizontal} + B_{Horizontal} \qquad (1)$$

Moreover, with attenuation by traversing through the atmosphere, one may express equation (1) in a relative form with the extra-terrestrial irradiance H_o. The ratio H/H_o is referred to as the clearness index. In a similar manner, the sun's direct beam reaching the surface is the cloudless index given by the ratio of instantaneous actual sunshine hours (N_a) to a

predicted possible sunshine hours (N_p). Equation (1) is then represented as:

$$\frac{H}{H_o} = D_{Horizontal} + \alpha \cdot \frac{N_a}{N_p} \tag{2}$$

The factor α is introduced to account for atmospheric scattering and attenuation by air particles such as water, dust, aerosols and gases. In its simplest form, with $D_{Horizontal}$ and α as regression parameters, Equation (2) leads to the basic Angström-Prescott model to predict solar irradiance at any location. So, estimation of the monthly average daily global solar radiation ($\overline{H}$, MJ/m^2) on a horizontal surface, at a particular site, is given by Angström-Prescott (Angstrom, 1924) and (Iqbal, 1983) as:

$$\frac{\overline{H}}{\overline{H_o}} = a + b \cdot \frac{\overline{N_a}}{\overline{N_p}} \tag{3}$$

where a and b are now the regression coefficients to be determined, $\overline{H_o}$ is the monthly average horizontal extra-terrestrial solar radiation. $\overline{N_a}$ and $\overline{N_p}$ are daily average duration of the actual and predicted possible sunshine hours respectively. The parameters $\overline{N_p}$ and $\overline{H_o}$ are computed according to the deterministic expressions given in, for example (Iqbal, 1983), (Garg *et al.*, 2002) and (Rai, 2009).

2.2 Data accessibility
It follows from equation (3) that a linear regression analysis between clearness index and cloudless index will directly lead to the estimation of the Angström-Prescott coefficients a and b. $\overline{H_o}$ and $\overline{N_p}$ have fixed values for a geographic location, as shown. $\overline{H}$ and $\overline{N_p}$ vary temporally and depend on the site specific climatic conditions. They are gotten from observation with pyranometers and sunshine recorders, respectively.

$\overline{N_a}$ for the stations in this study was obtained from the South African Weather Services (SAWS). The SAWS network uses Campbell Stokes sunshine recorders to measure the actual sunshine hours. $\overline{H}$ data is from the Agricultural Research Council (ARC). The ARC and SAWS networks do not overlap so 5 locations were selected such that the ARC/SAWS pairs are in close proximity to one another.

The data set considered is from 1 January to 31 December, 2010. This data set was quality checked to remove outliers, which included removing data points if the clearness index was greater than 1, as well as data points with little or no correlation between observed sunshine duration and global horizontal solar irradiance.

2.3 Statistical analysis
The generated regression equations for each site were used to calculate the monthly average daily global solar radiation. These values of global radiation were then compared against the measured data for the locations. The accuracy of the estimated horizontal global solar radiation data against the 'true data' is tested by calculating the mean bias error (MBE), root mean squared error (RMSE) and t-statistic. The formulas for the statistics indicators (Togrul, 1998) and (Stone, 1993) are given by equations (3), (4) and (5), where is the number of data pairs observed and the subscript i ranges from 1 to n. The use of the additional t-statistic is used to assess the reliability of the results. This is because it is possible to have a large RMSE value whilst at the same time a very small MBE. To determine whether the model's estimates are statistically significant, a critical *t-value* at $a = 95\%$ confidence level is taken from the standard statistical tables. That is, $t_{critical} = 1.96$. In order to analyse the model's estimates at $(1 - \alpha)$ confidence level the calculated t must be such that $-1.96 \le t \le 1.96$.

$$MBE = \frac{1}{n}\sum_i^n (\overline{H}_{estimated_i} - \overline{H}_{observed_i}) \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_i^n (\overline{H}_{estimated_i} - \overline{H}_{observed_i})^2}{n}} \tag{5}$$

$$t = \sqrt{\frac{(n-1).MBE^2}{RMSE^2 - MBE^2}} \tag{6}$$

3. Study area
The Vhembe Region and selected study points are shown in Figures 1 and 2, respectively.
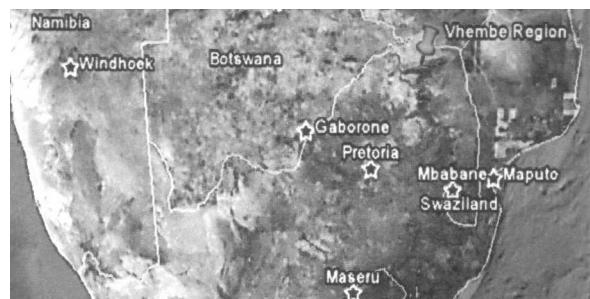

Figure 1: Vhembe Region in South Africa

Figure 2: Locations of the five study points

The study points are: Alldays, Rabali, Mutale, Mhinga and Mulima (see Table 1 for their geographical location). Alldays, which is just outside the Vhembe Region, is included for comparison.

Table 1: Geographical locations of the selected stations

Station	Latitude	Longitude	Altitude (m)
Alldays	22.68264°S	29.10466°E	842
Mutale	22.73461°S	30.52188°E	550
Mhinga	22.79675°S	30.84219°E	460
Rabali	22.87038°S	30.08112°E	720
Mulima	23.35612°S	29.98923°E	1032

4 Results and discussions

4.1 Actual versus possible sunshine hours

Figure 3 shows $\overline{N}_a$ and $\overline{N}_p$ for the Vhembe Region. In the Vhembe region, we generally observe winter season during the months of May, June and July. During this period we normally expect lower actual sunshine hours than the other seasons. But the actual sunshine hours is rather more during the year under investigation in winter and this may be attributed to limited absolute clear sky days for that particular year. In addition to that it is always preferable to calibrate the measuring instruments on regular basis in order to get a reliable set of data. Unfortunately, the calibration records are not available for the area under study.

4.2 Station specific linear regressions

The ratios of clearness and cloudless indices for the accumulated data sets (that is, data for an annum) are plotted in Figure 4 for the five sites in this study. The figure also shows the coefficient of determination (R^2) for the 1st order linear curve fits for each of the sites. $R^2 \geq 0.9$ for over 80 % of the linear regressions. One may assume $R^2 < 0.8$ as an indication that the investigated data set is questionable in terms of reliability. The reliability of the data set will depend on the implementation of calibration of the pyranometers and general maintenance of the observing instruments. It is noted that only data with $R^2 < 0.5$ have been excluded from the analysis in this work in an effort to investigate a larger data set.

The annual regression coefficients for all the sites have been determined by the linear regression curves to the data sets for the respective sites. The derived annual regression coefficients are given in Table 2. The table also shows the calculated linear regression coefficients for all stations combined as well as the respective standard deviations. The standard deviations suggest a combined uncertainty of $a \pm 20\%$ and $b \pm 11\%$, which may be acceptable given the uncertainties of raw data. It follows that the annual average linear regression coefficients may be considered representative coefficients for the Vhembe Region. That is, the Angström-Prescott linear regression model for the Vhembe Region in the Limpopo Province may be represented as:
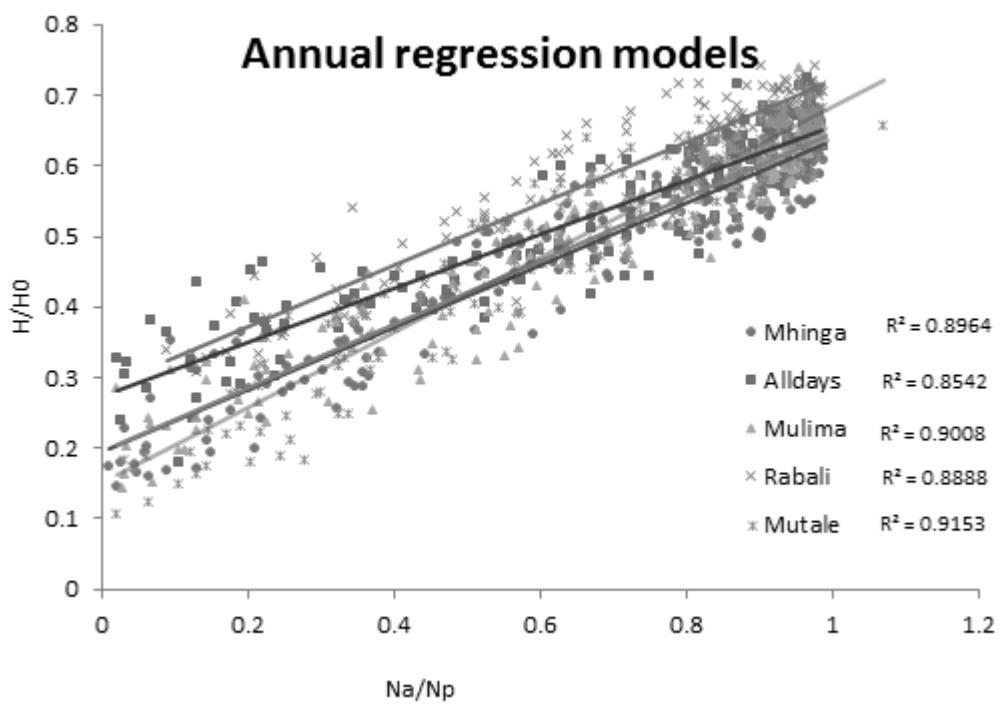
$$\frac{\overline{H}}{\overline{H}_o} = 0.2 + 0.5 \cdot \frac{\overline{N}_a}{\overline{N}_p} \tag{7}$$

Table 2: Annual regression coefficients

Station	a_{annual}	b_{annual}
Mutale	0.2	0.5
Mhinga	0.2	0.4
Rabali	0.3	0.4
Alldays	0.2	0.5
Mulima	0.2	0.5
Average, stations combined	0.22	0.46
Standard deviation, stations combined	0.04	0.05
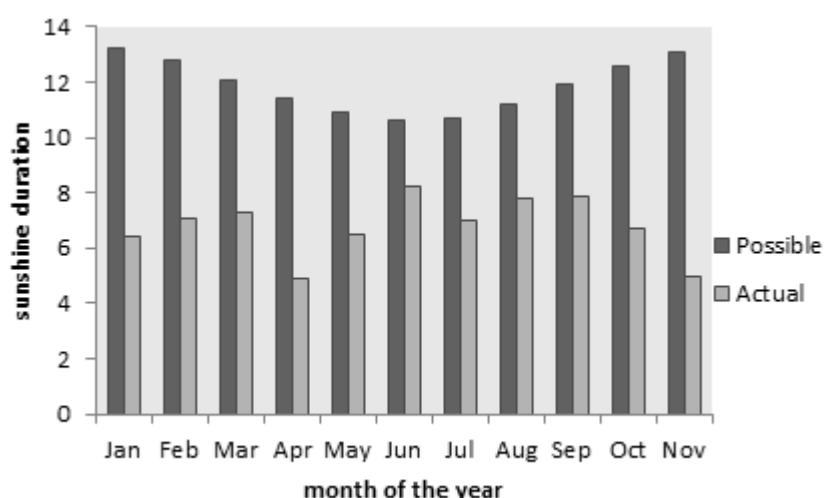
Figure 3: Possible and actual sunshine duration for Vhembe district

Figure 4: Annual regression models
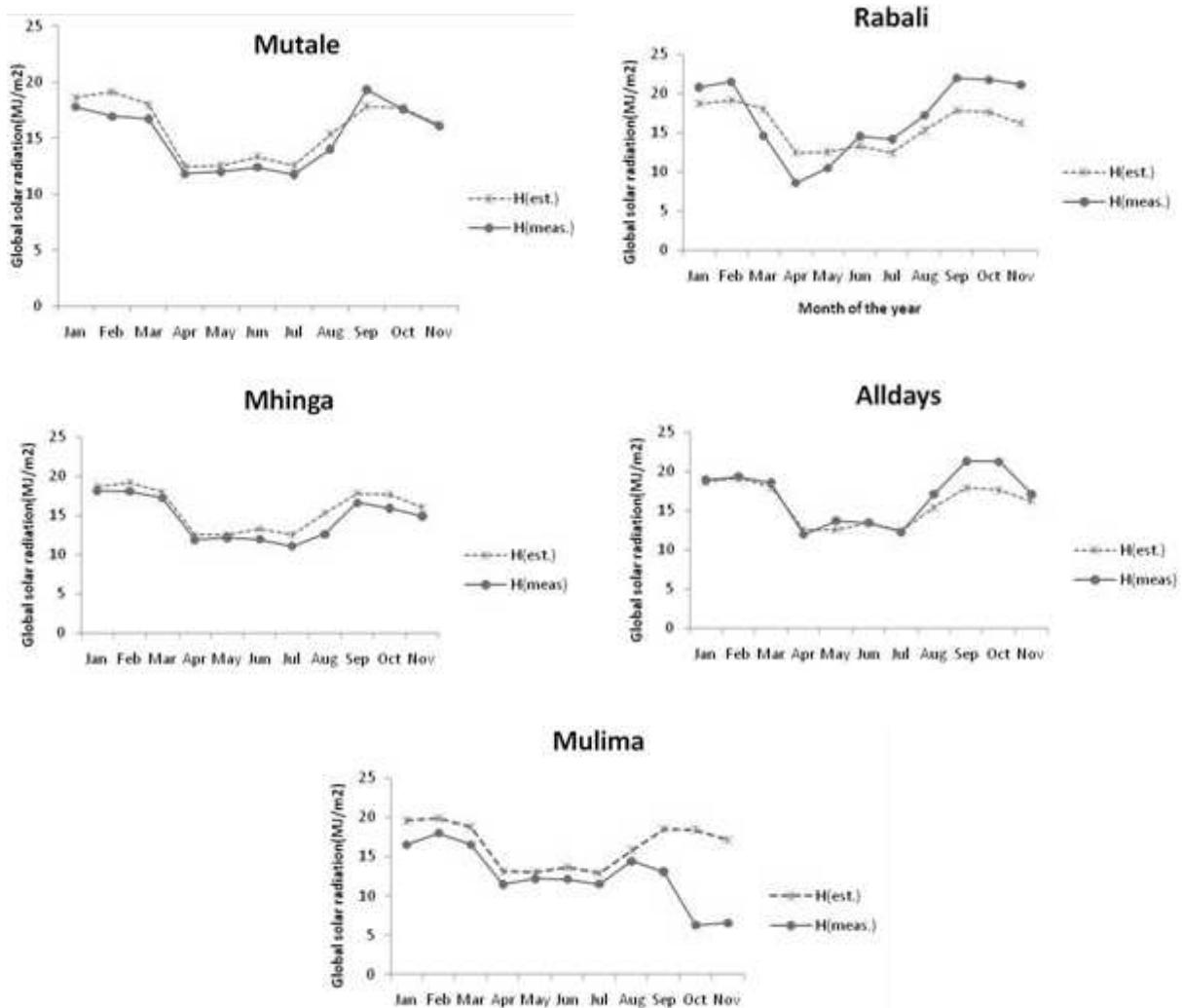
Figure 5: Measured versus estimated irradiances

4.3 Validation of results

The developed model shown as equation (7) is used to estimate the horizontal global solar radiation for the five stations in this study. The respective estimated horizontal global solar radiation is then compared with the measured values. The $\overline{H}_{estimated}$, $\overline{H}_{measured}$, MBE, RMSE and t-statistic for each of the stations is given in Table 3 (with $t_{critical}$ at 5 % = 1.96). The comparison of observed to estimated horizontal global solar radiation for the five locations is given by the plots in Figure 5. It is assumed that the larger variations (see, for example, Mulima from September to November) may be due to discrepancies in the raw data sets and subsequent filtering process. Also, one may expect larger deviations between $\overline{H}_{estimated}$ and $\overline{H}_{measured}$ when the cloudless indices are very low.

Table 3: Goodness of fit indicators

Mutale	$\overline{H}_{measured}$	$\overline{H}_{estimated}$	MBE	RMSE	t	Mhinga	$\overline{H}_{measured}$	$\overline{H}_{estimated}$	MBE	RMSE	t
Jan	18.60	17.80	0.03	0.14	1.00	Jan	18.68	18.15	0.02	0.10	1.00
Feb	19.11	16.96	0.08	0.41	1.00	Feb	19.20	18.06	0.04	0.22	1.05
Mar	17.99	16.71	0.04	0.23	1.00	Mar	17.99	17.24	0.02	0.13	1.00
Apr	12.49	11.82	0.02	0.12	1.00	Apr	12.48	11.91	0.02	0.10	1.02
May	12.56	11.98	0.02	0.10	1.00	May	12.54	12.16	0.01	0.07	1.00
Jun	13.34	12.39	0.03	0.17	1.00	Jun	13.32	11.95	0.05	0.25	1.02
Jul	12.53	11.75	0.03	0.14	1.00	Jul	12.52	11.10	0.05	0.25	1.00
Aug	15.34	14.01	0.04	0.24	1.00	Aug	15.32	12.64	0.09	0.48	1.00
Sep	17.84	19.33	-0.05	0.27	1.00	Sep	17.83	16.62	0.04	0.22	1.02
Oct	17.68	17.59	0.00	0.02	1.00	Oct	17.68	15.92	0.06	0.32	1.00
Nov	16.21	16.05	0.01	0.03	1.00	Nov	16.04	14.91	0.04	0.21	1.02

Rabali	$\overline{H}_{measured}$	$\overline{H}_{estimated}$	MBE	RMSE	t	Alldays	$\overline{H}_{measured}$	$\overline{H}_{estimated}$	MBE	RMSE	t
Jan	18.68	20.77	-0.07	0.37	1.00	Jan	18.67	18.92	-0.01	0.04	1.00
Feb	19.11	21.47	-0.08	0.45	1.00	Feb	19.11	19.32	-0.01	0.04	1.04
Mar	17.98	14.62	0.11	0.60	1.00	Mar	18.00	18.59	-0.02	0.11	1.00
Apr	12.47	8.61	0.13	0.70	1.00	Apr	12.50	11.94	0.02	0.10	1.00
May	12.52	10.50	0.07	0.36	1.00	May	12.57	13.69	-0.04	0.20	1.00
Jun	13.29	14.54	-0.04	0.23	1.00	Jun	13.35	13.47	0.00	0.02	1.00
Jul	12.42	14.16	-0.06	0.31	1.00	Jul	12.47	12.26	0.01	0.04	1.00
Aug	15.30	17.20	-0.06	0.34	1.00	Aug	15.35	17.06	-0.06	0.31	1.00
Sep	17.82	21.94	-0.14	0.75	1.00	Sep	17.84	21.26	-0.11	0.62	1.00
Oct	17.60	21.74	-0.13	0.74	1.00	Oct	17.60	21.23	-0.12	0.65	1.00
Nov	16.22	21.11	-0.16	0.89	1.00	Nov	16.21	17.10	-0.03	0.16	1.00

Mulima	$\overline{H}_{measured}$	$\overline{H}_{estimated}$	MBE	RMSE	t
Jan	19.57	16.53	0.10	0.31	1.00
Feb	19.90	17.97	0.07	0.26	1.00
Mar	18.71	16.52	0.07	0.27	1.00
Apr	13.11	11.44	0.06	0.24	1.00
May	12.95	12.18	0.02	0.16	1.00
Jun	13.63	12.10	0.05	0.23	1.00
Jul	12.88	11.51	0.04	0.21	1.00
Aug	15.80	14.39	0.05	0.21	1.00
Sep	18.48	13.08	0.18	0.42	1.00
Oct	18.42	6.26	0.39	0.63	1.00
Nov	17.15	6.56	0.34	0.58	0.97

The goodness of fit statistics, shown in Table 3, demonstrates that the derived linear correlation coefficients may be used in the Angström-Prescott model to estimate monthly average daily global solar radiation on a horizontal surface, for the respective locations in this study. All regressions are within the range of $t_{critical} = 1.96$.

Overall, the determined annual regression coefficients for the Vhembe Region appear consistent with the reports such as for the locations: Toledo (Almorox *et al.,* 2004) (where $a = 0.2170$ and $b = 0.5453$), Lesotho (Safari et al, 2009) (where $a = 0.266$ and $b = 0.512$), and Bulawayo (Gopinathan, 1988) (where $a = 0.304$ and $b = 0.440$). Moreover, Pereira (Pereira, 1988) has recommended general values of a and b to be 0.250 and 0.50 respectively, for any location and climate.

5. Summary

It may be concluded from the presented work that the monthly average daily global solar radiation on a horizontal surface for any location in the Vhembe Region may be estimated using the re-parameterized Angström-Prescott linear regression model as given by equation (7). Further investigation on the quality of the raw data and consideration of a longer observation period is needed to improve the confidence of the model.

References

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M., (1988). Crop Evapotranspiration. Guidelines for computing crop water requirement. FAO Irrigation and drainage paper 56. Rome.

Almorox, J., and Hontoria, C., (2004). Global solar radiation estimation using sunshine duration in Spain. *Energy conservation and management*, Vol. 45: 1529 – 1535. Elsevier.

Angstrom, A., (1924). Solar and terrestrial radiation. *Q.J.R.Meteorol. Soc,* Vol. 50: 121-125.

El-Sebaii, A. A., and Trabea, A. A., (2005). Estimation of global solar radiation on horizontal surfaces over Egypt. *J.Solids*, Vol. 28 (1).

Garg, H. P. and Prakash, J., (2002). Solar Energy fundamentals and applications. Tata McGraw-Hill Publishing Company Limited, New Delhi.

Gopinathan, K. K., (1988). A general formula for computing the coefficients of the correlation connecting global solar radiation to sunshine duration. *Solar Energy*, Vol. 41(6): 499-50

Iqbal, M., (1983). Introduction to solar radiation. New York: Academic Press.

Rai, G. D., (2009). Solar Energy Utilization. Khanna Publishers

Safari, B., and Gasore, J., (2009). Estimation of global solar radiation in Rwanda using empirical models. *Asian Journal of Scientific Research*, Vol. 2 (2): 68-75.

Statistics South Africa, (2001). Census 2001: Primary tables Limpopo: Census '96 and 2001 compared, Report No. 03-02-11.

Stone, R. J., (1993). Improved statistical procedure for the evaluation of solar radiation estimation models. *Solar Energy* Vol. 51: 289.

Togrul, I. T., (1998). Comparison of statistical performance of seven sunshine-based models for Elazig, Turkey, *Chemica Acta Turuca*, 26, 37.

Togrul, I. T., (1998). Comparison of statistical performance of seven sunshine-based models for Elazig, Turkey, *Chemica Acta Turuca*, 26, 37.