## Estimation of Global Radiation using Clear-Sky Model at Selected Sites in Kenya


Ronno Cosmas
*School of Science, Department of Physics, University of Eldoret, P.O. BOX 1125, Eldoret*
*ronno2132@gmail.com*

Makau Nicholas
*School of Science, Department of Physics, University of Eldoret, P.O. BOX 1125, Eldoret*

Amolo George
*School of Science, Department of Physics, University of Eldoret, P.O. BOX 1125, Eldoret*

Rotich Samuel
*School of Science, Department of Physics, Moi University*



### ABSTRACT

*World energy demand has been outstripping supply since the 1973 Arab/OPEC oil crisis. This has led to a wide interest in, and development of, renewable and sustainable energy including solar energy. Solar radiation data is an important input parameter in the design and implementation of solar energy systems. Only a few scattered meteorological stations in Kenya measure solar radiation on a continuous basis. The use of models to estimate this parameter can alleviate the problem. A clear sky model can be easily obtained from measured global solar radiation on a horizontal surface situate on Earth's surface. Using solar radiation and duration of sunshine data from 11 Kenyan meteorological stations, this study tested seven Angstrom-Prescott type regression models for their suitability to estimate clearness index for clear skies* $(K_c)$*. The Angstrom-Prescott type models were obtained by regressing sunshine duration against clearness index and obtaining curves of best fit. Linear, quadratic, exponential, power and logarithmic fits were obtained. Model performance was measured using goodness of fit statistics including Pearson correlation coefficient (r), coefficient of determination (*$R^2$*), Mean Bias Error (MBE), Root Mean Square Error (RMSE), Students- t-statistic, and the t-test. Out of the 11 stations considered data from Dagoretti, Eldoret, J K Airport, and Voi meteorological stations showed high* $R^2$ *values and these were used to produce modified Angstrom-Prescott models whose long-term, short-term, and overall performances were measured using MBE, RMSE, and t-statistic respectively. For each of the four stations with high* $R^2$ *values 10 pairs of equations, one each for* $K_c$ *presented. These 20 equations are recommended for use in estimating clear-sky clearness indices at the four stations (and in the immediate neighbourhood) using measured fraction of duration of sunshine as the only input. The correlation coefficients for each model were determined and these were found to be site dependent. For example the correlation coefficients for the linear model of the Angstrom type for Dagoretti and Eldoret were 0.316, 0.706 and 0.421, 0.650 respectively while those for the quadratic model for the same two stations were respectively 0.348, 0.579, 0.118 and 0.264, 1.175, -0.426. A number of recommendations on the use of these models are given.*



**Keywords**: Clear Sky Model, Global Radiation, Goodness of Fit Statistics, Angstrom Model, Regression


### INTRODUCTION

In order to convert solar radiation into either thermal or electrical power knowledge of solar radiation received on a surface (along with other essential meteorological parameters) is of primary importance. This knowledge is required in many other diverse applications where solar radiation is a key input. For example, architects and engineers use this information to design homes, offices, factories, etc, for thermal comfort, illumination and fenestration (windows), cooling and air-conditioning. Another example is in agriculture and forestry. The productivity and quality of plant growth is influenced by the diurnal and seasonal variation in solar radiation and illuminance. In horticulture, floriculture, and the emerging field of Urban Agriculture, where greenhouses are used to control growing conditions solar radiation data is a key parameter in optimising plant growth. For energy conversion systems the designer must have detailed knowledge of the temporal and spatial variation of solar radiation in order to size and model the system for optimum performance. Of greatest practical value is continuous long-term measured solar radiation data on an hourly (or even sub hourly) and monthly basis. However, such data is rarely available as equipment and personnel costs are prohibitive. Such impediments have historically led to the development of a number of ways of estimating, modelling and predicting solar radiation at any given location. Most of such approaches are based on correlations between solar radiation and meteorological/geographical parameters such as sunshine duration, cloud cover, atmospheric turbidity, ambient temperature, altitude, latitude and relative humidity on the one hand and modelling of optical-atmospheric interaction on the other.

The need for solar resource data base is increasing every day due to numerous current and emerging solar energy conversion systems covering such diverse areas as heating and cooling of buildings, agricultural and process heat systems, photovoltaic systems technology, solar thermal electric systems, biomass production, photochemistry, solar detoxification and destruction of hazardous wastes, as well as water desalination and distillation. Several empirical formulae have been developed to calculate the solar radiation using various parameters.

The aim of this study was to develop models that estimate the long-term mean monthly global irradiance over Kenya using a Clear-Sky based model. To achieve this the following three objectives were set out. One: establish the nature of correlation between global solar radiation and sunshine duration. Two: Fit the data into various models based on regression of fractional sunshine duration ($\bar{S}/\bar{S}_0$) and Clear-Sky clearness index ($K_c = \bar{H}/\bar{H}_c$) so as to identify the best models to estimate global radiation. Three: Evaluate the reliability and usability of the models.

**Clear-Sky Model**

Hottel (1976) has presented a simple model for the estimation of the transmittance of beam radiation in clear sky conditions. The inputs needed are the altitude of the location (in km above mean sea level), day number of the year ($n$) and the zenith angle ($\theta_z$). Combined with the Liu and Jordan (1960) model for the transmittance of diffuse radiation through clear skies, the clear sky beam and diffuse radiation can be easily computed. The instantaneous clear sky beam radiation on a horizontal surface is:

$$G_b = G_n \tau_b \cos\theta_z \qquad (1)$$

The instantaneous extraterrestrial irradiance $G_n$ is given by:

$$G_n = I_{sc}\left[1 + 0.033\cos\left(\frac{360n}{365}\right)\right] \qquad (2)$$

The atmospheric transmittance for beam radiation $\tau_b$ is given in the form:

$$\tau_b = a_0 + a_1 \exp\left(\frac{-k}{\cos\theta_z}\right) \qquad (3)$$

The constants $a_0$, $a_1$ and $k$ for the standard atmosphere with 23 km visibility are found from $a_0^*$, $a_1^*$ and $k^*$ where:

$$r_0 = \frac{a_0}{a_0^*}, r_1 = \frac{a_1}{a_1^*}, r_k = \frac{k}{k^*} \qquad (4)$$

Eqn.( 4) leads to the following simple result for altitudes less than 2.5 km:

$$a_0^* = 0.4237 - 0.00821(6 - A)^2, \qquad a_1^* = 0.5055 - 0.00595(6.5 - A)^2 \qquad (5)$$

$$k^* = 0.2711 + 0.01858(2.5 - A)^2 \qquad (6)$$

The parameter $A$ is the altitude of the observer in kilometers.

The parameters $r_0$, $r_1$, and $r_k$ in Eqn. (4) are called correction factors and have been calculated for various climate types (Hottel, 1976). The following values are for tropical regions:

$$r_0 = 0.95,\ r_1 = 0.98,\ r_k = 1.02 \qquad (7)$$

The transmittance of diffuse radiation through clear skies can be estimated from the beam radiation transmittance based on the study by Liu and Jordan (1960):

$$\tau_d = 0.271 - 0.2939\tau_b \qquad (8)$$

The instantaneous clear sky diffuse irradiance is given as:

$$G_d = G_n \tau_d \cos\theta_z \qquad (9)$$

Therefore, the instantaneous global horizontal irradiance is given by:

$$G = G_b + G_d = G_n[\tau_b + \tau_d]\cos\theta_z = G_n[0.271 + 0.7066\tau_b]\cos\theta_z \qquad (10)$$

From the above expressions one can integrate $G$ over the hour angle ( $\omega_h$ ) of interest to obtain hourly totals of global radiation by considering the midpoint of the hour angle in the following expression:

$$I_0 = G_{SC}\left[1+0.033\cos\frac{360n}{365}\right]\left[\cos\phi\cos\delta\cos\omega_h + \sin\phi\sin\delta\right] \qquad (11)$$

These calculations are repeated for each hour of the day, to obtain a standard clear day's mean radiation $\overline{H}_c$, and this can be extended to the monthly average of daily values, either by picking an average day in the month or by obtaining the arithmetic mean.

Clear sky models have been validated for, inter alia, measured values (Ineichen, 2016, Engerer and Mills, 2015, Lefevre et al, 2013, , Annear and Wells, 2007) and spectral baseline dependence (Bird and Hulstrom, 1981). Even though clear sky models are widely used in solar radiation research and application no such study has been carried out in connection with solar radiation in Kenya.

## MATERIALS AND METHODS

The irradiance and sunshine data used in this study is archived at the World Radiation Data Centre (WRDC) and retrievable online at the National Renewable Energy Laboratory (NREL) at Golden, Colorado, USA. Detailed instructions for submitting data files are found in Krichak (1987), a WMO document that also serves as a comprehensive booklet for those using the data.

Data available from Kenya's meteorological service has at most two parameters, namely daily sums and monthly mean of daily sums of global solar radiation in units of 0.01 $MJ/m^2$ or $J/cm^2$ and, the monthly sum, as well as monthly mean of daily sums of bright sunshine duration in units of 0.1 hr. For the purpose of this study the meteorological stations selected were those with the longest possible record and also with valid sets of data, namely containing both solar radiation and sunshine duration for the same month. These stations were, Dagoretti (DAG), Eldoret (ELD), Garissa (GAR), JK Airport (JKA), Kisumu (KIS), Kitale (KIT), Lamu (LAM), Lodwar (LOD), Malindi (MAL), Moi Airport Mombasa MOI), and Voi (VOI). The WRDC record for Kenya has radiation data covering various periods with the earliest going back to 1964. However the majority of sunshine data go back to 1969. All WRDC data ends in June as there has been no update since July, 1993.

For each station the clearness index $K_0$ data was used to produce Clear-Sky clearness index $K_c$ which in turn was regressed against fraction of bright sunshine duration and obtaining curves of best fit. The result of this type of regression is Angstrom-type models. An overview of the development of the Angstrom-Prescott-Page equation is found in Martinez-Lozano, Tena, Onrubia, and Rubia (1984).. For Clear-Sky case the linear Angstrom-type models have the form:

$$\overline{H}/\overline{H}_c = a + b(\overline{S}/\overline{S}_0) \qquad (12)$$

Here $\overline{H}$ is the measured monthly mean of daily global (total) radiation falling on a horizontal collector located on Earth's surface while $\overline{H}_c$ is the corresponding value for a surface located at the top of the atmosphere at same latitude and longitude. $\overline{S}$ is the measured monthly mean of daily bright sunshine duration while $\overline{S}_0$ is the corresponding monthly mean of daily daylength (time from sunrise to sunset). The regression constants $a$ and $b$ are in general site dependent (Baker and Haines, 1969; Panoras and Mavroudis, 1994). The Angstrom-type model in this work was extended to include quadratic, exponential, power and logarithmic forms.
The resulting five models were subjected to the following goodness of fit statistics:

**Pearson Correlation Coefficient**: measures the strength of linear dependence between two variables, and can be given as,

$$r_{xy} = \frac{N\sum x_i y_i - \sum x_i y_i}{\sqrt{(N\sum x_i^2 - (\sum x_i)^2).(N\sum y_i^2 - (\sum y_i)^2)}}. \qquad (13)$$

Here $x$ is the independent variable (fraction of bright sunshine $\overline{S}/\overline{S}_0$ for example) and $y$ is the dependent variation (for example Clear-Sky clearness index $\overline{H}/\overline{H}_c$ ). This parameter detects the nature and magnitude of the relationship between the measured values and those predicted by the model.

**Coefficient of Determination $R^2$**: This parameter is just the square of the Pearson coefficient and sometimes goes by the name R-Squared. It measures the success of the regression in predicting the values of the dependent variable within the sample. $R^2$ =1 if the regression fits perfectly, and zero if it fits no better than the sample mean of the dependent variable. $R^2$ is the fraction of the variance of the dependent variable explained by the independent variables. It can be negative if the regression does not have an intercept, or constant, if two-stage least squares is used.

**The Root Mean Square Error (RMSE):** Is defined as

$$RMSE = \left( \frac{1}{N} \sum_{i=1}^{N} d_i^2 \right)^{\frac{1}{2}}. \tag{14}$$

Here $N$ is the number of data pairs and $d_i$ is the difference between $i^{th}$ estimated and $i^{th}$ measured values. This test provides information on the short-term performance of the model by comparing measured and estimated values term by term. A smaller RMSE indicates better model performance. However, a few large errors in the sum can produce an undesirable increase in the RMSE.

**The Mean Bias Error (MBE):** Is given by

$$MBE = \left( \frac{1}{N} \sum_{i=1}^{N} d_I \right). \tag{15}$$

This test measures the long-term performance of a model. A low MBE is desired. A positive value gives the average amount of overestimation in the calculated value and vice-versa. A shortcoming of this test is that overestimation of an individual observation will cancel underestimation in another observation.

**The t-statistic**: Allows the comparison between the performances of various models and in addition may be used to calculate the model's statistical significance to any stated confidence level. In terms of the MBE and RMSE the $t$ -statistic for $N$ data points, is given as:

$$t_{stat} = \left( \frac{(N-1)MBE^2}{RMSE^2 - MBE^2} \right)^{0.5}. \tag{16}$$

The smaller the value of $t$ -statistic, the better the overall performance of the model.

**The Students t-Distribution**: The statistical significance of the correlation coefficient was determined using the 2-tailed test of Student's $t$ distribution given by

$$t = r_{xy} \sqrt{\frac{N-2}{1-R^2}}. \tag{17}$$

**Results and Discussion**

Only stations with high correlation ( $r > .800$) between Clear-Sky clearness index and fraction of bright sunshine were chosen for further consideration. These were J K Airport, Eldoret, Dagoretti, and Voi.

**Linear Models**: For a true linear relationship, a high $R^2$ is desirable. This condition holds for Dagoretti ($R^2$ = .734), Eldoret ($R^2$ = .769), J K Airport ($R^2$ = .823), and Voi ($R^2$ = .755) which means there is little difference between measured and estimated values for these stations. The resulting model equations for the stations with high $R^2$ values are given in a later section below. These equations may be used to estimate clearness index $K_c$ given the corresponding input sunshine fraction $\overline{S}/\overline{S}_0$. A low MBE is desirable and is an indication of long-term performance of the model. All four stations returned zero MBE and no particular best long-term model could be selected on this basis. A low RMSE value is desirable and indicates good short-term model performance. Eldoret and JK Airport had the lowest (*RMSE* = .039) followed, Voi, and Dagoretti. This makes both Eldoret and JK

Airport the stations with the best short-term performance for Linear Model. All four stations registered zero (*t-stat = .000*) and the overall best performing model could not be selected on the basis of this test.

**Quadratic Models**: The MBE for Dagoretti and Voi was positive and for Eldoret and JL Airport it was negative. Voi had the lowest value (MBE = .00011) and thus the Quadratic Model for Voi was the best long-term performer. The RMSE test returned the lowest value (RMSE = .038) for Eldoret followed by JK Airport, Voi and Dagoretti. The Quadratic Model for Eldoret was therefore the best short-term performer. The t-statistic results were Voi (*t-stat = .030*), JK Airport (*t-stat = .067*), Eldoret (*t-stat =.103*), and Dagoretti (t-stat = .129). Therefore Voi had the best overall performing Quadratic Model.

**Exponential Models**: Results of the MBE test shows that models for all four stations underestimated the calculated value (negative MBE), with the highest underestimate being returned for Dagoretti. Taking absolute values of MBE showed JK Airport had the minimum MBE thus making it the station with the best long-term performing Exponential Model. The lowest RMSE was obtained for Eldoret (RMSE = .039) making it the best performing short-term performing Exponential Model. This was followed by JK Airport, Voi, and Dagoretti in order of increasing RMSE. The t-statistic test indicated that JK Airport had the minimum value (*t-stat =.026*) making it the best overall performing Exponential Model.

**Power Model**s: The Power Model for both JK Airport and Dagoretti had the lowest MBE (MBE = .00000). These two stations thus produced the best long-term performing Power Models. Eldoret returned the lowest RMSE value (RMSE = .038) making it the best short-term performing Power Model. The best overall performing Power Model was noted to be JK Airport (*t-stat = .000*) followed by Dagoretti, Voi, and Eldoret.

**Logarithmic Model**: Results of the MBE test showed that Eldoret, having the lowest MBE (*MBE = .00000*) for both Models would have produced the best performing long-term Logarithmic Model. However the other three stations namely, Voi, JK Airport, and Dagoretti (in decreasing order of performance) had positive values of MBE indicating a slight overestimation of the calculated value. Results for RMSE still show that Eldoret had the best performing short-term model. This is a contradiction as a model cannot be simultaneously the best long-term and short-term performer. This contradiction extends to the remaining three stations. Thus both MBE and RMSE failed to discriminate between the long and short term performing model. The t-statistic however indicates that Eldoret (*t-stat = .000*) has the best overall performing Logarithmic Model.

The equations for the Linear, Quadratic, Exponential, Power, and Logarithmic Angstrom-Type Models are given below:

**Linear Models**

Dagoretti: $$K_c = 0.316 + 0.706\left(\bar{S}/\bar{S}_0\right). \quad (18)$$

Eldoret: $$K_c = 0.421 + 0.650\left(\bar{S}/\bar{S}_0\right). \quad (19)$$

J K Airport: $$K_c = 0.373 + 0.489\left(\bar{S}/\bar{S}_0\right). \quad (20)$$

Voi: $$K_c = 0.310 + 0.690\left(\bar{S}/\bar{S}_0\right). \quad (21)$$

**Quadratic Models**

Dagoretti $$K_c = 1.1 - 0.98\left(\bar{S}/\bar{S}_0\right) + 0.82\left(\bar{S}/\bar{S}_0\right)^2. \quad (22)$$

Eldoret& $$K_c = .26 + 1.2\left(\bar{S}/\bar{S}_0\right) - 0.43\left(\bar{S}/\bar{S}_0\right)^2. \quad (23)$$

J K Airport $$K_c = .35 + 0.59\left(\bar{S}/\bar{S}_0\right) - 0.096\left(\bar{S}/\bar{S}_0\right)^2. \quad (24)$$

Voi $$K_c = 0.099 + 1.4\left(\bar{S}/\bar{S}_0\right) - 0.61\left(\bar{S}/\bar{S}_0\right)^2. \quad (25)$$

**Exponential Models**

Dagoretti $$K_c = 0.397\exp\left[1.017\left(\bar{S}/\bar{S}_0\right)\right]. \quad (26)$$

Eldoret $$K_c = 0.503\exp\left[0.788\left(\bar{S}/\bar{S}_0\right)\right]. \quad (27)$$

J K Airport $$K_c = 0.418\exp\left[0.765\left(\bar{S}/\bar{S}_0\right)\right]. \quad (28)$$

Voi: $$K_c = 0.410\exp\left[0.942\left(\bar{S}/\bar{S}_0\right)\right]. \quad (29)$$

**Power Models**

Dagoretti $$K_c = 0.961\left(\bar{S}/\bar{S}_0\right)^{0.502}. \quad (30)$$

Eldoret $$K_c = 1.041\left(\bar{S}/\bar{S}_0\right)^{0.479}. \quad (31)$$

J K Airport $K_c = 0.813\left(\bar{S} / \bar{S}_0\right)^{0.375}$. (32)

Voi $K_c = 0.970\left(\bar{S} / \bar{S}_0\right)^{0.565}$. (33)

**Logarithmic Models**

Dagoretti $K_c = 0.925 + 0.789\log\left(\bar{S} / \bar{S}_0\right)$. (34)

Eldoret $K_c = 1.016 + 0.893\log\left(\bar{S} / \bar{S}_0\right)$. (35)

JK Airport $K_c = 0.794 + 0.535\log\left(\bar{S} / \bar{S}_0\right)$. (36)

Voi $K_c = 0.936 + 0.929\log\left(\bar{S} / \bar{S}_0\right)$. (37)

## CONCLUSIONS

The overall aim of this study was to develop models that estimate the long-term mean monthly Clear-Sky global irradiance over Kenya. The study was conducted using available archived meteorological data collected and collated by the Kenya Government's Meteorological Department at its network of meteorological stations. The data was retrieved from an online archival facility run (on behalf of WMO) by the National Renewable Energy Laboratory (NREL) at Boulder, Colorado. The data was checked for consistency, missing entries, and the radiation data had to be synchronized with sunshine data. Regression analysis was then carried out in order to fit the data into the various selected models outlined above.

Data from 11meteorological stations in Kenya was used to generate a set of clearness indices which was then correlated with a sunshine fraction. The Pearson correlation coefficient test indicated that there was a positive and significant correlation between Clear-Sky clearness index and fraction of sunshine duration. The strength of the correlation was measured using the coefficient of correlation $R^2$ and data from J K Airport, Eldoret, Voi, and Dagoretti stations showed high correlation. Thus only stations whose data exhibited high correlation were considered most suited for model fitting.

The main task was to fit this data into global radiation-based and sunshine duration-based models for estimating Clear-Sky global radiation. The first model presented and discussed was Linear Model clearness index $K_c$. The goodness of fit statistics indicated that this model may be reliably applied to estimate $K_c$ using only sunshine data from the four stations. The MBE test failed to identify the corresponding best performing long-term model for $K_c$. The best short-term performing model for $K_c$ was JK Airport. The statistical tests failed to select the best overall performing model for $K_c$. For the Quadratic Model the goodness of fit statistics picked out the data from the four stations as usable in the Quadratic Models for estimating $K_c$. Data from Voi station produced the best long-term model for estimating $K_c$ while that for Eldoret produced the best short-term model. Data from Voi also produced the best overall performing model. The four pairs of exponential models obtained with data from the four stations all passed the tests for usability. The best long-term performing model for $K_c$ was obtained with data from JK Airport. Data from Eldoret produced the best short-term model $K_c$ while JK Airport data gave the best overall performing model. Data from the four stations fitted into the power model to produce usable models. The long-term best performer in respect of $K_c$ was obtained with data from JK Airport. JK Airport data also produced the best overall performing model. Taken together, both MBE and RMSE tests failed to select either the best long-term or short-term performing Logarithmic Model. However the t-statistic test did select the best overall performing model for $K_c$ which was obtained with Eldoret data.

## RECOMMENDATIONS

Based on the findings of this study and the conclusions outlined above the following recommendations can be made: All models derived from data from Dagoretti, Eldoret, JK Airport, and Voi and expressed in equation form (Eqn. 18 to Eqn. 37) may be used to estimate the clearness indices $K_c$ for sites at or near the four locations.

**Acknowledgements**

The authors are indebted to the University of Eldoret for use of computing facilities

## REFERENCES

Annear, R.L., and Wells, S.A. (2007). A Comparison of Five Models for Estimating Clear-Sky Solar Radiation. Civil and Environmental Engineering Faculty Publications and Presentations. Portland State University PDX Scholar. 10-1-2007.

Engerer, N.A., and Mills, F.P. (2015). Validating Nine Clear Sky Radiation Models in Australia. *Solar Energy.* **120,** 9-24

Baker, D. G., and Haines, D. A. (1969). Solar Radiation and Sunshine Duration Relationship in the North-Central Region of Alaska. Minnesota Agricultural Experimental Station. Tech. Bull. 262

Bird, R.E., and Hulstrom, R.L. (1981). Simplified clear-sky model for direct and diffuse insolation on horizontal surface. SERI/TR 642-761.

Hottel, H. C. (1976). A simple model for estimating the transmittance of direct solar radiation through clear atmospheres. *Solar Energy*. **18,**129.

Ineichen, P. (2016). Validation of models that estimate the clear sky global and beam solar irradiance. *Solar Energy*. **132**, 332-344

Krichak, M. O. (1987). Input Format Guidelines for World Radiometric Network Data. Leningrad. Downloadable from http://wrdc-mgo.nrel.gov/html/formats.html

Lefevre *et al*, (2013). McClear: a new model estimating down-welling solar radiation at ground level in clear-sky conditions. *Atmospheric Measurement Techniques* **6,** 2403-2418.

Liu, B.Y.H., and Jordan, R. C. (1960). The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Solar Energy* **4**, 1-19.

Martinez-Lozano, J. A., Tena, F., Onrubia, J. E., and De la Rubia, J. (1984). The Historical Evolution of the Angstrom formula and its Modifications: Review and Bibliography. Agricultural and Forest Meteorology. **33**.

Panoras, A. G., and Mavroudis, I. G. (1994). Relation between duration of sunshine and global solar radiation in Thessaloniki Plain. *Hydrotechnica: J. Hellenic Hydrotechnical Association*. **4** (1), 3-15.