

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Generating Artificial Weather Data Sequences for Solar Distillation Numerical Simulations

*Bao The Nguyen*

## Abstract

According to the natural geographical distribution, developing countries are concentrated in tropical climates, where radiation is abundant. So the use of solar energy is a sustainable solution for developing countries. However, daily or hourly measured solar irradiance data for designing or running simulations for solar systems in these countries is not always available. Therefore, this chapter presents a model to calculate the daily and hourly radiation data from the monthly average daily radiation. First, the chapter describes the application of Aguiar's model to the calculation of daily radiation from average daily radiation data. Next, the chapter presents an improved Graham model to generate hourly radiation data series from monthly radiation. The above two models were used to generate daily and hourly radiation data series for Ho Chi Minh City and Da Nang, two cities representing two different tropical climates. The generated data series are tested by comparing the statistical parameters with the measured data series. Statistical comparison results show that the generated data series have acceptable statistical accuracy. After that, the generated radiation data series continue to be used to run the simulation program to calculate the solar water distillation system and compare the simulation results with the radiation data. Measuring radiation. The comparison results once again confirm the accuracy of the solar irradiance data generation model in this study. Especially, the model to generate the sequences of hourly solar radiation values proposed in this study is much simpler in comparison to the original model of Graham. In addition, a model to generate hourly ambient temperature data from monthly average daily ambient temperature is also presented and tested. Then, both generated hourly solar radiation and ambient temperature sequences are used to run a solar distillation simulation program to give the outputs as monthly average daily distillate productivities. Finally, the outputs of the simulation program running with the generated solar radiation and ambient temperature data are compared with those running with measured data. The errors of predicted monthly average daily distillate productivities between measured and generated weather data for all cases are acceptably low. Therefore, it can be concluded that the model to generate artificial weather data sequences in this study can be used to run any solar distillation simulation programs with acceptable accuracy.

**Keywords:** solar stills, solar distillation numerical simulations, monthly average daily solar radiation, hourly solar radiation values, monthly average daily temperature, hourly monthly-average ambient temperature, Markov Transition Matrix

## **1. Introduction**

With the development of computers, simulation programs are increasingly developed and become useful and indispensable tools for researchers and designers. It helps users to optimize design and system parameters without having to spend money to build experimental models and waste time to conduct experiments. Prophylactic programs in the field of solar water distillation are no exception. To run solar water distillation simulations, users need to provide weather data such as solar irradiance and ambient temperature measured in days, hours or smaller time periods. According to the requirements of the software. However, hourly weather data is not always available, especially in developing countries because measuring hourly weather data requires equipment, time and money. According to Duffie and Beckman [1], these weather data must be collected for at least 8 years to get the average value to remove the anomalies of the weather such as El Nino, La Nina phenomena, etc.

Another solution is to use typical meteorological year (TMY) data. In fact, the concept of TMY is derived from long-term weather data, which is determined in the correlation and statistical distribution to determine the characteristic indexes to produce the average value [2]. These data are then extracted from the selection criteria to produce month-by-month data from 23 years' data. TMY data were established for 26 Canadian sites, and were applied to the concept of a similar test reference year (TRY) for Europe [2]. While this approach reduces computational effort and the data base required to run simulations, the metric is also based on long-term data, something not available in most places in the world, especially in developing countries.

As pointed out by Nguyen and Hoang [3], the shortage of weather data, especially solar radiation data is very serious in developing countries. For example, in Vietnam, out of a total of 171 hydro-meteorological stations, only 12 have total solar radiation data, of which only 9 have continuous measurements. The remaining meteorological stations only record the number of hours of sunshine. Furthermore, radiation metrics are manually measured by humans every 3 hours instead of hourly. Therefore, hourly radiation data at hydro-meteorological stations in this country are not reliable enough to be used for simulation programs using solar energy systems.

There are two ways to solve the problem of lack of measurement data at the survey site: (i) using extrapolation to process data from hydrometeorological sites adjacent to similar climate features, and (ii) use aggregate generation to generate a series of weather data from the data requiring at least monthly averages. However, the first method can lead to large errors, moreover, very few developing countries have such data available [2]. Therefore, the following method has been studied and developed by many researchers.

Many researchers have proposed mathematical models to calculate the complete series of weather data. Fernandez-Peruchena et al. [4] and Boland [2] used numerical methods to probabilistically simulate daily and hourly solar irradiance data series. Brecl and Topic [5] used a similar approach to generate daily and hourly solar irradiance data from average daily irradiance values. Bright et al. [6] and Hofmann et al. [7] also apply statistical probabilistic techniques to generate a series of solar irradiance values per minute or every 5 minutes from hourly solar irradiance data. Soubdhan and Emilion [8] even used a random method to generate a sequence of solar radiation in seconds. Magnano et al. [9] applied the same technique to generate a synthetic sequence of half an hour's temperature. A common feature of the aforementioned studies is the use of a probability distribution function (PDF) of the data to normalize random variables to bring them to a Gaussian distribution [10].

Gafurov et al. [11] another approach was to incorporate the spatial correlation of solar irradiance (SCSR) into random solar irradiance data generation models to generate monthly solar irradiance time series and daily.

Recently, several researchers have used different types of artificial neural networks (ANNs) to model the values of total solar irradiance on horizontal surfaces, such as [12–16]. Wu and Chan [17] used a new combined model of ARMA (Automatic Recovery and Moving Average) and TDNN (Time Delayed Neural Network) to predict hourly solar irradiance in Singapore. However, Mora-Lopez [10] pointed out that the limitation of these methods is that they are “black boxes” for outputting and analyzing averages of daily global irradiance, resulting in no important information can be obtained from these methods. Mora-Lopez et al. [18] proposed to use machine learning theory with a combination of probability finite automata (PFA) to calculate the values of total daily solar irradiance. The limitation of this method is that the use of PFA is complex and the method has not been shown to be universally applicable.

The results of the above review and analysis show that the stochastic methods are still globally applicable, simple and require minimal input data. Therefore, in this study, randomization technique was chosen to generate series of weather data, including solar irradiance and daily and hourly ambient temperature from monthly averages. These are important weather metrics for running the numerical simulations of solar distillation systems. First, a stochastic model is used to generate a composite of daily irradiance from monthly average daily solar irradiance values. The generated daily radiation sequences are then used to generate the hourly solar radiation sequences. Similarly, a model for generating hourly temperature series from monthly mean temperature values is also presented.

## **2. Stochastic model for daily radiation data generation**

### **2.1 Model of generating daily solar radiation sequence**

When analyzing data of 300 months of solar radiation taken from 9 hydro-meteorological stations with different weather characteristics, Aguiar et al. [19] discovered the analyzed solar irradiance values in For any time period, there is a probability distribution function that seems to be related to the monthly mean clearness index,  $\bar{K}_T$ , for that time period. Furthermore, they also found that the daily radiation value of any given day is statistically related to the value of the previous day. Based on this finding, Aguiar and colleagues built up 10 Markov matrices (called MTM library) from the data analysis of 300 months of solar radiation mentioned above. 10 subdivided matrices include: 1 matrix with  $\bar{K}_T \leq 0.3$  typical for months with very low direct radiation components; The next 8 matrices are for the months where  $\bar{K}_T$  varies from 0.3 to 0.7 with the increment of  $\bar{K}_T$  increasing by 0.05 for the next matrix; the final matrix with  $\bar{K}_T > 0.7$  is for months with a high direct radiation component. The MTM library was then used to generate the daily radiation series from the average daily irradiance values for the locations in the United States for which the irradiance data were not used to generate the aforementioned MTM library. This simulation result was compared with the measured radiation results and compared with the simulation results from Graham's model [20]. When comparing statistical parameters such as mean, variance, and probability density functions as well as statistic characteristics (e.g. autocorrelation functions), Aguiar's model produced more accurate data series than with the Graham model. Furthermore, Aguiar's model is computationally simpler than Graham's model [21].

To calculate the  $K_T$  values from the monthly mean  $\overline{K}_T$ , in this study, the Aguiar method was chosen for the following reasons:

- The calculation expressions given by Graham are based on results that are built from the data of the United States, so certainly not suitable for tropical climates [22]. This was confirmed by the Nguyen and Pryor [21] in their study, and even in high  $K_T$  regions, the above expressions do not fit the curve due to Liu and Jordan [21].
- The above disadvantage of Graham's method can be overcome by adding the expressions for tropical climates developed by Aguiar group [19], but the use of this expression has not been fully verified by scientists and will create a complex, climate-dependent computational model.
- The locations where the Aguiar group used the measured data to build the MTM matrices include many characteristic climate zones, in these locations there is one location with a central tropical rainforest ( $C_{aw}$ ) climate, which is Macao. and one location with a rainforest climate ( $A_w$ ) is Polana, Mozambique. The use of this method would therefore be suitable for the study's objective of tropical climates, where most developing countries are often located.

2.2 The data are used to evaluate the accuracy of the model

Solar irradiance data were measured in 2 cities representing 2 tropical climates to evaluate the accuracy of daily solar radiation data generated from the calculated model selected in this study. They are Ho Chi Minh City representing the tropical forest climate ( $A_w$ ) and Da Nang representing the tropical monsoon climate ( $A_m$ ). Pyranometers were used to measure total irradiance on a horizontal plane in the two cities every 5 minutes, measuring continuously from 5:30 a.m. to 6:30 p.m. Since these two cities have low latitudes ( $10^0N$  and  $16^0N$  respectively), the day length does not change much during the year, so there is no need to extend the seasonal solar irradiance measurement period of the year [3]. Then, solar irradiance by hour, by day and by day of month average is calculated from the measured data. **Table 1** presents the average daily solar irradiance of the two cities mentioned above, used to run the program to generate date and time irradiance data in this study.

From the series of daily radiation values of 365 days of the year, the series of 365 values of the daily clearness index is calculated according to the following Eq. (1):

$$K_{Tmea.} = \frac{H}{H_0} \tag{1}$$

where H is the total daily solar radiation measured on a horizontal plane and  $H_0$  is the daily radiation outside the atmosphere, calculated by the equation:

	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
HCM	13.0	18.0	18.1	18.7	16,7	17,4	17.3	17.6	15.9	15.0	14.7	13.9
Danang	10.3	18.8	18.6	22.1	22.9	23.9	20.3	18.7	17.2	14.8	11.9	8.4

**Table 1.**  
Total monthly average solar irradiance measured on a horizontal plane in Ho Chi Minh City and Da Nang (MJ/m<sup>2</sup>. day).



$$H_0 = \frac{24}{\pi} G_{sc} 3600 \left\{ \left[ 1 + 0,033 \cdot \cos \left( \frac{360n}{365} \right) \right] \times \left[ \cos \phi \cos \delta \sin \omega_s + \frac{\pi}{180} \omega_s \sin \phi \sin \delta \right] \right\} \tag{2}$$

with  $G_{sc}$ ,  $n$ ,  $\phi$ ,  $\delta$  and  $\omega_s$  respectively are solar constance, day of the year, latitude of the investigated location, declination angle and sunset hour angle, defined in [1].  
From the values of daily clearness index  $K_T$ , the monthly average daily values of clearness index  $\overline{K_T}$  for 12 months of the year are calculated:

$$\overline{K_T} = \frac{\overline{H}}{\overline{H_0}} \tag{3}$$

where the monthly average daily irradiance values  $\overline{H}$  are taken in **Table 1** and monthly average daily irradiance values outside the atmosphere  $\overline{H_0}$  are calculated by Eq. (2) with day  $n$  being the average day of the month, given in [1]. **Table 2** presents the  $\overline{K_T}$  values of the 2 investigated cities.

2.3 Applying Aguiar’s model

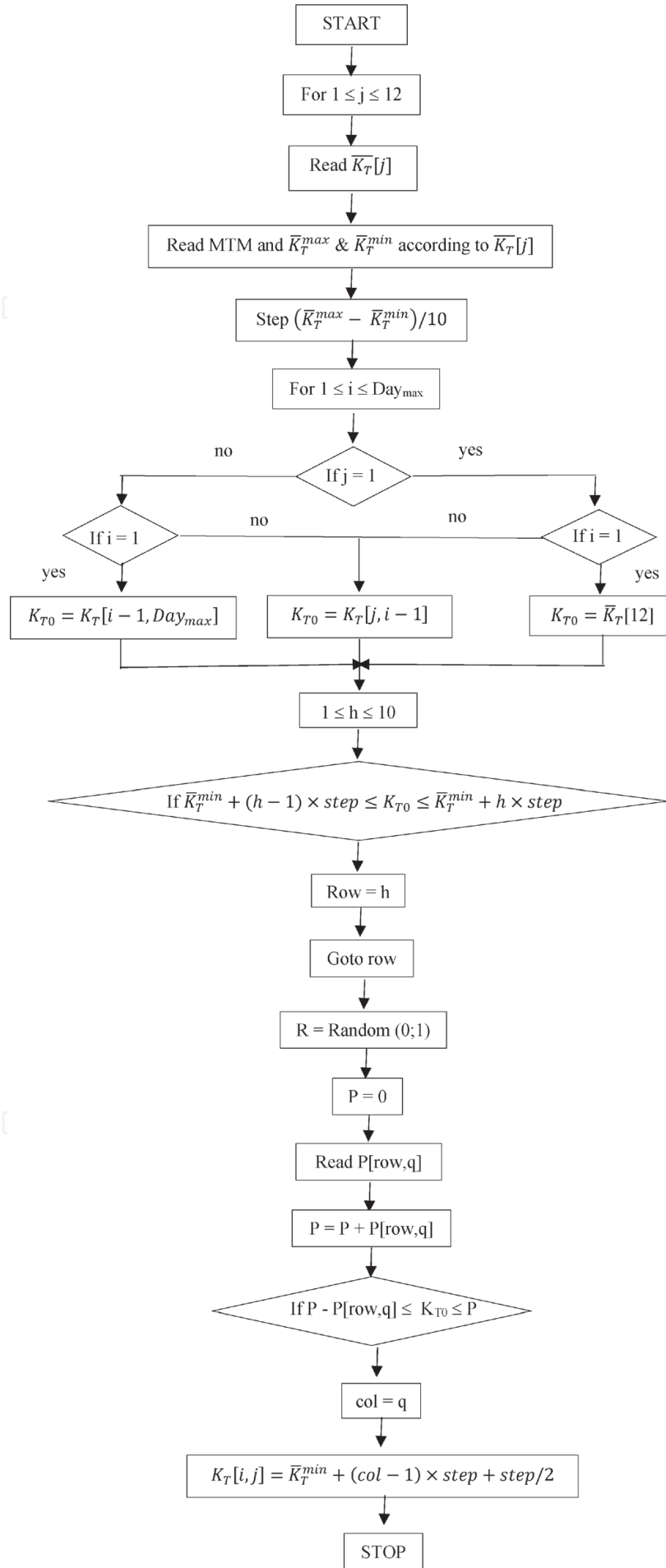
**Figure 1** presents the procedure for calculating the series of daily clearness index from the monthly average daily values.

After 365 values of the daily photometric index are calculated for each location, these value series are compared with the measurement series through statistical functions such as cumulative distribution function (CDF), density function, etc. probability (PDF). **Figures 2** and **3** present the cumulative distribution function of CDF of the calculated and measured  $K_T$  in Ho Chi Minh City and Da Nang while **Figures 4** and **5** represent the probability density function PDF for these two cities. Statistical parameters including mean, median, minimum, maximum, standard deviation, mean absolute error (MAE) and mean square error (RMSE) were also compared between the  $K_T$  series. Calculated and measured, as shown in **Tables 3** and **4**.

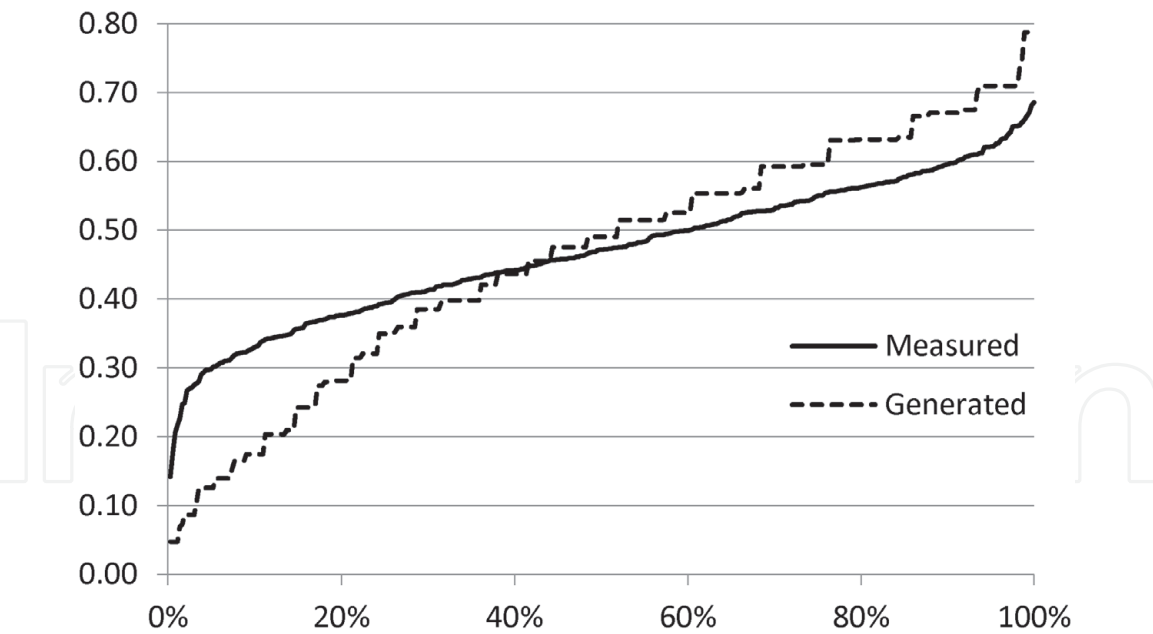
The results shown in **Figures 2–5** show that the Aguiar model produced a  $K_T$  daily value series with an acceptable level of accuracy compared with the measured series. Similarly, the statistical parameters in **Tables 3** and **4** also show that the statistical error between the calculated and measured series is relatively small. Specifically, the mean and median error percentages of generated chains are 1% and – 4% for Ho Chi Minh City and 6% and 14% for Da Nang, respectively. Therefore, this model is expected to be able to be used to generate a series of daily cloud optical coefficients for any location because the Aguiar model has been proven to be universally applicable in the world [2, 4, 5, 19, 22]. As shown above, this model only requires input of 12 average daily solar irradiance values at the location to be calculated.

	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
HCM	0.42	0.53	0.50	0.50	0.45	0.47	0.47	0.47	0.44	0.42	0.47	0.46
Danang	0.36	0.59	0.53	0.59	0.60	0.63	0.53	0.50	0.48	0.45	0.40	0.30

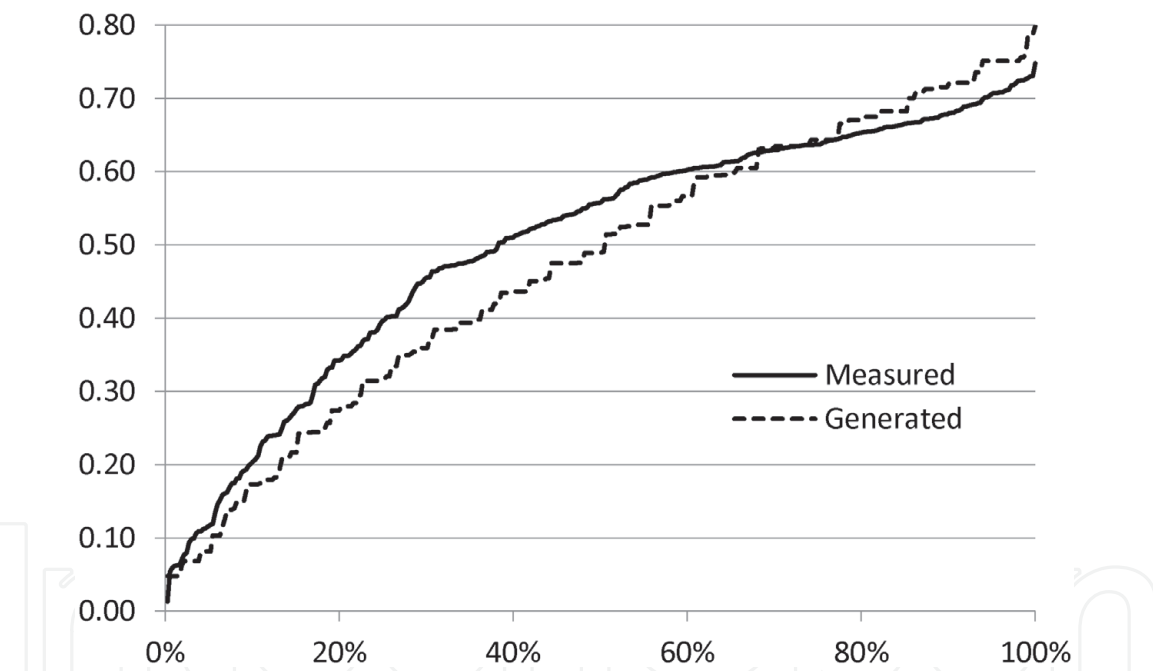
**Table 2.**  
*Monthly average daily values of clearness index in Ho Chi Minh City and Da Nang.*



**Figure 1.**  
The procedure for calculating the series of  $K_T$  from  $\bar{K}_T$ .



**Figure 2.**  
*Cumulative distribution function of  $K_T$  for Ho Chi Minh City.*



**Figure 3.**  
*Cumulative distribution function of  $K_T$  for Da Nang.*

### 3. Model of generating hourly solar radiation

#### 3.1 Choosing a model to create hourly solar radiation chains

There are many studies in the world on establishing mathematical models to generate daily and hourly radiation data series [2, 4, 5]. Basically, these studies are based on the approach of Aguiar [19] and Graham [20].

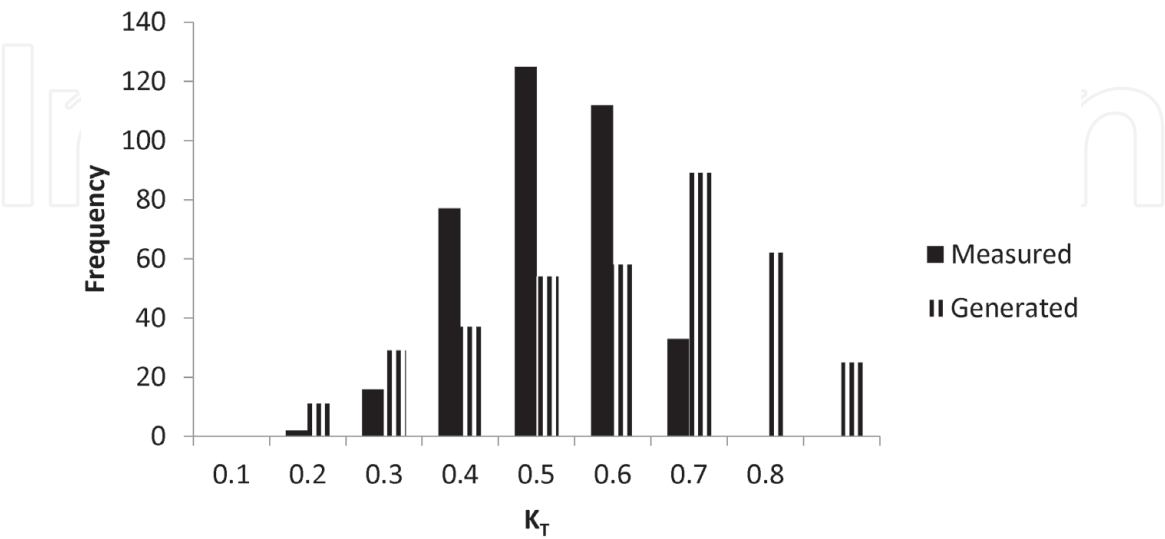
With the assumption that the sky clarity index  $k_t$  depends only on the cloudiness coefficient of the  $K_T$  day, Graham et al. [20] analyzed  $k_t$  into two components: an average (or trend) component and a random component.



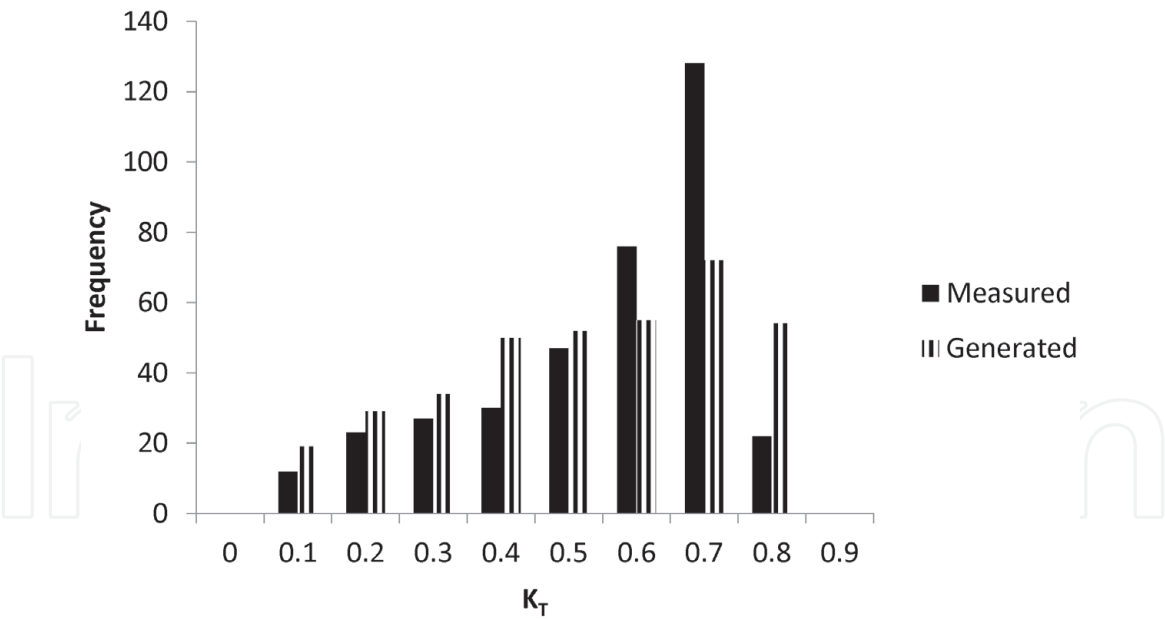
$$k_t = k_{tm} + \alpha \tag{4}$$

The formula for calculating the trend component or the regular component:

$$k_{tm} = \lambda + \rho \exp .(-\kappa m) \tag{5}$$



**Figure 4.**  
*Probability density function of  $K_T$  for Ho Chi Minh City.*



**Figure 5.**  
*Probability density function of  $K_T$  for Da Nang.*

	Mean	Median	Min	Max	Std. dev.	MAE (%)	RMSE (%)
$K_{Tmea}$	0.47	0.47	0.14	0.69	0.10		
$K_{Tgen.}$	0.46	0.49	0.05	0.78	0.18		
Error (%)	1	−4				16.8	8.1

**Table 3.**  
*Statistic parameters of  $K_T$  series of Ho Chi Minh City.*

	Mean	Median	Min	Max	Std. dev.	MAE (%)	RMSE (%)
$K_{Tmea}$	0.50	0.56	0.01	0.75	0.18		
$K_{Tgen.}$	0.47	0.49	0.05	0.80	0.20		
Error (%)	6	14				4.5	5.1

**Table 4.**  
 Statistic parameters of  $K_T$  series of Da Nang.

where  $m$  is the air mass, the value calculated at the time of the middle of the hour. The parameters  $\lambda$ ,  $\rho$  and  $\kappa$  are the identity function of  $K_T$ :

$$\lambda(K_T) = K_T-1.167K_T^3(1-K_T) \tag{6}$$

$$\rho(K_T) = 0.979(1-K_T) \tag{7}$$

$$\kappa(K_T) = 1.141(1-K_T)/K_T \tag{8}$$

The standard deviation  $\sigma_\alpha$  of the variable  $\alpha$  is expressed as:

$$\sigma_\alpha(K_t) = 0.16 \sin (\pi K_t/0.90) \tag{9}$$

Then use the Gaussian normalization technique to transform this variable  $\alpha$  into a Gaussian variable  $\beta$  with the relation between  $\alpha$  and  $\beta$  as follows:

$$\alpha = -\frac{1}{1.158} \ln \left\{ \frac{1}{0.5 \left[ 1 + \operatorname{erf} \left( \frac{\beta}{\sqrt{2}} \right) \right]} - 1 \right\} \tag{10}$$

Then apply ARMA models to the data series  $\beta$  and determine that  $\beta$  follows the model as AR(1):

$$\beta_t = \phi \beta_{t-1} + \vartheta_t \tag{11}$$

where:

$\beta_{t-1}$  is the value of the variable at  $t-1$ .

$\Phi$  is the automatic regression coefficient.

$\vartheta_t$  is a random number from a normal distribution with zero mean and a standard deviation  $\sqrt{1-\phi}$

As in the case of using date values, the coefficient  $\Phi$  varies slightly by locality but the value 0.54 can be chosen as the value to use in the model for all localities.

Aguiar’s method [23] to generate  $k_T$  is similar to Graham’s method but has some differences as follows:

- First, the standard deviation  $\sigma_\alpha$  depends not only on  $K_T$  but also on the altitude angle of the sun  $h_s$ :

$$\sigma_\alpha(K_T, h_s) = A \times \exp \{ B \times [1 - \sin (h_s)] \} \tag{12}$$

with:

$$A = 0.14 \times \exp \left[ -20.0(K_T - 0.32)^2 \right] \tag{13}$$

$$B = 3.0(K_T - 0.45)^2 + 16.0K_T^5 \tag{14}$$

- Second, the coefficient  $\Phi$  depends on  $K_T$  according to the following expression:

$$\Phi = 0.38 + 0.06 \cos (7.4K_T - 2.5) \quad (15)$$

- Third, the calculated  $k_t$  values are limited by the clear sky clearness index  $k_{cs}$ .

$$k_{cs}(t) = 0.88 \times \cos [\pi(t - 12.5)/30] \quad (16)$$

With  $t$  is the hour considered.

Aguiar's model has been successfully applied to generate hourly solar irradiance in Spain and Slovenia [4, 5], while Graham's approach to generate hourly solar irradiance sequences at different locations has been successfully applied for locations in the United States in particular and many parts of the world in general [24]. On the other hand, Aguiar and Graham's models were used in generating and comparing hourly solar irradiance for 6 locations in Australia and the results showed that Graham's model produces a global solar irradiance sequence which better fit for all six sites [21]. Therefore, Graham's model was chosen to generate hourly solar radiation in this study.

However, Graham's model also has some disadvantages as follows:

- The use of Gaussian mapping technique to process the random component of  $k_t$  values is very complicated and time consuming.
- The error of the hourly series of transparency index generated from Graham's model in comparison to the measured series is larger than the error from the model suggested in this study.

To solve the complicated and time consuming problem of the Graham model, the  $\beta$  values given in Eq. (10) are transformed to a non-standard distribution  $h$  by using Norminv function in MATLAB. The random components of the  $k_t$  value are then calculated by:

$$\alpha = h \times \sigma_\alpha (K_T) \quad (17)$$

with  $\sigma_\alpha (K_T)$  is the standard deviation computed by using Eq. (9).

This suggested model are not only much simpler than Graham's method, but also produces more accurate results. **Table 5** demonstrates the hourly solar irradiance sequence error generated by Graham's model and this study's modified model compared with the measured solar irradiance sequence for Ho Chi Minh City over 20 times run a program written in MATLAB.

### 3.2 Modified Graham model to create a series of hourly clearness indices

**Figure 6** shows the process of creating a series of hourly clearness index series. This procedure is modified from the Graham model, as analyzed above. In this Figure:

$\phi$  is the investigated location's latitude

$L_{st}$  and  $L_{loc}$  are respectively the longitude of the standard meridians and the considered location.

$j$  is the month of the year.

$i$  is the day of a month

$\omega_s$  is the angle of sunset for the calculated day

$K_T [i][j]$  is the daily clearness index of the  $i^{th}$  day in the  $j^{th}$  month

Generating program runtime	Errors of the generated versus measured hourly solar radiation values (%)	
	Graham's model	Modified model of this study
1	2.87%	0.74%
2	5.31%	2.70%
3	5.75%	2.49%
4	3.81%	1.46%
5	11.15%	9.47%
6	5.04%	2.84%
7	4.22%	2.06%
8	5.19%	3.60%
9	4.12%	2.27%
10	4.70%	2.18%
11	6.40%	4.40%
12	7.33%	5.29%
13	6.32%	4.22%
14	5.07%	2.60%
15	5.03%	2.78%
16	7.52%	4.67%
17	4.88%	2.52%
18	3.24%	1.48%
19	8.15%	6.44%
20	8.68%	5.36%

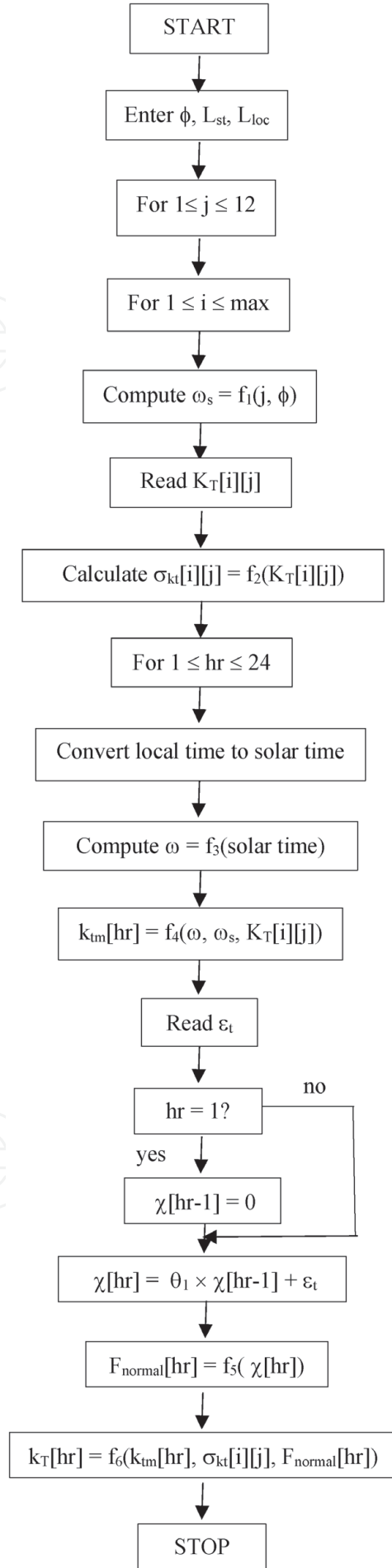
**Table 5.**  
*Errors of the generated versus measured hourly solar radiation values in Ho Chi Minh City.*

$\omega$  is the calculated hour angle.  
 $k_{tm}$  is the “long-term” average value of  $k_t$   
 $\sigma_{kt}$  is the standard deviation of  $k_t$  toward the values of the “long-term” average value  
 $\varepsilon_t$  is a Gaussian distribution’s random number  
hr. is the investigated hour.  
 $\chi$  is a Gaussian distribution’s random variable with “0” mean and “1” variance.  
 $\theta_1$  is the parameter of the AR1 model.  
 $F_{normal}$  is a function to convert a Gaussian variable into a non-normally distributed variable  
MATLAB is used to write generation program for hourly  $k_t$  series.

3.3 Validate generated hourly clearness index strings

The daily generated transparency index values were used as input to the hourly  $k_t$  series generation program. The calculated  $k_t$  values are then compared with  $k_{tmea}$  values, where  $k_{tmea}$  is the measured hourly clearness index values given by:

$$k_{tmea.} = \frac{I}{I_0} \tag{18}$$



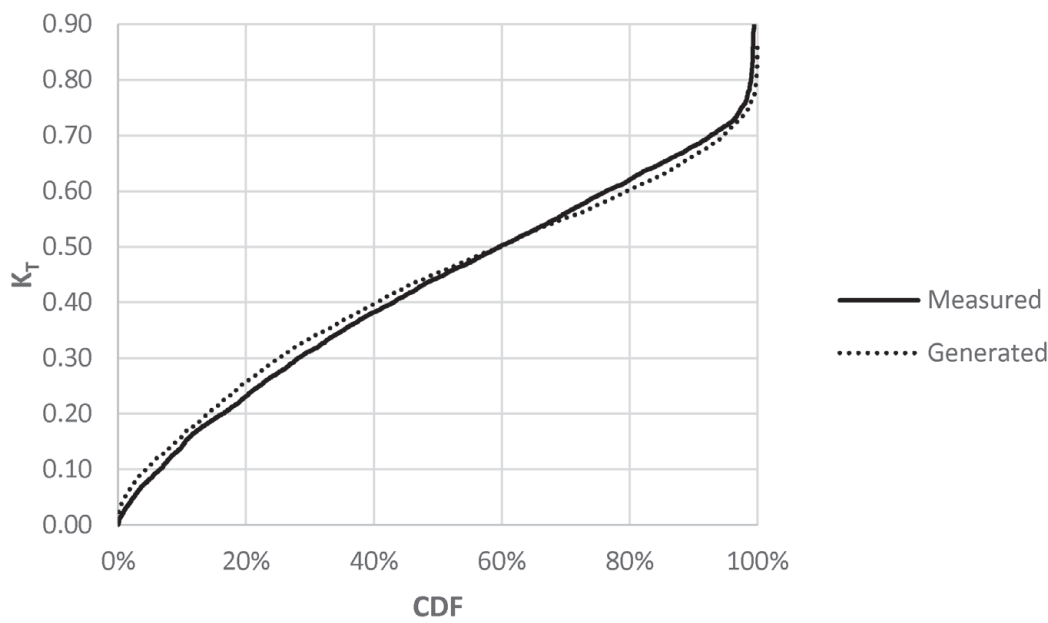
**Figure 6.**  
Flow chart for generating hourly  $k_t$  series from daily optical clearness index series  $K_T$ .



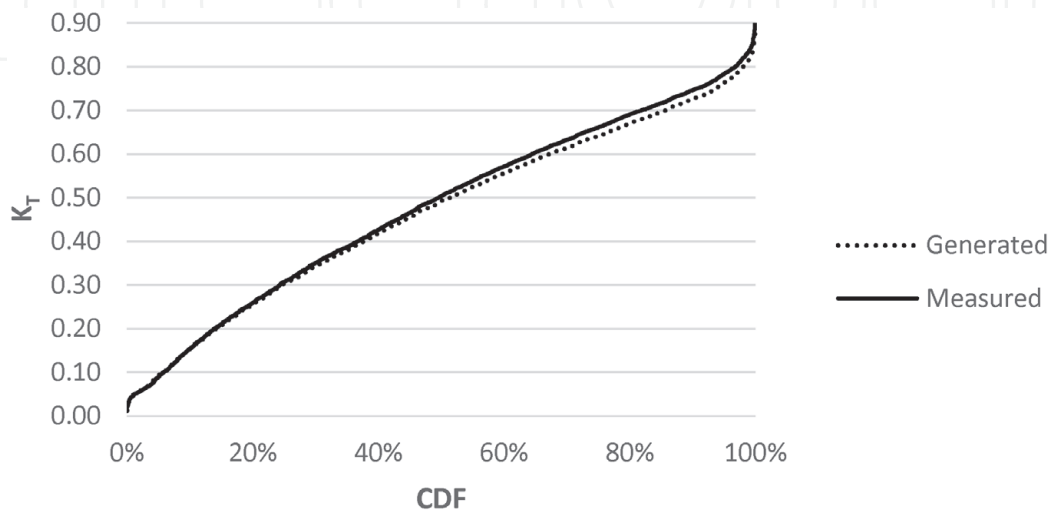
where  $I$  is the horizontal total solar irradiance measured from  $\omega_1$  to  $\omega_2$  in Ho Chi Minh City and Da Nang;  $I_0$  is the solar radiation outside the atmosphere from hour angle  $\omega_1$  to  $\omega_2$ , given by [1]:

$$I_0 = \frac{12}{\pi} G_{SC} \times 3600 \left\{ \left[ 1 + 0,033 \cdot \cos \left( \frac{360n}{365} \right) \right] \times \left[ \cos \phi \cos \delta (\sin \omega_2 - \sin \omega_1) + \frac{\pi}{180} (\omega_2 - \omega_1) \sin \phi \sin \delta \right] \right\} \quad (19)$$

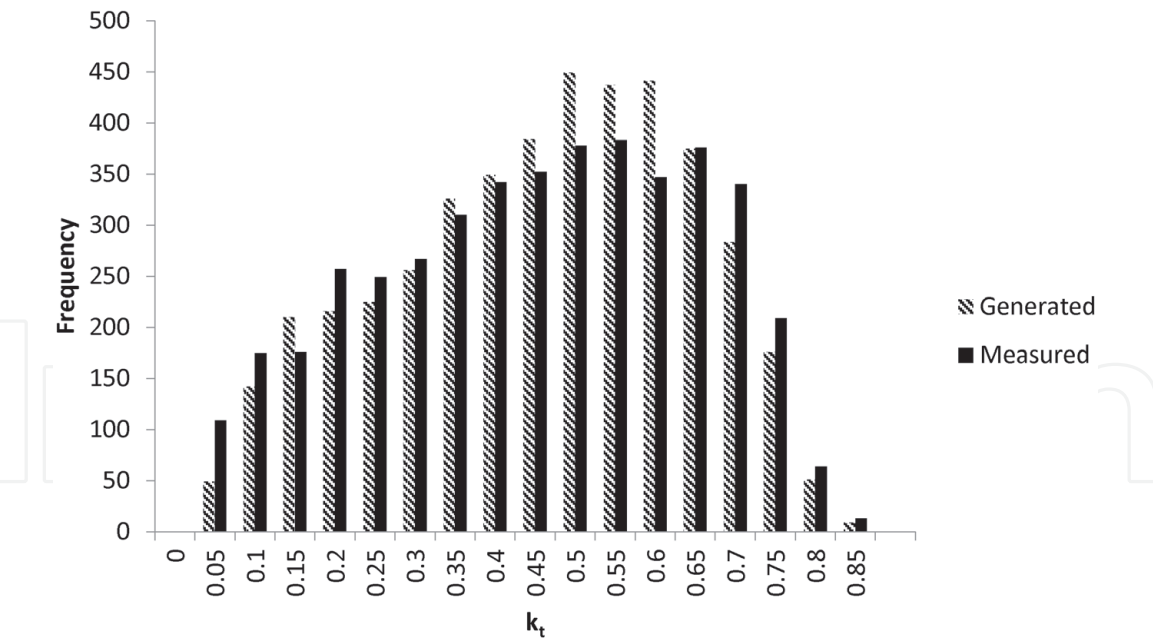
The cumulative distribution function (CDF) graphs of  $k_t$  over the hours for Ho Chi Minh City and Da Nang are shown in **Figures 7** and **8** respectively while the probability density function (PDF) of  $k_t$  over the hour for these cities are shown in **Figures 9** and **10**. Additionally, several statistical parameters, including mean, median, minimum, maximum, standard deviation, mean absolute error (MAE) and mean square error (RMSE) of the measured and generated  $k_t$  series in Ho Chi Minh City and Da Nang are also shown in **Tables 6** and **7**, respectively.



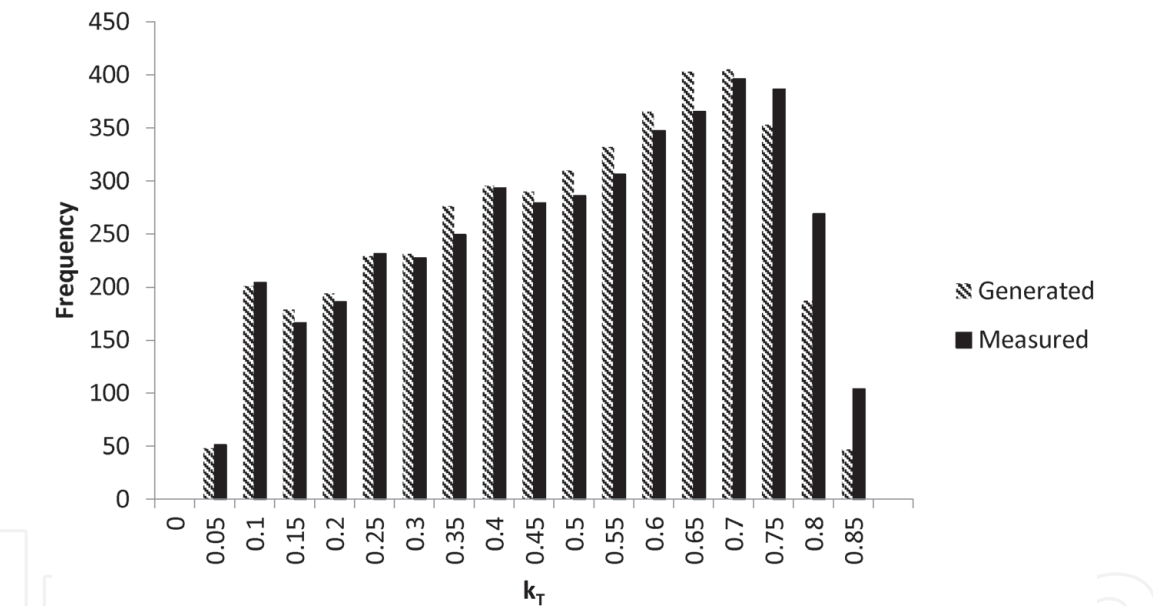
**Figure 7.**  
*Cumulative distribution function (CDF) of hourly  $k_t$  for Ho Chi Minh City.*



**Figure 8.**  
*Cumulative distribution function (CDF) of hourly  $k_t$  for Da Nang.*



**Figure 9.**  
Probability density function (PDF) of hourly  $k_t$  for Ho Chi Minh City.



**Figure 10.**  
Probability density function (PDF) of hourly  $k_t$  for Da Nang.

	Mean	Median	Min	Max	Std. dev.	MAE (%)	RMSE (%)
$k_{tmea}$	0.426	0.443	0.001	0.899	0.197		
$k_{tgen.}$	0.433	0.453	0.007	0.858	0.191		
Error (%)	-1.5	-2.4				2.0	0.03

**Table 6.**  
Statistical parameters of the  $k_t$  series of Ho Chi Minh City.

As presented in **Figures 7–10** and **Tables 6** and **7**, the hourly  $k_t$  series for the two investigated cities have been successfully generated by the suggested model with very high accuracy. The mean and median error percentages of generated sequences were  $-1.5\%$  and  $-2.4\%$  for Ho Chi Minh City and  $-1.3\%$  and  $0.3\%$  for Da Nang,

	Mean	Median	Min	Max	Std. dev.	MAE (%)	RMSE (%)
$k_{tmea}$	0.459	0.491	0.019	0.905	0.210		
$k_{tgen.}$	0.465	0.492	0.012	0.884	0.217		
Error (%)	-1.3	-0.3				3.3	0.03

**Table 7.**  
*Statistical parameters of the  $k_t$  series of Da Nang.*

respectively. Since stochastic models have been approved to have universal characteristics, as mentioned above, the model in this study is expected to be applicable to any location in the world.

**4. Model to generate hourly ambient temperature sequences**

The procedure to generate hourly ambient temperature sequences from monthly mean ambient temperature,  $\overline{T}_a$ , and monthly mean daily clearness index,  $\overline{K}_t$ , was described by Knight et al. [25]. The model to generate artificial hourly ambient temperature sequences for Australia was developed by Nguyen and Pryor [21].

First, to generate the deterministic component of the hourly ambient temperatures series, the concept of the average normalized diurnal temperature variation, developed by Erbs et al. [26], is applied. Hourly measured ambient temperature data from two locations (Ho Chi Minh city and Da Nang) are used to calculate the hourly monthly-average ambient temperature,  $\overline{T}_{a,h}$ , at each hour of the day for each month. These curves are standardized by subtracting the monthly-average daily temperature,  $\overline{T}_a$ , from each of the hourly values and then dividing by the amplitude of the curve (defined as the difference between the maximum and minimum hourly average temperatures over the day),  $A$ . Subsequently twelve cosine curves are derived for each location.

The average of these 24 (ie., 12 curves \* 2 locations) standardized curves are calculated. Interestingly, the equation originally derived by Erbs and his colleagues is found to fit the average standardized curve in this study. The equation is expressed:

$$\frac{\overline{T}_{a,h} - \overline{T}_a}{A} = 0.4632 \cos(t^* - 3.805) + 0.0984 \cos(2 t^* - 0.360) + 0.0168 \cos(3 t^* - 0.822) + 0.0138 \cos(4 t^* - 3.513) \tag{20}$$

$t^*$  is given by:  $t^* = \frac{2\pi(t-1)}{24}$  where temperature is in hours with 1 and 24 corresponding to 1 am and midnight, respectively.

The relation between the amplitude  $A$  and the monthly mean clearness index,  $\overline{K}_t$ , is calculated as [26]:

$$A = 20.231 \overline{K}_t - 3.103 \tag{21}$$

After the trend components are removed from the ambient temperature values, the variable component of the data is converted into a normal distribution and then tested with many ARMA models. The AR2 model is finally selected:

$$\chi_t = \Phi_1 \chi_{t-1} + \Phi_2 \chi_{t-2} + \varepsilon_t \tag{22}$$

here,  $\chi_{t-1}$  and  $\chi_{t-2}$  are the values of the weather data variables at t-1 and t-2, respectively, and  $\Phi_1$  and  $\Phi_2$  are calculated from the available ambient temperature data and are found to be 0.9072 and  $-0.1430$ , respectively.  $\varepsilon_t$  is a random number from a normal distribution with zero mean and a standard deviation  $\sqrt{1 - \Phi_1 r_1 - \Phi_2 r_2}$ .  $r_1$  and  $r_2$  are the corresponding autocorrelation coefficients.

The generated  $\chi$  is transformed to an hourly temperature by equating the cumulative function of  $\chi$ ,  $F_{\text{normal}}$ , and hourly ambient temperature,  $F_{\text{temp}}$ .  $F_{\text{normal}}$  is given by:

$$F_{\text{normal}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{1}{2}t^2\right) dt = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{\chi}{\sqrt{2}}\right)\right] \quad (23)$$

whereas  $F_{\text{temp}}$  is calculated as follows:

$$F_{\text{temp}} = \frac{1}{1 + \exp(-3.396h)} \quad (24)$$

with:

$$h = (T - \bar{T}_{a,h}) \left( \sigma_m \sqrt{N_m/24} \right) \quad (25)$$

in which  $N_m$  is the number of hours in the months and  $\sigma_m$  is the standard deviation of the monthly-average daily temperature  $\bar{T}_a$  given by:

$$\sigma_m = 1.45 + 0.029\bar{T}_a + 0.0664\sigma_{yr} \quad (26)$$

$\sigma_{yr}$  is the standard deviation of the yearly average ambient temperature. Solving these equations, hourly ambient temperature is given by:

$$T = \bar{T}_a - \frac{\sigma_m \sqrt{\frac{N_m}{24}}}{3.369} \times \ln \left\{ \frac{1}{0.5 \left[ 1 + \operatorname{erf}\left(\frac{\chi}{\sqrt{2}}\right) \right]} - 1 \right\} \quad (27)$$

**Figure 11** shows the schematic diagram of the procedure to generate hourly ambient temperature sequences.

In this figure:

$L_{\text{st}}$  and  $L_{\text{loc}}$  are the standard meridian for the local time zone and the longitude of the location considered, respectively.

$T_a[j]$  and  $K[j]$  are the monthly mean ambient temperature and monthly mean daily radiation of  $j^{\text{th}}$  month, respectively.

$T_{a,yr}$  is the year average ambient temperature, calculated from the 12 monthly values.

$\sigma_m[j]$  is the monthly standard deviation of the  $j^{\text{th}}$  month, obtained from the yearly average value and the monthly average temperature for that month.

$A[j]$  is the amplitude of the diurnal variation (peak to peak) of ambient temperature, and is a function of monthly average daily clearness index.

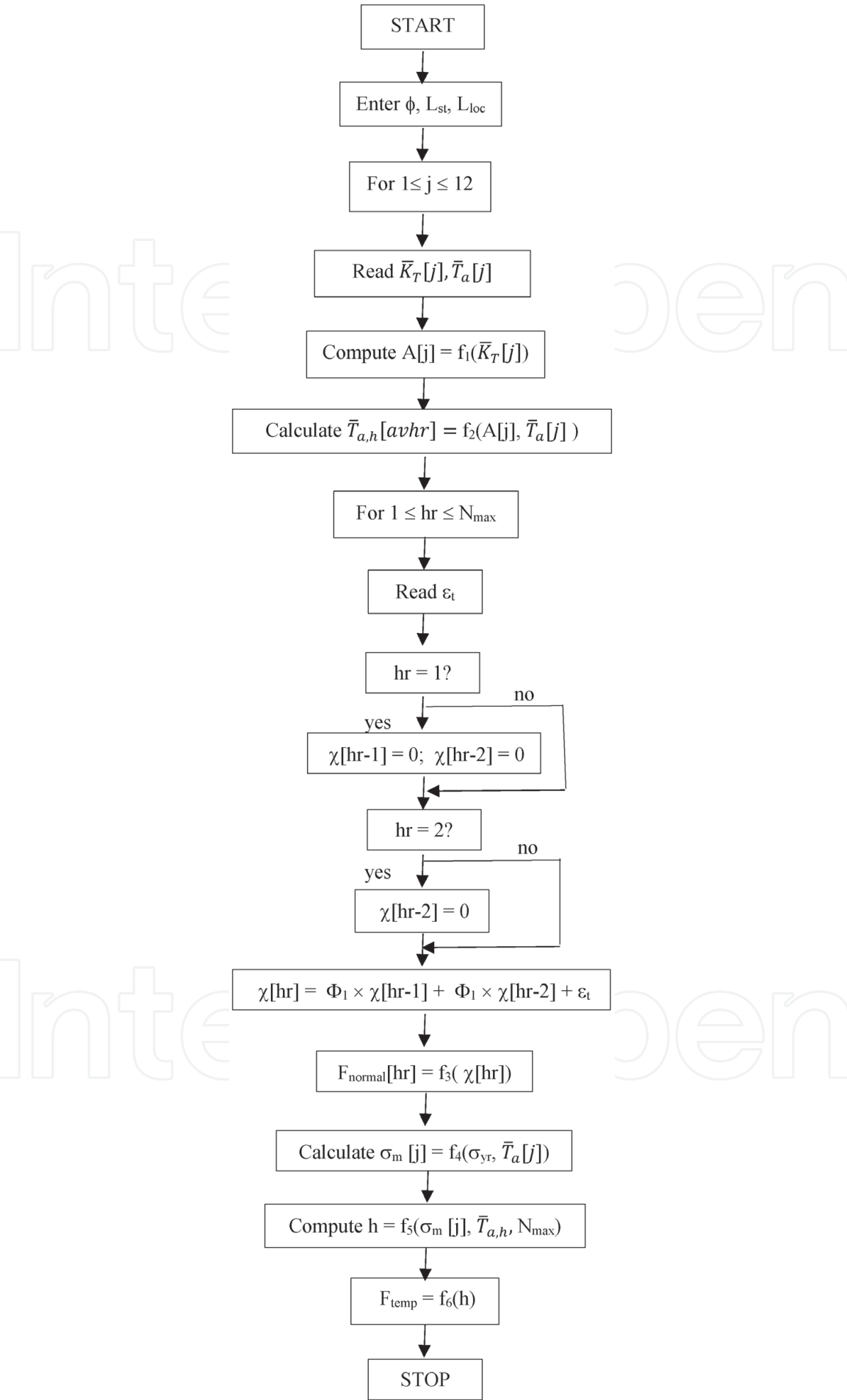
$T_{a,h}[\text{avhr}]$  is the hourly monthly-average ambient temperature; the subscript “avhr” indicates the calculated monthly-average hour (avhr = 1 to 24).

$\varepsilon_t$  is a random number from a Gaussian distribution.

hr is the hour considered.

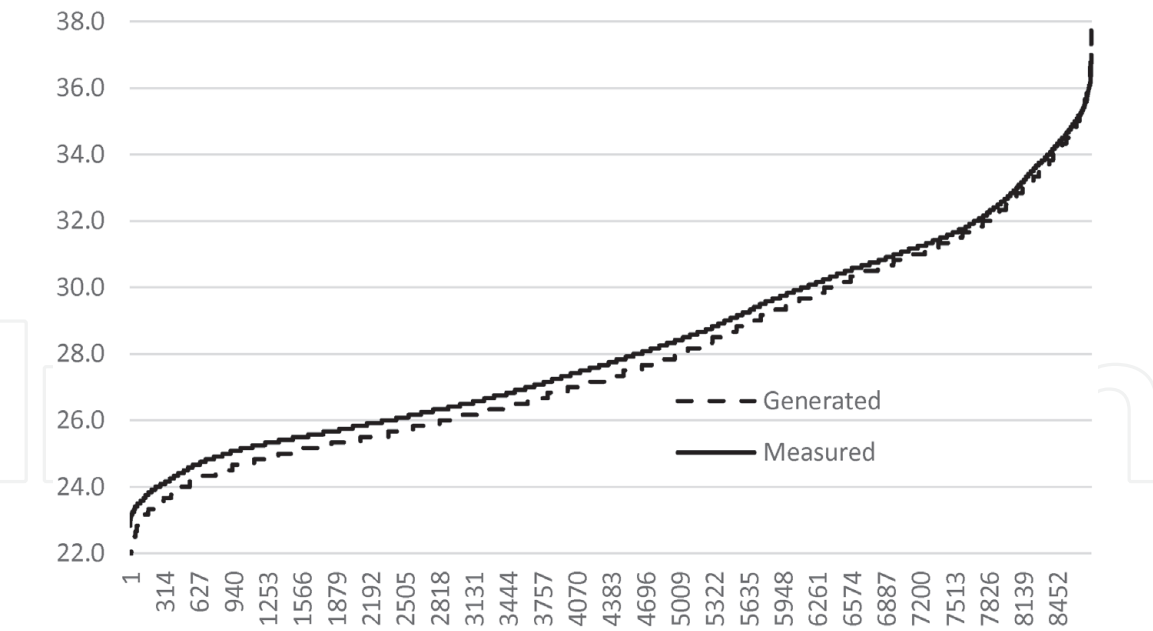
$N_{\text{max}}$  is the number of hours in the respective month

$\chi$  is the normally distributed stochastic variable with a mean of 0 & a variable of 1.



**Figure 11.**  
 Schematic diagram of the procedure to generate hourly ambient temperature sequences.





**Figure 12.**  
*Cumulative distribution function (CDF) of hourly ambient temperature for Ho Chi Minh City.*

$\Phi_1$  and  $\Phi_2$  are the first and second parameters of the AR2 model.

$F_{\text{normal}}$  is the cumulative distribution of a normally distributed variable.

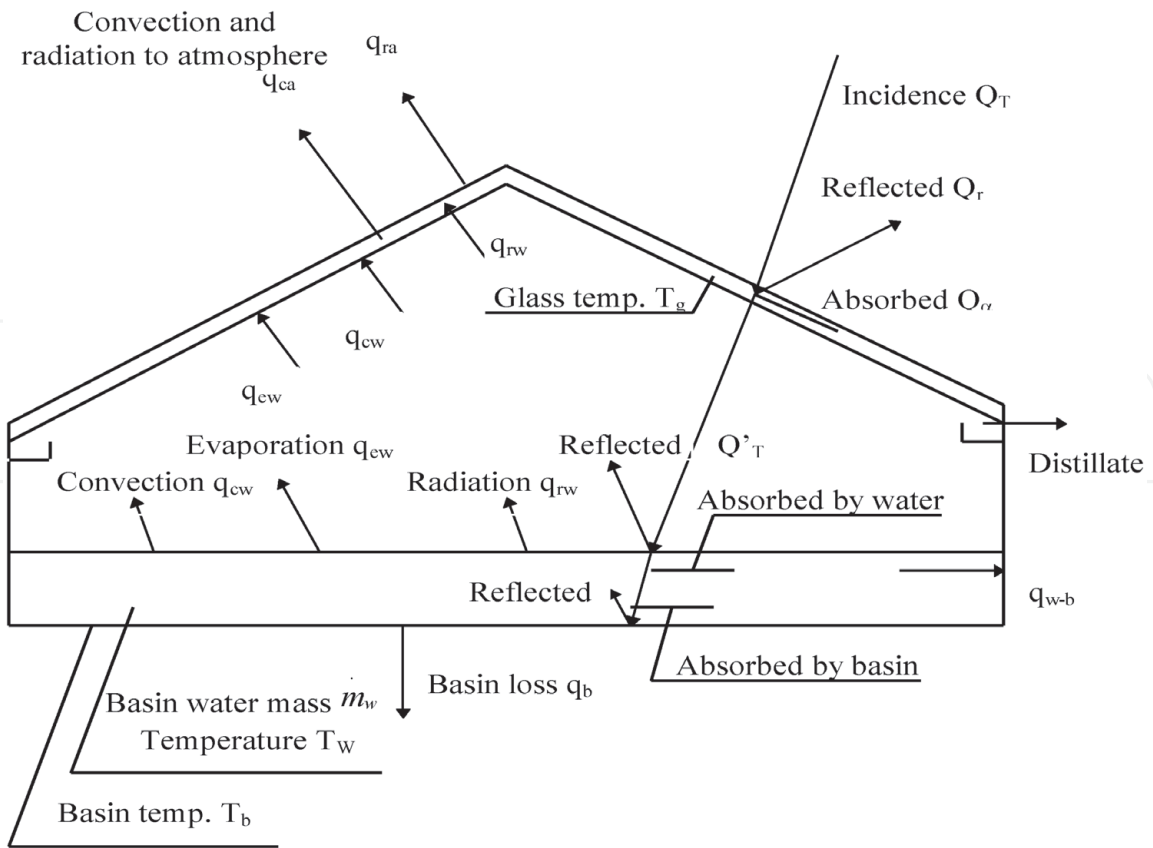
**Figure 12** shows the cumulative distribution of hourly ambient temperature sequences for Ho Chi Minh City. The figure compares the results using measured data and artificial data generated from the equations described in this section. As shown, the model presented here produced accurate hourly ambient temperature in comparison with measured data.

### 5. Validating generated versus measured weather data by running a solar distillation simulation program

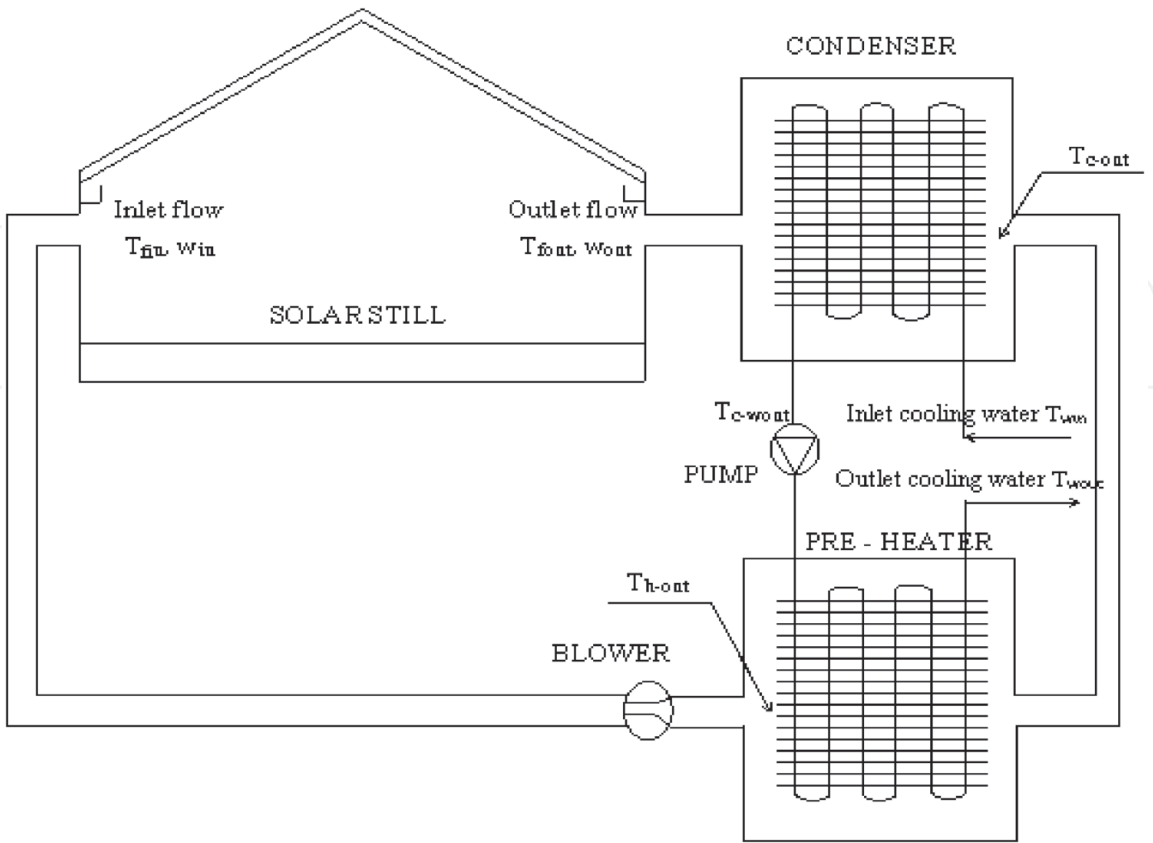
The main objective of this study is to build a model to generate weather data, including daily and hourly solar radiation sequences and ambient temperature series; then these weather data chains must be used to run simulation programs. Therefore, the generated weather data is used as input for SOLSTILL – a simulation program for solar distillation systems [27]. This simulation program was designed to enable to simulate both passive solar stills and active solar distillation systems.

**Figure 13** presents the heat and mass diagrams in a passive solar still whereas **Figures 14** and **15** respectively shows the schematic diagram and heat and mass transfer process in a forced circulation solar still.

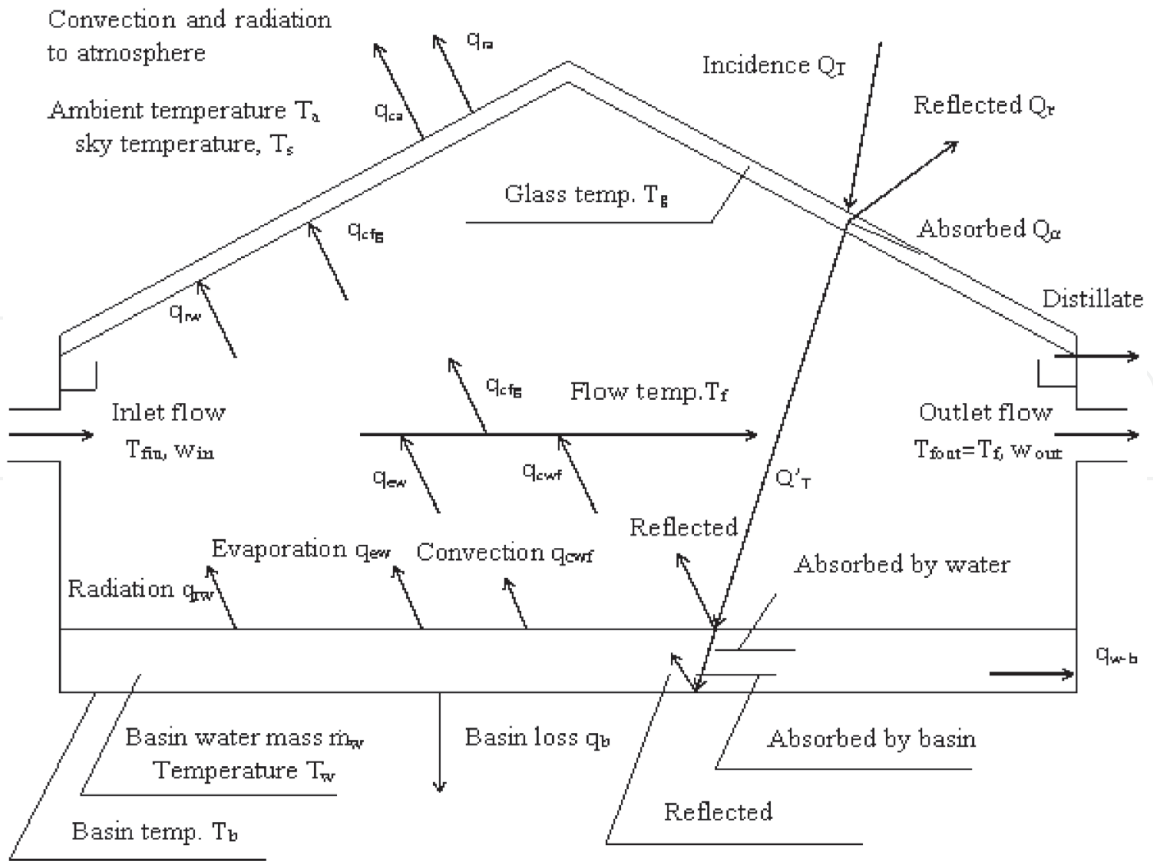
The inputs for SOLSTILL can be in form of hourly weather data if the measured hourly solar radiation and ambient temperature are available. If not, the weather data generation function in SOLSTILL can be called to generate weather sequences with input as monthly average daily solar radiation and ambient temperature values of 12 months [27]. Both modes of weather data in SOLSTILL (i.e., input hourly weather data and generation mode) are used in this study. For Mode 1, hourly measured weather values, achieved from National Center for Hydro-Meteorological Forecasting [28], are input the program. For Mode 2, the weather data generation function in SOLSTILL does its job. The outputs of SOLSTILL consist of hourly amounts of distillate water, hourly temperatures of the cover, basin water and the basin, etc. In this study, only hourly amounts of distillate water are considered.



**Figure 13.**  
The heat and mass transfer processes in a conventional solar still.



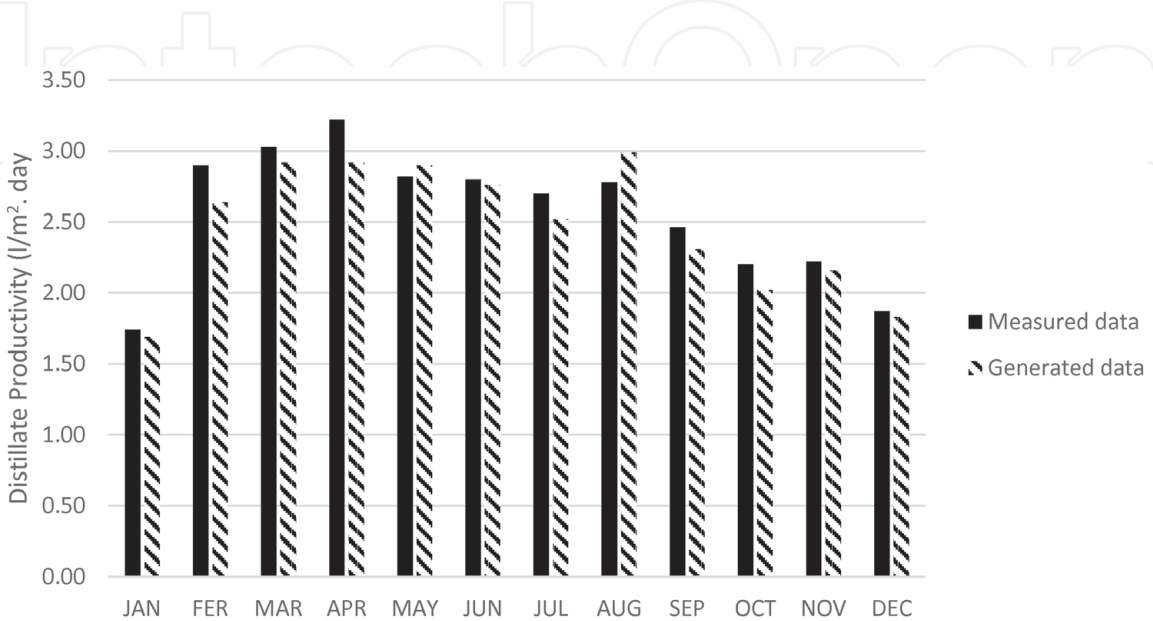
**Figure 14.**  
Schematic diagram of a forced circulation solar still with enhanced water recovery.



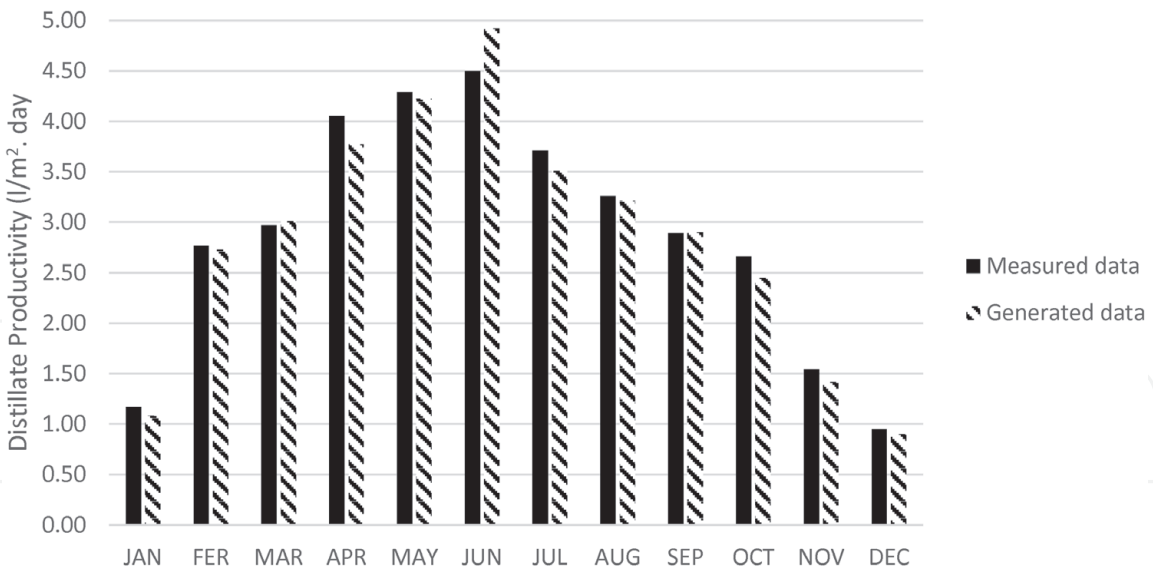
**Figure 15.**  
The heat and mass transfer process in a forced circulation solar still.

Then, daily and monthly average daily amounts of distillate water are achieved. **Figures 16** and **17** show the monthly average daily distilled water of a conventional solar still for Ho Chi Minh City and Da Nang whereas **Figures 18** and **19** show those of a forced circulation solar still with enhanced water recovery, respectively.

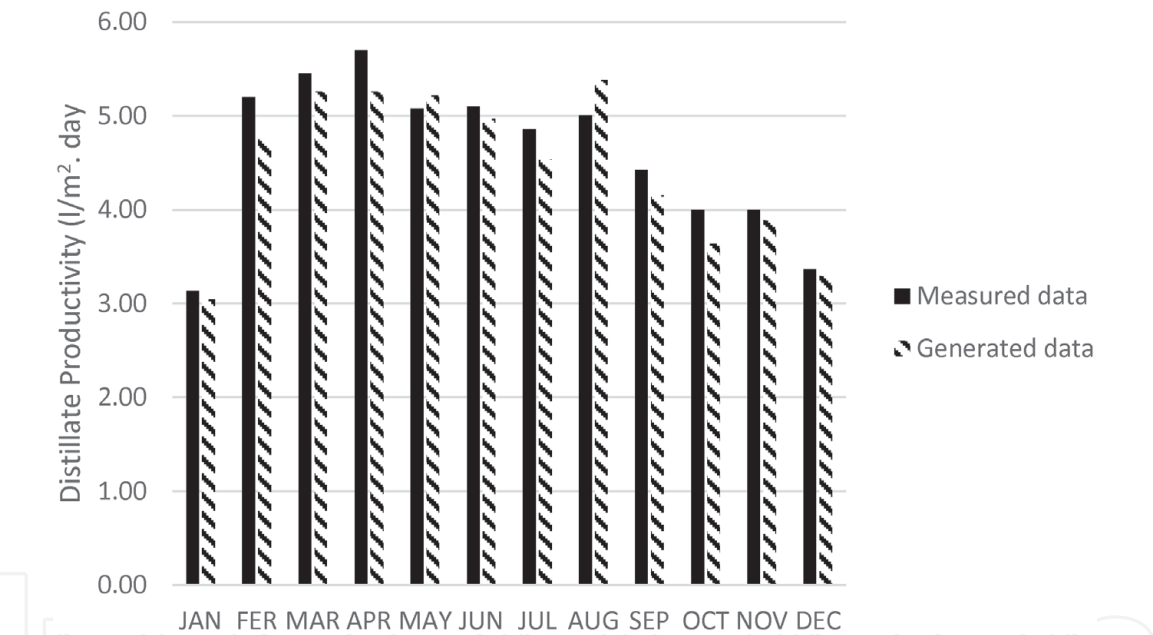
As shown in **Figures 16–19**, the errors of the predicted monthly average daily distillate productivity of both a conventional passive solar still and a forced circulation solar still with measured and generated weather series as input data are very



**Figure 16.**  
Monthly average daily distillate productivity of a conventional solar still in Ho Chi Minh City.



**Figure 17.**  
*Monthly average daily distillate productivity of a conventional solar still in Da Nang.*

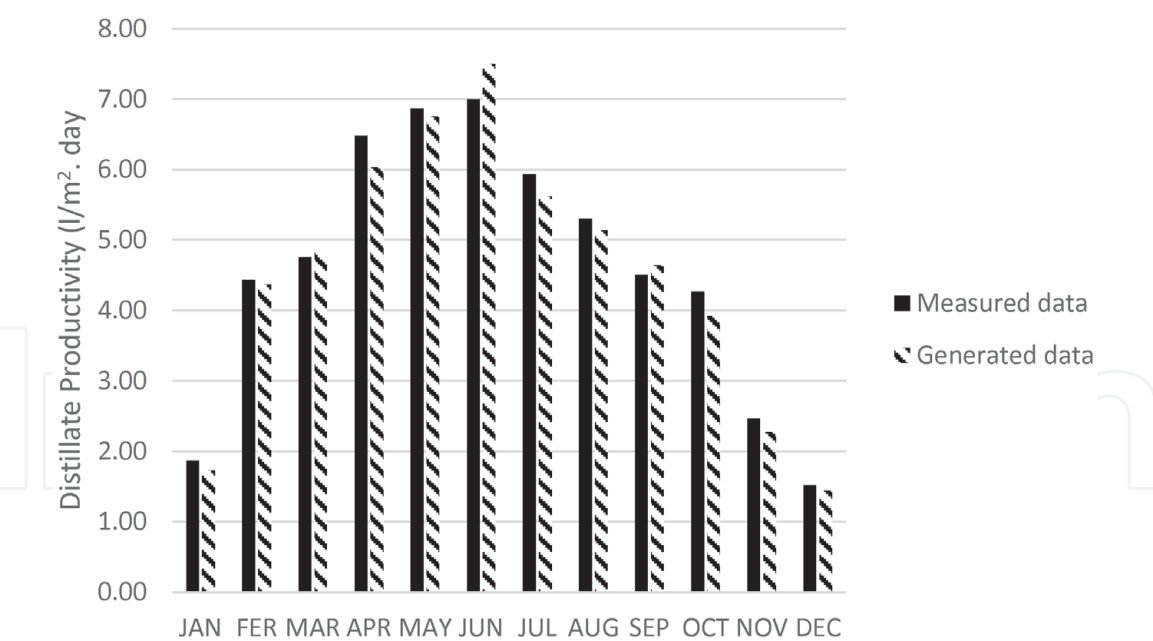


**Figure 18.**  
*Monthly average daily distillate productivity of a forced circulation solar still in Ho Chi Minh City.*

small. The largest error is 9.3%, occurred in April in Ho Chi Minh City for a conventional solar still. The errors of predicted yearly average daily distillate productivities are less than 5%. Therefore, it can be expected that the weather data generated from the proposed models can be used to run any simulation programs for solar distillation systems.

6. Conclusion

In this study, to generate daily clearness index sequences for Ho Chi Minh City and Da Nang, two cities presenting for two climate types in tropical region, Aguiar’s model was chosen. Then a modified model of Graham was proposed to generate hourly clearness index sequences from generate daily clearness index series for these



**Figure 19.**  
*Monthly average daily distillate productivity of a forced circulation solar still in Da Nang.*

two locations. After that, a model to generate hourly ambient temperature sequences from monthly average daily ambient temperatures was presented. Having been proved by some statistic configurations and the predicted distillate productivities of solar still simulations, the models in this study are accurate in predicting daily and hourly irradiances and ambient temperature sequences. Especially, the model proposed in this study to generate the hourly solar radiation values is much simpler compared with Graham model. Therefore, both solar radiation and ambient temperature generating models in this work are believed to be used to calculate daily and hourly weather data for any numerical simulation programs of solar distillation systems with very limited input parameters, including the latitude, monthly average daily clearness index and ambient temperature values of the investigated locations.

### Author details

Bao The Nguyen<sup>1,2</sup>

1 Ho Chi Minh City University of Technology, Vietnam National University,  
Ho Chi Minh City, Vietnam

2 Institute of Sustainable Development – ISED, Ho Chi Minh, Vietnam

\*Address all correspondence to: thebao@hcmut.edu.vn

### IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 



## References

- [1] J. Duffie, and W. Beckman, *Solar Engineering for Thermal Processes - 4th Edition*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2013.
- [2] J. Boland, "Time series modelling of solar radiation". *Modelling Solar Radiation at the Earth's Surface*, V. Badescu (Ed.), Springer-Verlag Berlin Heidelberg, 2008, pp. 283-312.
- [3] B. Nguyen, and V. Hoang, "The study of direct solar radiation data in the project of mapping the solar resource and potential in Vietnam". *The 5th International Conference on Sustainable Energy 2017*, Ho Chi Minh City, Vietnam.
- [4] C. Fernandez-Peruchena, L. Ramirez, I. Pagola, and M. Gaston, "Assessment of models for estimation of hourly irradiation series from monthly mean values". *15th SolarPACES Conference*, Berlin, Germany, pp.121-126. Hal-00919043.
- [5] K. Brecl and M. Topic, "Development of a stochastic hourly solar irradiation model". *International Journal of Photoenergy* Volume 2014, DOI: 10.1155/2014/376504, 7 pages.
- [6] J. Bright, C. Smith, P. Taylor, and R. Crook, "Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data". *Solar Energy* 115, pp. 229-242.
- [7] M. Hofmann, S. Riechelmann, C. Crisosto, R. Mubarak, and G. Seckmeyer, "Improved synthesis of global irradiance with one-minute resolution for PV system simulations". *International Journal of Photoenergy* Volume 2014, DOI: 10.1155/2014/808509. 10 pages.
- [8] T. Soubdhan, and R. Emilion, "Stochastic differential equation for modelling global solar radiation sequences". *Modelling, Identification and Control*, GOSIER, France. DOI: 10.2316/P.2010.702-099. Hal-01823269.
- [9] L. Magnano, J. Boland, and R. Hyndman, "Generation of synthetic sequences of half hourly temperature". *Environmetrics*, Vol. 19, No. 8, pp. 818-835.
- [10] L. Mora-Lopez, "A new procedure to generate solar radiation time series from machine learning theory". *Modelling Solar Radiation at the Earth's Surface*, V. Badescu (Ed.), Springer-Verlag Berlin Heidelberg, 2008, pp. 313-326.
- [11] T. Gafurov, J. Usaola, and M. Prodanovic, "Incorporating spatial correlation into stochastic generation of solar radiation data". *Solar Energy* 115, pp. 74-84.
- [12] W. Hammed, B. Sawadi, S. Al-Kamil, M. Al-Radhi, and R. Abd-Alhameed, "Prediction of solar irradiance based on artificial neural networks". *Invention* Vol. 4, No. 45. DOI: 10.3390/inventions40300045.
- [13] M. Vakili, S. Sabbagh-Yazdi, and K. Kalhor, "Using artificial neural networks for prediction of global solar radiation in Tehran considering particulate matter air pollution". *Energy Procedia*. Vol. 74, pp. 1205-1212.
- [14] S. Priya, and M. Idbal, "Solar radiation prediction using artificial neural networks". *International Journal of Computer Applications* Vol. 116, No. 6.
- [15] F. Tymvios, S. Michaelides, and C. Skouteli, "Estimation of surface solar radiation with artificial neural networks". *Modelling Solar Radiation at the Earth's Surface*, V. Badescu (Ed.), Springer-Verlag Berlin Heidelberg, 2008, pp. 221-246.

- [16] A. Sfetsos, and A. Coonick, "Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques". *Solar Energy* 68, pp. 169-178.
- [17] J. Wu, and C. Chan, "Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN". *Solar Energy*, vol. 85, pp. 808-817.
- [18] L. Mora-Lopez, J. Mora, R. Morales-Bueno, and M. Sidrach-de-Cardona, "Modelling time series of climatic parameters with probabilistic finite automata". *Environmental Modelling & Software*, Vol. 20, pp. 753-760.
- [19] R. Aguiar, M. Collares-Pereira, and J. Conde, "Simple procedure for generating sequences of daily radiation values using a library of Markov transition matrices". *Solar Energy* 40, pp. 269-279.
- [20] V. Graham and K. Hollands, "A method to generate synthetic hourly solar radiation globally". *Solar Energy*, Vol. 44, No. 6, pp. 333-341.
- [21] B. Nguyen, and T. Pryor, "Generating artificial weather data sequences for Australian conditions". *The 34th Annual Conference of ANZSES – Solar'96: Energy for Life*, Darwin, Australia, pp. 101-108.
- [22] C. Mustacchi, V. Cena, and M. Rocchi, M. "Stochastic simulation of hourly global radiation sequences". *Solar Energy*, 23, pp. 47-51.
- [23] R. Aguiar and M. Collares-Pereira, TAG: a time dependent-, autoregressive, Gaussian model for generating synthetic hourly radiation. *Solar energy* 49, pp. 167-174.
- [24] HOMER Energy LLC. Hybrid Optimization of Multiple Energy Resources, Ver 3.14.0. HOMER Energy, Boulder, CO, USA, 2020.
- [25] K.M.Knight, S.A. Klein and J.A. A methodology for the synthesis of hourly weather data. *Solar Energy* 46, 1991, pp. 109-120.
- [26] D.G.Erb, S.A. Klein and W.A. Beckman. Estimation of degree-days and ambient temperature bin data from monthly average temperatures. *ASHRAE Journal* 25, 1983, pp. 60-65.
- [27] B. Nguyen. The Mathematical Model of Basin-Type Solar Distillation Systems, Distillation - Modelling, Simulation and Optimization, Vilmar Steffen, IntechOpen, 2019, DOI: 10.5772/intechopen.83228. Available from: <https://www.intechopen.com/books/distillation-modelling-simulation-and-optimization/the-mathematical-model-of-basin-type-solar-distillation-systems>
- [28] National Center for Hydro-Meteorological Forecasting. <http://www.nchmf.gov.vn>