Can Tho University Journal of Science
website: sj.ctu.edu.vn

STOCHASTIC MODELING FOR DAILY CLEARNESS INDEX SEQUENCE IN CAN THO CITY

Tran Van Ly
College of Natural Sciences, Can Tho University, Vietnam

ARTICLE INFO

Received date: 08/08/2015
Accepted date: 19/02/2016

KEYWORDS

Hidden Markov model, maximum likelihood estimation, expectation maximization algorithm, filter, clear index, solar radiation

ABSTRACT

A stochastic model for daily clearness index sequence in Can Tho city has been proposed. This model based on a pair of stochastic processes, being called the state process and the observation process. The random dynamic of meteorological regimes in random medium was modellized by the state process, a hidden homogeneous Markov chain. The observation process, which represents the daily clearness index sequences, was formed by a real value function whose values are corrupted by Gaussian noise. Parameters of the model were estimated from the real data using Maximum Likelihood estimation via Expectation Maximization algorithm. The simulated data were used to estimate the experimental distribution of daily clearness index sequences.

Cited as: Ly, T.V. 2016., Stochastic modeling for daily clearness index sequence in Can Tho city. Can Tho University Journal of Science. Vol 2: 90-99.

1 INTRODUCTION

The predicting short-term average energy delivery of solar collectors can be based on the precise knowledge of statistic (or physique) models of the global solar radiation G_t or the frequency distribution of its dimensionless form, the clearness index:

$$k_t = \frac{G_t}{I_t}, \qquad (1)$$

where I_t is the extraterrestrial solar radiation.

For the long-term predictions, the clearness index are often considered over a given time interval Δt :

$$K_h^{\Delta t} = \frac{\int_{\Delta t} G_t dt}{\int_{\Delta t} I_t dt} . \qquad (2)$$

The usual used integration periods are the day and the hour, termed daily clearness index and hourly clearness index, respectively.

Sahin and Sen (2008) stated that daily clearness index denoted as K_h, based on the well known Angstrom-type correlation between K_h and sunshine duration, some authors applied the regression technique to develop the linear or non-linear statistic models for K_h which can be used to estimate the daily, monthly or annual global radiation from simple measurements of sunshine duration. All these models are essentially the outcome of considering deterministic components of solar radiation sequences; stochastic characteristics are considered less powerful.

In order to understand a better model of the behaviour of solar radiation and clearness, which is also ruled by the stochastic parameters (frequency and height of the clouds and their optical properties, atmospheric aerosols, ground albedo, water vapour and atmospheric turbidity), we propose a new approach for modeling the Clearness Index Sequences (CIS). That is a stochastic model of Hidden Markov Model (HMM) type, which can represent the CIS under the random effects of meteorological events. Then, a simulated data application which will be considered at our model, which is estimating experimental distribution of CIS. This is very useful in predicting long-term average energy delivery of solar collectors.

For the problem of parameter estimation, considering the relation between complete data and incomplete data, the Expectation Maximization (EM) algorithm will be applied, where the stationary and converging properties were evaluated by (Dembo and Zeitouni, 1986; Dempster *et al.*, 1977). The used equations of filter processes for updating parameters are referred to the results in (Elliott *et al.*, 2010).

In the numerical application, model parameters were estimated from daily CISs having the same monthly characteristic and its simulated data were used to estimate the experimental probability density function (PDF) of K_h for this characteristic month.

This paper is organized as follows. In section 2, we present the establishment of the proposed model for CIS. We describe the EM algorithm and the estimating parameters on real data in Section 3. The applying simulated data for estimating experimental PDF of daily CISs is presented in Section 4. Finally, in Section 5, we conclude with some notes.

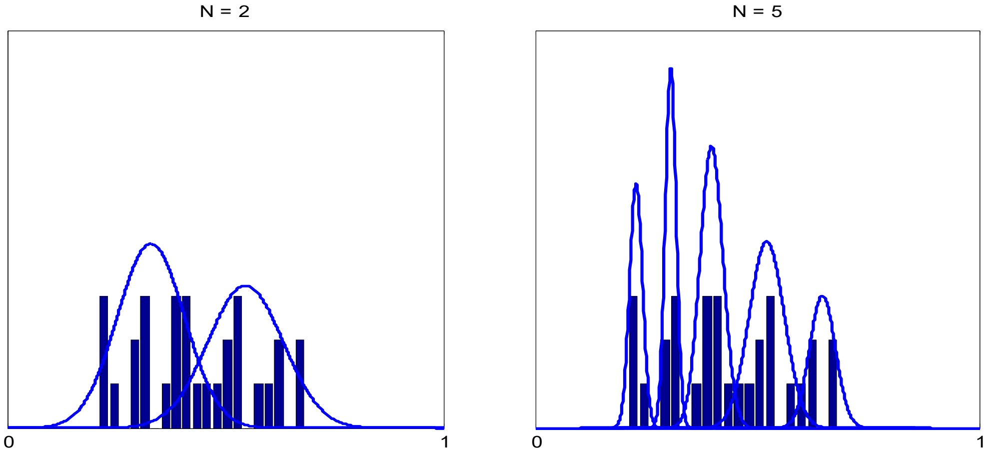

Fig. 1: Histogram of the daily CIS during June 2014 in Can Tho city

2 THE MODEL

The empirical distribution of a daily CIS during a period suggests that the daily CIS distribution could be a Gaussian mixture (for instance, see the histogram of daily CIS during June 2014 in Can Tho city is shown in the Figure 1), each Gaussian component corresponding, may be, to some specific meteorological regime. This has lead to modelize the dynamic of the sequence by a discrete-time HMM, where:

(i) the unobserved *state process* is a Markov chain representing the dynamic of regimes, each daily index belonging to a regime, several daily index belonging eventually to a same regime.

(ii) the *observation process* is such that, given (or within) regime *i*, the various observed daily clearness index are outcomes of a Gaussian distribution whose mean μ_i and standard deviation σ_i depend on regime $i, i = 1, 2, \ldots, N$.

Actually, each regime corresponds to a Gaussian component of the suggested Gaussian mixture, and in terms of probabilistic classification, each regime corresponds to a (Gaussian) class. The advantage of considering a HMM is that it provides a parametric description of the random dynamic of the regimes, which is not the case in a classification setting.

2.1 State process

We assume that there are $N \geq 1$ meteorological regimes, regime *i* being represented by e_i, the unit column vector of R^n with 1 at position $i, i = 1, 2, ..., N$.

The random dynamic of meteorological regimes will be modellized by an unobserved or hidden homogeneous Markov chain $(X_h)_{h=0,1,2,...}$, called the *state process*, with state space $\mathbf{S} = \{e_1, e_2, \ldots, e_N\}$ and probability transition matrix $A = (a_{ji})$, where

$$a_{ji} = P(X_{h+1} = e_j \mid X_h = e_i), \quad i, j = 1, 2, \ldots, N.$$

Note that $a_{ii} = 1 - \sum_{j \neq i} a_{ji}, i = 1, 2, \ldots, N.$

We assume that the distribution of X_0 is π_0 known.

2.2 Observation process and model parameters

The random values of a daily CIS (K_h) are modelled by the so-called *observation process* as follows. In regime i, that is when the Markov chain is in state e_i $(i = 1, 2, \ldots, N)$, the daily clearness index K_h will be considered as an outcome of a Gaussian distribution $\mathrm{N}(\mu_i, \sigma_i^2)$ depending on regime i. In other words:

$$\mathbf{1}_{(X_h = e_i)} K_h = \mathbf{1}_{(X_h = e_i)} (\mu_i + \sigma_i w_h), \quad h = 1, 2, \ldots$$

where w_h are independent random variables having $\mathrm{N}(0,1)$ and μ_i, σ_i are estimated parameters.

The model proposed for daily CIS under the random effects of meteorological events will be the HMM with the state process (X_h) and the observation process (K_h) defined by

$$K_h = \sum_{i=1}^{N} \mathbf{1}_{(X_h = e_i)} K_h = \sum_{i=1}^{N} \mathbf{1}_{(X_h = e_i)} (\mu_i + \sigma_i w_h), \quad h = 1, 2, \ldots$$

The prime symbol denotes transpose, let

$$\mu = (\mu_1, \mu_2, \ldots, \mu_N)',$$
$$\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_N)',$$

we have equivalently

$$K_h = \langle X_h, \mu \rangle + \langle X_h, \sigma \rangle w_h, \quad h = 1, 2, \ldots, \qquad (3)$$

where $\langle \cdot , \cdot \rangle$ denoting the inner product.

The parameter set of the proposed model is $\theta = \{a_{ji}, 1 \le i \ne j \le N; \mu_1, \mu_2, \ldots, \mu_N; \sigma_1, \sigma_2, \ldots, \sigma_N\}$.

2.3 Some notations

In order to estimate parameters of the model, we represent some necessary notions listed below:

(i) Number of jumps of the state process from e_i to e_j:

$$J_h^{ij} = \sum_{l=1}^{h} \langle X_{l-1}, e_i \rangle \langle X_l, e_j \rangle.$$

(ii) Occupation time of the state process in state e_i:

$$O_h^i = \sum_{l=1}^{h} \langle X_{l-1}, e_i \rangle.$$

(iii) Level sums of the observation process in state e_i:

$$T_h^i(g) = \sum_{l=1}^{h} g(K_l) \langle X_{l-1}, e_i \rangle.$$

(iv) Filtration of incomplete data:

$$\mathcal{K}_h = \sigma\{K_1, K_2, \ldots, K_h\}.$$

(v) Filtration of complete data:

$$\mathcal{G}_h = \sigma\{X_0, X_1, X_2, \ldots, X_h, K_1, K_2, \ldots, K_h\}.$$

(vi) With $H_h \equiv J_h^{ij}, O_h^i$ *or* $T_h^i(g)$, the normalized filter of proces H_h:

$$\pi(H_h) = E(H_h \mid \mathcal{K}_h).$$

Where $\sigma\{U\}$ denotes the σ-algebra generated by the set *U* and $g(K_l) = K_l$ or $g(K_l) = K_l^2$.

3 PARAMETER ESTIMATION

In this Section we represent the results of updating ML estimates for parameters using EM algorithm.

3.1 EM Algorithm

We wish to determine a new parameter set $\hat{\theta}$, which maximizes the complete data log-likelihood function via EM algorithm.

The complete data log-likelihood function is defined by:

$$Q(\hat{\theta},\theta) = E_\theta\left(\log \Lambda_h^{\theta\hat{\theta}} \mid K_h\right), \tag{4}$$

where $\Lambda_h^{\theta\hat{\theta}} = \dfrac{dP_{\hat{\theta}}}{dP_\theta} \Big| G_h$ and P_θ denote the probability measure depending on the parameter set θ.

Starting from an initial value $\hat{\theta}^{(0)}$, iterations of EM algorithm will generate a sequence $\left\{\hat{\theta}^{(p)}, p \ge 1\right\}$ of estimates for θ. Each iteration consists of the two following steps:

E-Step (*Expectation Step*): Set and $\theta = \hat{\theta}^{(p)}$ compute

$$Q(\hat{\theta},\hat{\theta}^{(p)}) = E_{\hat{\theta}^{(p)}}\left(\log \Lambda_h^{\hat{\theta}^{(p)}\hat{\theta}} \mid K_h\right).$$

M-Step (*Maximization Step*): Find $\hat{\theta}^{(p+1)} = \arg\max_{\hat{\theta}\in\Theta} Q\left(\hat{\theta},\hat{\theta}^{(p)}\right)$.

We repeat from **E-Step** with *p = p + 1*, unless a stopping test is satisfied. The stationary and converging properties of the EM algorithm had been evaluated by (Dembo and Zeitouni, 1986; Dempster *et al.*, 1977).

3.2 Updating Parameter

In the each iteration of EM algorithm, updating the transition probabilities a_{ji} is as follows (Elliott *et al.*, 2010):

$$\hat{a}_{ji} = \frac{\pi\left(J_h^{ij}\right)}{\pi\left(O_h^i\right)}, \quad 1 \le i \ne j \le N, \tag{5}$$

where $\pi\left(J_h^{ij}\right)$ and $\pi\left(O_h^i\right)$ are the normalized filters of the number of jumps and the occupation time, respectively.

We now consider the update from μ and σ to $\hat{\mu}$ and $\hat{\sigma}$, respectively.

We have

$$\Lambda_h^{\theta\hat{\theta}} = \prod_{l=1}^{h} \frac{\langle X_l,\sigma\rangle \varphi\left(\dfrac{K_l - \langle X_l,\hat{\mu}\rangle}{\langle X_l,\hat{\sigma}\rangle}\right)}{\langle X_l,\hat{\sigma}\rangle \varphi\left(\dfrac{K_l - \langle X_l,\mu\rangle}{\langle X_l,\sigma\rangle}\right)}, \tag{6}$$

where $\varphi(\cdot)$ denotes $\mathrm{N}\left(0,1\right)$ density function.

From (4) and (6), we get:

$$Q(\hat{\theta},\theta) = E_\theta\left\{\sum_{l=1}^{h}\left[\log\frac{1}{\langle X_l,\sigma\rangle} - \frac{1}{2}\left(\frac{K_l - \langle X_l,\hat{\mu}\rangle}{\langle X_l,\hat{\sigma}\rangle}\right)^2\right] \mid K_h\right\} + R\left(\theta,K_h\right), \tag{7}$$

where the function $R\left(\theta,K_h\right)$ does not depend on $\hat{\theta}$.

E-Step: Set $\theta = \hat{\theta}^{(p)}$ and rewrite (7) as

$$Q\left(\hat{\theta},\hat{\theta}^p\right) = \sum_{i=1}^{N} E_{\hat{\theta}}\left\{\sum_{l=1}^{h}\left[\langle X_l,e_i\rangle \log\frac{1}{\hat{\sigma}_i} - \frac{1}{2}\frac{\langle X_l,e_i\rangle}{\hat{\sigma}_i^2}\left(K_l-\mu_i\right)^2 \mid K_h\right]\right\} + R\left(\hat{\theta}^p,K_h\right)$$

$$= \sum_{i=1}^{N}\left\{\pi\left(O_h^i\right)\log\frac{1}{\hat{\sigma}_i} - \frac{1}{2\hat{\sigma}_i^2}\left[\pi\left(T_h^i\left(K_h^2\right)\right) - 2\hat{\mu}_i\,\pi\left(T_h^i\left(K_h\right)\right) + \hat{\mu}_i^2\,\pi\left(O_h^i\right)\right]\right\} + R\left(\hat{\theta}^p,K_h\right).$$

M-Step: Let us find now $\hat{\theta}^{(p+1)} = \arg\max_{\hat{\theta}\in\Theta} Q\left(\hat{\theta},\hat{\theta}^{(p)}\right)$:

Taking derivative of $Q\left(\hat{\theta},\hat{\theta}^p\right)$ with respect to $\hat{\mu}_i, i = 1, 2, \ldots, N$, we obtain

$$\frac{\partial}{\partial\hat{\mu}_i} Q\left(\hat{\theta},\hat{\theta}^p\right) = -\frac{1}{2\hat{\sigma}_i^2}\left[-2\,\pi\left(T_h^i\left(K_h\right)\right) + 2\hat{\mu}_i\,\pi\left(O_h^i\right)\right].$$

Now $\dfrac{\partial}{\partial\hat{\mu}_i} Q\left(\hat{\theta},\hat{\theta}^p\right) = 0$ yields

$$\hat{\mu}_i = \frac{\pi\left(T_h^i\left(K_h\right)\right)}{\pi\left(O_h^i\right)}. \tag{8}$$

(i) Similarly, for $i = 1, 2, \ldots, N$, $\dfrac{\partial}{\partial\hat{\sigma}_i} Q\left(\hat{\theta},\hat{\theta}^p\right) = 0$ yields

$$\hat{\sigma}_i^2 = \frac{1}{\pi\left(O_h^i\right)}\left[\pi\left(T_h^i\left(K_h^2\right)\right) - 2\hat{\mu}_i\,\pi\left(T_h^i\left(K_h\right)\right) + 2\hat{\mu}_i^2\pi\left(O_h^i\right)\right]. \tag{9}$$

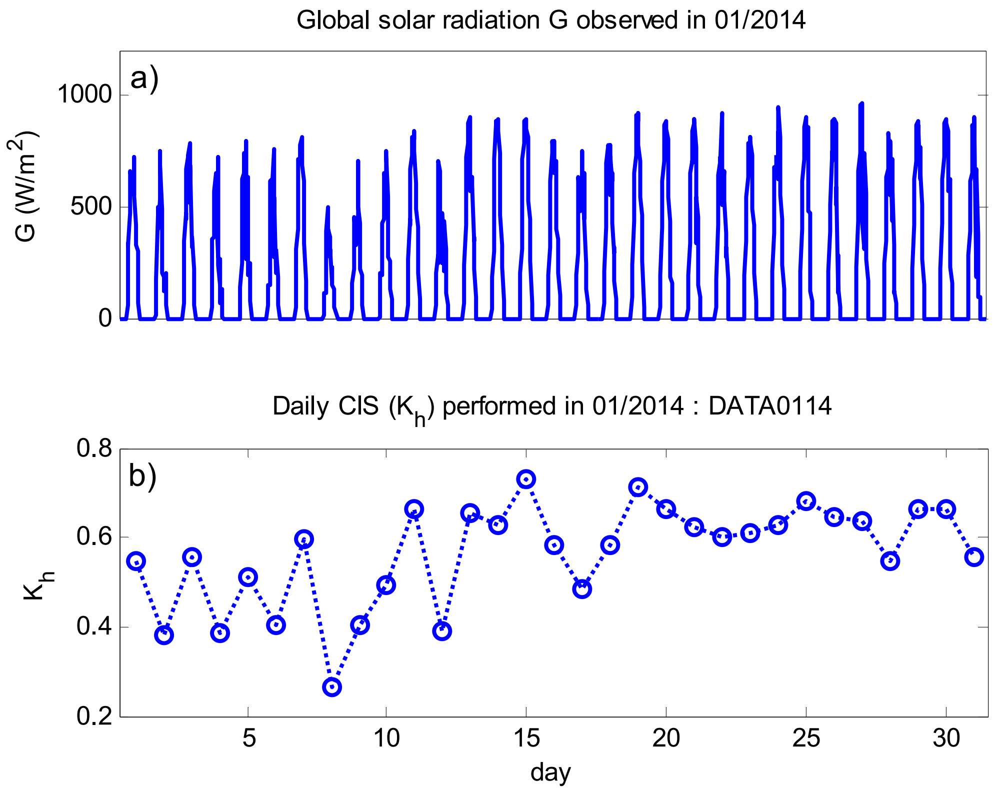

Fig. 2: Global solar radiation and Daily CIS performed in 01/2014, Can Tho city

Table 1: Daily CIS K_h performed in January and June 2014, Can Tho city

	Day	K_h	Day	K_h	Day	K_h	Day	K_h	Day	K_h
	1	0.5469	7	0.5977	13	0.6559	19	0.7118	25	0.6833
	2	0.3801	8	0.2651	14	0.6298	20	0.6655	26	0.6469
	3	0.5563	9	0.4016	15	0.7304	21	0.6250	27	0.6386
January	4	0.3858	10	0.4932	16	0.5841	22	0.5992	28	0.5488
	5	0.5118	11	0.6631	17	0.4861	23	0.6092	29	0.6616
	6	0.4058	12	0.3904	18	0.5834	24	0.6281	30	0.6630
									31	0.5570
	1	0.3878	7	0.5937	13	0.3805	19	0.4777	25	0.2100
	2	0.3115	8	0.6693	14	0.3184	20	0.2499	26	0.6296
June	3	0.4515	9	0.5029	15	0.5155	21	0.2270	27	0.5306
	4	0.4185	10	0.3569	16	0.2228	22	0.2934	28	0.5299
	5	0.3154	11	0.4009	17	0.5317	23	0.6827	29	0.3788
	6	0.6309	12	0.2884	18	0.4341	24	0.4017	30	0.5681

3.3 Experiments with real data

Using (5), (8), and (9), the model parameters will be estimated from the observed data via the EM algorithm. The number of states being chosen after examining the data histograms and the Akaike information criterion (AIC) (Scott, 1992). We deal with data coming from a tropical area, but our method can also be tested on other types of climate.

3.3.1 Real data

Using standard formulas of the extraterrestrial radiation reported by Liu. and Jordan (1960), we computed the daily CISs from the global solar radiation measurements performed in the Can Tho city (latitude 10°2′0″N, longitude 105°47′0″E), which is a tropical and monsoonal area with two seasons: rainy, from May to November; and dry, from December to April. Average annual humidity is 83% and temperature 27°C [9, 13].

Our numerical application were carried on the two typical months (see Table 1):

(i) DATA0114 (Figure 2b): a daily CIS K observed in 01/2014, a month of the rainy.

(ii) DATA0614 (Figure 7a): a daily CIS K observed in 06/ 2014, a month of the dry.

(iii) Observing the histograms and examining the AIC (selecting the model with the smallest AIC value) of these data (Figure 1 and Figure 3), we will apply the models with $N=2$ states.

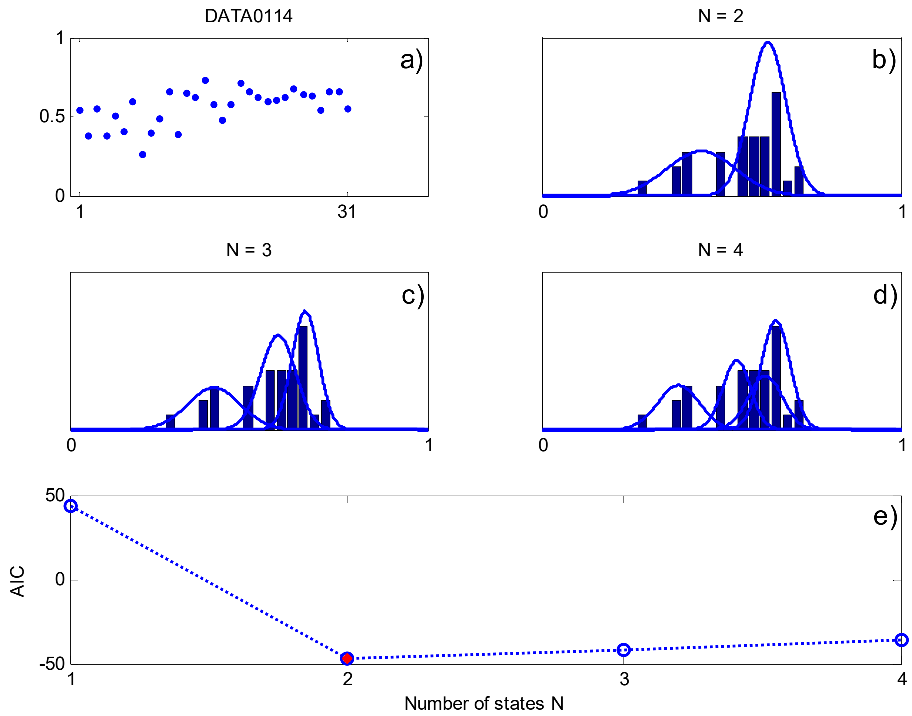

Fig. 3: Selecting the number of states by observing the histograms (Figure 3b, 3c, 3d) and examining the AIC (Figure 3e) of the DATA0114 (Figure 3a)

3.3.2 Estimating model parameters from DATA0114

With the number of states $N=2$, the model is determined by the parameter set $\theta=\{A, \mu, \sigma\}$, where $\mu=(\mu_1, \mu_2)'$, $\sigma=(\sigma_1, \sigma_2)'$ and the transition probability matrix:

$$A=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Initial parameters are given by:

$$A=\begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix},$$

$$\mu=(0.7475, \quad 0.5845)',$$

$$\sigma=(0.1144, \quad 0.1144)'.$$

After 100 iterations of the EM algorithm, we obtain the following estimates:

$$A=\begin{pmatrix} 0.4803 & 0.3085 \\ 0.5197 & 0.6915 \end{pmatrix},$$

$$\mu=(0.6431, \quad 0.5236)',$$

$$\sigma=(0.0421, \quad 0.1194)'.$$

The graphs in Figure 4a, Figure 5a and Figure 6a show the evolution of these estimates.

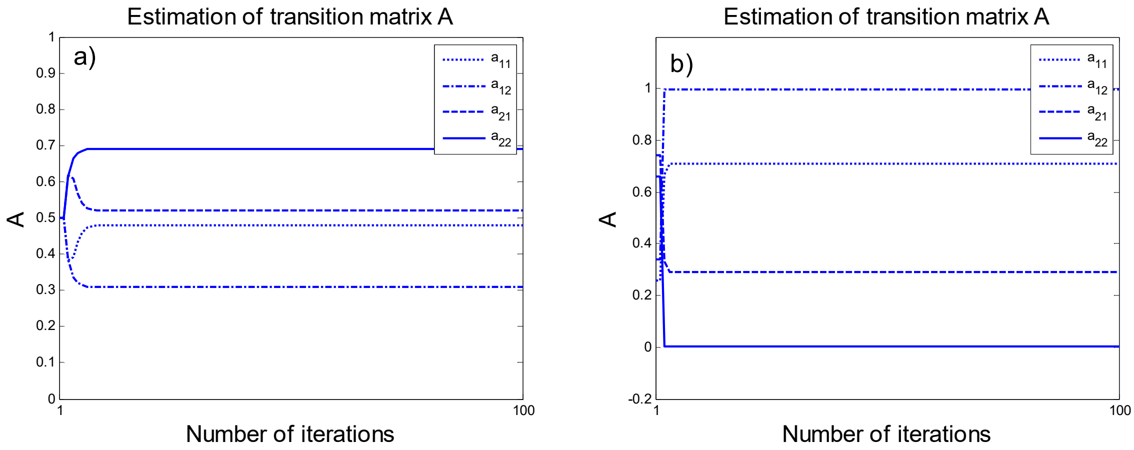

Fig. 4: Estimation of transition probability matrix A: a) From DATA0114; b) From DATA0614

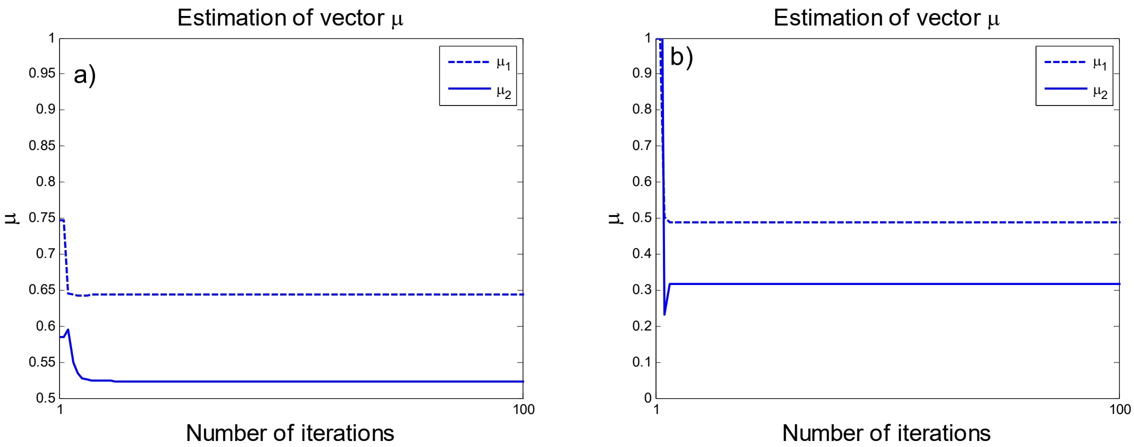

Fig. 5: Estimation of the vector μ: a) From DATA0114; b) From DATA0614

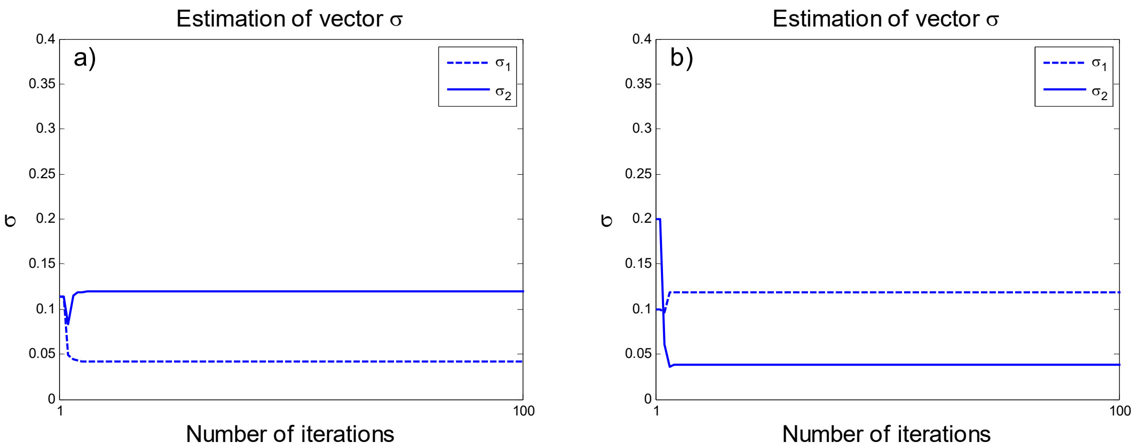

Fig. 6: Estimation of the vector σ : a) From DATA0114; b) From DATA0614

3.3.3 Estimating model parameters from DATA0614

Using DATA0614, with the number of state $N = 2$, model parameters are estimated from the following initial parameter set:

$$A = \begin{pmatrix} 0.2568 & 0.3389 \\ 0.7432 & 0.6611 \end{pmatrix},$$

$$\mu = (1, \quad 2)', \ \sigma = (0.1, \quad 0.2)'.$$

The obtained estimates after 100 iterations of the EM algorithm (evolutions of these estimates are showed in Figure 4b, Figure 5b and Figure 6b):

$$A = \begin{pmatrix} 0.7110 & 0.9961 \\ 0.2890 & 0.0039 \end{pmatrix},$$

$$\mu = (0.4870, \quad 0.3176)',$$

$$\sigma = (0.1185, \quad 0.0378)'.$$

4 APPLICATION

This section presents an application using paths simulated by our models for improvement of the PDF of daily CISs.

Estimating the PDF of daily CIS over a month or over a specific period can be of interest in deciding whether our model estimated over this period still works for a longer period or not. It can also be used for clustering daily CISs observed on various periods.

Indeed, using the model with its parameter estimated from a sample of daily CISs, of one-month-length say, we can simulate a much larger *n*-sample of K_h, say $\{K_1^*, K_2^*, \ldots, K_n^*\}$, over this period and get a smooth estimation of the PDF over this month. Doing the same with another month and getting another *n*-sample, a *KS* (Kolmogorov-Smirnov) test can be performed to reject or not the hypothesis that both PDF are the same. If the hypothesis is rejected (w.r.t. a *p*-value), we can reject the hypothesis that both models are the same. On the other hand, *KS* distance between two sequences, computed from the two *n*-samples, can be used for clustering CISs by performing some standard clustering methods.

4.1 Kernel estimators

The Gaussian kernel estimator of the density is the function $\hat{f}_\delta$ defined as (Scott, 1992):

$$\hat{f}_\delta(x) = \frac{1}{\delta n} \sum_{h=1}^{n} \varphi\left(\frac{x - K_h^*}{\delta} \right),$$

where $\delta > 0$ is a bandwidth (a smoothing parameter) and $\varphi(\cdot)$ denotes the $\mathrm{N}(0,1)$ density function kernel.

This estimator is of course much smoother than the uniform kernel estimator (histogram estimation), that is the empirical PDF $\hat{f}$, defined as follows: divide [0, 1] interval (the range of K_h) into *L* subintervals $(x_{l-1}, x_l]$ of equal length $\Delta x = \frac{1}{L}$ with $x_0 = 0$ and $x_l = l\Delta x$, $l = 1, 2, \ldots, L$, then

$$\hat{f}(x) = \frac{n_l}{n} \frac{1}{\Delta x},$$

where n_l is the number of observed values in the interval $(x_{l-1}, x_l]$, $l = 1, 2, \ldots, L$.

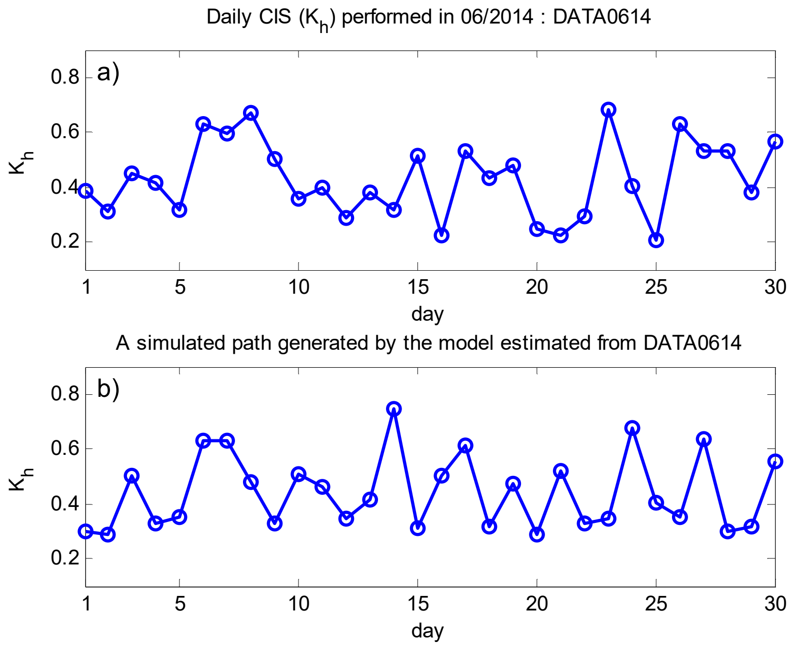

Fig. 7: a) Daily CIS performed in 01/2014, Can Tho city (DATA0614); b) A simulated path for DATA0614

4.2 Experiments

From DATA0614, we have estimated the parameters. We have generated 5000 simulated paths of 30 values from the estimated model (for instance, a simulated path showed in Figure 7b). These simulated paths have the same distribution with DATA0614, evaluated by *KS* test (Joaquim, 2007). Then, from these $n = 5000 \times 30$ simulated values, we have estimated the PDF of K_h for June in Can Tho city (shown in Figure 8). Note that this is an estimation obtained from DATA0614 (daily CIS K_h performed in June 2014); if the model were estimated with more data, for example with adding up the data in 06/2015, 06/2013, 06/2012, ... then the PDF estimation will be better.

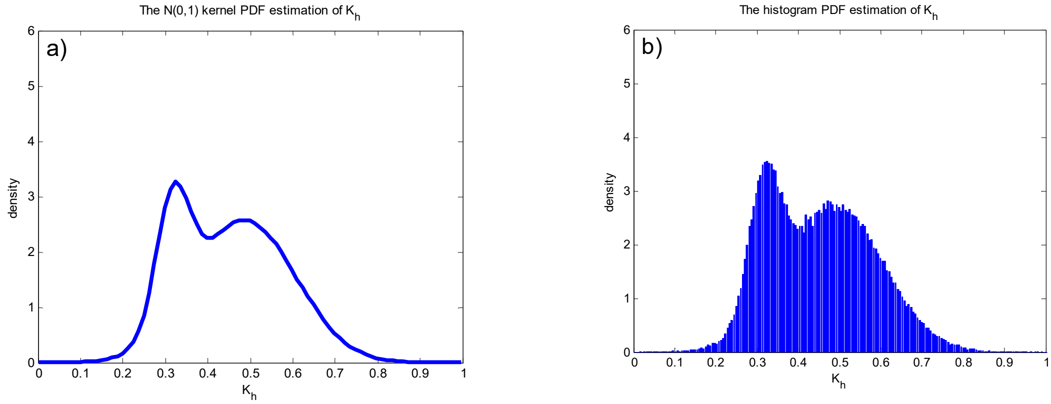

Fig. 8: PDF of (K_h) in June (Can Tho city):

a) N $(0,1)$ kernel estimation; b) Histogram estimation

PDF of K_h were obtained similarly in January from DATA0114 (see Figure 9).

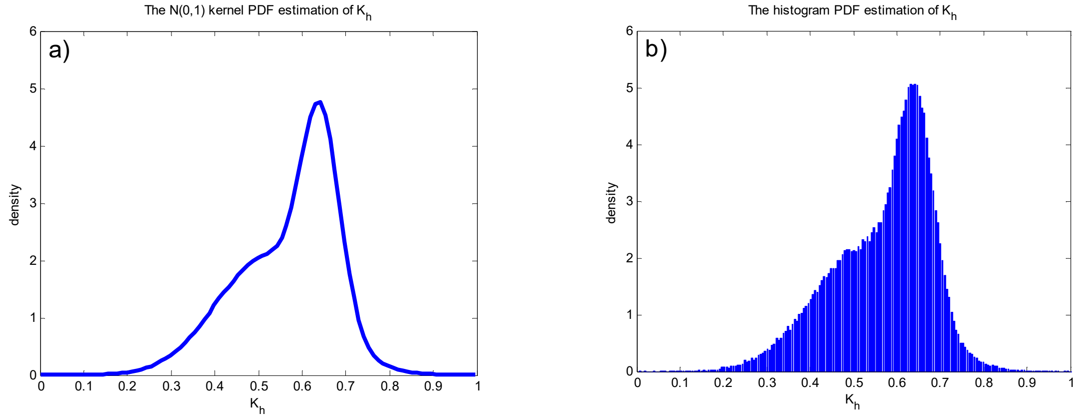

Fig. 9: PDF of $\left(K_h\right)$ in January (Can Tho city):

a) $\mathcal{N}\left(0,1\right)$ kernel estimation; b) Histogram estimation

5 CONCLUSION

Clearness index sequences under the random effects of meteorological events are modellized by the HMM-type, a modelling-type plays a prominent role in a range of application areas. The parameters of model obtained from the ML estimation method via the celebrated EM algorithm. The methodology was tested on real data.

Using estimated parameters, the model will generate the simulated data having the same distribution characteristic of observation data, because it enjoys properties of EM algorithm used in the estimating technique. From this, if the model established from daily CISs observed in the months having the same distribution characteristic then we can use it to generate a large number of simulated paths having this monthly distribution characteristic. Using this large number of simulated values, the obtained estimates of experimental PDF of daily clearness index are very smoothing. This will be very useful for predicting the short-term or long-term average energy delivery of solar systems.

ACKNOWLEDGMENTS

The author is grateful thank to Mr. Phan Thanh Hai, Director of the Meteorological station of Can Tho city, for providing the data to realize the numerical application.

REFERENCES

Bendt, P., Collares-Peraeia, M., Rabl, A., 1981. The frequency distribution of daily insolation values. Solar Energy. 27: 1-5.

Dembo, A., Zeitouni, O., 1986. Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm. Stochastic Processes and their Applications. 23: 91-113.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society, Series B (Methodological). 39(1): 1-38.

Elliott, J.R., Aggoun, L., Moore, J.B., 2010. Hidden Markov Models: Estimation and control. Springer. 377 pages.

Feuillard, T., Abillon, J.M., Martias, C., 1989. The brobability density function of the clearness index: a new approach. Solar Energy. 43(6): 363-372.

James, M.R., Krishnamurthy, V., Le Gland, F., 1996. Time Discretization of Continuous-Time Filters and Smoothers for HMM Parameter Estimation. IEEE transactions on information theory. 42(2): 593-604.

Joaquim, P.M.S., 2007. Applied Statistics Using SPSS, STATISTICA, MATLAB and R. Springer. 505 pages.

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

Nguyen, B.T., Pryor, T.L, 1996. A computer model to estimate solar radiation in Vietnam. Proceedings of WREC, 26-27 May 1996, Murdoch, Australia, 19-25.

Sahin, A.D., Sen, Z., 2008. Solar Irradiation Estimation Methods from Sunshine and Cloud Cover Data. In: Badescu, V. (Ed.). Modeling Solar Radiation at the Earth's Surface. Springer. pp. 246-279.

Scott, D.W., 1992. Mutivariate density estimation: Theory, practice and visualization

Visualization. John Wiley & Son, New York. 430 pages.

Tong, H., 1975. Determination of the order of a Markov chain by Akaike's Information Criterion. Journal of Application Probability. 12: 488-497.

Wikipedia.org/wiki/Can_Tho_City.