Energy and Buildings 43 (2011) 2371–2377

Contents lists available at ScienceDirect

Energy and Buildings

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

First order multivariate Markov chain model for generating annual weather data for Hong Kong

Hongxing Yang*, Yutong Li, Lin Lu, Ronghui Qi

Renewable Energy Research Group (RERG), Department of Building Services Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

ARTICLE INFO

Article history:
Received 26 January 2011
Received in revised form 3 May 2011
Accepted 30 May 2011

Keywords:
Markov chain model
Annual weather data
Building simulation
Transition probability matrix

ABSTRACT

It becomes popular to use computer simulation technique to evaluate the performance of energy utilizations in buildings. The hourly weather data as simulation input is a crucial factor for the successful energy system simulation, and obtaining an accurate set of weather data is necessary to represent the long-term typical weather conditions throughout a year. This paper introduces a stochastic approach, which is called the first order multivariate Markov chain model, to generate the annual weather data for better evaluating and sizing different energy systems. The process for generating the weather data time series from the multivariate Markov transition probability matrices is described using 15-years actual hourly weather data of Hong Kong. The ability of this new model for retaining the statistical properties of the generated weather data series is examined and compared with the current existing TMY and TRY data. The main statistical properties used for this purpose are mean value, standard deviation, maximum value, minimum value, frequency distribution probability and persistency probability of the weather data sequence. The comparison between the observed weather data and the synthetically generated ones shows that the statistical characteristics of the developed set of weather data are satisfactorily preserved and the developed set of weather data can predict and evaluate different energy systems more accurately.

1. Introduction

With more concerns on energy shortage and environmental problems, there are growing interests for researchers, architects and engineers in Hong Kong to use energy simulation techniques to estimate renewable energy application systems in buildings. The simulation tools could help to achieve better energy efficiency and sustainable building development. One key element in building energy simulations is local weather data, i.e., the 8760 hourly records of weather data representative of the prevailing weather conditions as most energy simulation programs require weather data input to drive the thermal models within the simulation tools. This paper is to propose a stochastic approach, i.e., the first order multivariate Markov chain model, to generate the annual weather data for better evaluating and sizing different renewable energy systems.

For annual energy performance prediction and energy system design, there are two kinds of annual weather data used in Hong Kong: one is the Typical Meteorological Year (TMY) and another is the Example Weather Year (EWY) or Test Reference Year (TRY). The EWY or TRY of Hong Kong was developed by Wong and Ngan [1] based on 13-years real hourly weather data (from 1979 to 1991). The selection criterion of TRY is based on the smallest deviation from the long term historical values and the year 1989 was selected as the example year. The TRY selection tends to be rather mild, as extremely high or low mean temperatures were progressively eliminated so that it may not reflect the prevailing "average" weather conditions. To address the shortcomings of the TRY, more sophisticated TMY was developed by Chow and Chan [2] using the hourly measured weather data of a 25 year period (1979–2003) in Hong Kong. Unlike the TRY, the derived TMY is not an actual past year. The TMY is composed of 12 typical meteorological months (TMMs), which are selected by comparing their cumulative distribution function (CDF) with the long-term CDF. By considering the weighting factors of annual and monthly mean values of major weather indices, the month with the smallest deviation from the mean value is taken as 'standard' month and is used for formulating the TMY. The selection of the TMY highly relied on the weighting factors of different metrological variables. Because of strict statistical significance, the simulations for different energy systems should have different TMY inputs. The previous studies only focused on building energy simulation and may not be appropriate for renewable energy application systems such as wind and solar energy applications.

Based on their different weather parameters and weighting factors, Yang and Lu [3] studied the impact of different TMYs and EWYs, and concluded that right selection of the weighting factors of meteorological parameters play an important role in the process

* Corresponding author. Tel.: +852 27665863; fax: +852 27746146.
E-mail address: behxyang@polyu.edu.hk (H. Yang).

0378-7788/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.enbuild.2011.05.035

Nomenclature

B	stationary matrix defined in Eq. (10)
d	humidity ratio of moist air [kg/kg DA]
dec	declination angle in degrees
f, F	transition frequency and its matrix form defined in Eq. (4)
hour	hour of the day
i	initial state
G_{SC}	the solar constant [1366 W/m^2]
G	hourly global radiation on horizontal surface [W/m^2]
G_0	global extraterrestrial solar radiation on horizontal surface [W/m^2]
lati	local latitude in degrees
L	length of the sequence [h]
N_{day}	number of days in a year
P	the one-step transition probability matrix
Q	the one step transition probability matrix of the multivariate model
s	number of the categorical sequences
T	the number of sub-intervals
Ta	dry-bulb air temperature [°C]
TMY	Typical Meteorological Year
TRY	Test Reference Year
EWY	Example Weather Year
V	actual value of weather variable
X	state probability distribution
$\hat{X}$	stationary vector

Greek letters

ε	uniform random number between (0, 1)
θ	Hour angle in degrees
α	statistics in Eq. (18)
γ	statistics in Eq. (19)
λ	model parameter in Eq. (1)

Subscripts

avg	average value
dem	deterministic component of a weather variable
j, k	state of the weather state
h	high boundary of a weather state
l	Lower boundary of a weather state
ram	random component of a weather variable
V	Actual value of weather variable
X	state probability distribution

for generating the TMYs for renewable energy application systems. In addition, the persistent structure of meteorological parameters is also important for estimating the reliability and initial cost of the renewable energy systems with energy storage. Argiriou et al. [4] presented a more detailed examination of 17 major TMY selection methodologies reported in literatures. Compared with their different influences on performance simulations for different conditions (flat plate solar collector, photovoltaic, large scale solar heating system with seasonal storage, annual heating and cooling loads of a typical building), those derived TMYs were also evaluated. They found that the optimal TMY varies from system to system. The difference of the TMYs may lead to erroneous different conclusions of the system performance for sizing design and feasibility study of a certain energy system.

The previous studies reveal that the selection of the TMY or TRY greatly depends on the weighting factors of weather variables. Different types of energy systems require different TMYs or TRYs for performance evaluation and energy prediction. It is necessary to generate the TMY in a different way. This paper introduces a new approach to generate the annual weather data, namely the First Order Multivariate Markov Chain Model, for different energy systems. Since this method does not require pre-assumed relative importance of different weather variables, the obtained weather data does not rely on the weighting factors and can be applied to simulate any energy systems with more than one dominant weather variables. The proposed weather data series were obtained and compared with existing TMYs and TRYs based on 15-years actual hourly weather data of Hong Kong.

2. Description of the first order multivariate Markov chain model

The weather is a multi-dimensional random process, and therefore could be modeled using the stochastic process theories. The Markov chain model is one of these theories, representing a system of elements making transition from one state to another over time. The order of the chain gives the number of time steps in the past influencing the probability distribution of the present state. The use of the Markov chain to simulate the meteorological variable distribution was reported by researchers. Muselli et al. used the first order Markov chain to generate synthetic “typical days” series of global irradiation for sizing stand-alone photovoltaic systems [5]. Maafi and Adane analyzed the performance of the first order two state Markov chain model based on solar radiation properties [6]. With long-term wind speed data, Shamshad et al. [7] compared the prediction accuracy between the first and second order Markov chains. All these researches proved that the Markov chain is a very effective way to describe weather variables, such as solar radiation, temperature and wind speed. However, the previous studies were limited to one single weather variable. The interrelationship between different climate phenomena is omitted in these models.

The first-order multivariate Markov chain model was proposed by Ching to model the behavior of multiple categorical sequences generated by similar sources [8]. In this study, this method is used to generate a series of weather data which involves solar radiation, air dry bulb temperature and absolute humidity. The development of this model is briefly introduced as follow.

Assuming there are s categorical sequences of weather variables and each of them has m possible states, the state probability distribution of the jth weather variable data sequence at time $t=r+1$ depends on the state probabilities of all the weather variables (including jth itself) at time $t=r$. Therefore, in this proposed first-order multivariate Markov chain model, the following relationship could be gained:

$$X_{r+1}^{(j)} = \sum_{k=1}^{s} \lambda_{jk} P^{(jk)} X_r^{(k)}, \quad \text{for } j = 1, 2, \ldots, s \text{ and } r = 0, 1, \ldots, \tag{1}$$

where

$$\lambda_{jk} \geq 0,\ j \geq 1,\ s \geq k \text{ and } \sum_{k=1}^{s} \lambda_{jk} = 1, \quad \text{for } j = 1, 2, \ldots, s \tag{2}$$

and $X_0^{(j)}$ is the initial probability distribution of the jth weather variable. The state probability distribution of the jth weather variable data sequence at the time $r+1$, $X_{r+1}^{(j)}$, depends on the weighted average of $P^{(jk)} X_r^{(k)}$ where $P^{(jk)}$ is the one-step transition probability matrix from the states at time t in the kth weather variable data sequence to the states in the jth weather variable data sequence at time $r+1$, and $X_r^{(k)}$ is the state probability distribution of the kth weather variable at the time r. In matrix form, it can be described by:

$$X_{r+1}=\begin{pmatrix}X_{r+1}^{(1)}\\X_{r+1}^{(2)}\\\vdots\\X_{r+1}^{(s)}\end{pmatrix}=\begin{pmatrix}\lambda_{11}P^{(11)} & \lambda_{12}P^{(12)} & \cdots & \lambda_{1s}P^{(1s)}\\\lambda_{21}P^{(21)} & \lambda_{22}P^{(22)} & \cdots & \lambda_{2s}P^{(2s)}\\\vdots & \vdots & \vdots & \vdots\\\lambda_{s1}P^{(s1)} & \lambda_{s2}P^{(s2)} & \cdots & \lambda_{ss}P^{(ss)}\end{pmatrix}\begin{pmatrix}X_{r}^{(1)}\\X_{r}^{(2)}\\\vdots\\X_{r}^{(s)}\end{pmatrix}\tag{3}$$

or

$$X_{r+1}=QX_r$$

Q is the one step transition probability matrix of the multivariate model that determines the probability for making a transition depending on the current state. These matrices become the basis of future likely weather state.

The Markov chains are stochastic processes that can be parameterized by empirically estimating transition probabilities between discrete states in the observed weather data sequences. A numerical algorithm based on the linear programming is proposed to solve the model parameters $P^{(jk)}$ and λ_{jk}. When the weather data sequence is given, the transition frequency from the states of the kth weather variable data sequence to the states of the jth weather variable data sequence can be counted. Hence, the transition frequency matrix of the data sequence can be constructed. After making normalization, the estimates of the transition probability matrices can also be obtained. More precisely, the transition frequency $f_{i_j i_k}^{(jk)}$ from the state i_k in the weather variable sequence $\{X(k)\}$ to the state i_j in the sequence $\{X(i)\}$ can be counted and therefore the transition frequency matrix for the weather data sequences can be constructed as follows:

$$F^{(jk)}=\begin{pmatrix}f_{11}^{(jk)} & \cdots & f_{m1}^{(jk)}\\f_{12}^{(jk)} & \cdots & f_{m2}^{(jk)}\\\vdots & \vdots & \vdots\\f_{1m}^{(jk)} & \cdots & f_{mm}^{(jk)}\end{pmatrix}\tag{4}$$

From $F^{(jk)}$, $P^{(jk)}$ can be estimated as follows:

$$P^{(jk)}=\begin{pmatrix}p_{11}^{(jk)} & \cdots & p_{m1}^{(jk)}\\p_{12}^{(jk)} & \cdots & p_{m2}^{(jk)}\\\vdots & \vdots & \vdots\\p_{1m}^{(jk)} & \cdots & p_{mm}^{(jk)}\end{pmatrix}\tag{5}$$

where

$$p_{i_j i_k}^{(jk)}=\begin{cases}s\dfrac{f_{i_j i_k}^{(jk)}}{\sum_{i_k=1}^{m}f_{i_j i_k}^{(jk)}} & if\displaystyle\sum_{i_k=1}^{m}f_{i_j i_k}^{(jk)}\neq 0\\ 0 & \text{otherwise}\end{cases}\tag{6}$$

Beside the estimation of $P^{(jk)}$, the parameters λ_{jk} also need to be estimated. Assuming that the multivariate Markov chain model has a stationary vector $\hat{X}$, the vector $\hat{X}$ can be estimated from the data sequences by computing the proportion of the occurrence of each state in each of the weather data sequences:

$$\hat{X}=\left(\hat{X}^{(1)},\hat{X}^{(2)},\ldots,\hat{X}^{(s)}\right)\tag{7}$$

One would expect that:

$$\hat{X}\approx\begin{pmatrix}\lambda_{11}P^{(11)} & \lambda_{12}P^{(12)} & \cdots & \lambda_{1s}P^{(1s)}\\\lambda_{21}P^{(21)} & \lambda_{22}P^{(22)} & \cdots & \lambda_{2s}P^{(2s)}\\\vdots & \vdots & \vdots & \vdots\\\lambda_{s1}P^{(s1)} & \lambda_{s2}P^{(s2)} & \cdots & \lambda_{ss}P^{(ss)}\end{pmatrix}\hat{X}\tag{8}$$

Ching et al. proposed the following linear programming problems to estimate the parameters $\lambda=\{\lambda_{jk}\}$ [8]. For each j,

$$\left\{\begin{array}{l}\min\limits_{\lambda}\omega_j\\ \text{subject to}\\ \begin{pmatrix}\omega_j\\\omega_j\\\vdots\\\omega_j\end{pmatrix}\geq\hat{X}^{(j)}-B\begin{pmatrix}\lambda_{j1}\\\lambda_{j2}\\\vdots\\\lambda_{js}\end{pmatrix},\\ \begin{pmatrix}\omega_j\\\omega_j\\\vdots\\\omega_j\end{pmatrix}\geq-\hat{X}^{(j)}+B\begin{pmatrix}\lambda_{j1}\\\lambda_{j2}\\\vdots\\\lambda_{js}\end{pmatrix},\\ \omega_j\geq 0,\\ \displaystyle\sum_{k=1}^{s}\lambda_{jk}=1,\quad\lambda_{jk}\geq 0,\quad\forall k.\end{array}\right.\tag{9}$$

where

$$B=\left[\hat{P}^{(j1)}\hat{X}^{(1)}|\hat{P}^{(j2)}\hat{X}^{(2)}|\cdots|\hat{P}^{(js)}\hat{X}^{(s)}\right]\tag{10}$$

3. Weather data analysis

The analysis of hourly weather data sequences was carried out by two steps: (1) remove the time dependency of the weather variables; and (2) convert the weather data time series to weather states. The essence of the first step is the stochastic identification which divides weather variables into two elements, namely random and deterministic components:

$$V=V_{\text{dem}}+V_{\text{ram}}\tag{11}$$

The deterministic components have been found by taking the hourly averages within a day of 15 years for every month separately using auto-correlation method. The following two equations are used to calculate the deterministic components for air dry-bulb temperature and absolute humidity and to remove time dependency of the weather variables:

$$\begin{aligned}Ta_{\text{dem}}&=Ta_{\text{avg}}+7.3384\times\cos\left(\frac{2\pi}{365}\times N_{\text{day}}-2.6575\right)\\&\quad+1.5124\times\cos\left(\frac{2\pi}{24}\times\text{hour}+5.3689\right)\end{aligned}\tag{12}$$

$$\begin{aligned}d_{\text{dem}}&=d_{\text{avg}}-0.0062\times\cos\left(\frac{2\pi}{365}\times N_{\text{day}}-0.3163\right)\\&\quad+2.3311\times10^{-4}\times\cos\left(\frac{2\pi}{24}\times\text{hour}+0.7390\right)\end{aligned}\tag{13}$$

Ta_{avg} and d_{avg} are the average values over 15 years, N_{day} is the number of days in a year, and hour is the hour of the day.

Regarding the global solar radiation, its hourly amounts were described by the time series of a clearness index. The clearness index is defined as the ratio of the hourly global radiation on horizontal surface G to the global extraterrestrial solar radiation on horizontal surface G_0 which is calculated by the well-known empirical equations:

$$\text{dec}=23.45\times\sin\left(\frac{360\times(284+N_{\text{day}})}{365}\times\frac{\pi}{180}\right)\tag{14}$$

$$\theta=(\text{hour}-12)\times 15\tag{15}$$

$$\begin{aligned}G_0&=G_{\text{SC}}\times\left[1+0.033\times\cos\left(\frac{2\pi}{365}\times N_{\text{day}}\right)\right]\times[\cos(\text{lati})\\&\quad\times\cos(\text{dec})\times\cos(\theta)+\sin(\text{lati})\times\sin(\text{dec})]\end{aligned}\tag{16}$$

G_{SC} is the solar constant, i.e., 1366 W/m^2.

To convert the random part of weather data into weather states which contain weather variables between certain values, the histogram of these three weather variables were examined and 10 states of each variable were adopted with an upper and lower limited difference of 0.4 times of the standard deviation of each weather data sequence. The effectiveness of this process is tested by the following methods.

3.1. Testing of the Markov chain properties

The Markov chain properties of the obtained weather data states can be tested statistically by checking if the successive events are independent of each other (the null hypothesis) or dependent (the alternative hypothesis). If being dependent, they can form a first order Markov chain. If successive events are independent, then the statistic α is defined as:

$$\alpha = 2\sum_{i,j}^{m} n_{ij} \ \ln\left(\frac{p_{ij}}{p_j}\right) \tag{17}$$

α is distributed asymptotically as χ^2 with $(m-1)^2$ degrees of freedom, where m is the total number of states. n_{ij} is the number of transitions and p_j the marginal probabilities for the jth column of the transition probabilities matrix. The results obtained from the statistics α are larger than the χ^2 value of 103 at the 5% level with 81 degrees of freedom, so we can reject the null hypothesis that the successive transitions are independent, and conclude that the transition of the hourly weather data occurrence could possess the first order Markov chain property.

3.2. Testing of the Markov chain stationary

A Markov process is stationary if its transition probabilities are independent with respect to time. A convenient way to check the stationary is to divide the whole sequence of events into several subintervals, and then compute and compare the transition probability matrix amongst each interval. For a stationary process, all these matrices should be approximately equal to each other. The statistic test is defined as follows:

$$\gamma = 2\sum_{t}^{T}\sum_{i,j}^{m} n_{ij}(t) \ \ln\left(\frac{p_{ij}(t)}{p_{ij}}\right) \tag{18}$$

where T is the number of subintervals, and $n_{ij}(t)$ and $p_{ij}(t)$ represent the number of transitions and the transition probabilities for each subinterval, respectively. If the Markov chain is stationary, the statistics γ has χ^2 distribution with $(T-1)m(m-1)$ degrees of freedom. In this case, the statistics, including solar radiation, air temperature and humidity, are smaller than the χ^2 value of 16,200 at the 5% level with 16,110 degrees of freedom, which means that the Markov chain property of the global irradiation typical day occurrence is stationary.

4. Synthetic generation

To generate the sequences of weather states, the initial state, say i, was selected randomly. Then, the random values between 0 and 1 were produced by a uniform random number generator. For the next state in the first order Markov process, the random number value was compared with the elements of ith row of the cumulative probability transition matrix. If the random number value is greater than the cumulative probability of the previous state but less than or equal to the cumulative probability of the following state, the following state is adopted. The weather states have been converted to the actual weather variable using the following relationship:

$$V = V_l + \varepsilon_i(V_h - V_l) \tag{19}$$

where V_l and V_h are lower boundary and upper boundary of the weather states, respectively and ε_i is the uniform random number between (0, 1).

The procedure was repeated for other hours. However, if two consecutive hours, $t=r-1$ and $t=r$, belong to the same state k, it is considered that the sequences of hours with the same state k began at hour $r-1$. Thus, to determine the length of this sequence (hours of state k), the following procedure was used: a random number ε_i, between 0 and 1, was randomly chosen; and elements of persistence probabilities ($P(r-1:k)$, $P(r-2:k)$, ..., $P(r-L:k)$) were added until their sum is greater than ε_i. Thus, the length of the sequence is L. In this case, the hours $r-1$ to $r+L-1$ belong to the same state k. The generation process is shown in Fig. 1.

5. Results and validation

Using this method, the one-year hourly total solar radiation, humidity and ambient air temperature can be found. The synthetically generated data, by the first order multivariate Markov chain

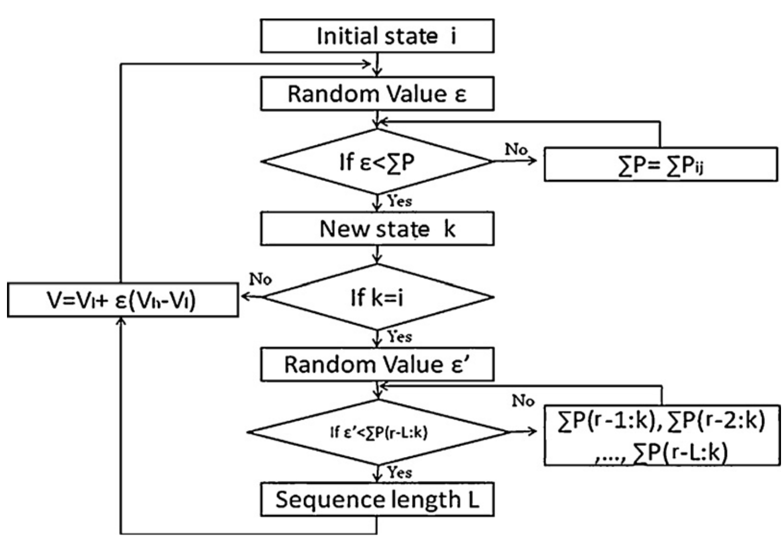

Fig. 1. Flowchart of the weather data synthetic generation process.

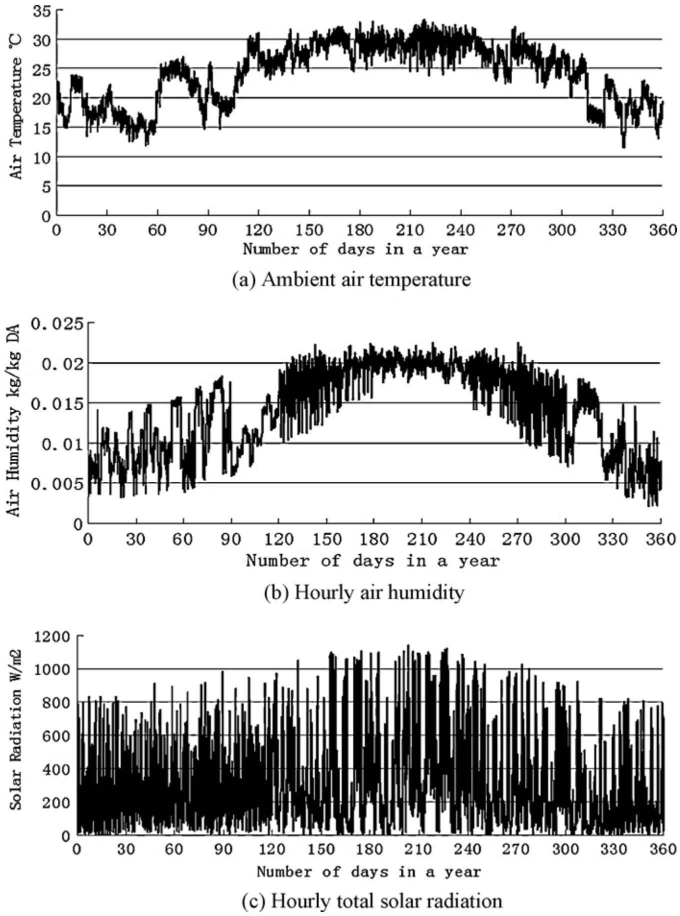

Fig. 2. Simulated hourly total solar radiation, humidity and ambient air temperature.

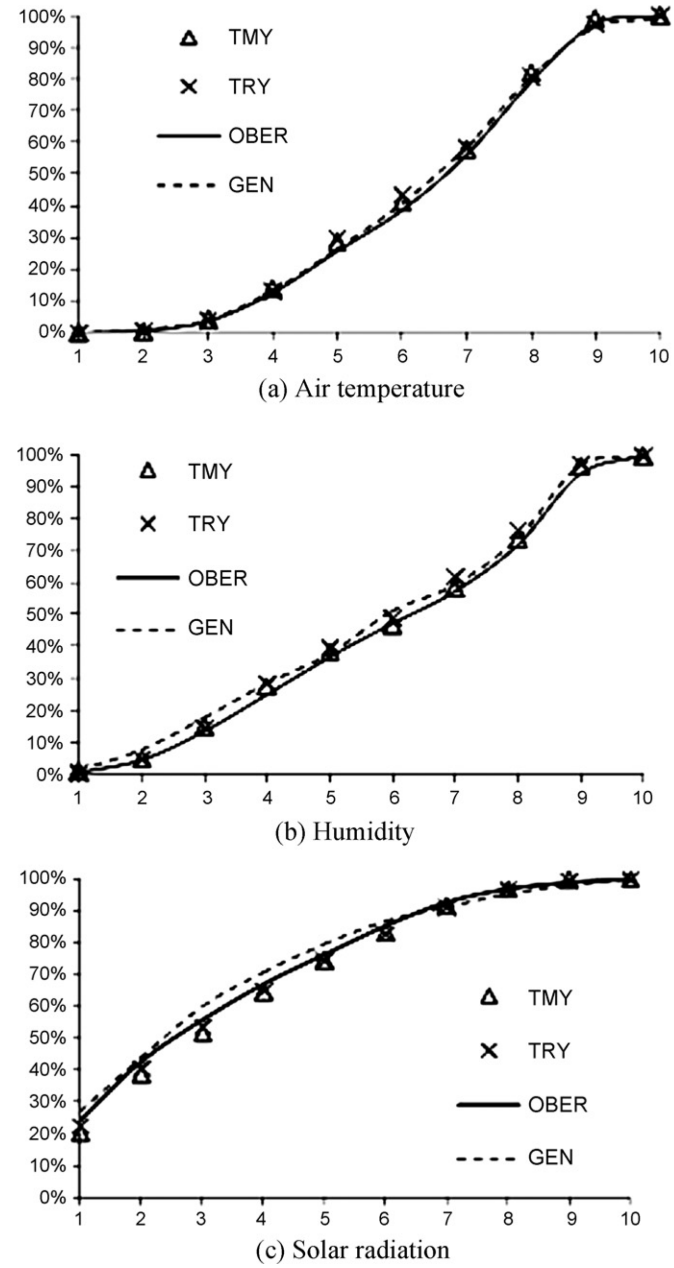

Fig. 3. Probability distributions of solar radiation, humidity and air temperature.

model, has been compared qualitatively and quantitatively with those of the observed values in terms of probability distribution. Fig. 2 shows a set of the simulated weather data. The statistical characteristics of the generated weather data were compared with the long-term average real weather data, the local TMY and TRY. As listed in Table 1, the comparisons include the annual average value, standard deviation, yearly maximum value and minimum value. The results show that the simulated weather data can well preserve the characteristics of the real weather data and the relative errors are all within 5% magnitude. The largest relative error happens in the minimum value of air dry-bulb temperature. Compared with TMY and TRY, the results of the generated weather data are also satisfactory. The relative errors of the generated weather data are within the same magnitude of those of the TMY and TRYs. Actually, the generated weather data performs better than the TMY and TRY. The largest relative error of the TMY was 15%, and 6.94% for the TRY, which are both more larger than that of the generated weather data.

For further assessment, the frequency distributions of the observed weather data, such as the generated time series, the TMY and TRY, were examined. The weather variables were classified into 10 states and the cumulative probability distributions of solar radiation, humidity and air temperature were calculated, as shown in Fig. 3. The figure shows that the cumulative probability distributions of the generated weather data, TMY and TRY at different states were close to the 15 year values. However, the deviation of the cumulative probability distributions of the solar radiation seems to be more obvious than the one of other two weather variables. The cumulative probability distribution curve of the generated solar radiation data is above the long-term one, which means the solar radiation of the generated data is lower than the actual value, but those from the TMY and TRY are lower than the long-term cumulative probability distribution curve.

Besides those statistic characteristics and frequency distribution, the persistent structure of the weather conditions is also an important factor that influences the accuracy of the energy system performance simulations, especially for those renewable energy application systems with energy storage subsystem, such as hot water tank and battery bank. Fig. 4 illustrates the comparison of the cumulative persistence probability between the actual weather data and generated weather data of the TMY and TRY. To simplify the discussion, only comparisons of the extreme weather conditions in Hong Kong (dry-bulb air temperature: 35–32 °C; absolute air humidity: 2.30E-4 ~ 2.09E-4 kg/kg DA; global solar radiation: 0–105 W/m^2 and 940–1050 W/m^2) were presented. As shown in Fig. 4a, the high air dry-bulb tempera-

Table 1
Comparison between real weather data and synthetically generated weather data.

Parameters	Observed					
	Global solar radiation (W/m^2)		Air dry-bulb temperature (°C)		Air absolute humidity (kg/kg DA)	
Average	326.17		23.94		1.44E−02	
Stdev	250.31		5.34		4.91E−03	
Max	998.52		33.70		2.29E−02	
Min	0.00		7.99		2.83E−03	
Parameters	Value (W/m^2)	Relative errors (%)	Value (°C)	Relative errors (%)	Value (kg/kg DA)	Relative errors (%)
Generated						
Average	336.90	3.29	23.42	2.19	1.44E−02	0.01
Stdev	258.78	3.38	5.50	3.02	4.88E−03	0.50
Max	1030.12	3.16	33.59	0.33	2.27E−02	1.22
Min	0.00	0.00	8.32	4.17	2.72E−03	3.74
TMY						
Average	344.88	5.74	23.85	0.38	1.44E−02	0.10
Stdev	246.34	1.59	5.35	0.12	4.88E−03	0.55
Max	972.00	2.66	32.80	2.67	2.34E−02	2.18
Min	0.00	0.00	9.20	15.19	2.96E−03	4.61
TRY89						
Average	337.23	3.39	23.81	0.56	1.42E−02	1.81
Stdev	251.17	0.34	5.38	0.77	4.85E−03	1.28
Max	1015.00	1.65	33.90	0.59	2.31E−02	0.51
Min	0.00	0.00	7.80	2.34	3.02E−03	6.94

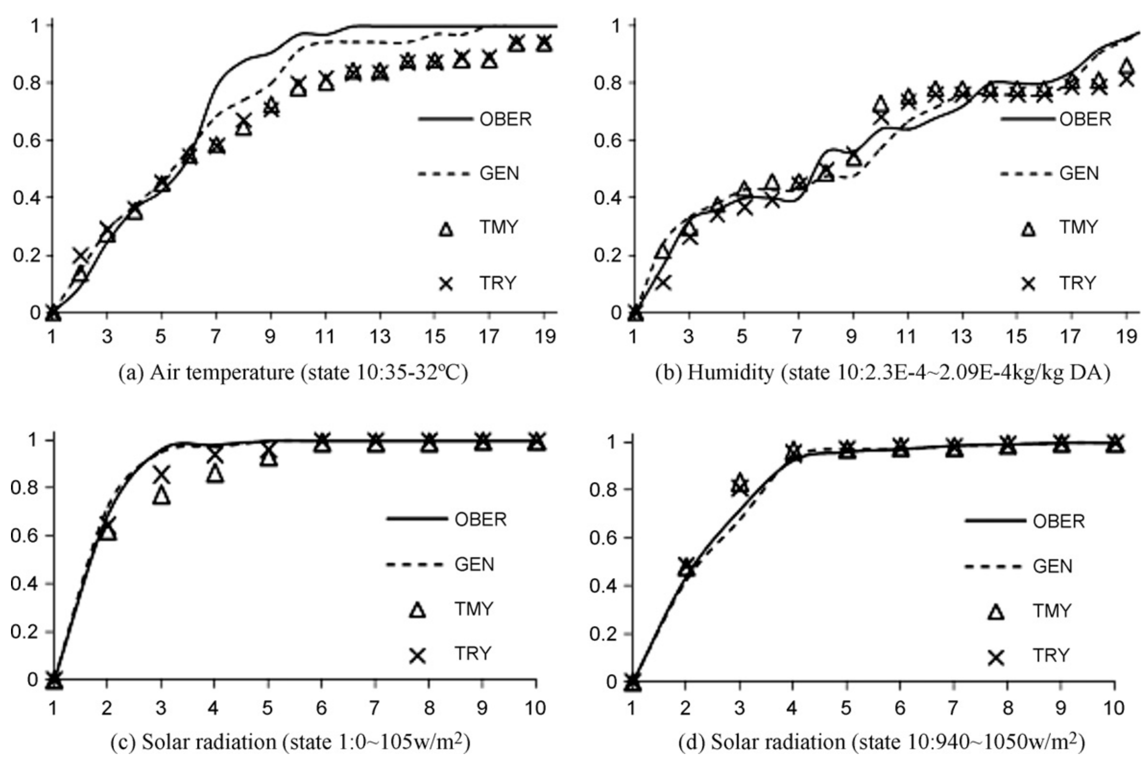

Fig. 4. Comparison of the cumulative persistence probabilities.

ture of the actual weather data will not last for more than 11 h and the accumulative persistence probability reaches 100% after 11 h, whereas those cumulative persistence probabilities of the TMY and TRY can last for more than 20 h. The generated weather data is close to the trends of the long-term actual weather data, whose cumulative persistence probabilities reach 100% in 16 h. Regarding the cumulative persistence probabilities of the high absolute air humidity, as shown in Fig. 4b, there is no obvious difference amongst the generated weather data, TMY and TRY. Because of the nature of solar radiation, it cannot stay in one state for more than 10 h. The cumulative persistence probabilities of the solar radiation at high level and lower level were demonstrated in Fig. 4c and d correspondingly. The dash line showed that the generated solar radiation almost overlaps with the solid line which stands for actual long-term weather data. Generally, the generated weather data kept the actual solar radiation characteristics very well.

6. Conclusions

This article introduced a new method to generate the annual weather data by using the first order multivariate Markov chain model. The weather variables were described in a stochastic way, and multiple categorical sequences were generated by similar sources. Unlike those methods used to select TMYs and TRYs, this new method does not need to pre-assume weighting factors of different weather variables, so the generated annual weather data can be employed as weather data input for evaluating or sizing those energy systems which involve more than one dominant weather variables. To validate this model, 15-years actual hourly weather data of Hong Kong were analyzed for generating the annual hourly data. The generated weather data were compared with the existing developed TMY and TRY data of Hong Kong and long-term actual weather data. The comparison considered several important criteria, including general statistical characteristics, distribution probabilities and persistence probabilities. Although there are some differences between the actual weather data and generated one when the distribution probabilities and persistence probabilities are considered, the generated weather data are still acceptable when the data is compared with the TMY and TRY. It is concluded that the newly developed set of annual weather data in this study has several advantages than the existing TMY and TRY of Hong Kong, i.e., no requiring weighting factors, better statistical characteristics and persistent structure. So the new generated data is more suitable for sizing and evaluating different renewable energy systems, especially mixed renewable ones. Further study is recommended to investigate the accuracy of the generated weather data when they are applied for different types of energy systems.

Acknowledgements

This work was financially supported by The Hong Kong SAR Government through Environmental Conservation Fund (ECF) Project 9/2008 and Innovation & Technology Fund (ITF) Project (GHP/040/08GD).

References

[1] W.L. Wong, K.H. Ngan, Selection of an ‘example weather year’ for Hong Kong, Energy and Buildings 19 (44) (1993) 313–316.
[2] L.S. Chan, T.T. Chow, K.F. Fong, Z. Lin, Generation of a typical meteorological year for Hong Kong, Energy Conversion and Management 47 (2006) 87–96.
[3] H.X. Yang, L. Lu, The development and comparisons of typical meteorological years for building energy simulation and renewable energy applications, ASHRAE Transactions 110 (2) (2004) 424–431.
[4] A. Argiriou, S. Lykoudis, S. Kontoyiannidis, C.A. Balaras, et al., Comparison of methodologies for TMY generation using 20 years data for Athens, Greece, Solar Energy 66 (1) (1999) 33–45.
[5] M. Muselli, P. Poggi, G. Notton, A. Louche, First order Markov chain model for generating synthetic “typical days” series of global irradiation in order to design photovoltaic stand-alone systems, Energy Conversion and Management 42 (6) (2001) 675–687.
[6] A. Maafi, A. Adane, Analysis of the performances of the first-order two-state Markov model using solar radiation properties, Renewable Energy 13 (2) (1998) 175–193 (19).
[7] A. Shamshad, M.A. Bawadi, W.M.A. Wan Hussin, T.A. Majid, et al., First and second order Markov chain models for synthetic generation of wind speed time series, Energy 30 (5) (2005) 693–7085.
[8] W. Ching, E. Fung, M. Ng, A multivariate Markov chain model for categorical data sequences and its applications in demand predictions, IMA Journal of Management Mathematics 13 (2002) 87–199.