STOCHASTIC GENERATION OF ARTIFICIAL WEATHER DATA FOR SUBTROPICAL CLIMATES USING HIGHER-ORDER MULTIVARIATE MARKOV CHAIN MODEL

Andrew Y. A. Oyieke* and Freddie L. Inambao

Discipline of Mechanical Engineering, University of KwaZulu-Natal, Mazisi Kunene Road, Glenwood, Durban 4041, South Africa
*Corresponding Author.

ABSTRACT

Liquid desiccant air conditioning systems provide an efficient and less energy-intensive alternative to conventional vapour compression systems due to their ability to use low-grade energy provided by a hybrid photovoltaic and thermal solar power module. Air conditioning systems are major energy consumers in buildings especially in extreme climatic conditions and are therefore primary targets in so far as energy efficiency is concerned. Building energy performance has traditionally been simulated using typical meteorological year (TMY) and test reference year (TRY) weather tools. In both cases, the value allocation is pegged on the least nonconformity from the long-range data of the past 29 years. The extreme low and high points are successively disregarded which means that the actual prevailing hourly mean settings are not precisely represented. The multivariate Markov chain provides flexibility for use in circumstances where dynamic sequential and categorical weather data for a given region is required. This study presents a simplified higher order multivariate Markov chain analysis founded on a combination of a mixture-transition and a stochastic technique to project the solar radiation, air humidity, ambient temperatures as well as wind speeds and their interrelationships in sub-tropical climates, typically the coastal regions of South Africa. The generic simulation of weather parameters is produced from 20 years of actual weather conditions using a stochastic technique. The series of weather parameters developed are then implemented in the simulation of solar powered air dehumidification and regeneration processes. The outcomes indicate that the model is devoid of constraints and more accurate in the estimation of variable parameters implying that a properly designed solar-powered liquid desiccant air conditioning system is capable of supplying the majority of the latent cooling load.

Keywords: Weather data, higher-order, multivariate Markov chain, liquid desiccant, sub-tropical climate.
1. INTRODUCTION

Due to tremendous advances in the sector of the built environment in recent years, there has been a positive movement towards the use of renewable energy owing to the ever-increasing danger posed to the environment by conventional sources. Thus, the current trend among building specialists, architects and researchers is towards using renewable sources of energy such as solar to alleviate the high cost of conventional energy. One area of such application is the liquid desiccant air conditioning system, especially at dehumidification and regeneration stages.

The climatic characteristics in sub-tropical regions have not made it easy to include renewable energy into buildings. South Africa is one such country, with a sub-tropical climate along the coastal strip adjacent to the Indian Ocean which is where the city of Durban is located. This region experiences moderately hot and highly humid weather so ideally the air needs to be conditioned before delivery into occupied spaces. The air conditioning process is often a very costly and energy-intensive process. The use of solar energy in air conditioning has been demonstrated to curb energy costs by up to 70% [1]. This form of energy is highly dynamic and unpredictable and its application is often dependent on hourly variations.

Typical meteorological year 2 (TMY2) provides hourly solar data sets in terms of radiation and other weather-related data distributed over a period of one year. The use of TMY2 is limited to simulations of solar energy conversion systems and applications in hypothetical building performance assessments in terms of architectural configurations and geographical positioning in various parts of the world. However, they only characterize archetypical other than real circumstances which may in some occasions be deemed extreme and therefore not suitable for designing structures to withstand worst-case scenarios especially in the regions along coastal strips [4, 5].

In an attempt to solve this shortcoming, another sophisticated set of meteorological data was generated by Chan et al. [6] which involved measured data in an hourly interval for 25 years spanning from 1979 to 2003, dubbed typical meteorological months (TMM). The cumulative distribution function (CDF) for the 12 months in a year were weighed against the long term CFD. The CFD technique involves choosing the month with the least deviation as the standard month and applying this in the composition of TMY. However, since various TMYs possess different statistical inferences, the application is restricted to energy simulation in the built environment and not necessarily suitable for inclusion of other forms of energy such as wind and solar.

Various weighting parameters formed the basis of comparison of TMY and extreme weather year (EWY) weather parameters by Yang et al. [7]. The researchers concluded that weighting factors are useful in the generation of TMY data especially where renewable energy systems are concerned. For reliability and least capital costs of such energy systems, a consistent pattern of meteorological factors needs to be considered. An even more comprehensive analysis of different methods of choosing TMY available in the literature versus those from TRY has been presented by [8]. The effects of both TMY and TRY on functional capabilities of solar thermal collectors and photovoltaic (PV) cells with periodic energy retention systems in buildings was considered.
The effect of solar radiation and ambient air conditions cannot be underestimated since solar radiation provides the regeneration heat while the ambient temperature and humidity of air dictate the ability to either expel or absorb more water vapour. Wind speed determines the intensity of solar radiation reaching a surface. Existing annual solar water heating and photovoltaic weather data have less potential in the present case. The situation is compounded by the fact that it may be difficult to distinguish the effects of solar radiation and ambient air conditions on the performance of a solar-powered liquid desiccant dehumidifier and regenerator. Moreover, the weighting factor assignment to generate the corresponding TMY and TRY weather data can be tedious. It is evident that various TRY files may be needed for building simulations case by case and TMY greatly relies on weighting factors which often overestimate or understate the data points. Based on the identified gaps, a new set of reliable and more realistic and dynamic weather data needs to be developed for air conditioning applications, particularly for closed-direct liquid desiccant (LD) dehumidification and regeneration systems.

In this paper, a stochastic methodology is used to develop a higher-order multivariate Markov chain model incorporating a feedback loop with the capability of linking the observed states and the interdependence of parameters to generate yearly weather data for application in simulation of energy systems comprising multiple weather characteristics over a long period of time. Higher-order MMC models find better application in the modelling of sequential data with the capability of incorporating the long-term dependence on a sequentially measured variable such as solar LD dehumidification and regeneration systems. The main aim of this study was to formulate the governing probability matrices of the measured and predicted parameters as higher-order stochastic equations that provide a more effective method of estimation of model parameters.

2. MULTIVARIATE MARKOV CHAIN THEORY

The Markov chain (MC) is a stochastic process theory that can be applied in modelling multivariable processes which occur unpredictably. Based on probability distribution theory, MC takes care of parameters that change form with a change in time, and therefore a time step variation of a past state is dictated by the assigned order. In other words, the rules of stochastic systems probability generally rely on the previous phases but only up to a given level. The multivariate Markov chain (MMC) model variables accept a modest probabilistic interpolation capable of fitting into an iterative algorithm.

A first-order MC was evidently applied in Yang et al. [9], where a generic weather data set "typical days" was generated and used in the design of an independent photovoltaic system. A bi-state MC model fed with solar radiation parameters was developed and used by Maafi and Adane [10]. A comparison between first and second MC was provided by Shamshald et al. [11] in terms of accuracy based on the wind velocity over a long period of time. Ching et al. [12] put forward a first order multivariate MC model for analysis of various sequential signals produced by a common source. Ching and Fung [13] conducted a study on demand forecast by applying an MMC model based on sequential data series and categories. Yutong and Yang [14] developed a first-order multivariate MC model and used it to generate a series of synthetic weather data for Hong-Kong based on 15-year actual weather data for simulation of solar-powered desiccant air conditioning systems. More recently, Wang et al. [15] formulated a new MMC model for the addition of first-hand data series.

A higher-order MMC model was developed and recommended by Raftery [16], however, the solution using this approach gets complicated due to the non-linear optimization problem and a global maximum and local convergence is not guaranteed. Later a mixture transition technique was used by Raftery et al. [17] to estimate and model recurring patterns. Ching et al.
[18] used the stochastic difference equation technique to generalize Raftery\'s models to create adequate circumstance for sequential convergence to a static distribution. Zhu et al. [19] presented a higher-order interactive hidden MMC and highlighted areas of its application. However, there are minimal meteorological applications of higher-order MMC from the available literature even though the Markov chain technique can give efficient and reliable models capable of generating changing meteorological data. The idea of higher-order MMC is introduced by extending the first-order MMC to a higher order so as to include long-term dependence on actual as well as predicted variables with the possibility of including a feedback mechanism for reliability. The majority of the applications point to the fact that higher-order MMC provides superior outcomes than the first order instances especially in the accurate estimation of unpredictable variables such as solar irradiance, ambient temperature and relative humidity. A lean and efficient non-negative matrix factorization and two-level optimization iterative algorithm that predicts unknown parameters based on scarce information is now outlined.

A multivariate MC theory can be developed in the following manner: taking n sequential categories of meteorological variables each with k possible combinations; and assuming the probability distribution sequence i at time \( t = k + 1 \), to rely on sequential probabilities of all states at time \( t = k \), it can be deduced that:

\[
\chi_{k+1}^{(i)} = \sum_{j=1}^{k} Y_{ij} \chi_{k}^{(j)} \quad \text{for } i = 1, 2, \ldots, n \text{ and } k = 1, 2, \ldots, n
\]  

(1)

For \( i = 1, 2, 3, \ldots, n \) and \( k = 1, 2, 3, \ldots, n \); \( \gamma_{ij} \geq 0 \), \( 1 \leq i, j \leq n \); \( \sum_{j=1}^{k} Y_{ij} \) for \( i = 1, 2, 3, \ldots, n \); \( \chi_{o}^{(j)} \) is the probability distribution at the origin of sequence i. It, therefore, follows that the probability distribution of sequence i is greatly dictated by the weighted mean of \( p^{(i,j)} \chi_{k}^{(j)} \). A matrix can then be generated by considering a single-step probability combination \( p^{(i,j)} \) at state t and sequence j and k at time \( t = k + 1 \), with \( \chi^{(j)} \) probability distribution of sequence j at time k. This relationship yields the matrix in equation 2:

\[
\begin{bmatrix}
\chi_{k+1}^{(1)} \\
\chi_{k+1}^{(2)} \\
\chi_{k+1}^{(3)} \\
\vdots \\
\chi_{k+1}^{(n)}
\end{bmatrix} =
\begin{bmatrix}
Y_{11}p^{(11)} & Y_{12}p^{(12)} & \cdots & Y_{1n}p^{(1n)} \\
Y_{21}p^{(21)} & Y_{22}p^{(22)} & \cdots & Y_{2n}p^{(2n)} \\
Y_{31}p^{(31)} & Y_{32}p^{(32)} & \cdots & Y_{3n}p^{(3n)} \\
\vdots & \vdots & \ddots & \vdots \\
Y_{n1}p^{(n1)} & Y_{n2}p^{(n2)} & \cdots & Y_{nn}p^{(nn)}
\end{bmatrix}
\chi_{k}^{(1)}
\]

(2)

The above matrix in equation 2 can be solved by a simple linear algorithm to obtain the underlying parameters after performing a normalization procedure. In order to estimate \( p^{(i,j)} \) and \( \gamma_{ij} \), we consider a transition probability matrix encompassing the state at \( i \) and \( j \) sequences respectively and generate a state matrix for data frequency. Hence, it is vital that an \( m \times m \) transitional matrix for MMC model be estimated prior to solution.

Therefore, if we take the frequency \( f^{(i,j)} \) in phases \( l_i \) and \( l_j \) within sequences \( \chi^{(i)} \) and \( \chi^{(j)} \) we can develop a transition frequency matrix in equation 3.

\[
f^{(i,j)} =
\begin{bmatrix}
f_{11} & f_{12} & \cdots & f_{1n} \\
f_{21} & f_{22} & \cdots & f_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
f_{n1} & f_{2n} & \cdots & f_{nn}
\end{bmatrix}
\]

(3)
The probability matrix $\hat{P}^{(i,j)}$ can then be estimated from $f^{(i,j)}$ as:

$$\hat{P}^{(i,j)} = \begin{bmatrix} p_{11}^{(i,j)} & p_{21}^{(i,j)} & p_{31}^{(i,j)} & \cdots & p_{1n}^{(i,j)} \\
p_{12}^{(i,j)} & p_{22}^{(i,j)} & p_{32}^{(i,j)} & \cdots & p_{2n}^{(i,j)} \\
p_{13}^{(i,j)} & p_{23}^{(i,j)} & p_{33}^{(i,j)} & \cdots & p_{3n}^{(i,j)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
p_{1n}^{(i,j)} & p_{2n}^{(i,j)} & p_{3n}^{(i,j)} & \cdots & p_{nn}^{(i,j)} \end{bmatrix}$$

(4)

From which the exact value of $p_{l_{ij}}^{(i,j)}$ can be evaluated for as long as the following condition is met:

$$p_{l_{ij}}^{(i,j)} = \begin{cases} \frac{n}{\sum_{l_{ij}=0} f_{l_{ij}}^{(i,j)}} & \text{when } \sum_{l_{ij}=0} f_{l_{ij}}^{(i,j)} \neq 0 \\ 0 & \text{other cases} \end{cases}$$

(5)

We then estimate the factors $\gamma_{ij}$ by assuming an MMC model with a static probability vector $\chi^{*}$ which can be evaluated by getting the occurrence probability of a state in a sequence denoted by a stable probability vector $\chi^{*}(\chi^{*}(1), \chi^{*}(2), \chi^{*}(3) \ldots \chi^{*}(n))$ which relates to matrices $\gamma_{ij}$ and $p_{ij}$ in the following configuration matrix:

$$\hat{\chi} \approx \hat{\chi} = \begin{bmatrix} y_{11}p_{(11)}^{(i)} & y_{12}p_{(12)}^{(i)} & y_{13}p_{(13)}^{(i)} & \cdots & y_{1n}p_{(1n)}^{(i)} \\
y_{21}p_{(21)}^{(i)} & y_{22}p_{(22)}^{(i)} & y_{23}p_{(23)}^{(i)} & \cdots & y_{2n}p_{(2n)}^{(i)} \\
y_{31}p_{(31)}^{(i)} & y_{32}p_{(32)}^{(i)} & y_{33}p_{(33)}^{(i)} & \cdots & y_{3n}p_{(3n)}^{(i)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
y_{n1}p_{(n1)}^{(i)} & y_{n2}p_{(n2)}^{(i)} & y_{n3}p_{(n3)}^{(i)} & \cdots & y_{nn}p_{(nn)}^{(i)} \end{bmatrix}$$

(6)

By using Ching et al.'s [13] optimization principle rule, a linear programming problem for each instance $l$ on the parameter $\gamma = \gamma_{ij}$ which satisfies the condition set out if equation 5, can be formulated as shown in equation 7:

$$\beta = | P^{(l,i)} \hat{\chi}^{(1)} | P^{(l,i)} \hat{\chi}^{(1)} | P^{(l,i)} \hat{\chi}^{(1)} | \ldots | P^{(l,i)} \hat{\chi}^{(1)} |$$

(7)

A combination of equations 6 and 7 can then be implemented in the objective function in equation 8 to give the outlined parameters:

$$\begin{align*}
\text{Min}_{\gamma} & \quad \text{Min}_{\tilde{l}} \left\| \sum_{i,j=1}^{m} \gamma_{ij} \tilde{p}^{(j)} \hat{\chi}^{(i,j)} - \hat{\chi}^{(i)} \right\|_l \\
\text{subject to} & \quad \sum_{i=1}^{m} \gamma_{ij} = 1 \text{ and } \gamma_{ij} \geq 0, \forall j
\end{align*}$$

(8)

However, the solution using this approach gets complicated due to the non-linear optimization problem and a global maximum and local convergence is not guaranteed. Therefore, a higher-order MMM can be introduced by considering distinct time-dependent parameter sets given that $\chi^{*}$ is the probability matrix of MMC i.e. $(\chi^{*} \in \mathbb{R}^{km})$ and $\beta$ gives the probability matrix of the observed parameter at time $t$ i.e. $(\beta \in \mathbb{R}^{lkkm})$. It follows that in the higher-order MMC, the observed and the predicted states interactively affect each other according to linear higher order stochastic difference equations:
Stochastic Generation of Artificial Weather Data for Subtropical Climates Using Higher-Order Multivariate Markov Chain Model

\[
\begin{align*}
\chi_t &= \sum_{j=1}^{a} \gamma_j p^{(j)} \beta_{(t-j)} \\
\beta_t &= \sum_{i=1}^{b} \mu_i M_i \chi_{(t-i+1)}
\end{align*}
\]

subject to \( \sum_{j=1}^{a} \gamma_j = \sum_{i=1}^{b} \mu_i = 1 \) and \( \gamma_j \geq 0, \mu_i \leq 1 \) \hspace{1cm} (9)

Where \( a \) and \( b \) are the orders of predicted and measured phases respectively and the matrices \( P^{(j)} \) and \( M_{i(j)} \) are the jth step change-over probability matrices from the observed to predicted states and vice versa. If \( a = b = 1 \), then the Yutong and Ching models are generated [14, 20]. However, if \( \chi' \) is substituted into the equation for \( \beta_t \), a higher-order MMC model \( \beta_t \) is obtained as follows:

\[
\beta_t = \sum_{j=1}^{a} \sum_{i=1}^{b} \mu_i \gamma_j M_i p^{(j)} \beta_{(t-i-j+1)}
\]

Equation 10 can be solved by applying the non-negative matrix factorization method from [20]. In another scenario, second order homogeneous MMC model with value \( a = b = 2 \) can be formulated as:

\[
\begin{align*}
\chi_t &= \gamma_j p^{(j)} \beta_{(t-1)} + (1-Q) \beta_{(t-2)} \\
\beta_t &= \mu_i M_i \chi_t + (1+\mu) N \chi_{t-1} \hspace{1cm} \text{where} \hspace{1cm} \gamma_j \geq 0, \hspace{1cm} \mu_i \leq 1
\end{align*}
\]

Substituting the values of \( \chi' \) in equation 12, we get:

\[
\beta_t = \gamma_j \mu_i M_i p^{(j)} \beta_{(t-1)} + [(1-\gamma)\mu MQ + \gamma(1-\mu) NP] \beta_{(t-2)} + (1-\gamma)(1-\mu) N Q \beta_{(t-3)}
\]

Therefore, a matrix \( H_t \) can be formulated as follows:

\[
H_t = \begin{bmatrix}
\beta_{(t-1)} \\
\beta_{(t-2)} \\
\beta_{(t-3)} \\
\vdots \\
\beta_{(t-n)}
\end{bmatrix} = \begin{bmatrix}
C_1 & C_2 & C_3 & \ldots & C_n \\
1 & 0 & 0 & \ldots & 0 \\
0 & 1 & 0 & \ldots & 0 \\
0 & 0 & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1
\end{bmatrix} \begin{bmatrix}
\beta_{(t-1)} \\
\beta_{(t-2)} \\
\beta_{(t-3)} \\
\beta_{(t-4)} \\
\vdots \\
\beta_{(t-n)}
\end{bmatrix}
\]

Equation 13

\[
\begin{align*}
C_1 &= \gamma \mu MP \\
C_2 &= (1-\gamma) \mu MQ + \gamma(1-\mu) NP \\
C_3 &= (1-\gamma)(1-\mu) NP \\
C_4 &= (1-\gamma)(1-\mu) (\gamma-\mu) QP
\end{align*}
\]

For proper estimation of parameters \( \gamma, \mu, M, N, P, Q \) and \( R \), then the matrix \( C \) must be estimated first by minimizing the normalization factor as follows:

\[
\min_C \left\{ \sum_{t=4}^{T} \| H_t - C H_{(t-1)} \| \right\}
\]

Based on the values of \( \gamma \) and \( \mu \), equation 14 represents the change-over matrix since the column sums of \( C_1, C_2, C_3 \) and \( C_4 \) are equal. The optimal estimates of the unknown parameter
can be achieved by a two-level optimization algorithm based on non-negative matrix factorization [20, 21, 22].

2.1. The Algorithm

1. Initialize W, N and h = 1.

2. Using the Lin sub-problem algorithm [21], solve for $P^{h}, Q^{h}$ and $R^{h}$ by minimizing $||y\mu W^{(h-1)} P^{(h)} - C_{1} ||^2_F, ||(1 - y)(1 - \mu)N^{(h-1)} Q^{(h)} - C_{3} ||^2_F$ and $||(1 - y)(1 - \mu)(y - \mu)Qh - R(h) - C_{4} ||^2_F$.

3. By minimizing $||C_{2} - (1 - y)\mu W(h)Q(h) - y(1 - \mu)N - 1)P(h)||^2_F$, solve W(h) and N (h) subject to $0 \leq W(h), N(h) \leq 1$ and the total sums of $W^{(h)}$ and $N^{(h)}$ being 1.

4. If $||W^{(h)} - W^{(h-1)}||^2_F + ||N^{(h)} - B^{(h-1)}||^2_F <$ the tolerance, then STOP otherwise $h = h+1$ and return to step 2.

5. Repeat steps 1 to 4 until the function in step 4 is greater than the tolerance.

Initial guesses should be chosen randomly so as to avoid attainment of the local minimum. When the normalization of P and Q values is done, the column sums of the probability matrix will be equal to 1 and hence only the parameters minimizing the function $\|\mu WP - C1\|_F^2 + \|C_{2} - (1 - y)\mu WQ - y(1 - \mu)NP\|_F^2 + \|(1 - y)(1 - \mu)NQ - C3\|_F^2 + \|(1 - y)(1 - \mu)(y - \mu)QR - C4\|_F^2$ are chosen.

3. APPLICATION OF THE MMC MODEL TO GENERIC WEATHER DATA GENERATION

The weather data can be decomposed into random and deterministic segments which are autocorrelated daily on an hourly basis for 20 years. A clearness index as defined in [14] and [3] is based on a time series which conveniently estimates solar radiation, temperature, humidity and wind speed which eliminates interdependencies among the weather variables. It then requires testing before any estimation procedure commences.

Before implementation of the formulated MMC model in the estimation of global solar radiation as well as humidity, wind speed and ambient temperatures, consideration must be given to that past studies have shown that MMC is best applicable to studying dependent variables as reported in [14]. Therefore, we then determined through tests whether the parameters in consideration for this study fell in this category by considering an asymptotically spread parameter $\sigma$ characterized by $\chi^2$ having $(n-1)^2$ degrees of freedom, $\{k\}$ and $n$ number of phases and transitions respectively, and a borderline probability for the kth column of changeover probability matrix. In the case of an independent relationship, then $\sigma$ must statistically satisfy the relationship defined as:

$$\sigma = 2 \sum_{j=i}^{k} n_{ij} \ln \left( \frac{\theta^{(ij)}}{\theta^{(ij)}} \right)$$  \hspace{1cm} (16)

A dependence test between variables showed that at 5 % level and 93 degrees of freedom, $\sigma$ was large enough above $\chi^2$ with a corresponding value of 115. This occurrence meant that transitional phases of hourly weather parameters were dependent and thus satisfied the MMC conditions.

A time phase change-over probability is independent if the MMC is stationary. To check the dependability, a complete cycle of occurrence was disintegrated into $k$ mini intervals whose transitional probability matrices were then calculated and compared. If the comparison returned the same results then it could be concluded that the MMC was stationary. For this test, we used...
the statistic $\lambda$, which combines the $n_{ij}(s)$ transitions with $p^{(ij)}(s)$ probability. The solution of equation 17 is expected to give the value of $\lambda$ with $\chi^2$ exhibiting $(k-1)n(n-1)$ degrees of freedom in order to satisfy the condition of a stationary MMC in the following manner:

$$\lambda = 2 \sum_{i,j} n_{ij} \ln \left( \frac{p_{ij}}{p^{(ij)}} \right)$$

(17)

Generally, during this test, the solar radiation, humidity wind speed ambient temperature gave lesser values than $\chi^2$ of 17 500 at 5% confidence level and 15 750 degrees of freedom. This evidently implies that the MMC is stationary for the mentioned weather parameters.

A pattern of weather data was then generated by considering an initial random phase j, then applying MMC to arbitrary values of 0 to 1 produced in a sequential generator. These values formed the basis of comparison with the components of the $j^{th}$ row in the change-over probability matrix. The ensuing states adopted were those of superior value to the total probability of the preceding phase and lower than the probability of the phases.

Taking the minimum and maximum boundary conditions of wind speeds as $V_0$ and $V_1$, and $Z_i$ random numbers falling within range 0-1, a relationship was adopted from [19] and formulated for wind speed, which was replicated hourly for all the t hours.

$$V_w = V_0 + Z_i(V_1 - V_0)$$

(18)

For consecutive time intervals i.e. $t_1$, $t_2$, .... and $t_n$ falling within the phase $i$, the former state was selected. On the other hand, the duration of phase $j$ was determined by selecting a random $Z_j$ whose value was less than the cumulative probabilities $p^j(t - 1; j)$, $p^j(t - 2; j)$ $p^j(t - n; j)$ provided that $(t - 1)$ to $(t + L - 1)$ fell within phase $j$.

The number of change-over probabilities to be predicted rises in an exponential pattern as the order of the model. This results in computational complexities which can only be overcome by incorporating more parameters. Using the data available from measurements at Mangosuthu University of Technology weather station, the hourly variation of solar radiation, wind speed, relative humidity and ambient air temperatures were used to generate a sequence of daily occurrences and subsequent changes. For the formation of data sequences for each parameter, the daily upward and downward variations were identified and marked within intervals to obtain two sets of finite discrete data sequences within the time range of $t = 8760$ hours as follows:

$$R_1 = [X_0X_2X_3X_1X_4X_5X_1 \ldots X_0X_1], \quad R_2 = [X_0X_1X_2X_3X_4X_5X_1 \ldots X_0X_1]$$

$$R_3 = [X_0X_2X_3X_1X_4X_5X_1 \ldots X_0X_1], \quad R_4 = [X_0X_2X_3X_4X_5X_1 \ldots X_0X_1]$$

Where the ranges $X_1 = 0 < X_4 < 3\gamma$ $X_2 = R_4< 3\gamma$, $X_3 = -3\gamma$ and $X_4 = -3\gamma \leq R_4 \leq 0$ represent normal upward change, maximum upward change, minimum decrease and normal decrease respectively in values of each parameter. $R_1$, $R_2$, $R_3$ and $R_4$ are the data sequences for the variations in solar radiation, ambient air temperature, wind speed and relative humidity respectively. From the above combinations, the respective changeover matrices for each parameter can be determined as follows:

$$W_{R_1} = \begin{bmatrix} 1773.42 & 1060.94 & 113.07 & 1050.7 \\ 2107.84 & 1463.13 & 17.50 & 1655.57 \\ 894.85 & 535.12 & 373.07 & 990.78 \\ 670.50 & 495.20 & 925.54 & 960.96 \end{bmatrix}.$$  

$$W_{R_2} = \begin{bmatrix} 23.75 & 28.85 & 25.95 & 27.75 \\ 30.85 & 26.15 & 27.75 & 23.25 \\ 28.75 & 18.40 & 17.54 & 26.15 \\ 27.55 & 25.45 & 26.15 & 19.75 \end{bmatrix}.$$
Andrew Y. A. Oyieke and Freddie L. Inambao

\[ W_{R3} = \begin{bmatrix} 7.95 & 4.5 & 9.9 & 4.65 \\ 7.35 & 1.6 & 6.75 & 8.5 \\ 4.0 & 3.3 & 2.65 & 7.1 \\ 3.4 & 6.95 & 7.1 & 1.6 \end{bmatrix}, \]

\[ W_{R4} = \begin{bmatrix} 77.5 & 71.5 & 66.5 & 89.5 \\ 91.5 & 88.5 & 86.5 & 79.5 \\ 56.0 & 58.5 & 94.5 & 82.5 \\ 58.5 & 86.0 & 58.0 & 61.0 \end{bmatrix} \]

Where \( W_{R1}, W_{R2}, W_{R3} \) and \( W_{R4} \) are the changeover frequency occurrences between phases of \( R_1, R_2, R_3 \) and \( R_4 \) respectively. The next procedure is to normalize the change-over occurrence matrices to determine the probability matrices as follows;

\[ P_{R1} = \begin{bmatrix} 0.0138 & 0.0138 & 0.0124 & 0.0117 \\ 0.0119 & 0.0121 & 0.0131 & 0.0115 \\ 0.0133 & 0.0098 & 0.0109 & 0.0114 \\ 0.0099 & 0.0107 & 0.0114 & 0.0098 \end{bmatrix}, \]

\[ P_{R2} = \begin{bmatrix} 0.0188 & 0.0154 & 0.0177 & 0.0155 \\ 0.0144 & 0.0157 & 0.0135 & 0.0114 \\ 0.0133 & 0.0142 & 0.0149 & 0.0137 \\ 0.0111 & 0.0125 & 0.0128 & 0.0118 \end{bmatrix}, \]

\[ \tilde{P}_{R3} = \begin{bmatrix} 0.0160 & 0.0114 & 0.0173 & 0.0151 \\ 0.0137 & 0.0129 & 0.0114 & 0.0148 \\ 0.0222 & 0.0130 & 0.0111 & 0.0104 \\ 0.0106 & 0.0133 & 0.0137 & 0.0149 \end{bmatrix}, \]

\[ P_{R4} = \begin{bmatrix} 0.0126 & 0.0101 & 0.0105 & 0.0198 \\ 0.0147 & 0.0147 & 0.0163 & 0.0170 \\ 0.0185 & 0.0167 & 0.0184 & 0.0138 \\ 0.0158 & 0.0146 & 0.0164 & 0.0145 \end{bmatrix}. \]

4. RESULTS AND DISCUSSION

After computation and solution of change-over probability matrices, the properties exhibited similar characteristics as those of the actual weather data based on a 20-year average. The actual weather data used as a basis of comparison in this study was obtained from one of the South African Universities Radiometric Networks (SAURAN) at Mangosuthu University of Technology STARLab station (STA), situated at Umlazi, Durban, South Africa on latitude, longitude and elevation of -29.97027° E, 30.91491° S and 95 m respectively. The annual hourly distribution of air ambient temperature is shown in Figure 1. The maximum and maximum values of temperature during the year are 5.7 °C and 32.2 °C. The former was experienced in the month of June while the latter occurred in February, these months corresponding to summer and winter seasons in South Africa respectively.

The relative humidity annual hourly distribution is shown in Figure 2. The predicted minimum and maximum values were 44 % and 100 % receptively for the Durban region which lies on the coastal line. The minimum RH corresponds to the winter month of July while the maximum range was recorded in the summer month of January. The annual hourly variation of wind speed was plotted as shown in Figure 3, from which the minimum and maximum values of 0.1 m/s and 13.55 m/s were recorded. The values corresponded with the winter and summer months of July and December respectively.

The total solar radiation distribution over a one-year period is shown in Figure 4. The lowest value was 0 W/m² during the non-sunshine hours while the highest value was 3391.57 W/m².
experienced between the months of May and August. All the maximum and minimum values of these parameters appeared to follow the same trend as the actual measured weather data.

Figure 1 Predicted annual ambient air temperature hourly distribution per year

Figure 2 Predicted annual relative humidity hourly distribution

Figure 3. Predicted annual wind speed hourly distribution
Figure 4 Predicted annual total solar radiation hourly distribution

The data was subsequently evaluated qualitatively in terms of frequency of occurrence of different phases of the parameters as measured and predicted. The MMC model outputs were then contrasted with the probability ranges for the actual measured values. Figure 5 presents the comparisons of MMC predicted and actual weather data statistics on long term range. The individual probability comparisons are presented in Figures 5a, 5b, 5c and 5d for solar radiation, ambient temperature, wind speed and relative humidity respectively. From the analysis, the deviations of the predicted probabilities from the actual were 0.53 %, 0.5 %, 0.2 % and 0.16 % respectively, thus depicting great accuracy of the model.

Figure 5 Comparison of measured and MMC predicted probability distributions for (a) Solar radiation (b) Ambient temperature (c) Wind speed and (d) Relative humidity.
Stochastic Generation of Artificial Weather Data for Subtropical Climates Using Higher-Order Multivariate Markov Chain Model

A more detailed comparison of annual data in terms of annual range, average deviation as well as maximum and minimum values is tabulated in Table 1. In general, the generated data displayed similar properties to the actual data in terms of range and magnitude with below 5% marginal error.

Table 1 Comparison of annual measured and MMC predicted weather data for Durban South Africa.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Predicted</th>
<th>Measured</th>
<th>Variation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solar rad. (W/m²)</td>
<td>AV 419.5</td>
<td>SD 248</td>
<td>Max 1030</td>
</tr>
<tr>
<td>Air temp. (°C)</td>
<td>AV 20.51</td>
<td>SD 5.78</td>
<td>Max 32.7</td>
</tr>
<tr>
<td>Wind speed(m/s)</td>
<td>AV 4.121</td>
<td>SD 2.62</td>
<td>Max 0.1</td>
</tr>
<tr>
<td>Air humidity (kg/kg dry)</td>
<td>AV 78.59</td>
<td>SD 10.14</td>
<td>Max 44</td>
</tr>
</tbody>
</table>

Since the extreme weather patterns under study displayed very compact curves over the annual span, it was prudent to use a shorter span to demonstrate the relationships between the measured and predicted data. For this reason, data from the summer month of February 2018 was extracted for comparison purposes. Figure 6 shows the curves predicted and the measured solar radiation distribution. There was a precise fit between the curves laid side by side with an average variation of 0.14%. On the air relative humidity, a side by side comparison of the predicted and measured data profiles are presented in Figure 7. The mean variation between the two sets of curves was -0.22%.

Figure 8 shows the variation of ambient air temperature as well as wind speeds for the said month. The MMC model precisely predicted the ambient temperature to an accuracy of -0.65% while the wind speed was estimated with a variation of -3.4%. In all these comparisons the values fall far below the 5% limit and similar proportionate variations were cascaded through the year. These low values demonstrated that the MMC predicted values of the weather parameters generally matched the corresponding measured data with great accuracy.

Figure 6 Comparison between predicted and actual solar radiation hourly distribution for the month of July
5. CONCLUSION

In this study, the theoretical modelling of MMC has been concisely presented and implemented in order to make the structure more understandable. A stochastic methodology was used to develop a higher-order multivariate Markov chain model incorporating a feedback loop with the capability of linking the observed states and the interdependence of parameters. The higher-order MMC model was then used to generate yearly weather data for application in simulation of energy systems comprising multiple weather characteristics over a long period of time.

The first order MMC was extended to a higher-order to include the long-range dependent sequences. Consequently, the number of categorical consequences were increased to obtain a distinctive and realistic estimation of weather patterns and how the parameters depend on each other. The number of phases in an occurrence was also increased to obtain as many data points as possible to the tune of the second interval to enhance accuracy.
The governing probability matrices of the measured and predicted parameters were developed as higher-order stochastic equations that provided a more effective method of estimation of model parameters. Since the MMC model is computationally simple and can reliably be implemented, the number of parameters to be estimated were equally increased with ease compared to the number earlier considered in [14]. An algorithm founded on non-negative matrix factorization and multilevel optimization was developed to estimate the unknown parameters depending on little information.

The hourly distribution of solar radiation, air temperature, air humidity and wind speeds were generated and compared to the actual measured data from a weather station. The predicted annual averages for solar radiation, air temperature, wind speed and air humidity were 419.5 W/m², 20.51 °C, 4.121 m/s and 78.59 kg/kg dry respectively. Subsequently, the respective mean measured values were 425.8 W/m², 21.3 °C, 4.324 m/s and 80.15 kg/kg dry. The percentage variations between the measured and predicted values were 1.48 %, 3.71 %, 2.67 % and 1.95 % respectively. All these values of percentage deviations were below 5 % implying that the higher-order MMC model estimated the corresponding parameters with great accuracy.

Higher-order MMC model in comparison to first-order MMC give a more precise prediction of categorical sequences due to computational accuracy and efficient application. The increased number of parameters and constants in this study helped to create a new MMC model of a higher order. The outcomes imply that higher-order MMC replicates the actual weather data with great accuracy. Stochastically generated artificial weather parameters can reliably be implemented in the study of the solar energized system of liquid desiccant dehumidification and regeneration. By increasing the number of sequential categories of data, it is possible to gain more explicit results and precise prediction of the sequences influencing each other. The higher-order MMC model is more truthful when the number of phases in a sequence is increased because more data is generated.

REFERENCES


