Available online at www.sciencedirect.com
ScienceDirect
Energy Reports 8 (2022) 76–81
www.elsevier.com/locate/egyr

The 8th International Conference on Energy and Environment Research ICEER 2021, 13–17 September

Electricity demand prediction for sustainable development in Cambodia using recurrent neural networks with ERA5 reanalysis climate variables

Karodine Chreng[a], Han Soo Lee[b,*], Soklin Tuy[a]

[a] Department of Development Technology, Graduate School for International Development and Cooperation (IDEC), Hiroshima University, 1-5-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8529, Japan
[b] Transdisciplinary Science and Engineering Program, Graduate School of Advanced Science and Engineering, Hiroshima University, 1-5-1 Kagamiyama, Higashi-Hiroshima 739-8529, Hiroshima, Japan

Received 18 December 2021; accepted 7 January 2022
Available online 2 February 2022

Abstract

Sustainable energy development plays a prominent role in energy planning to maintain natural resources and mitigate the usage of fossil fuels. The atmospheric factor is one of the main influencing factors that changed the electricity consumption's behavior due to global warming. In this study, the recurrent neural network (RNN) models were developed to examine the effects of 66 climate variables, collected from the European Center for Medium-Range Weather Forecast (ECMWF) ERA5 reanalysis, on power demand in Cambodia. The statistically significant climate variables were filtered by considering the cross-correlation between power demand and each climate variable. Moreover, the wide range of feedback delays was computed from the power demand dataset and was defined using the 95% confidence intervals. The comparison between a nonlinear autoregressive neural network with exogenous inputs (NARX) using historical power demand with the correlated climate variables and a nonlinear autoregressive neural network (NAR) using only historical power demand dataset was made. The various benchmarked models were evaluated and compared for their performances using statistical indices such as normalized root-mean-square error (NMSE) and coefficient of determination (R^2). The results showed the NARX model could perform better than the NAR model for predicting electricity demand time-series.

Peer-review under responsibility of the scientific committee of the 8th International Conference on Energy and Environment Research, ICEER, 2021.

Keywords: Electricity demand; NARX; NAR; ERA5 reanalysis; Feedback delays; Climate variables; Cambodia

1. Introduction

Energy planning on the demand side is the crucial task for addressing in terms of sustainable development against the global warming caused by fossil fuel generations. Demand management provides particular aspects of the future

* Corresponding author.
E-mail addresses: karodinec@gmail.com (K. Chreng), leehs@hiroshima-u.ac.jp (H.S. Lee), soklin90@gmail.com (S. Tuy).

https://doi.org/10.1016/j.egyr.2022.01.025

Peer-review under responsibility of the scientific committee of the 8th International Conference on Energy and Environment Research, ICEER, 2021.

perspective for decision-makers to manage the energy mix's green and clean energy generations. When the demand and supply are imbalanced, the fossil fuel generations would probably be prioritized for solving the emergency situation; however, those generations could definitely produce the CO2 emission with the unaffordable price in the global power market.

Goal number seven among 17 sustainable development goals (SDGs) describes the guaranteeing accessibility of affordable, reliable, sustainable, and modern energy for all [1]. Therefore, well-prepared electricity demand planning is entirely crucial for the reliability and sustainability of the energy supply. At the same time, numerous factors could affect electricity consumption, such as climate factors (temperature, dewpoint, and so on) and socio-economic factors (population, gross national product, imports and exports, occupants, family composition, household income, and so on) [2,3]. Therefore, the objectives of this study are described as follows:

- Development of the optimized recurrent neural network (RNN) models for future prediction of electricity demand.
- Identification of the correlated climate variables for using in the NARX model to improve prediction outcomes.
- Comparison of prediction techniques using (i) NAR model using historical data, and (ii) NARX model with historical electricity demand and climate variables, in terms of their performance.

2. Materials and methods

2.1. Study area and data

The locations of substations and data used in this study were presented in Fig. 1. The main criteria for site selection were based on the available data from 2013 to 2018 for power demand and the ECMWF-ERA5 Reanalysis climate variables in Phnom Penh city (GS1, GS2, GS3, and WPP site), as shown in Fig. 1. In this study, electricity demand data is collected for six years from January 01, 2013, to December 31, 2018, with 52,584 data points. Hourly electricity demand data in megawatts [MW] from each substation were converted to daily intervals by taking power demand value at 10:00 (peaking time in Cambodia) with 2191 data points.

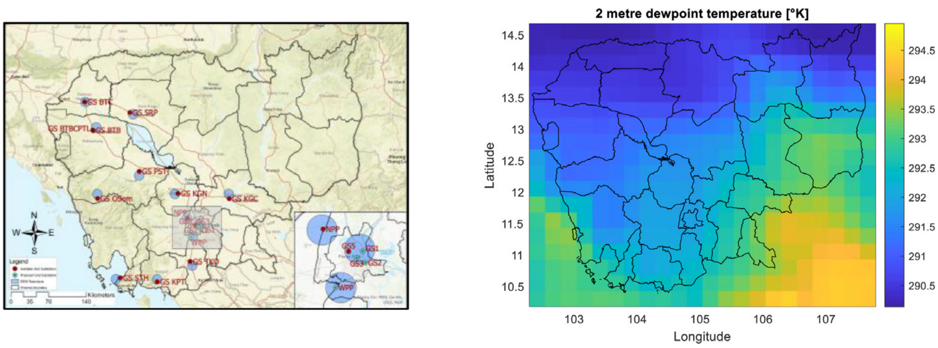

Fig. 1. Map of Cambodia with the locations of grid substation (GSs) and the coverage area of the nearest ECMWF-ERA5 Reanalysis grid points in blue-shaded circles (left) and the color shades indicates the 2-meter dewpoint temperature climate data of ERA5 Reanalysis at 00:00 on January 1, 2013 (right).

2.2. Artificial neural network - NAR and NARX neural networks

A nonlinear autoregressive model (NAR) and a nonlinear autoregressive model with exogenous inputs (NARX) are among the groups of recurrent neural networks (RNNs). Also, NARX and NAR are nonlinear dynamic models that fit the time series prediction [4]. NAR neural network model is used the historical dataset only as the input data to predict itself for future values [5]. The output function of the NAR model can be mathematically given in Eq. (1):

$$y(t) = f(y(t-1), y(t-2), \ldots, y(t-d_y)) \tag{1}$$

where y is the historical power demand dataset over (t) time, d_y is the feedback delay (FD) or lagged feedback output, and f is the activation function in the neural network model. The feedback delay impacts the closed-loop simulated production as well as multistep predictive outcome, and it was defined by determining the auto-correlation function of the input dataset (power demand). Consequently, sufficient feedback delay is crucially important, which permits the training process could visibly understand the characteristic of the historical data.

In this study, the critical configurations for the NAR model are described as follows:

- The FDs were determined by autocorrelation of the training dataset, and all significant values were used as feedback delays in the model (FD = [1:397,475:527,660:1271,1304:1332]), as shown in Fig. 3.
- For defining the number of hidden layer neurons, the value is set differently, based on case studies with ten neurons [6], 15 neurons [7], ranging between 3 to 10 [8], and ranging between 1 to 20 [9]. Therefore, the trial-and-error procedure is applied by investigating the number of hidden layer neurons ranging from 1 to 20.
- Transfer function: Since feedback delays were used many values, the performance of training time is technically the constraint for this model. Kumar and Murugan [10] suggested scaled conjugate gradient backpropagation (trainscg) which is appropriate to be used in this model for reducing training memory and time.
- Activation function of hidden and output layers: Sarkar et al. [11] found that tangent-sigmoidal (tansig) transfer function (Eq. (2)) could provide better results based on error evaluation in the training process and was accordingly considered as the activation function for the hidden layer and linear function (Purelin) (Eq. (3)) in the output layer in this study.
- The weights and bias of this model: Updating with random initialization of weights and biases were set from 1 to 10 neuron sizes as the trial-and-error approach uses the double loop in each number of hidden layer neurons that lead to 200 tests [12].

$$F\left(x\right)=\frac{2}{1+e^{-2x}}-1 \qquad \text{Tangent Sigmoid (tansig)} \tag{2}$$

$$F\left(x\right)=x \qquad \text{Positive Linear (purelin)} \tag{3}$$

The NARX has the same structure as the NAR model, but the NARX model has external inputs, as described in Fig. 2. Therefore, the equation of the NARX model is expressed as Eq. (4).

$$y(t)=f(y(t-1),y(t-2),\ldots,y(t-dy),x(t-1),x(t-2),\ldots,x(t-dx)) \tag{4}$$

where x is the exogenous input dataset, and d_x is the input delay (ID). The NARX's configurations are based on the NAR's configuration; nonetheless, ID and correlated input variables need to be explored. Since 66 ERA5 climate variables could be utilized as the inputs in the NARX model, statistically significant variables were evaluated from the total of ERA5 climate variables by employing the cross-correlation function between power demand and ERA5 climate variables. In site GS1, the most six correlated climate variables were found as the exogenous inputs of the NARX model, and the input delays were chosen from 0 to 2 as vector value ($d_x \geq 0$).

The performance of models was evaluated by using error indices: normalized root-mean-square error (NMSE) [13], coefficient of determination $\left(R^2\right)$, mean absolute error ($MAE,[MW]$), mean absolute percentage error ($MAPE,[\%]$), and root-mean square error ($RMSE,[MW]$), were expressed as given in Eq. (5) to Eq. (9).

$$NMSE=\frac{\sum_{i=1}^{N}\left(P^{i}_{predicted}-P^{i}_{actual}\right)^2}{\sum_{i=1}^{N}\left(P^{i}_{actual}-\overline{P}_{actual}\right)^2} \tag{5}$$

$$R^2=1-\frac{\sum_{i=1}^{N}\left(P^{i}_{predicted}-P^{i}_{actual}\right)^2}{\sum_{i=1}^{N}\left(P^{i}_{actual}-\overline{P}_{actual}\right)^2} \tag{6}$$

$$MAE=\frac{1}{N}\sum_{i=1}^{N}\left|P^{i}_{predicted}-P^{i}_{actual}\right| \tag{7}$$

$$MAPE=\frac{100}{N}\sum_{i=1}^{N}\left|\frac{P^{i}_{actual}-P^{i}_{predicted}}{P^{i}_{actual}}\right| \tag{8}$$