

TRƯỜNG ĐẠI HỌC QUY NHƠN
QUY NHON UNIVERSITY

TẠP CHÍ KHOA HỌC
TRƯỜNG ĐẠI HỌC QUY NHƠN

QUY NHON UNIVERSITY
JOURNAL OF SCIENCE

KHOA HỌC TỰ NHIÊN VÀ KỸ THUẬT
NATURAL SCIENCES AND ENGINEERING

18 (5)

2024

OCTOBER 2024

CONTENTS

1.	Temporal point processes for business process monitoring Nguyen Truong Thanh Hung, Tran Thi Cam Thanh, Truong Thi Cam Mai	5
2.	Investigating parameters affecting pile cap internal forces using the finite element method Hoang Cong Vu, Nguyen Thi Thao Nguyen, Tran Ba Canh	21
3.	A fault location method based on single-ended traveling wave using discrete wavelet transform Ngo Minh Khoa, Doan Duc Tung	29
4.	Evaluation of the agronomical traits and saponin content of wild bitter melon accessions (Momordica charantia L. var. abbreviata Ser.) collected from South Central region of Vietnam Bui Hong Hai, Nguyen Thi Y Thanh, Ho Tan	43
5.	Research and development of a monitoring and switching system of electrical parameters based on image processing and IoT Luong Ngoc Toan, Doan Duc Tung, Le Tuan Ho	55
6.	Research on influence of cement content on bearing capacity of soil-cement piles to treat soft clay in Ha Thanh river urban area, Quy Nhon city, Binh Dinh province, Vietnam Hoang Cong Duy, Pham Van Truc, Pham Thi Lan, Nguyen Thi Khanh Ngan	65
7.	Generating 3D interaction surfaces and assessing load-bearing capacity of rectangular reinforced concrete columns Le Thi Mai Trang, Vang Hieu Quang	71
8.	A study of SCADA Designer application for developing demand side management program Tran Tien Dat, Nguyen Vu Hoa, Ngo Minh Khoa	83
9.	Application of graph learning in travel recommender system Vu Son Lam, Le Quang Hung, Tran Dinh Luyen, Phan Thi Thuy, Tang Nhat Hung	95
10.	A classical simulation study for carbon dioxide adsorption and separation capacity of MIL-53(Cr) Nguyen Thi Xuan Huynh, Tran Thi Diem Thanh, Nguyen Le Bao Tran, Nguyen Ngoc Khoa Truong	109

Ứng dụng quá trình điểm thời gian trong giám sát quá trình kinh doanh

Nguyễn Trương Thành Hưng^{1,*}, Trần Thị Cẩm Thanh², Trương Thị Cẩm Mai³

¹*Analytics Everywhere Lab, Trường Đại học New Brunswick, Canada*

²*Khoa Kinh tế và Kế toán, Trường Đại học Quy Nhơn, Việt Nam*

³*Khoa Khoa học Tự nhiên, Trường Đại học Quy Nhơn, Việt Nam*

Ngày nhận bài: 26/04/2024; Ngày sửa bài: 14/06/2024

Ngày nhận đăng: 17/06/2024; Ngày xuất bản: 28/10/2024

TÓM TẮT

Dự báo quá trình giám sát kinh doanh là một nhiệm vụ dạng chuỗi thời gian đầy thách thức do bản chất phức tạp và biến thiên của các quy trình kinh doanh, liên quan đến việc dự đoán các trường hợp đang diễn ra như hoạt động tiếp theo, hậu tố hoạt động và dự đoán thời gian còn lại trong một quy trình kinh doanh. Quá trình điểm thời gian được sử dụng rộng rãi để mô hình hóa chuỗi các sự kiện xảy ra ở các khoảng thời gian không đồng đều, để mô hình hóa thời gian xảy ra và nắm bắt các phụ thuộc thời gian giữa các sự kiện. Với những tiến bộ gần đây trong mạng nơ-ron sâu, Quá trình điểm thời gian sâu đã nổi lên như một cách tiếp cận đầy hứa hẹn để nắm bắt các mẫu phức tạp trong chuỗi sự kiện với dấu thời gian. Do đó, Quá trình điểm thời gian sâu có thể là một cách tiếp cận tiềm năng để dự đoán quá trình giám sát kinh doanh. Trong bài báo này, chúng tôi thử nghiệm và xem xét hiệu quả của các nghiên cứu gần đây trong Quá trình điểm thời gian sâu đối với vấn đề giám sát quy trình kinh doanh dự đoán. Kết quả của chúng tôi cho thấy rằng các phương pháp Quá trình điểm thời gian sâu có tiềm năng trong hoạt động tiếp theo và dự đoán thời gian còn lại trong dự đoán quy trình giám sát kinh doanh. Những phát hiện này có thể hữu ích cho các chuyên gia và nhà nghiên cứu quan tâm đến việc phát triển các hệ thống dự đoán giám sát cho các quy trình kinh doanh.

Từ khóa: Giám sát quá trình kinh doanh, quá trình điểm thời gian, mạng nơ-ron sâu.

*Tác giả liên hệ chính.

Email: hung.ntt@unb.ca

Temporal point processes for business process monitoring

Nguyen Truong Thanh Hung^{1,*}, Tran Thi Cam Thanh²,
Truong Thi Cam Mai³

¹*Analytics Everywhere Lab, University of New Brunswick, Canada*

²*Faculty of Economics and Accounting, Quy Nhon University, Vietnam*

³*Faculty of Natural Sciences, Quy Nhon University, Vietnam*

Received: 26/04/2024; Revised: 14/06/2024

Accepted: 17/06/2024; Published: 28/10/2024

ABSTRACT

Predictive business process monitoring is a challenging time series task due to the complex and dynamic nature of business processes, which involves predicting the ongoing cases in terms of the next activity, activity suffix, and remaining time prediction on a business process. Temporal point processes (TPPs) are widely used to model sequences of events happening at irregular intervals, to model the occurrence times of events, and to capture the temporal dependencies among them. With the recent advances in deep neural networks, deep TPPs have emerged as a promising approach for capturing complex patterns in event sequences with occurrence timestamps. Hence, deep TPPs can be a potential approach to tackle business predictive monitoring tasks. In this paper, we experiment and review the effectiveness of recent research on deep TPPs on the predictive business process monitoring problem. Our results suggest that TPP methods have the potential in the next activity and remaining time prediction in the predictive business process monitoring problem. The findings can be helpful to practitioners and researchers interested in developing predictive monitoring systems for business processes.

Keywords: *Business process monitoring, temporal point process, deep neural network.*

1. INTRODUCTION

A business process is a collection of tasks performed asynchronously by various resources, such as humans, software, or hardware, to achieve a specific goal.¹ The execution of these tasks is tracked and documented in an event log, which records details such as the identifier of the case, the event performed, and the timestamp of the event.² There may also be optional case attributes, which are shared by events of the same case, or event attributes that are unique to each event. Business process mining is the discipline

concerned with the analysis of these logs, tackling it from different perspectives such as discovering the underlying process model from the log, checking that the executions registered in the log are conformant with the process model, or extracting or inferring analytics that enhances the description of what has happened in the process executions.³ Predictive monitoring is a process mining technique that predicts how an ongoing process case will unfold using the event log's information. The ability to make predictions is beneficial for anticipating issues

**Corresponding authors.*

Email: hung.ntt@unb.ca

before they arise, enabling the reallocation of resources before they are wasted, and providing recommendations.⁴

Studying the temporal distribution of events and discovering the relationships among different types of events is a great scientific approach for predictive monitoring and understanding the dynamics and mechanism of events occurrence.⁵⁻⁹ One of its choices is the Temporal Point Process (TPP), the stochastic process with marked events on the continuous domain of time, which can naturally capture the clustering or self-correcting phenomena of such sequences of events.^{10,11} Often, the rate of event occurrence, known as conditional intensity, is modeled as a function of time based on the prior observation of events to capture the dynamics of the process. Given that the conditional intensity function (CIF) entirely governs the distribution of such a process, statistical prediction, and inference can all be performed via the CIFs.

Despite significant advancements in TPP, especially in models based on deep neural networks (DNNs), most of these models use

Table 1. Excerpt of a Helpdesk’s business process log.

Case ID	Activity	Resource	Timestamp
Case 1	Assign seriousness	Value 1	2012/10/09 14:50:17
Case 1	Take in charge ticket	Value 1	2012/10/09 14:51:01
Case 1	Take in charge ticket	Value 2	2012/10/12 15:02:56
Case 1	Resolve ticket	Value 1	2012/10/25 11:54:26
Case 1	Closed	Value 3	2012/11/09 12:54:39
Case 2	Assign seriousness	Value 4	2012/04/03 08:55:38
Case 2	Take in charge ticket	Value 4	2012/04/03 08:55:53
Case 2	Resolve ticket	Value 4	2012/04/05 09:15:52
Case 2	Closed	Value 5	2012/05/19 09:00:28

The input of business process mining techniques is an event log, usually composed of events with at least a case identifier, an activity, and a timestamp, and, optionally, case attributes, which are values shared by all the events of the same case, and event attributes, which are specific of each event.^{19,20} A sample log from the Helpdesk data set is shown in Table 1, part

different history encoders to embed historical events and various forms of intensity functions that are parameterized by the embedded historical sequence of events.¹²⁻¹⁷ Also, to our knowledge, no experiment has been conducted on the efficiency of TPP in monitoring the business process.¹⁸

Hence, in this paper, we contribute to defining a data preprocessing procedure for the business process monitoring data set to suit the deep TPP models and evaluating different combinations of TPP methods regarding the history encoders and CIFs. We experiment to evaluate the capability and ability of deep TPP models on the predictive business process monitoring datasets.

2. BACKGROUND

This section provides an overview of the key concepts and techniques used in the study, including predictive business process monitoring and TPPs. It lays the foundation for understanding the experiments and results.

2.1. Predictive business process monitoring

of a real-life help desk event log from an Italian Company.²¹ This event log provides information about each event's case identifier, activity, timestamp, and resource.

Given a certain event prefix of a running case, predictive monitoring is concerned with forecasting how different aspects of the next event or sequence of events will unfold until

the end of the case. There are several prediction targets, such as next activity, next activity suffix, next timestamp, next remaining time, next outcome, next attributes, and next attribute suffix.²² In this paper, our interest is the next activity and the next timestamp prediction.

Formally, m_i , t_i , e_i represent the activity, timestamp, and event, respectively. Let $hd^j(\sigma)$ be an event prefix such as $hd^j(\sigma) = \langle e_1, \dots, e_j \rangle$. Two tackled problems can be defined as the following functions Ω using the newly predicted activities as new inputs for the next prediction until the dummy activity representing the end of the case (“[EOC]”) is reached:

- The next activity prediction problem:
 $\Omega_M(hd^j(\sigma)) = m'_{k+1}$.
- The next timestamp prediction problem:
 $\Omega_T(hd^j(\sigma)) = t'_{k+1}$.

2.2. Temporal point processes (TPPs)

In this section, we introduce the definition of Marked TPP and CIF.

2.2.1. Marked TPP

Marked TPP is a random process representing as an event sequence $X = \{(t_1, m_1), \dots, (t_N, m_N)\}$ with the increasing arrival times of events $\{t_i\}_{1 \leq i \leq N}$ and markers $\{m_i\}_{1 \leq i \leq N}$, such that $t_i \in [0, T)$, $t_i < t_{i+1}$, $\forall i \geq 1$ where N is the number of events.

The mark is equivalent to the event's activity within the context of the business process. Thus, both terms can be used interchangeably afterwards. The inter-event time $\tau_i = t_i - t_{i-1}$ is also considered due to their convenience in computing.¹⁵

Categorical marks $\mathcal{M} = \{1, 2, \dots, K\}$ occurring in the time interval $[0, t)$ of the type- k event. The history $\mathcal{H}(t) = \{(t_j, m_j), t_j < t\}$ which can be considered the event prefix $hd^j(\sigma)$ in the business monitoring context.

The task of TPP models is to parameterize the K conditional intensity function (CIF) $\lambda_k^*(t)$, which can be characterized as follows:

$$\lambda_k^*(t) = \lambda_k(t | \mathcal{H}(t))$$

$$= \lim_{\Delta t \rightarrow 0^+} \frac{Pr(\text{event of type } k \text{ in } [t, t + \Delta t) | \mathcal{H}_t)}{\Delta t}$$

which is defined as the expected instantaneous rate of happening events given the history. The $*$ symbol indicates the conditioning on the history $\mathcal{H}(t)$.

Due to the TPP modeling the distribution of the next timestamp t_i or inter-event τ_i time under the history $\mathcal{H}(t_i)$, the next timestamp prediction task is equivalent to considering the next timestamp t_i given $\mathcal{H}(t_i)$ denoted as follows:

$$\Omega_T(\mathcal{H}(t)) = P_i^*(t)$$

Given the CIF, the distribution $P_i^*(t)$ can be represented by any following functions:^{12,15,16,23,24}

1. Probability density function (PDF): $f_i^*(t)$

2. Cumulative distribution function (CDF):

$$F_i^*(t) = \int_0^t f_i^*(u) du$$

3. Survival function: $S_i^*(t) = 1 - F_i^*(t)$

4. Hazard function: $\phi_i^*(t) = f_i^*(t) / S_i^*(t)$

5. Cumulative hazard function (CHF):

$$\Phi_i^*(t) = \int_0^t \phi_i^*(u) du$$

Here, we pick the PDF of the type- k event at time t as the parametric form, which is defined as:

$$f_k^*(t) = \lambda_k^*(t) \exp\left(-\int_{t_{i-1}}^t \lambda_k^*(u) du\right) \quad (1)$$

where $\exp\left(-\int_{t_{i-1}}^t \lambda_k^*(u) du\right)$ is an exponential term where the exponent is the negative integral from t_{i-1} (the time of the last event before t , namely $i-1 = \arg \max_{j \leq n} \{t_j, t_j < t\}$) to t of the CIF. This integral represents the expected number of type- k events at time u , which happens in the time interval $(t_{i-1}, t]$. The exponential of the negative of this value penalizes the presence of other events in the interval, making it less likely for a new event to occur at time t . The product of these two parts gives the probability density of an event of

type k occurring strictly at time t . Note that we must integrate this density over that interval to obtain the probability of an event occurring within a specific time interval. By aiming to parameterize a model to fit the timestamp distribution, the TPP can infer PDF or CIF for timestamp prediction, including the next event's timestamp and activity prediction.

2.2.2. Conditional Intensity Formulation (CIF)

The CIF with parameters $\Theta_k(t)$ is written as $\lambda_k(t; \Theta_k(t) | \mathcal{H}(t))$. The parameter $\Theta_k(t)$ is considered a piece-wise function of t as:

$$\Theta_k(t) = \chi_k(h_i)$$

where $t \in [t_{i-1}, t_i)$. The formula means that the new occurrence of the type- k event changes the h_i and thus updates the $\Theta_k(t)$.

The choice of the family of CIF functions to approximate the target CIF is critical because the function's ability to approximate accurately determines the TPP's performance in fitting the distribution. Additionally, Equation (1) indicates that the integral term is unavoidable if we maximize the likelihood of the observed sequence of events. Hence, the challenge in computing the log-likelihood is the high computational cost due to the integral term. The closed form of this integral term, such as cumulative hazard function, can make the computation of likelihood feasible.^{15,23,24} In conclusion, the goal of approximating the target CIF is to choose a family of functions in the closed integral form with powerful expressivity.

2.2.3. Modeling the marks

In the business monitoring context, where multiple event types exist, the next activity prediction task is determining the most probable event type based on historical data, which can be dealt with as the categorical classification task. It is generally achieved by first converting the historical encoding to logit scores of a discrete distribution, as shown in the following equation:

$$\kappa(h_i) = \text{logit}(\widehat{m}_i) \quad (2)$$

where $\text{logit}(\widehat{m}_i) \in R^K$, $\kappa: R^D \rightarrow R^K$.

Then, we apply a softmax function to transform logit scores into the categorical distribution, which is the solution to the next activity prediction task as follows:

$$\begin{aligned} \Omega_M(\mathcal{H}(t)) &= \Pr(\widehat{m}_i = k | \mathcal{H}(t)) \\ &= \text{softmax}(\text{logit}(\widehat{m}_i))_k \end{aligned}$$

where $\text{softmax}(\text{logit}(\widehat{m}_i))_k$ is to choose the k -th mark from its output. By forming the loss of activities as the logit scores, the cross-entropy loss for categorical classification is added to the log-likelihood loss given the actual activity m of the i -th event to maximize the joint likelihood of the next timestamp and activity, which is considered independent. Several works on maximizing joint likelihood in conditional forms are proposed, such as time conditioned on marks^{25,26} and marks conditioned on time,¹⁵ which can capture the dependencies between timestamp and activity and leverage the TPP models performance in predicting the timestamp and activity simultaneously. In our experiment, we utilize the idea of using the joint negative log-likelihood (NLL) under the independence between the next timestamp and activity, with the type- k mark for a single sequence X for categorical marks computed as:

$$\begin{aligned} & -\log p(X) \\ &= \sum_{i=1}^N \sum_{k=1}^K -\Pr(\widehat{m}_i = k | \mathcal{H}(t)) \\ &+ \sum_{k=1}^K f_k^*(t) \end{aligned} \quad (3)$$

3. CLASSIFICATION OF TPP MODELS

This section introduces the classification of TPP models to solve the predictive business process monitoring problem. As illustrated in Figure 1, our procedure considers two essential parts of a deep TPP: the history encoder and the CIF.

Table 2. The classification of all options for each component by history encoder, mixture distribution, and prediction target.

History Encoder	Recurrent neural network (RNN), Gated recurrent unit (GRU), Long-short term memory (LSTM), Attention, Fourier Transformer (FNet)
Mixture Distribution	Log-Normal, Gompertz, Log-Cauchy, Exponential decay, Weibull
Prediction	Next timestamp, Next activity

3.1. Historical event encoders

Since CIF or PDF is a function of t and historical events before t , namely $\mathcal{H}(t)$, we have to encode the history sequence of each event (t_j, m_j) as a feature vector e_j to formulate the CIF or PDF of the occurrence of different events to model the process. For the i -th event's history, $\mathcal{H}(t_i)$, j -th event in the history set is embedded in a high-dimensional space including time and mark features, as follows:

$$e_j = [\omega(t_j); E^T m_j]$$

where:

- ω represents the time feature that transforms one-dimension temporal information t_j (or inter-event time τ_j) into a high-dimension vector directly or via its logarithm^{16,26} or trigonometric functions.^{12,27}
- E represents the mark feature, an embedding matrix for marks, and m_j is the one-hot encoding of mark m_j .
- A historical encoder H can be obtained via concatenation of the sequence of embedding $\{e_1, e_2, \dots, e_{(i-1)}\}$ into a vector space of dimension D under the following formula:

$$h_i = H(\{e_1; e_2; \dots; e_{i-1}\})$$

H can be chosen as Recurrent-based encoders, Attention-based encoders, or Fourier transform encoders, and h_i is utilized for the CIF parameterization.

3.1.1. Recurrent-based encoders

Recurrent-based encoders, including RNN units, GRU, and LSTM, can be used as history encoders.¹⁴⁻¹⁶ Their CIF can be formulated as follows:

$$h_0 = 0; \quad h_i = \text{RNN}(e_{i-1}, h_{i-1})$$

where the initial state of the history encoder, h_0 , is set as zero. For each subsequent time step i , the new state, h_i , is updated based on the previous state, h_{i-1} , and the previous event, e_{i-1} , through the RNN function. The RNN takes as inputs the previous state and the previous event and outputs the new state. This state represents the RNN's memory, encoding information about past events that it can use to predict future events.

The advantage of using recurrent-based encoders as the history encoder is that it requires low storage space due to the capability of serial computing. The states and events are processed one at a time, meaning that the RNN does not need to store all of them at once, which can be beneficial in situations where storage space is limited.

However, there are also disadvantages to this approach. The serial computing nature of RNNs can limit their computational speed in both the forward and backward processes.²⁸ Additionally, RNNs can suffer from issues such as the gradient vanishing effect, where the gradients used in learning become very small, making learning slow or even impossible.²⁹ They can also suffer from long-term memory loss, struggling to retain information about events that occurred long ago. These issues can potentially compromise the performance of the RNN history encoder.

3.1.2. Attention-based encoders

Attention-based encoders are part of encoder-decoder architectures that utilize the concept of attention. This mechanism allows models to focus on relevant parts of the input sequence when generating an output.³⁰ Self-attention is proposed as the history encoder in TPPs with fast

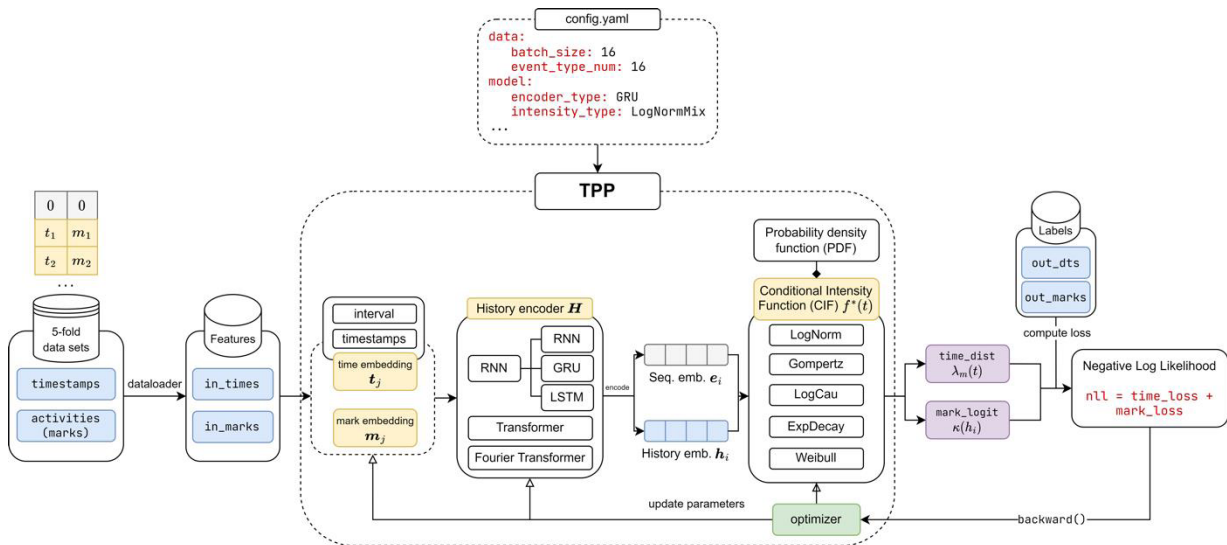


Figure 1. The procedure of TPPs framework for predictive business process monitoring.

parallel computing and the capability of encoding more long-term sequences than recurrent-based encoders.¹² The attention-based history encoders for CIF can be defined as follows:

$$h_i = \frac{\sum_{j=1}^{i-1} \phi(e_j, e_{i-1}) \psi(e_j)}{\sum_{j=1}^{i-1} \phi(e_j, e_{i-1})}$$

where h_i represents the encoded history at time i . This history is computed as a weighted sum of transformed event embeddings $\psi(e_j)$, where the attention mechanism determines the weights $\phi(e_j, e_{i-1})$. The attention mechanism $\phi(\cdot, \cdot)$ is a function that takes two event embeddings as inputs and outputs a scalar called the attention weight. This weight determines the importance or relevance of the event e_j when encoding the history at the time i . The transformation function ψ transforms the event embedding e_j into a series of D -dimensional vectors called values. These transformed embeddings are then used to compute the encoded history.

While attention-based encoders overcome some of the problems with RNNs, their space complexity of the attention matrix is $O(N^2)$, which can become problematic when dealing with very long sequences because the attention mechanism computes pairwise interactions between all events, leading to a quadratic increase in storage requirements as the number of events

increases.³¹ This problem can be temporarily resolved by limiting the encoder only to access the last L events $\{e_{i-L}, \dots, e_{i-1}\}$, which can reduce the time complexity to $O(NL)$.¹⁵

3.1.3. Fourier transform encoders

The Fast Fourier Transform (FFT) module was generally used in the natural language processing (NLP) field³² and recently adopted into the history encoder family under the TPP context called FNet, which aims to speed up the computation and replace the attention mechanism.

$$h_i = \text{Top}_p\{\text{FFT}([\text{FFT}(e_1); \dots; \text{FFT}(e_{i-1})])\}$$

where the $\text{FFT}(\cdot)$ represents the FFT, which operates on the events' embedding, then on the whole sequence. Then, $\text{Top}_p\{\cdot\}$ means choosing the highest p frequencies in the set as the history encoding. The dimension of the feature vector e_j has to equal D dimension. $\text{Top-}p$ needs to be chosen due to the unequal event embedding history sequence length, so the padding operation is required for batch processing as many sequences contain the same padding values, which contain useless information and lead to low-frequency values in spectra. Therefore, filtering the low frequency can capture and retain more information about the historical sequence in the high frequency.

FFT encoder inherits fast computational time complexity in $O(N \log N)$ and the ability to capture long-term patterns due to the global property of the sequences' spectrum.³³ However, one disadvantage of this approach is that the backward process of gradient propagation leads to significant memory complexity.

3.2. Mixture distribution

The mixture distribution family is the major component in TPP that approximates the target PDF under CIF.

3.2.1. Log-normal mixture

Log-normal Mixture is proposed to approximate any distribution due to the feasible computation of its PDF and cumulative distribution function (CDF) in the closed form of CIF and CHF, where the mixture form reads as follows:¹⁵

$$f_{\text{LNM}}^*(t) = \sum_{s=1}^S w_s \frac{1}{t \sigma_s \sqrt{2\pi}} \exp\left(-\frac{(\ln(t - t_{i-1}) - \mu_s)^2}{2\sigma_s^2}\right)$$

where $t \in [t_{i-1}, t_i]$. S are mixture distribution numbers, $\{w_s\}_{1 \leq s \leq S}$ are non-negative mixture weights and $\sum_{s=1}^S w_s = 1$, $\sigma_s > 0 \forall s$.

The distribution of the next timestamp $\lambda_{\text{LN}}^*(t)$ under the log-normal mixture can be modeled by different functions such as PDF, CDF, survival function, hazard function, or cumulative hazard function. Moreover, the preferable function to model the distribution $\lambda_{\text{LN}}^*(t)$ is the cumulative hazard function due to its ability to compute the NLL in the closed form without numerical integration, leading to the loss function in Eq. (3) replaces the $f_m^*(t)$ by the $\Phi_m^*(t)$. Although the CDF has no closed form, the approximation of the function has minor deviation and permits gradient back-propagation, allowing both the forward and backward processes.

3.2.2. Gompertz mixture

Gompertz Mixture is proposed to predict both timestamps and marks of future events without any prior knowledge about the hidden functional

forms of the latent temporal dynamics.¹⁴ The CIF of Gompertz distribution reads as:

$$\lambda(t) = \eta \exp(\beta t)$$

where $\eta, \beta > 0$. The corresponding PDF can be obtained as follows:

$$\lambda_{\text{GP}}^*(t) = \exp(\beta(t - t_{i-1}) + v^T h_i + b_t)$$

which its PDF reads:

$$f_{\text{GP}}^*(t) = \eta \exp\left(\beta(t - t_{i-1}) - \frac{\eta}{\beta} (\exp(\beta(t - t_{i-1})) - 1)\right)$$

for $t \in [t_{i-1}, t_i]$, where $\eta = \exp(v^T h_i + b_t)$, and $\beta > 0$. The process $f_{\text{GP}}^*(t)$ becomes the Poisson distribution when $\beta = 0$. The mixture can be formulated as follows:

$$f_{\text{GPM}}^*(t) = \sum_{s=1}^S w_s \eta_s \exp\left(\beta_s(t - t_{i-1}) - \frac{\eta_s}{\beta_s} (\exp(\beta_s(t - t_{i-1})) - 1)\right)$$

for $t \in [t_{i-1}, t_i]$, where $\beta_s > 0$ and $\eta_s > 0$ for any s . The parameters are obtained as a function of history encoding h_i for $t \in [t_{i-1}, t_i]$ as follows:

$$\Theta(t) = \{w_s(t), \beta_s(t), \eta_s(t)\}_{1 \leq s \leq S} = \chi(h_i)$$

3.2.3. Exp-decay mixture

Zhang et al.¹² extend the expressivity of the multivariate Hawkes Process by the Self-Attentive Hawkes Process (SAHP) by adapting the self-attention mechanism to fit the intensity function of the Hawkes processes. This allows the Exp-decay mixture to capture longer historical information and is more interpretable because the learned attention weight tensor shows the contributions of each historical event. It models the intensity function as the exponential-decaying form like the classical Exp-decay Hawkes Process and extends with a nonlinear transform *softplus* stacked after. This is also the cause of unmanageable computation of the integral term when dealing with long data sequences due to attention weights computation

for each pair of events in the sequence under the self-attention mechanism.

To solve the infeasible computation, the final transformation of non-linearity is removed, and the CIF of the defined Exp-decay distribution can be obtained as follows:

$$\lambda_{ED}^*(t) = \eta \exp(-\beta(t - t_{i-1})) + \alpha$$

where the first term indicates the impacts of historical events decay with an exponential ratio added with the α , which is the basic intensity.

By using the distribution as a component, the mixture of Exp-decay distribution reads the PDF as follows:

$$f_{EDM}^*(t) = \sum_{s=1}^S w_s (\eta_s \exp(-\beta_s(t - t_{i-1})) + \alpha_s) \exp\left(\left(\frac{\eta_s}{\beta_s} - 1\right) \exp(-\beta_s(t - t_{i-1})) - \alpha_s(t - t_{i-1})\right)$$

for $t \in [t_{i-1}, t_i)$, whose parameters are all positive, calculated by $\chi(h_i)$.

3.2.4. Weibull mixture

Weibull Mixture assumes a population of two or more subpopulations with different Weibull distributions.³⁴ The Weibull distribution has the advantage of high approximating ability due to no numerical instability, so the parameter range is not limited to a certain range. The parameters of the Weibull mixture model can be estimated using the maximum likelihood estimation (MLE) or Bayesian methods. Its CIF reads:

$$\lambda_{WB}^*(t) = \eta \beta (\eta(t - t_{i-1}))^{\beta-1}$$

where $\eta, \beta > 0$. CIF will increase when $\beta > 1$, decrease when $\beta < 1$, and be constant when $\beta = 1$.

And its PDF represents:

$$f_{WBM}^*(t) = \sum_{k=1}^S w_s \eta_s \beta_s (\eta_s(t - t_{i-1}))^{\beta-1} \exp\left(-(\eta_s(t - t_{i-1}))^\beta\right)$$

3.2.5. Log-Cauchy mixture

Log-Cauchy Mixture is utilized because the Log-Cauchy distribution can model a wide range of data due to its flexibility by handling both symmetric and asymmetric data and super-heavy-tailed distributions with no given mean or standard deviation.³⁵ The Log-Cauchy distribution is also robust to outliers, which is helpful in monitoring the business.³⁶ The Log-Cauchy mixture with the PDF is written as:

$$f_{LCM}^*(t) = \sum_{s=1}^S \frac{w_s}{(t - t_i) \pi} \frac{\sigma}{(\ln(t - t_i) - \mu)^2 + \sigma^2}$$

4. EXPERIMENTS

This section details the experimental setup, including the datasets used, the procedure followed, and the evaluation metrics employed for next activity prediction and next timestamp prediction tasks. This section also provides the necessary context for interpreting the results.

4.1. Dataset

We evaluate two real-time event logs extracted from the *4TU Center for Research Data* to evaluate different combinations of history encoders and conditional intensities: Helpdesk and BPI2012.^{21,37} Table 3 shows relevant statistics from these logs, namely, the number of cases, the number of different activities, the number of events, the average and maximum case length, the maximum and mean event duration in days, the mean and maximum case duration in days and the number of different variants.

Table 3. Statistics of the event logs used for benchmarking. Time-related measures are shown in days.

Statistics	Helpdesk ²¹	BPI 2012 ³⁷
Number of cases	4580	13087
Number of event types	14	36
Number of events	21348	262200
Mean case length	4.66	20.04
Maximum case length	15	175
Mean event duration	11.16	0.45
Maximum event duration	59.92	102.85
Mean case duration	40.86	8.62
Maximum case duration	59.99	137.22
Variants	266	4366

4.2. Procedure

Figure 1 shows the procedure of our framework from preprocessing the data sets, data splitting, and training the history encoder and CIF until inferring the timestamps and activities. We perform data splitting and evaluate all TPP combinations in identical conditions to ensure comparable results. The goal is to simulate a scenario where past knowledge is utilized for training a predictive model, which is then used to predict the future. To achieve this, we perform a 5-fold cross-validation, where every approach is tested once per fold. The event log traces are sorted by their initial event timestamp and split into training, validation, and test sets with a distribution of 64%, 16%, and 20%, respectively. Timestamps and activities are extracted from each fold data set and then encoded into history embeddings h_i and sequence embeddings e_i by a chosen history encoder H . These embeddings are modeled under a specific CIF $f^*(t)$ to extract the time distribution and mark logit as the next timestamp and activity predictions. Later, we utilize the ground truth to compute the joint NLL (Equation (3)) and optimize the parameters for the embeddings, the encoder, and the CIF.

4.3. Metrics

We use the following metrics to evaluate the TPP combinations' performance on the next activity prediction and next timestamp prediction tasks, along with the goodness-of-fit by NLL as Eq. (3). The results are reported as the mean performance of each TPP combination on five folds. Additionally, our evaluation appends a dummy event as an ("*EOC*") token to the end of every log trace, which can reduce the process state and provide a clear stopping point for activity prediction.

4.3.1. Next activity prediction

We use the accuracy metric since the next activity prediction task is a classic classification problem. The accuracy measures the proportion of correct classifications in relation to the number of predictions done, which is implemented as follows:

$$\begin{aligned} &\text{Top-}q \text{ ACC}(\{\widehat{m}_i\}_{1 \leq i \leq N}, \{m_i\}_{1 \leq i \leq N}) \\ &= \frac{|\{m_i \in \text{Top}_q\{\text{logit}(\widehat{m}_i)\}: 1 \leq i \leq N\}|}{N} \end{aligned}$$

where $\text{logit}(\widehat{m}_i) \in R^K$ is obtained by Eq. (2) to measure the predicted discrete probability.

4.3.2. Next timestamp prediction

Since the time prediction problem is a regression task, the metric chosen for measuring the TPP performance in the next timestamp prediction task is the Mean Absolute Error (MAE). Instead of evaluating TPP performance based on the normalized value taken directly from the distribution under the Mean Absolute Percentage Error (MAPE), we alter the normalization step and postprocessing step to return the next timestamp prediction in days to have a fair comparison with our benchmarks. MAE metric has the advantage of not over-penalizing the variability in the observations, which is important in the time prediction in predictive process monitoring, where the time between two events in a trace can be potentially large. The MAE is defined as follows:

$$\text{MAE}(\{\widehat{t}_i\}_{1 \leq i \leq N}, \{t_i\}_{1 \leq i \leq N}) = \frac{\sum_{i=1}^N |\widehat{t}_i - t_i|}{N}$$

where $\widehat{t}_i$ is the i -th predicted timestamp.

Table 4. Experimental results of the next timestamp prediction MAE in days and the next activity accuracy of modeling the overall CIF with different combinations of history encoder and family of distribution. The arrows ↑/↓ indicate that the higher/lower results, the better. The metrics are computed as the mean of the 5-fold cross-validation. The metrics in **bold** mean that the model achieves the top-5 performance in the column.

Methods	Helpdesk				BPI 2012			
	NLL ↓	MAE ↓	Top-1 ACC ↑	Top-3 ACC ↑	NLL ↓	MAE ↓	Top-1 ACC ↑	Top-3 ACC ↑
LogNormMix+RNN	-2.046201	279.991608	0.695686	0.814908	-4.419353	19.107306	0.809411	0.934645
LogNormMix+GRU	-2.049259	269.787903	0.696841	0.814715	-4.498922	19.963001	0.811243	0.935909
LogNormMix+LSTM	-1.965748	265.744049	0.698613	0.814715	-4.805172	17.676338	0.810400	0.935561
LogNormMix+Attention	-2.025522	243.183641	0.698927	0.814715	-4.970017	19.699347	0.804081	0.942155
LogNormMix+FNet	-2.088489	354.522430	0.697804	0.814715	-3.997465	17.554281	0.700791	0.924937
GomptMix+RNN	0.862758	26.262592	0.694530	0.815293	-1.943585	3.082433	0.811463	0.935542
GomptMix+GRU	0.756814	36.176449	0.698190	0.814522	-1.989259	3.715980	0.814430	0.935689
GomptMix+LSTM	0.803448	33.733837	0.694915	0.814909	-1.718078	1.786893	0.811939	0.935103
GomptMix+Attention	0.198063	24.308924	0.798151	0.974923	-2.651078	2.053919	0.810492	0.936513
GomptMix+FNet	0.922608	26.557230	0.695300	0.812982	-1.338134	1.192549	0.658058	0.919295
LogCauMix+RNN	-0.245449	24.777443	0.697612	0.815100	-2.425044	1.674708	0.809997	0.933234
LogCauMix+GRU	-0.238866	25.448597	0.693374	0.812789	-2.447077	4.090276	0.809778	0.935616
LogCauMix+LSTM	-0.213667	9.497505	0.697034	0.811633	-2.442514	2.642548	0.810345	0.935744
LogCauMix+Attention	-0.572755	20.529713	0.697612	0.812982	-2.426620	1.027533	0.806151	0.928967
LogCauMix+FNet	-0.229975	24.389563	0.698960	0.814137	-2.372896	2.651052	0.698978	0.924735
WeibMix+RNN	-1.920710	354.471039	0.690100	0.814522	-4.106110	17.718061	0.807818	0.933308
WeibMix+GRU	-1.807741	35.020641	0.689137	0.811633	-4.190582	14.316733	0.808514	0.933930
WeibMix+LSTM	-1.856559	295.800934	0.686248	0.811633	-4.111487	15.760619	0.812818	0.935542
WeibMix+Attention	-1.986018	123.920174	0.687982	0.812982	-4.050585	19.363649	0.796498	0.931842
WeibMix+FNet	-1.750751	308.825226	0.683166	0.808166	-3.867040	19.018917	0.694252	0.923929
ExpDecayMix+RNN	1.778744	67.937225	0.696263	0.814908	1.221636	20.735715	0.811005	0.935011
ExpDecayMix+GRU	1.889372	117.146721	0.697612	0.815293	-1.230995	20.580959	0.814375	0.935689
ExpDecayMix+LSTM	1.963200	85.079498	0.696456	0.813367	-1.205069	20.735718	0.812086	0.935231
ExpDecayMix+Attention	1.447936	354.209351	0.696841	0.814522	-1.521961	20.748582	0.808111	0.934700
ExpDecayMix+FNet	1.800676	351.875275	0.697997	0.814522	-1.036910	20.735718	0.699857	0.924918

5. RESULTS

This section presents the findings from the experiments, highlighting the performance of different TPP models based on their combinations of history sequence encoders and formulations of conditional intensity functions. This section summarizes the key observations and insights gained from the experiments.

Table 4 evaluates different combinations of history encoders and overall conditional

intensities on two real-world datasets, namely Helpdesk and BPI 2012.

• **Goodness-of-fit** is typically evaluated via the NLL result. The choice of history encoders such as RNN-based, Attention-based, and FNet-based methods usually does not affect the overall performance of TPP models regarding the goodness-of-fit. Meanwhile, the intensity functions used for CIF approximation matter most. ExpDecayMix shows the worst fitting

ability. Besides, LogNormMix and WeibMix usually fit the data best due to the ability to fit the distribution via the NLL.

- **Next timestamp prediction** is evaluated according to the MAE. The choice of intensity function is also crucial, where LogCauMix and GompMix usually predict significantly better than others. In the Helpdesk dataset, interestingly, the LSTM with LogCauMix performs far better than any other combinations of TPP models.

- **Next activity prediction** is evaluated via the Top-1 ACC and Top-3 ACC. The results show that the history encoder is critical because the prediction depends on its encodings. Attention-based encoders usually have good predictive performance because they can capture long-term features from historical events. Besides, GRU and LSTM also achieve high results due to their ability to capture long memory.

To sum up, the NLL and MAE calculated by timestamps are predominantly influenced by the formulation of intensity and short-term influences, which the five history encoders can adequately capture. Though FNet is a new proposed approach and does not achieve high results compared with other history encoders, it still shows potential when pairing with suitable intensity functions such as LogCauMix. All history encoders can sufficiently model the dynamics of arrival time, given their minor differences. In contrast, Attention-based encoders usually surpass other history encoders to model the dynamics of the next activity due to the capability of capturing the long-term impacts of historical events.

6. DISCUSSION

In this section, we present our key findings which are concluded from the experimental results, and outline our future study.

6.1. Key findings

The experimental results presented in this study provide valuable insights into the performance

of different TPP models for predictive business process monitoring tasks. The findings highlight the importance of selecting an appropriate combination of history sequence encoders and CIFs to achieve optimal results. One key observation is that the choice of CIF plays a crucial role in the next timestamp prediction task. The LogCauMix and GompMix intensity functions consistently outperform other options, indicating their suitability for capturing the temporal dynamics of business processes. This suggests that the formulation of the intensity function should be carefully considered when designing TPP models for timestamp prediction. Another notable finding is the impact of the history encoder on the next activity prediction task. Attention-based encoders, such as the self-attention mechanism, demonstrate superior performance compared to other encoders. This can be attributed to their ability to capture long-term dependencies and selectively focus on relevant historical events. The results underscore the importance of leveraging attention mechanisms to effectively model the complex relationships between past activities and future predictions. The experiments also reveal that the FNet encoder, despite being a relatively new approach, shows potential when paired with suitable intensity functions like LogCauMix. While its performance may not surpass other established encoders, the FNet's ability to capture temporal patterns efficiently makes it a promising direction for future research in TPP models for business process monitoring. Conclusively, the performance of TPP models can vary depending on the characteristics of the dataset and the specific business process being monitored. The Helpdesk and BPI 2012 datasets used in this study represent real-world scenarios, but the generalizability of the findings to other domains and processes should be further investigated. Future research could explore the application of TPP models to a wider range of business processes and datasets to validate the robustness of the observed trends.

6.2. Future study on explainability

Another important aspect to consider when applying TPP models in predictive business process monitoring is explainability.³⁸ As businesses increasingly rely on automated decision-making systems, the ability to interpret and understand the predictions made by these models becomes crucial. Explainability helps to build trust in the model's outputs, facilitates debugging and error analysis, and enables stakeholders to gain insights into the factors influencing the predictions.³⁹⁻⁴¹ However, it is important to note that achieving explainability in TPP models is not without challenges. The complexity of the models, the high-dimensional nature of the input data, and the temporal dependencies can make it difficult to provide simple and intuitive explanations. Striking a balance between model performance and interpretability is an ongoing research challenge. Future work in this area could focus on developing novel explainability techniques tailored to TPP models, as well as conducting user studies to assess the effectiveness and usability of these explanations in real-world business settings.

7. CONCLUSION

In this paper, we evaluate the performance of different TPP models via their combinations of the history sequence encoder and formulation of CIF on the predictive business process monitoring data sets. The results show that the formulation of intensity influences the next timestamp prediction and can be captured by any of the history encoders. The next event prediction is dominated by the ability to capture long-term impacts from historical events, especially attention-based encoders. In our future work, we plan to conduct a more profound experiment around several aspects of TPP models, such as loss computation, history embedding's normalization, the relational discovery of events, and optimizations. In our future work, we extend the capability of TPP models on other prediction

problems such as activity suffix and remaining time prediction, and continue research on explainability to bridge the gap between model performance and interpretability, ultimately leading to more effective and user-friendly monitoring systems.

REFERENCES

1. M. Kirchmer. *High performance through business process management*, Springer, USA, 2017.
2. W. V. D. Aalst. *Process mining: discovery, conformance and enhancement of business processes*, Springer, New York, 2011.
3. W. V. D. Aalst. Process mining: overview and opportunities, *ACM Transactions on Management Information Systems*, **2012**, 3(2), 1-17.
4. F. M. Maggi, C. D. Francescomarino, M. Dumas, C. Ghidini. *Predictive monitoring of business processes*, Advanced Information Systems Engineering: 26th International Conference, CAiSE 2014, Thessaloniki, Greece, 2014.
5. D. Luengo, M. Sepúlveda. *Applying clustering in process mining to find different versions of a business process that changes over time*, Proceedings of the 9th Business Process Management Workshops, BPM 2011 International Workshops, Clermont-Ferrand, France, 2011.
6. H. R'bigui, C. Cho. The state-of-the-art of business process mining challenges, *International Journal of Business Process Integration and Management*, **2017**, 8(4), 285-303.
7. Y. Bertrand, S. Veneruso, F. Leotta, M. Mecella, E. Serral. *NICE: the Native IoT-centric event log model for process mining*, Proceedings of the 5th International Conference on Process Mining, ICPM 2023, Rome, Italy, 2023.
8. M. Vidgof, S. Bachhofner, J. Mendling. *Large language models for business process management: opportunities and challenges*,

- Proceedings of the 5th International Conference on Business Process Management, Utrecht, The Netherlands, 2023.
9. A. Saha, P. Agarwal, S. Ghosh, N. Gantayat, R. Sindhgatta. *Towards business process observability*, Proceedings of the 7th Joint International Conference on Data Science & Management of Data (11th ACM IKDD CODS and 29th COMAD), Bangalore, India, 2024.
 10. A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes, *Biometrika*, **1971**, 58(1), 83-90.
 11. V. Isham, M. Westcott. A self-correcting point process, *Stochastic Processes and their Applications*, **1979**, 8(3), 335-347.
 12. Q. Zhang, A. Lipani, O. Kirnap, E. Yilmaz. *Self-attentive hawkes process*, Proceedings of the 37th International Conference on Machine Learning, ICML 2020, Virtual event, 2020.
 13. J. Yan, X. Liu, L. Shi, C. Li, H. Zha. *Improving maximum likelihood estimation of temporal point process via discriminative and adversarial learning*, Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI-18, Stockholm, 2018.
 14. N. Du, H. Dai, R. Trivedi, U. Upadhyay, M. G. Rodriguez, L. Song. *Recurrent marked temporal point processes: embedding event history to vector*, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, USA, 2016.
 15. O. Shchur, M. Biloš, S. Günnemann. *Intensity-free learning of temporal point processes*, Proceedings of the 8th International Conference on Learning Representations, ICLR 2020, Germany, 2020.
 16. T. Omi, N. Ueda, K. Aihara. *Fully neural network based model for general temporal point processes*, Proceedings of the 33rd Conference on Neural Information Processing Systems, NeurIPS 2019, Vancouver, Canada, 2019.
 17. H. Mei, J. Eisner. *The neural hawkes process: a neurally self-modulating multivariate point process*, Proceedings of the 31st Conference on Neural Information Processing Systems, USA, 2017.
 18. P. G. Fernandez, E. R. Maneiro, J. C. Vidal, M. Lama. VERONA: a python library for benchmarking deep learning in business process monitoring, *SoftwareX*, **2024**, 26, 101734.
 19. W. M. V. D. Aalst, H. A. Reijers, A. J. M. M. Weijters, B. F. V. Dongen, A. K. A. D. Medeiros, M. Song, H. M. W. Verbeek. Business process mining: an industrial application, *Information Systems*, **2007**, 32, 713-732.
 20. D. Dakic, D. Stefanovic, I. Cosic, T. Lolic, M. Medojevic. *Business process mining application: a literature review*, Proceedings of the 29th International DAAAM Symposium, Croatia, 2018.
 21. M. Polato. *Dataset belonging to the help desk log of an Italian company*, University of Padova, Italy, 2017.
 22. E. R. Maneiro, J. Vidal, M. Lama. Deep learning for predictive business process monitoring: review and benchmark, *IEEE Transactions on Services Computing*, **2021**, 16(1), 1-20.
 23. M. Okawa, T. Iwata, T. Kurashima, Y. Tanaka, H. Toda, N. Ueda. *Deep mixture point processes: Spatio-temporal event prediction with rich contextual information*, Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, USA, 2019.
 24. A. Soen, A. Mathews, D. G. Cheng, L. Xie. *UNIPoint: universally approximating point processes intensities*, Proceedings of the 35th AAAI Conference on Artificial Intelligence, AAAI 21, USA, 2021.
 25. J. Enguehard, D. Busbridge, A. Bozson, C. Woodcock, N. Hammerla. Neural temporal point processes for modelling electronic health records, *Machine Learning for Health*, **2020**, 136, 85-113.
 26. S. Zuo, H. Jiang, Z. Li, T. Zhao, H. Zha. *Transformer hawkes process*, Proceedings of the 37th International Conference on Machine Learning, ICML 2020, Virtual event, 2020.
 27. S. Zhu, S. Li, Z. Peng, Y. Xie. Imitation learning of neural spatio-temporal point processes, *IEEE Transactions on Knowledge and Data Engineering*, **2021**, 34(11), 5391-5402.

28. Y. Su, C. C. J. Kuo. Recurrent neural networks and their memory behavior: a survey, *APSIPA Transactions on Signal and Information Processing*, **2022**, 11(1), e26.
29. P. Le, W. Zuidema. *Quantifying the vanishing gradient and long distance dependency problem in recursive neural networks and recursive LSTMs*, Proceedings of the 1st Workshop on Representation Learning for NLP, Germany, 2016.
30. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin. *Attention is all you need*, Proceedings of the 31st Conference on Advances in Neural Information Processing Systems, NIPS 2017, USA, 2017.
31. K. Choromanski, V. Likhoshesterov, D. Dohan, X. Song, A. Gane, T. Sarlos, P. Hawkins, J. Davis, A. Mohiuddin, L. Kaiser, D. Belanger, L. Colwell, A. Weller. *Rethinking attention with performers*, Proceedings of the 9th International Conference on Learning Representations, ICLR 2021, Virtual conference, 2020.
32. J. L. Thorp, J. Ainslie, I. Eckstein, S. Ontanon. *FNet: mixing tokens with fourier transforms*, Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022, USA, 2022.
33. H. Lange, S. L. Brunton, J. N. Kutz. From fourier to koopman: spectral methods for long-term time series prediction, *Journal of Machine Learning Research*, **2021**, 22, 1881-1918.
34. J. M. Marín, M. T. R. Bernal, M. P. Wiper. Using weibull mixture distributions to model heterogeneous survival data, *Communications in Statistics - Simulation and Computation*, **2005**, 34(3), 673-684.
35. T. Ruzgas, M. Lukauskas, G. Čepkauskas. Nonparametric multivariate density estimation: case study of cauchy mixture model, *Mathematics*, **2021**, 9(21), 2717.
36. Z. I. Kalantan, J. Einbeck. Quantile-based estimation of the finite cauchy mixture model, *Symmetry*, **2019**, 11(9), 1186.
37. B. V. Dongen. *BPI challenge 2012*, Eindhoven University of Technology, The Netherlands, 2012.
38. J. N. Adams, S. J. V. Zelst, T. Rose, W. M. P. V. D. Aalst. Explainable concept drift in process mining, *Information Systems*, **2023**, 114, 102177.
39. N. Mehdiyev, P. Fettke. Explainable artificial intelligence for process mining: a general overview and application of a novel local explanation approach for predictive process monitoring. *Interpretable artificial intelligence: a perspective of granular computing*, Springer Cham, Switzerland, 2021.
40. H. T. T. Nguyen, H. Q. Cao, K. V. T. Nguyen, N. D. K. Pham. *Evaluation of explainable artificial intelligence: shap, lime, and cam*, Proceedings of the FPT AI Conference, Hanoi, Vietnam, 2021.
41. T. T. H. Nguyen, V. B. Truong, V. T. K. Nguyen, Q. H. Cao, Q. K. Nguyen. *Towards trust of explainable AI in thyroid nodule diagnosis*, Proceedings of the 7th International Workshop on Health Intelligence, AAAI 2023, Springer Cham, Switzerland, 2023.



© 2024 by the authors. This Open Access Article is licensed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license (<https://creativecommons.org/licenses/by-nc/4.0/>).

Khảo sát tham số ảnh hưởng đến nội lực đài cọc bằng phương pháp phần tử hữu hạn

Hoàng Công Vũ*, Nguyễn Thị Thảo Nguyên, Trần Bá Cảnh

Khoa Kỹ thuật và Công nghệ, Trường Đại học Quy Nhơn, Việt Nam

*Ngày nhận bài: 27/03/2024; Ngày sửa bài: 17/06/2024;
Ngày nhận đăng: 01/07/2024; Ngày xuất bản: 28/10/2024*

TÓM TẮT

Móng cọc ngày càng được áp dụng rộng rãi, đặc biệt là cho nhà cao tầng. Theo một số quan điểm thiết kế móng cọc hiện nay, người ta xem đài móng là tuyệt đối cứng và chưa xem xét đến sự làm việc của đất nền dưới đáy đài nên chưa phản ánh đúng thực tế. Bài báo này phân tích ảnh hưởng của cách bố trí cọc, chiều dài cọc và đất nền dưới đáy đài đối với các thông số áp lực xuống cọc và độ lún của nền đất dưới mũi cọc bằng phần mềm sử dụng phương pháp phần tử hữu hạn Plaxis 3D Foundation.

Từ khóa: *Móng cọc, phân tích phần tử hữu hạn, cách bố trí cọc, chiều dài cọc.*

*Tác giả liên hệ chính.

Email: hoangcongvu@qnu.edu.vn

Investigating parameters affecting pile cap internal forces using the finite element method

Hoang Cong Vu*, Nguyen Thi Thao Nguyen, Tran Ba Canh

Faculty of Engineering and Technology, Quy Nhon University, Vietnam

Received: 27/03/2024; Revised: 17/06/2024;

Accepted: 01/07/2024; Published: 28/10/2024

ABSTRACT

Nowadays pile foundation is more and more widely applied, especially for high-rise buildings. According to some current perspectives on pile foundation design, pile caps are often considered to be rigid, and the behavior of the soil beneath the pile cap is not fully taken into account, leading to an incomplete reflection of reality. This paper analyzes the influence of pile arrangement, pile length and the geology at the bottom of the pile foundation on the pressure on the pile and foundation settlement under the pile tip using Plaxis 3D software.

Keywords: *Pile foundation, finite element analysis, pile arrangement, pile length.*

1. INTRODUCTION

The utilization of pile foundations is increasingly pervasive, notably in the construction of high-rise structures. However, prevailing viewpoints in pile foundation design often depict pile caps as rigid entities, overlooking the dynamic interactions with the underlying soil, hence presenting an incomplete portrayal of real-world conditions.¹ This study utilizes Plaxis 3D software to investigate the effects of pile arrangement, pile length, and subsurface geology on pile load distribution and foundation settlement beneath the pile tip. The numerical results reveal that pile arrangement significantly influences internal forces within the pile group and settlement, with a square pile configuration emerging as the optimal choice over triangular arrangements. Moreover, the findings indicate that the subsoil beneath the raft contributes to bearing the load alongside the piles.

2. THEORETICAL FORMULATION

2.1. Calculation of low-cap pile foundations

The meaning of the issue of low-cap pile foundations is that the pile foundation is constructed with the pile cap located in the soil at a lower depth, which must satisfy the following working conditions:^{1,2}

- The horizontal load must be balanced with the passive pressure of the soil so that the pile is not affected by horizontal force and only works in compression.

- The external moment is balanced by the reaction forces at the pile head with the coordinates (x_i , y_i) of the pile.

- Particularly for foundations with only one pile placed at the center, it is necessary to consider that the pile can withstand moment and horizontal load. Therefore, the low-cap pile foundation must have more than 2 piles

*Corresponding author.

Email: hoangcongvu@qnu.edu.vn

to resist the moment, and the depth of the foundation must be lower than the depth of undisturbed soil to resist horizontal forces. The reaction force on the pile head has coordinates (xi, yi) as:

$$P_i = \frac{N}{n} \pm \frac{M_y}{\sum x_i^2} \times x_i \pm \frac{M_x}{\sum y_i^2} \times y_i \quad (1)$$

in which:

My: moment in the y-axis direction

Mx: moment in the x axis direction

xi: x coordinate of ith pile compared to load position

yi: y coordinate of ith pile compared to load position

Conditions that need to be checked to satisfy the maximum compression (x_{max}) are:

$$P_{max} \leq Q_a^{tk} \quad (2)$$

Allowable load capacity of piles:

$$Q_a^{tk} = \frac{\gamma_0}{\gamma_n} \times \frac{R_{c,k}}{\gamma_k} \quad (3)$$

with

γ_0 is the working condition coefficient, taken as 1 for single piles and 1.15 in multi-pile foundations

γ_n is the reliability coefficient on the importance of the project, equal to 1.2; 1.15; 1.1 corresponds to the importance of level I, II and III buildings (TCVN 10304:2014: Pile foundations - Design standards)

γ_k is the reliability coefficient according to soil (TCVN 10304:2014: Pile foundations - Design standards).

Table 1. Subsoil parameters.

Model	Type	Thickness (m)	γ_{unsat} (kN/m ³)	γ_{sat} (kN/m ³)	E(kN/m ²)	$\varphi(^{\circ})$	G_s	Groundwater level (m)
M-C	Drained	30	16.8	18.99	22.000	36°03	2.66	-2.8

The 4-pile foundation model in the Plaxis 3D Foundation software (Figure 2) with a mesh size is 0.1 m

2.2. Mohr - Coulomb model in Plaxis

The model is based on the idea of the law of elastic-plastic equilibrium with a fixed threshold not affected by plastic deformation and in the stress state of a point located in the purely elastic threshold surface (Figure 1).³ Re-strengthening or softening rules are not required with the Mohr - Coulomb model. This model is relatively simple and easy to use, often used to approximate the behavior in the early stages of soil with 5 basic parameters.

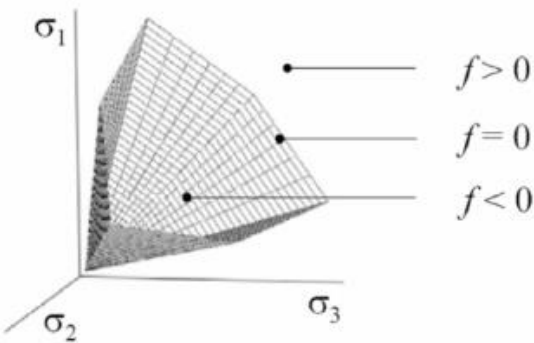


Figure 1. Mohr-Coulomb failure envelope in stress space.

3. NUMERICAL EXAMPLES

3.1. Problem

Input parameters: bored reinforced concrete piles, pile diameter D=0.5 m, pile length 17m; concrete grade B25 with $E_p=27.10^6$ kPa, the subsoil consists of medium coarse sand with the following parameters: Unit weight γ_{unsat} , saturated unit weight γ_{sat} , elastic Modulus E, internal friction angle φ , Specific gravity G_s ,... (Table 1).

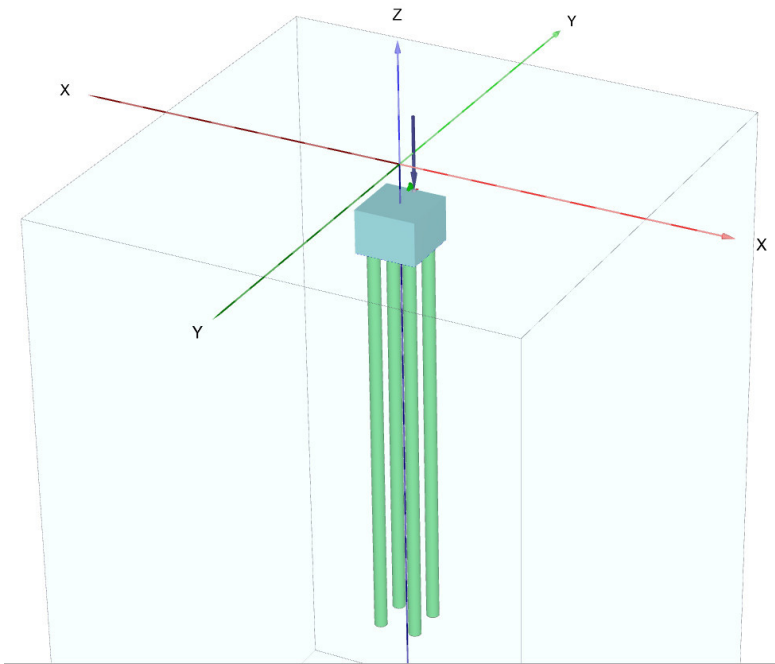


Figure 2. Model of pile foundation using Plaxis 3D Foundation V.22.

Using Plaxis 3D Foundation to analyze a foundation with group of 4 piles with the same length and varying pile spacing to calculate the group efficiency factor. The results are compared with the empirical formula of Converse-Labarre, TCN 272-05 (22TCN 272 – 05: Bridge design standards), and the experimental results of Professor Al-Mhaidib, A.I,⁴ as shown in Figure 3.

The Converse-Labarre formula for calculating the group efficiency factor is:

$$\eta = 1 - 2 \cdot \frac{\text{Arctg}(D / S)}{\pi} \cdot \left(2 - \frac{1}{m} - \frac{1}{n} \right) \tag{4}$$

where m and n are the number of piles in a row and the number of rows of piles, respectively; S is the spacing between piles; D is the diameter of the piles.

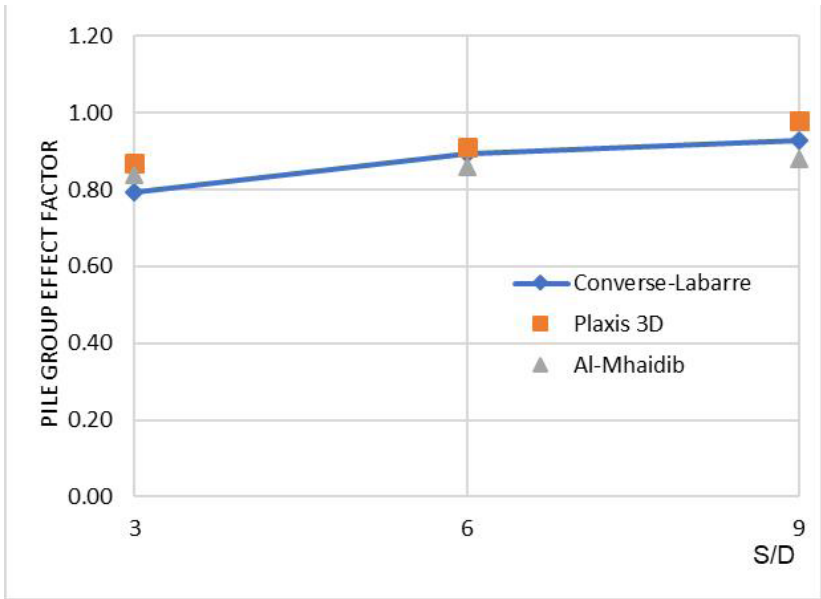


Figure 3. Comparison with published results.

Comments: Finite element analysis using Plaxis 3D Foundation provides results that closely match the empirical formula, TCN 272-05, and experimental results, demonstrating the reliability of the calculation method. Based on this, Plaxis 3D analysis is continued to investigate the following cases:

The authors sequentially varied the pile arrangements, as shown in Figure 4, to analyze the pile head reaction (P_i) and the settlement of the foundation (S). The pile cap height $h=1.8$ m. The calculated data is as follows: $N=2040.36$ kN,

$Q_x=128.85$ kN, $Q_y=16.76$ kN, $M_x=35.07$ kN.m, $M_y=339.21$ kN.m.

The authors continued to alter the subsurface geology beneath the pile cap with layers of clay and silty clay to analyze the pile head reaction (Figure 5). Then, we investigated scenarios involving changes in pile lengths. In this case, the pile cap height $h=1.8$ m, the foundation is only affected by the vertical force P , with P changing the following times: 2000 kN, 4000 kN, 8000 kN, 12000 kN, 16000 kN.

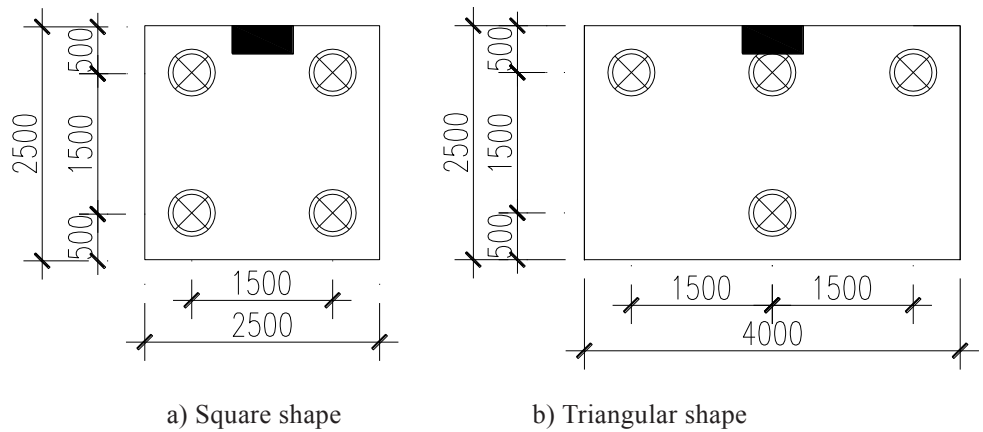


Figure 4. Pile arrangement.

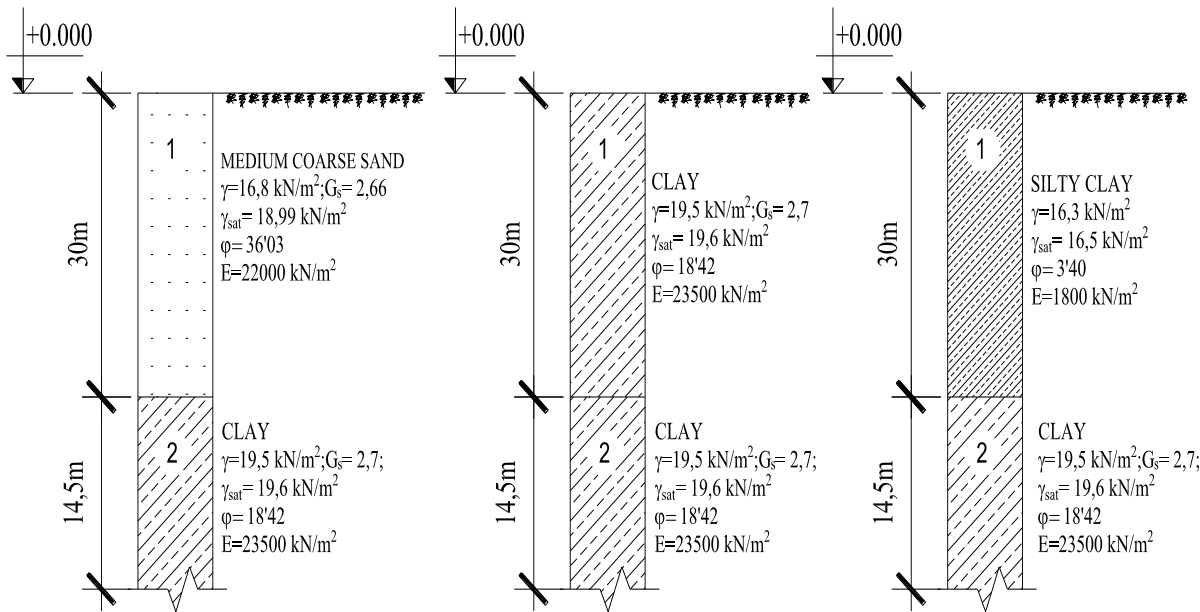
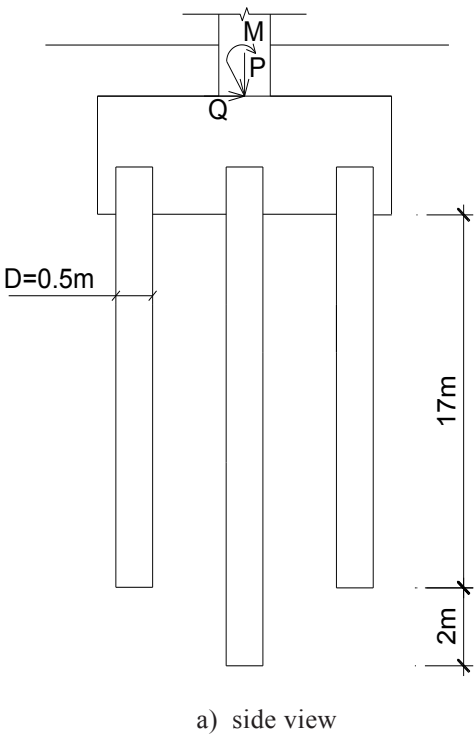


Figure 5. Geological parameters when changing the soil at the bottom of the pile cap.

This paper examined 3 different cases for a group of 4 piles:

- a) All piles have the same length $L=17\text{ m}$
- b) 2 boundary piles are 17 m long, while the remaining 2 piles are 19 m long
- c) All piles have the same length $L=19\text{ m}$



This paper examined 3 different cases for a group of 9 piles:

- a) All piles have the same length $L=17\text{ m}$
- b) Middle pile is 17 m long, while boundary piles are 19 m long
- c) All piles have the same length $L=19\text{ m}$

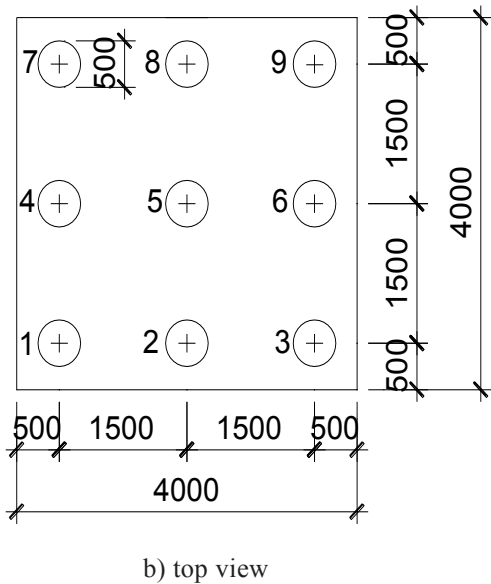


Figure 6. Layout of 3x3 pile group.

3.2. Calculation results

Table 2. Calculation results for settlement and pile head reaction when varying pile arrangements.

Pile arrangement	Square shape	Triangular shape
Settlement (cm)	1.106	0.940
P1 (kN)	407.85	267.56
P2 (kN)	152.72	632.97
P3 (kN)	739.85	518.48
P4 (kN)	856.00	634.87
$\sum P_i$ (kN)	2.156.42	2.053.88

Table 3. Calculation results for settlement and pile head reaction when varying subsurface geology beneath the pile cap.

Geology under the pile cap	Sandy soil	Clay soil	Silty clay soil
Settlement (cm)	1.106	1.105	1.120
P1 (kN)	407.85	407.53	407.86
P2 (kN)	152.72	160.09	155.49
P3 (kN)	739.85	746.04	749.13
P4 (kN)	856.00	852.40	856.87
$\sum P_i$ (kN)	2.156.42	2.166.07	2.169.34

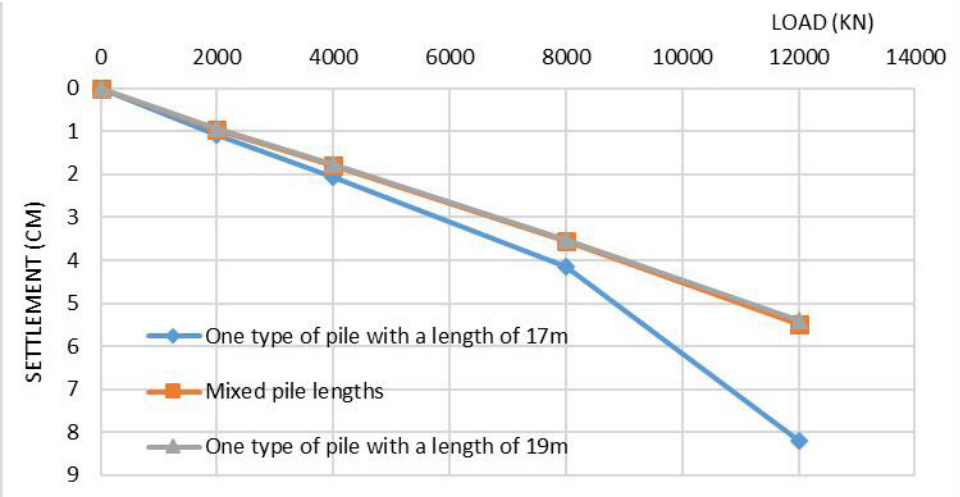


Figure 7. Load-settlement relationship for the 4-pile foundation option.

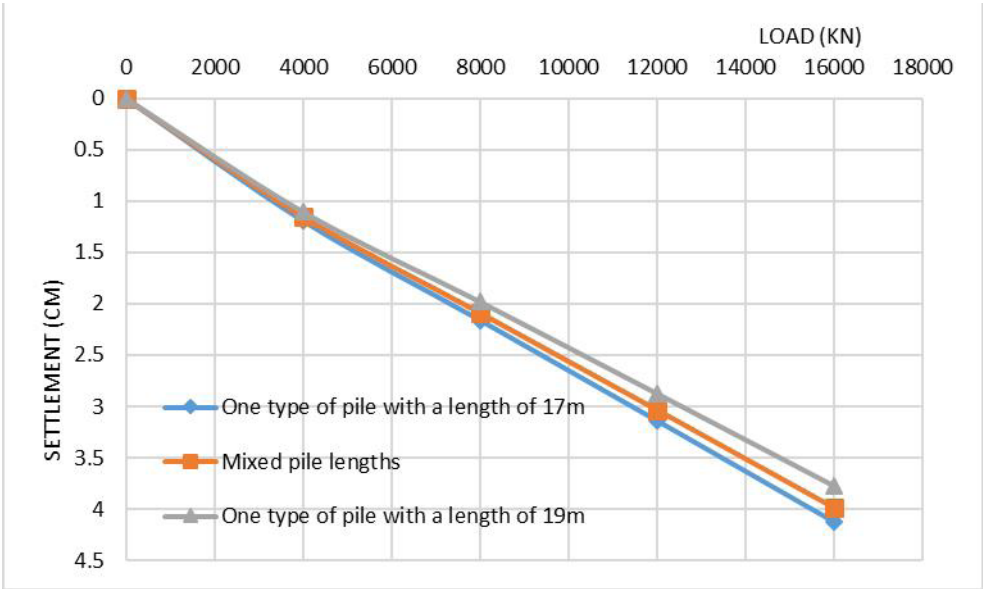


Figure 8. Load-settlement relationship for the 9-pile foundation option.

The analysis results obtained from the Plaxis 3D Foundation software show that when the arrangement of piles varies, the settlement and the value of the reaction force at the pile head in triangular configuration are slightly lower (about 5%) compared to the square configuration (Table 2).

The analysis results of pile head reactions using Plaxis 3D Foundation software show that, with the same applied load, the pile head reaction is affected when the subsoil beneath the raft changes. Among the cases where the subsoil beneath the raft changes, the case with the subsoil being clay yields the highest total pile

head reaction value (2169.34 kN), whereas the case with the subsoil being sandy clay results in the smallest calculated value (2156.42 kN) (Table 3).

From the results presented in Figure 7 and Figure 8, we can see that the stiffness of the two boundary piles (4-pile foundation) and the three middle piles (9-pile foundation) is significantly higher than that of the remaining piles due to the increased length of these piles. However, the 9-pile foundation, with a higher number of piles, exhibits weaker performance compared to a similar arrangement in the 4-pile foundation case. The analysis results indicate

that the effectiveness of a mixed foundation scenario with varying pile lengths lies between the other two cases. This suggests the feasibility of utilizing this type of foundation. However, merely increasing the length of two piles does not significantly enhance the overall load-bearing capacity as in the case of a 3-pile foundation.

4. CONCLUSION

The arrangement of piles affects the internal forces within the pile group and settlement. When changing the pile arrangement to a triangular layout, the internal forces within the pile group decrease. However, a larger pile cap width also leads to material waste. Therefore, to ensure material efficiency while maintaining load-bearing capacity, a square pile arrangement may be chosen.

The pile head reaction is influenced by the subsoil beneath the raft; when the subsoil is good, it yields a smaller total pile head reaction, whereas weak subsoil results in a larger pile head reaction. This calculation result indicates that the subsoil beneath the raft contributes to bearing the load along with the piles.

The research results enable the analysis of the performance of pile foundations in cases where the piles have different lengths. This aids

engineers in practical applications, especially during the construction process, where situations may arise where soil compression leads to some piles not achieving their intended lengths as per the design.

Acknowledgment

This research is conducted within the framework of science and technology projects at institutional level of Quy Nhon University under the project code T2023.804.14.

REFERENCES

1. L. A. Hoang. *Foundation and footing*, Construction Publishing Company, Hanoi, 2010.
2. T. V. Lan. *Foundations and foundations for students majoring in civil and industrial construction*, Construction Publishing Company, Hanoi, 2018.
3. D. V. De. *Plaxis 3D Foundation software applied to foundation and underground construction calculations*, Construction Publishing Company, Hanoi, 2012.
4. A. I. A. Mhaidib. Loading rate effecton piles in clay from laboratory model tests, *Journal of King Saud University*, **2001**, 13(1), 39-55.



© 2024 by the authors. This Open Access Article is licensed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license (<https://creativecommons.org/licenses/by-nc/4.0/>).

Phương pháp định vị sự cố dựa trên truyền sóng một đầu sử dụng biến đổi wavelet rời rạc

Ngô Minh Khoa*, Đoàn Đức Tùng

Khoa Kỹ thuật và Công nghệ, Trường Đại học Quy Nhơn, Việt Nam

*Ngày nhận bài: 22/4/2024; Ngày sửa bài: 28/06/2024;
Ngày nhận đăng: 01/07/2024; Ngày xuất bản: 28/10/2024*

TÓM TẮT

Định vị sự cố là một trong những phương pháp hiệu quả để nâng cao độ tin cậy và giảm thiểu thời gian khắc phục sự cố của hệ thống điện. Bài báo này phát triển một phương pháp định vị sự cố mới dựa trên kỹ thuật truyền sóng sử dụng biến đổi wavelet rời rạc (DWT). Dữ liệu đo lường dòng điện sự cố tại các đầu đường dây sẽ được chuyển đổi sang hệ $\alpha\beta 0$ bằng biến đổi Clarke, thành phần dòng điện alpha sẽ được phân tích bởi biến đổi DWT để trích xuất các hệ số chi tiết có hàm chứa các thành phần tần số bậc cao. Do vậy quá trình lan truyền sóng sẽ xuất hiện trên đường dây sau khi ngắn mạch xảy ra ở bất kỳ vị trí nào dọc theo chiều dài đường dây. Kỹ thuật định vị sự cố dựa trên phương pháp truyền sóng một đầu được nghiên cứu trong bài báo này cho một hệ thống điện có một đường dây mạch đơn. Sự lựa chọn tần số lấy mẫu phù hợp để định vị sự cố trên đường dây truyền tải cũng được khảo sát để đạt được độ chính xác cao hơn. Các kết quả nghiên cứu được mô phỏng và kiểm chứng bằng phần mềm MATLAB/Simulink.

Từ khóa: Đường dây truyền tải, sai số ước lượng, định vị sự cố, phương pháp truyền sóng, biến đổi wavelet rời rạc.

**Tác giả liên hệ chính.*

Email: ngominhkhoa@qnu.edu.vn

A fault location method based on single-ended traveling wave using discrete wavelet transform

Ngo Minh Khoa*, Doan Duc Tung

Faculty of Engineering and Technology, Quy Nhon University, Vietnam

Received: 22/04/2024; Revised: 28/06/2024;

Accepted: 01/07/2024; Published: 28/10/2024

ABSTRACT

Fault location is one of the most efficient methods to improve reliability and reduce downtime of a power system. This paper develops a new fault location method based on the traveling wave using the discrete wavelet transform (DWT). The fault current measurement data at the two line ends is converted to the $\alpha\beta 0$ frame via the Clarke transform, the alpha current component is then decomposed by the DWT to extract the detail coefficients containing the high frequency levels. Thus, the traveling wave propagation will appear on the transmission line after a fault occurs at any location along the line. The single-ended technique is considered in this paper for a test system based on a single-circuit power transmission line model. The choice of the most suitable sampling frequency for locating faults in transmission lines was also investigated to achieve a high fault location accuracy. The research results are simulated and verified in MATLAB/Simulink software.

Keywords: *Transmission line, estimation error, fault location, traveling wave method, discrete wavelet transform.*

1. INTRODUCTION

In power system operations, short circuits are extremely serious; they occur randomly and people cannot diagnose their occurrence. Additionally, because a power transmission line often spreads on wide areas and crosses over hilly regions with a high density of lightning strikes, the transmission line is an element which suffers most short circuits in a power system.¹ With the complicated topography of hilly regions, the fault detection, identification and location technique requires a high accuracy to create quick and convenient conditions in finding and restoring the power

system. As a result, the power system can be improved in terms of reliability and reduced downtime.²

In power systems, four typical types of faults consist of the three-phase-to-ground (ABCG), single-phase-to-ground (AG, BG, CG), phase-to-phase (AB, BC, AC), and double-phase-to-ground (ABG, BCG, ACG) faults as shown in Figure 1. The fault location system needs to detect and to identify accurately the fault type before it can locate the fault in the transmission line. Therefore, the fault locator uses the three-phase fault current characteristics to identify exactly the fault type.³

*Corresponding author.

Email: ngominhkhoa@qnu.edu.vn

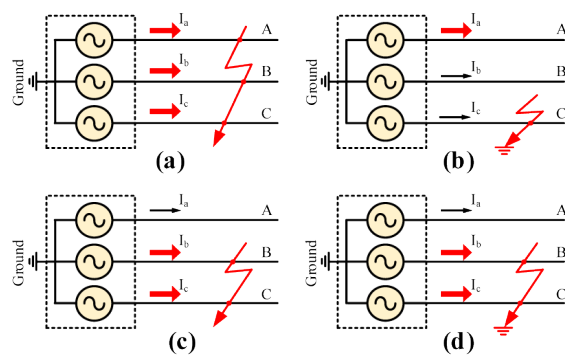


Figure 1. Fault types: (a) Three-phase-to-ground fault, (b) Single-phase-to-ground fault, (c) Phase-to-phase fault, (d) Double-phase-to-ground fault.

For the fault location on power transmission lines, distance information from the relay to the fault location recorded by the digital distance relay can be applied to determine the fault location. However, the information has a high estimation error, the manual methods need to be applied to supporting on the whole power transmission line. Thus, this method cannot find and restore quickly and accurately the fault on the transmission line as well as not have economical save.⁴ From the aforementioned issues, it can be necessarily recognized that the fault detection, identification and location on power transmission lines is an extremely important problem to bring a high effectiveness in power system operation.

The studies on the fault location on power transmission lines can be mainly divided into two groups including the impedance and traveling wave methods as shown in Figure 2. The impedance-based fault location methods have some advantages such as the simple integration into the central control unit of power systems.⁵⁻⁸ However, its disadvantage is the significant dependence on the fault resistance while the traveling wave-based methods have a high accuracy and do not depend on the fault resistance.⁹⁻¹⁴

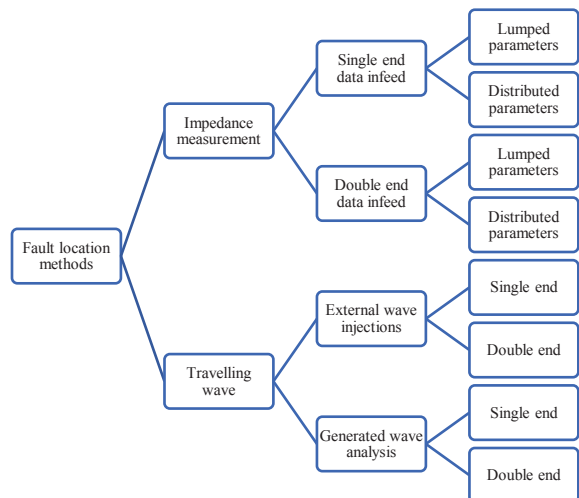


Figure 2. Two main fault location methods.³

In the published work,¹¹ the fault location based on traveling wave time-frequency characteristics was proposed to determine the fault location from a single end of the transmission line. The novel traveling wave-based fault location method for overhead power transmission lines which formed the complete theoretical analysis of the distortion phenomenon of traveling wave signal and the improved traveling wave fault location method considering distortion phenomenon was studied in the publication.¹² The fault locator based on the traveling wave method for high impedance events was implemented for the Brazilian transmission companies.¹³ The frequency-dependent characteristics of high-voltage alternating current (HVAC) transmission lines were researched to propose a traveling wave-based fault location system.¹⁴ As a result, the paper focuses on the fault location methods based on traveling wave. The hybrid method based on wavelet transform, GoogLeNet, convolutional neural network, and probabilistic neural network was proposed to identify, classify, and locate faults on transmission lines.^{15,16}

Two main traveling wave-based fault location methods consist of the single- and double-ended methods. For single-ended methods, voltage and current waves measured and recorded at the single end of the transmission line will be used to determine fault location.^{17,18}

For double-ended methods, both voltage and current waves measured and recorded at two ends of the transmission line will be applied in the fault location system, therefore the measurement data need to be synchronized.¹⁹⁻²² A wavelet transform (WT) is the decomposition of a signal into a set of basic functions consisting of contraction, expansion, and translation of a mother wavelet. Wavelet transforms are mathematical tools for analyzing data where features vary over different time-frequency scales, therefore they have been widely applied in many traveling wave-based fault location systems on power transmission lines.^{23,24}

From the above analysis, it can be observed that there are still gaps in the traveling wave-based fault location method; therefore, this paper focuses on studying and proposing a new fault location method based on traveling wave using the discrete wavelet transform (DWT). Some different wavelet mothers were considered to evaluate the estimation effectiveness in fault location. The arrival current times at the ends of the transmission line were extracted from the detail coefficients by using the DWT. Additionally, single-ended method was applied for a single-circuit transmission line which was modeled and simulated in MATLAB/Simulink software to record the voltage and current waves at two ends of the line. Besides, the different fault resistances and locations were varied along the transmission line to create many case studies in this paper. The main outstanding contributions of this paper are summarized as follows: (i) Developing the traveling wave-based fault location method using the DWT to extract the arrival times at the line ends and determine the fault location on the transmission line. Additionally, some mother wavelets including 'db4', 'coif4', and 'sym4' are considered to verify their effectiveness for fault location, (ii) Changing the sampling frequency of the current signal as well as the fault resistance to record the measurement data and applying the proposed method to calculate the fault location on the power transmission line, and (iii) Applying

MATLAB/Simulink software for modeling and simulating the typical single-circuit transmission line to assume many fault locations along the line.

The next parts of this paper are structured as follows: the proposed method for fault location based on traveling wave using wavelet transform is presented in Section 2. In Section 3, some case studies for the single-circuit transmission line are modeled and simulated in MATLAB/Simulink software to verify and evaluate the proposed method, followed by conclusions in Section 4.

2. PROPOSED METHOD

2.1. Traveling wave technique

In general, when a fault occurs on the transmission line at the time the voltage is not equal zero, it will form propagation waves along the line that will propagate from the fault location to the two line ends at the propagation speed approximately the light speed.⁴ A transmission line can be modeled by its parameters including the series resistance R , the series inductance L , the shunt conductance G , and the shunt capacitance C . For a dx -segment of a transmission line, the constant parameters are Rdx , Gdx , Ldx and Cdx as shown in Figure 3.

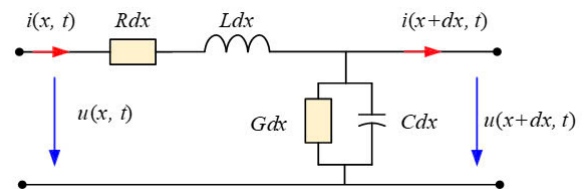


Figure 3. Single-phase transmission line model of distributed parameters.

The traveling wave can be determined as the solutions of linear differential equations for the transmission line. For the general transmission line, the partial differential equations represent for the voltage and current of the line as follows:¹⁰

$$\frac{\partial u(x, t)}{\partial x} = -Ri(x, t) - L \frac{\partial i(x, t)}{\partial t} \quad (1)$$

$$\frac{\partial i(x, t)}{\partial x} = -Gu(x, t) - C \frac{\partial u(x, t)}{\partial t} \quad (2)$$

For the lossless transmission line, the partial differential equations that describe the voltage and current of the line can be rewritten as follows:¹⁰

$$\frac{\partial u(x, t)}{\partial x} = -L \frac{\partial i(x, t)}{\partial t} \quad (3)$$

$$\frac{\partial i(x, t)}{\partial x} = -C \frac{\partial u(x, t)}{\partial t} \quad (4)$$

The voltage and current solutions of the above partial differential equations in the time domain can be satisfied by the general forms as follows:¹⁰

$$u(x, t) = A_1 \left(t + \frac{x}{v} \right) + A_2 \left(t - \frac{x}{v} \right) \quad (5)$$

$$i(x, t) = \frac{1}{Z_0} \left[A_1 \left(t + \frac{x}{v} \right) - A_2 \left(t - \frac{x}{v} \right) \right] \quad (6)$$

where $Z_0 = \sqrt{L/C}$ is the characteristic impedance of the lossless line; $A_1 \left(t + \frac{x}{v} \right)$ and $A_2 \left(t - \frac{x}{v} \right)$ is the reflected and forward traveling wave, respectively.

The propagation speed of the traveling wave along the transmission line is impacted by the line characteristics of inductance (L) and capacitance (C). This relationship can be seen in the following equation:¹⁰

$$v_{prop} = \frac{1}{\sqrt{LC}} \quad (7)$$

The traveling wave methods use the timing to locate the fault in the transmission line. The two traveling wave methods including single- and double- ended methods are applied as shown in Figure 4.

The single-ended traveling wave-based fault location method uses the propagation velocity (v_{prop}), but not the line's length (L_{line}). This method is based on the traveling wave's reflections from the lattice diagram in Figure 4(a). From this figure, the fault location can be determined by using the difference between the fault's arrival time (t_A) and the first reflection's

arrival time (t_{A2}). For this method, the fault location (X) is calculated as follows:

$$X = \frac{1}{2} [(t_{A2} - t_A) v_{prop}] \quad (8)$$

The double-ended traveling wave-based fault location records the arrival time of the traveling wave at each line end and also use the propagation velocity (v_{prop}), the line's length (L_{line}) and the difference time between the arrival times to locate the fault as shown in Figure 4(b). Therefore, this method requires the communication between the two relays at two line ends. According to this method, the fault location (X) is determined as follows:

$$X = \frac{1}{2} [L_{line} + (t_A - t_B) v_{prop}] \quad (9)$$

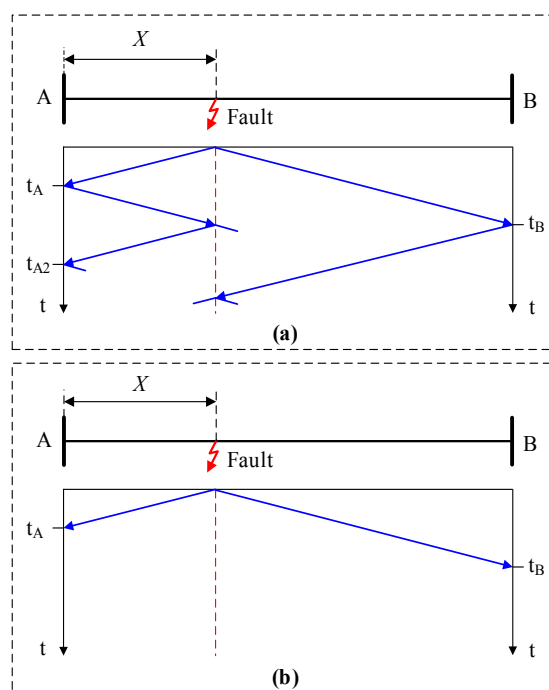


Figure 4. Basic visualization of traveling wave-based fault location methods: (a) Single-ended method, (b) Double-ended method.⁹

2.2. Discrete wavelet transform

DWT is a discrete form of continuous wavelet transform (CWT).²⁵ The CWT of a continuous signal in time domain is defined as follows:

$$CWT(a, b) = \int_{-\infty}^{+\infty} f(t) \psi_{a,b}^*(t) dt, \quad (10)$$

$$a, b \in R, a \neq 0$$

$$\text{with: } \psi_{a,b}^*(t) = \frac{1}{\sqrt{|a|}} \psi^*\left(\frac{t-b}{a}\right) \quad (11)$$

where $\psi(t)$ is the mother wavelet; the symbol “*” represents for a complex number; a and b are scale and transition coefficients, respectively. The scale coefficient a is used to establish the frequency variation and wavelet length; the transition coefficient b is used for depicting the signal transiting location.

In practical applications, the DWT is used to replace the CWT. This way is performed using the discrete values $a = a_0^m$ and $b = nb_0a_0^m$ to replace the scale and transition coefficients. Thus, the following equation is established:

$$\psi_{m,n}(k) = a_0^{-\frac{m}{2}} \psi(a_0^{-m}k - nb_0), \quad m, n \in Z \quad (12)$$

where m and n represent for the frequency and time locations. In general, $a_0 = 2$ and $b_0 = 1$ are set to create an orthogonal wavelet transform and multi-resolution analysis.

In multi-resolution analysis, the signal $f(k)$ is decomposed into the approximation and detail coefficients which are performed by the scale functions $\phi_{m,n}(k)$ and the wavelets $\psi_{m,n}(k)$:

$$\phi_{m,n}(k) = 2^{-(m/2)} \phi(2^{-m}k - n) \quad (13)$$

$$\psi_{m,n}(k) = 2^{-(m/2)} \psi(2^{-m}k - n) \quad (14)$$

The scale function is related to the low-pass filters with the filtering factors $\{g(n)\}$ and the wavelet is related to the high-pass filters with the filtering factors $\{h(n)\}$. These relationships are displayed as follows

$$\phi(k) = \sqrt{2} \sum_n g(n) \phi(2k - n) \quad (15)$$

$$\psi(k) = \sqrt{2} \sum_n h(n) \phi(2k - n) \quad (16)$$

The above filters have some following important characteristics:

$$\sum_n g^2(n) = 1 \quad (17)$$

$$\sum_n h^2(n) = 1 \quad (18)$$

$$\sum_n g(n) = \sqrt{2} \quad (19)$$

$$\sum_n h(n) = \sqrt{2} \quad (20)$$

The filter $h(n)$ is other type of the filter $g(n)$ and there is an odd integer N which satisfies the following equation:

$$h(n) = (-1)^n g(N - 1 - n) \quad (21)$$

The significant advantage of the DWT is to have the changeable window length, therefore it has a high resolution in time-frequency domain for all frequency ranges.²³⁻²⁵ As a result, the DWT can be applied for the traveling wave-based fault location method. To process signals using the DWT, this work applies some available functions in MATLAB toolbox. The function **wavedec** will give the wavelet decompositions of the input signal x at the level n with the mother wavelet **wname** as follows:

$$[c, l] = \text{wavedec}(x, n, 'wname') \quad (22)$$

The function **detcoef** is applied to determine the detail coefficients from the outputs c and l at the level n as follows:

$$D = \text{detcoef}(c, l, n) \quad (23)$$

The detail coefficient characteristics in Eq. (23) as shown in Figure 5 are used to detect and calculate the fault location on the transmission line based on the traveling wave theory.

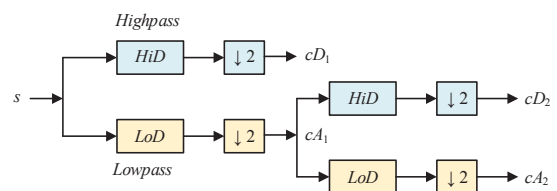


Figure 5. Filter bank using **wavedec** function.

2.3. Clarke transform

The Clarke transform is a signal processing technique to convert the initial signal in the abc

frame into the $\alpha\beta 0$ frame as follows.²⁶ In this paper, the alpha aerial mode is used for fault location on a transmission line in this paper.

$$\begin{bmatrix} I_\alpha \\ I_\beta \\ I_0 \end{bmatrix} = \frac{2}{3} \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix} \quad (24)$$

3. SIMULATED CASE STUDIES

3.1. Test system description

The proposed method is verified and evaluated in this section by applying to two case studies including the single-circuit transmission line. The three-phase transmission line is modeled and simulated in MATLAB/Simulink software to validate the effectiveness of the fault location method based on traveling wave using the DWT. Many different fault resistances and locations along the transmission line are created to record the voltage and current waveforms at the two line ends. The sampling frequency is also set at many levels to get the simulation data. Additionally, the estimation error for comparing the effectiveness of the case studies is calculated as the following equation.⁷

$$\text{Error}(\%) = \left| \frac{L_{\text{estimate}} - L_{\text{actual}}}{L} \right| 100 \quad (25)$$

where L_{actual} is the actual line segment length from the line end to the fault location, L_{estimate} is the distance from the line end to the fault location which is estimated by the proposed method; and L is the line length.

The case studies are simulated via MATLAB/Simulink software according to the test system single-line diagram as shown in Figure 6. In the test system, the main element parameters are as follows. The three-phase source A has the phase-to-phase nominal voltage of 220 kV, the nominal frequency of 50 Hz; the inner resistance of 0.8929 Ω ; and the inner inductance of 16.58 mH. The three-phase source B has the phase-to-phase nominal voltage of 220 kV, the

nominal frequency of 50 Hz; the inner resistance of 0.9375 Ω ; and the inner inductance of 17.41 mH; and the 100 km distributed parameters line has the positive- and zero- sequence resistances of 0.01273 Ω/km and 0.3864 Ω/km , respectively; the positive- and zero- sequence inductances of 0.9337 mH/km and 4.1264 mH/km, respectively; the positive- and zero- sequence capacitances of 0.01274 $\mu\text{F}/\text{km}$ and 0.007751 $\mu\text{F}/\text{km}$, respectively. Additionally, the fault block can be added to the middle position of the distributed parameter line to create different types of faults on the power transmission line.

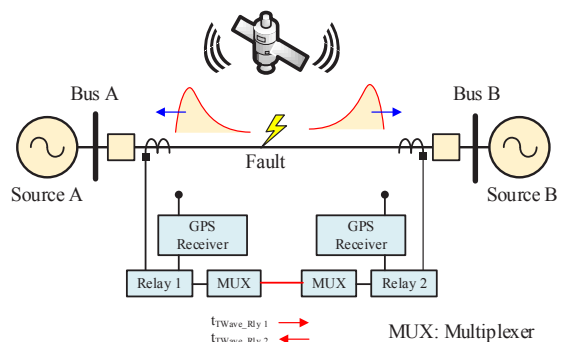


Figure 6. Single-line diagram of test system.

The proposed algorithm flowchart for performing the simulation results in this paper is presented in Figure 7. As shown in this figure, there are two main periods including the steady state simulation and fault mode simulation for the test system. The steps for this algorithm are summarized as follows:

After establishing all the initial parameters for the test system, the algorithm will run power flow at the steady state to get the voltage and current measurements at the power transmission line ends. A short circuit is created on the line and the test system is then simulated to record the fault current waveforms at the line ends. The Clarke transformation is applied using Eq. (24) to transform the fault current from the abc framework to $\alpha\beta 0$ framework. The alpha aerial mode will be sampled at various sampling frequency to perform the DWT using *wavedec* function Eq. (22). After that, Equation (23) is used to retrieve the detail coefficients using the function *detcoef*. The level-1 detail coefficient

is applied to extract the maxima time-frequency indices and calculate the fault location using Eq. (8)-(9) and the estimation error using Eq. (25). In the proposed algorithm flowchart, the fault type, resistance, and location can be randomly changed to verify the effectiveness of the proposed fault location method based on traveling wave using the DWT.

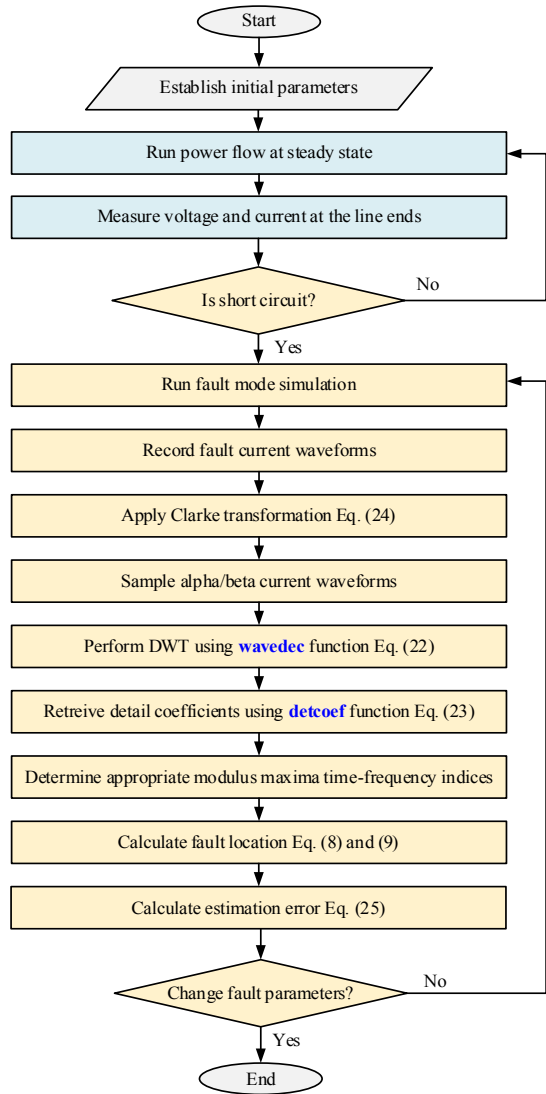


Figure 7. Proposed flowchart based on DWT.

The test power system modeled in MATLAB/Simulink software is shown in Figure 8. The continuous simulation mode is performed for all case studies in this paper. The fault current waveforms at Source A are sampled at various sampling frequency to create simulation data for verifying the proposed fault location method.

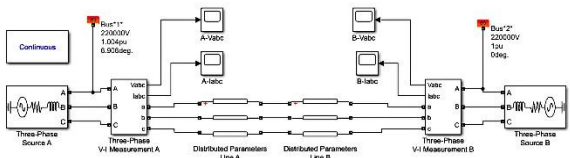


Figure 8. Single-circuit transmission line model in MATLAB/Simulink.

3.2. Pre-fault mode

At the pre-fault steady state, it is assumed that the three-phase source A is generating an active power of 200 MW transmitting through the transmission line to the three-phase source B which are considered as the swing bus in the test system. Power flow problem is performed at this mode to collect the initial measurement data including the voltage of Source A ($\dot{U}_A = 1.004 \angle 6.9^\circ$ pu) and the voltage of Source B ($\dot{U}_B = 1.00 \angle 0^\circ$ pu). The three-phase voltage and current waveforms at the beginning of the line are simulated in the total simulation time of 10 cycles (or 0.2 seconds) as shown in Figure 9. From these simulation results, it can be observed that at the pre-fault steady state the three-phase voltage phasors as well as the three-phase current phasors are balance. Additionally, the current measurement of 524 A can be determined from these current waveforms. This steady state is considered the initial mode for investigating various fault modes which occur at various fault resistances and locations along the power transmission line.

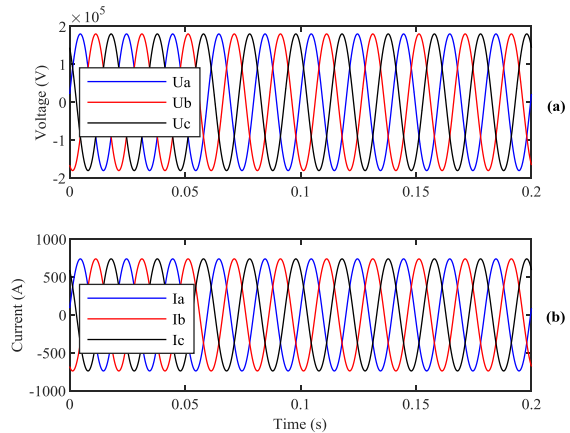


Figure 9. Pre-fault simulation results at Source A: (a) Voltage waveform, (b) Current waveform.

3.3. Fault mode

A single-phase-to-ground fault (AG) is assumed that it occurs at the time $t = 0.1$ sec. The fault appears at the location far away Source A of 20 km with a fault resistance of 1Ω . The total simulation time is set by 0.2 sec. The three-phase abc voltage waveform (Figure 10a) and three-phase abc current waveform (Figure 10b), and the $\lambda\beta 0$ current waveform (Figure 10c). It can be observed obviously that after occurring the AG fault, the phase-A voltage is decreased and the phase-A current is significantly increased. The voltage and current of the phase-B and phase-C are also maintained at the values same as the ones at the steady state. For the current in $\lambda\beta 0$ frame, the alpha and zero components are increased but the beta component is not changed.

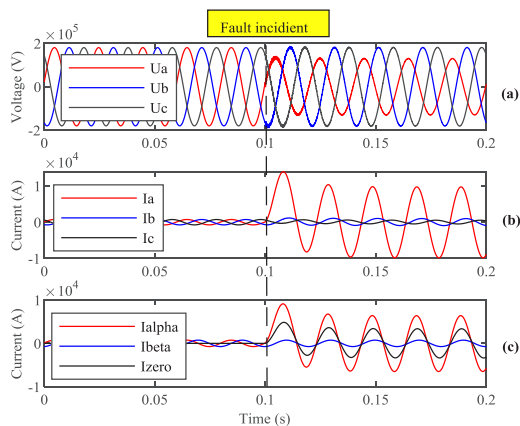


Figure 10. Fault simulation results at Source A: (a) Voltage waveform, (b) Current waveform in abc frame, (c) Current waveform in $\alpha\beta 0$ frame.

The proposed fault location method based on traveling wave using the DWT is performed by sampling the alpha component of the fault current at many frequencies to create the input data of the DWT. The level-1 detail coefficient of the DWT is then extracted to determine the arrival times of the reflected and forward waves at the beginning of the transmission line. The wavelet 1-D toolbox in MATLAB apps is used to analyze the alpha component of AG fault current in this case. The GUI of this toolbox is shown in Figure 11. These analysis results are an example for an AG fault at the location of 20 km. The wavelet 'db4' is applied in this case. The level-1 detail and the detail coefficients are displayed

on the second and third plots in Figure 11. All these coefficients can be exported to MATLAB workspace. Besides, the wavelet selection is performed by change it on the GUI. Therefore, this paper applies this wavelet toolbox to decompose the alpha component of the fault current for locating faults in a transmission line.

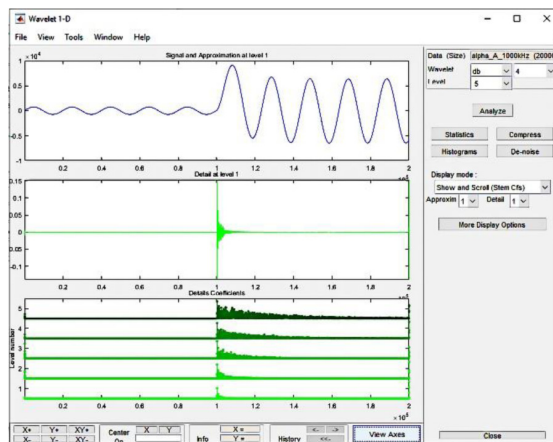


Figure 11. GUI of the wavelet 1-D toolbox.

To calculate the fault location, the propagation velocity of the traveling wave must be calculated based on the line parameters. The positive sequence line inductance and capacitance are used to determine the velocity $v_{prop} = 1/\sqrt{LC} = 289942$ km/s.

For the 20km-AG fault occurring at the time of 0.1 seconds, the level-1 detail coefficient of the DWT is analyzed as Figure 12(a) for the sampling frequency of 1000 kHz. From this figure, the index for the traveling wave from the fault location to the bus A is equal to 50036 samples and the reflection wave from the fault location is equal to 50105 samples. These indexes are multiplied by two and then plugging them into Eq. (8):

$$X = \frac{1}{2}[(t_{A2} - t_A)v_{prop}]$$

$$= \frac{1}{2}[(2 \times 50036 - 2 \times 50105) \times 289942 \times \frac{1}{1000000}] = 20.01 \text{ km}$$

Calculating the estimate error yields according to Eq. (25):

$$\begin{aligned} \text{Error}(\%) &= \left| \frac{L_{\text{estimate}} - L_{\text{actual}}}{L} \right| \times 100 \\ &= \left| \frac{20.01 - 20}{100} \right| \times 100 = 0.01\% \end{aligned}$$

The level-1 detail coefficients of the DWT for the 30km-AB, 40km-ABG, and 50km-ABCG faults for the sampling frequency of 1000 kHz are shown in Figures 12(b), (c), and (d), respectively. The indexes are clearly pointed on these figures to calculate the fault location on the transmission line. The estimate locations of three cases are 30.15 km, 40.01 km, and 49.87 km, respectively. As a result, the estimate errors of three cases are 0.15%, 0.01%, and 0.13%, respectively.

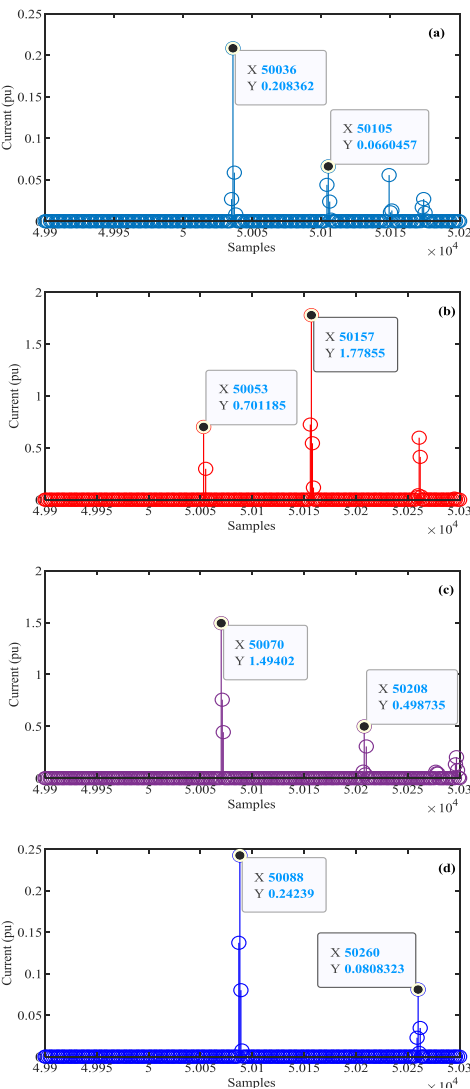


Figure 12. Level-1 detail coefficient, (a) 20km-AG fault, (b) 30km-AB fault, (c) 40km-ABG fault, (d) 50km-ABCG fault.

According to the above theory, the proposed fault location method was applied for many case studies in this paper to evaluate the effectiveness of the proposed method. The fault location was changed along the power transmission line. Additionally, the three wavelets including ‘db4’, ‘coif4’, and ‘sym4’ were applied to decompose the alpha component of the fault current. The estimate errors of the wavelets were compared as shown in Figure 13 for a single-phase-to-ground fault (AG), Figure 14 for a phase-to-phase fault (AB), Figure 15 for a double-phase-to-ground fault (ABG), and Figure 16 for a three-phase-to-ground fault.

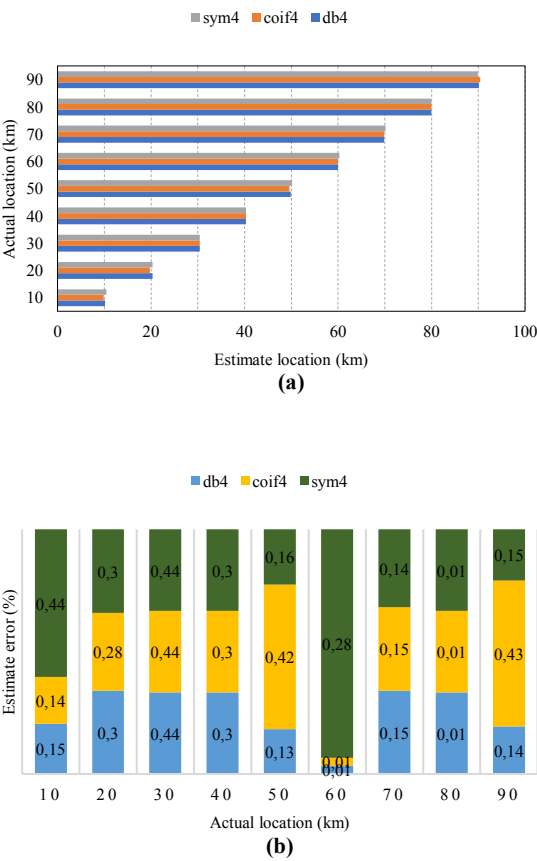


Figure 13. Simulation results of single-phase-to-ground fault, (a) Estimate location, (b) Estimate error.

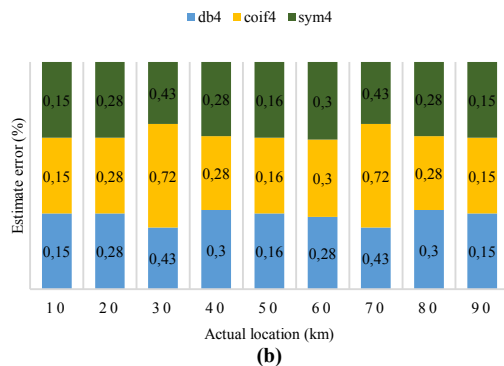
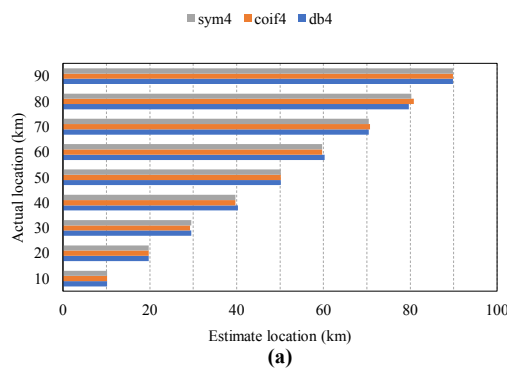


Figure 14. Simulation results of phase-to-phase fault, (a) Estimate location, (b) Estimate error.

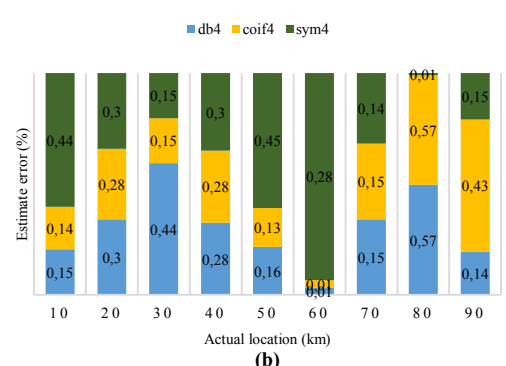
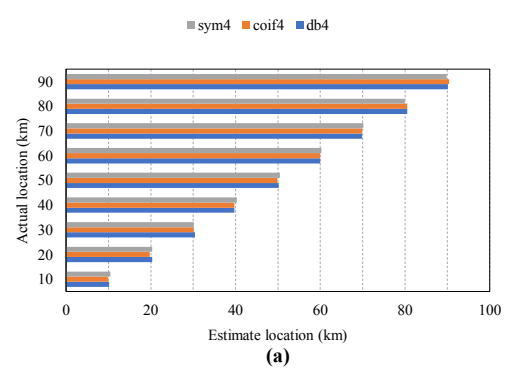


Figure 16. Simulation results of three-phase-to-ground fault, (a) Estimate location, (b) Estimate error.

3.4. Sampling frequency comparison

In this case, the AG fault current waveforms at three locations (20, 40, and 60 km) are sampled at five different sampling frequencies (10 kHz, 100 kHz, 300 kHz, 500 kHz, and 1000 kHz) to evaluate the performance of the proposed fault location method in this paper. These simulation results are presented in Figure 17. As shown in this figure, the dependences of the estimate error according to the sampling frequency which the wavelets ‘db4’, ‘coif4’, and ‘sym4’ are applied to extract the level-1 detail coefficient are shown in Figures 17(a), (b), and (c), respectively. In general, the estimate error will be decreased if the sampling frequency is increased.

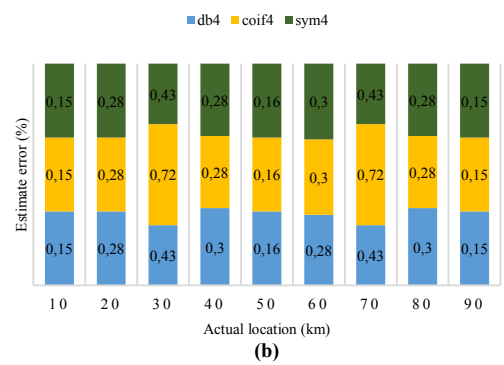
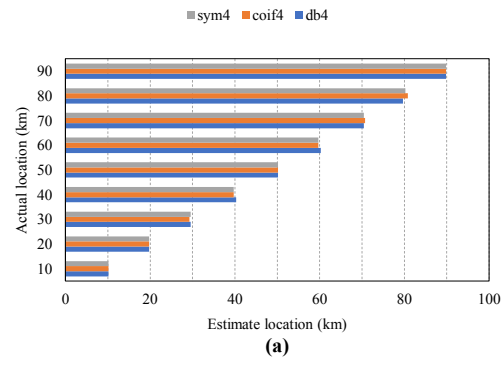


Figure 15. Simulation results of double-phase-to-ground fault, (a) Estimate location, (b) Estimate error.

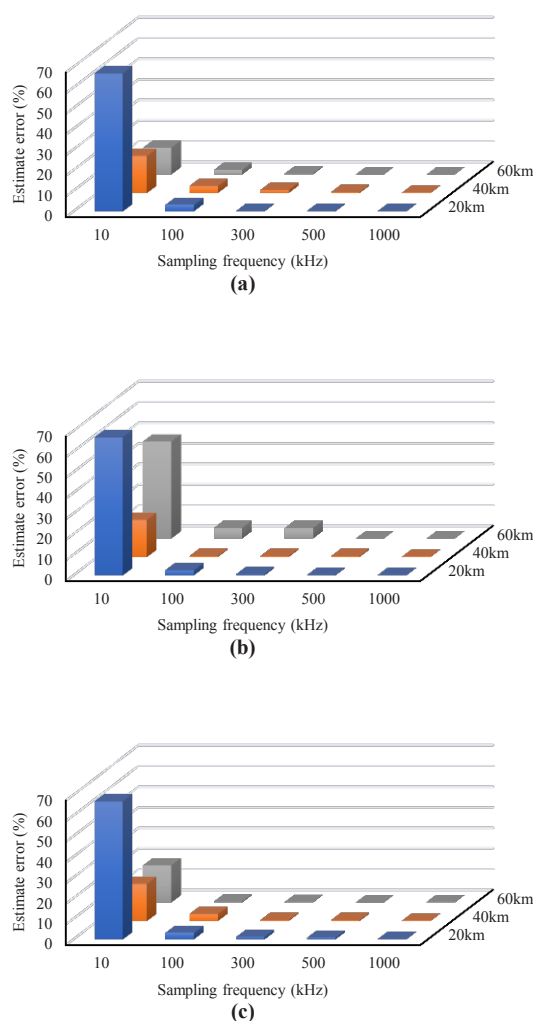


Figure 17. Estimate error according to sampling frequency, (a) wavelet 'db4', (b) wavelet 'coif4', (c) wavelet 'sym4'.

Figure 17 compares the fault location results of the three wavelets using the single-ended method for an AG fault sampled at 10 kHz, 100 kHz, 300 kHz, 500 kHz, and 1000 kHz. The error of the 10 kHz sample frequency goes as high as 75.21% indicating an inadequate sampling frequency for fault location. The sampling frequency requirement to get within 1000 m of the actual fault location begins at 300 kHz. The sampling frequency requirement to get within 100 m of the actual fault location requires at least 500 kHz. The sampling frequency requirement to get with 10 m of actual fault is seen only for 500 kHz and 1000 kHz for faults less than halfway of the total line length. It is obvious that the sampling

frequency of 1000 kHz has the smallest estimate error in this study. Therefore, practical fault location systems are recommended to initialize their sampling frequency of 1000 kHz to achieve a high estimate error.

3.5. Fault resistance comparison

As analyzed in Introduction, the fault resistance will have impact on the estimate error of the impedance-based fault location methods. However, the traveling wave-based fault location methods are not affected by the fault resistance. To demonstrate this issue, five different fault resistances including 1 Ω , 5 Ω , 10 Ω , 50 Ω and 100 Ω are assumed for an AG fault occurring at the locations (30 km, 50 km and 70 km). These simulation results are proved in Figure 18. It can be seen that Figure 18 compares the fault estimate error results of the three wavelets ('db4', 'coif4', 'sym4') using the single-ended method for the AG fault with various fault resistances. As a result, the estimate errors remain unchanged when increasing the fault resistance.

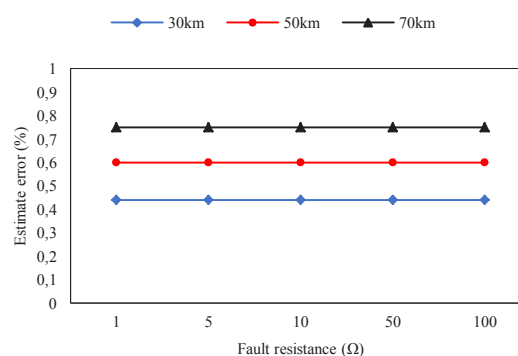


Figure 18. Estimate error according to fault resistance.

4. CONCLUSION

In this paper, a proposed method to locate faults accurately in a power transmission line was presented. The fault location method is based on single-ended traveling wave using the discrete wavelet transform to extract the level-1 detail coefficient from the fault current. The fault current at a single end of the transmission line is recorded and processed by the Clarke transform and discrete wavelet transform to obtain the

traveling wave signals and determine the forward wave propagation and reflected waves generated by the fault.

The proposed method was tested using the test system of a single-circuit transmission line, and it was implemented in MATLAB/Simulink software. Several fault resistances and locations along the line were simulated to generate case studies and to evaluate the effectiveness of the proposed method. The simulation results show that the proposed method produces small errors, around 0.18% using the wavelet 'db4', around 0.24% using the wavelet 'coif4', and around 0.25% using the wavelet 'sym4'. It should be recommended that the traveling wave-based fault location method does not depend on the fault resistance.

The choice of the most suitable sampling frequency for locating faults in transmission lines was also investigated in this paper. It can be obviously observed from the mentioned simulation results that the estimated error results depend on the sampling frequency. Based on the obtained results, it is possible to design a fault location system with a high sampling frequency of 1000 kHz to achieve a high accuracy.

Acknowledgment

This research is conducted within the framework of science and technology projects at institutional level of Quy Nhon University under the project code T2024.845.16.

REFERENCES

1. A. Mukherjee, P. K. Kundu, A. Das. Transmission line faults in power system and the different algorithms for identification, classification and localization: a brief review of methods, *Journal of The Institution of Engineers (India): Series B*, **2021**, 102, 855-877.
2. P. O. K. Anane, Q. Huang, O. Bamisile, P. N. Ayimbire. Fault location in overhead transmission line: a novel non-contact measurement approach for traveling wave-based scheme, *International Journal of Electrical Power and Energy Systems*, **2021**, 133, 107233.
3. A. Prasad, J. B. Edward, K. Ravi. A review on fault classification methodologies in power transmission systems: part-I, *Journal of Electrical Systems and Information Technology*, **2018**, 5(1), 48-60.
4. H. A. A. E. Ghany, A. M. Azmy, A. M. Abeid. A general travelling-wave-based scheme for locating simultaneous faults in transmission lines, *IEEE Transactions on Power Delivery*, **2019**, 35(1), 130-139.
5. S. Das, S. Santoso, A. Gaikwad, M. Patel. Impedance-based fault location in transmission networks: theory and application, *IEEE Access*, **2014**, 2, 537-557.
6. M. M. Hashim, H. W. Ping, V. K. Ramachandramurthy. *Impedance-based fault location techniques for transmission lines*, IEEE Region 10 Conference, Singapore, 2009.
7. N. M. Khoa, M. V. Cuong, H. Q. Cuong, N. T. T. Hieu. Performance comparison of impedance-based fault location methods for transmission line, *International Journal of Electrical and Electronic Engineering and Telecommunications*, **2022**, 11(3), 234-241.
8. J. M. Florez, J. Meléndez, G. C. Caicedo. Comparison of impedance based fault location methods for power distribution systems, *Electric Power Systems Research*, **2008**, 78(4), 657-666.
9. G. Ma, L. Jiang, K. Zhou, G. Xu. A method of line fault location based on traveling wave theory, *International Journal of Control and Automation*, **2016**, 9(2), 261-270.
10. A. M. Elhaffar. *Power transmission line fault location based on current traveling waves*, Doctoral thesis, Helsinki University of Technology, 2008.
11. S. Lin, Z. Y. He, X. P. Li, Q. Q. Qian. Travelling wave time-frequency characteristic-based fault location method for transmission lines, *IET Generation, Transmission and Distribution*, **2012**, 6(8), 764-772.
12. D. Wang, J. Liu, M. Hou. Novel travelling wave fault location approach for overhead transmission lines, *International Journal of Electrical Power and Energy Systems*, **2024**, 155, 109617.

13. S. L. Zimath, C. A. Dutra, R. R. Matos, L. B. de Oliveira, J. H. M. D. Resende, J. A. P. Moutinho. *Traveling wave fault location applied to high impedance events*, The 12th IET International Conference on Developments in Power System Protection, Copenhagen, Denmark, 2014.
14. D. Wang, M. Hou, Y. Guo. Travelling wave fault location of HVAC transmission line based on frequency-dependent characteristic, *IEEE Transactions on Power Delivery*, **2020**, 36(6), 3496-3505.
15. N. N. Bon, L. V. Dai. Fault identification, classification, and location on transmission lines using combined machine learning methods, *International Journal of Engineering and Technology Innovation*, **2022**, 12(2), 91-109.
16. L. V. Dai, N. N. Bon, L. C. Quyen. Deep learning method for fault diagnosis in high voltage transmission lines: a case of the Vietnam 220 kV transmission line, *International Journal on Electrical Engineering and Informatics*, **2022**, 14(2), 254-275.
17. M. Aurangzeb, P. A. Crossley, P. Gale. *Fault location on a transmission line using high frequency travelling waves measured at a single line end*, The 2000 IEEE Power Engineering Society Winter Meeting, Singapore, 2000.
18. O. D. Naidu, A. K. Pradhan. Precise traveling wave-based transmission line fault location method using single-ended data, *IEEE Transactions on Industrial Informatics*, **2020**, 17(8), 5197-5207.
19. F. V. Lopes, K. M. Silva, F. B. Costa, W. L. A. Neves, D. Fernandes. Real-time traveling-wave-based fault location using two-terminal unsynchronized data, *IEEE Transactions on Power Delivery*, **2016**, 30(3), 1067-1076.
20. F. V. Lopes, K. M. Dantas, K. M. Silva, F. B. Costa. Accurate two-terminal transmission line fault location using traveling waves, *IEEE Transactions on Power Delivery*, **2017**, 33(2), 873-880.
21. M. Korkali, H. L. Ari, A. Abur. Traveling-wave-based fault-location technique for transmission grids via wide-area synchronized voltage measurements, *IEEE Transactions on Power Systems*, **2012**, 27(2), 1003-1011.
22. O. D. Naidu, A. K. Pradhan. A traveling wave-based fault location method using unsynchronized current measurements, *IEEE Transactions on Power Delivery*, **2019**, 34(2), 505-513.
23. Q. Jian, C. Xiangxun, Z. Jianchao. *Travelling wave fault location of transmission line using wavelet transform*, International Conference on Power System Technology, Beijing, China, 1998.
24. V. H. G. Sanchez, V. T. García, D. Guillen. Fault location on transmission lines based on travelling waves using correlation and MODWT, *Electric Power Systems Research*, **2021**, 197, 107308.
25. D. T. Viet, N. H. Hieu, N. M. Khoa. A method for monitoring voltage disturbances based on discrete wavelet transform and adaptive linear neural network, *International Review of Electrical Engineering*, **2016**, 11(3), 314-322.
26. L. Zhan, Y. Liu, Y. Liu. A Clarke transformation-based DFT phasor and frequency algorithm for wide frequency range, *IEEE Transactions on Smart Grid*, **2016**, 9(1), 67-77.



© 2024 by the authors. This Open Access Article is licensed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license (<https://creativecommons.org/licenses/by-nc/4.0/>).

Đánh giá đặc điểm nông học và hàm lượng saponin của một số mẫu giống mướp đắng rừng (*Momordica charantia* L. var. *abbreviata* Ser.) thu thập ở khu vực Nam Trung Bộ, Việt Nam

Bùi Hồng Hải*, Nguyễn Thị Y Thanh, Hồ Tân

Khoa Khoa học Tự nhiên, Trường Đại học Quy Nhơn, Việt Nam

Ngày nhận bài: 30/04/2024; Ngày sửa bài: 08/07/2024

Ngày nhận đăng: 15/07/2024; Ngày xuất bản: 28/10/2024

TÓM TẮT

Cây mướp đắng (*Momordica charantia* L.) được trồng phổ biến và sử dụng làm rau, dược liệu nhằm chữa trị tiêu đường, viêm gan, sỏi thận, ung thư, ... do chứa nhiều dược chất như glycosides, alkaloids, triterpenes, đặc biệt hàm lượng saponin cao. Mướp đắng rừng (*M. charantia* L. var. *abbreviata* Ser.) có hàm lượng dược chất và tính chống chịu cao hơn nhiều so với các giống thương mại. Tuy nhiên, công tác thu thập, đánh giá các mẫu giống mướp đắng rừng còn nhiều hạn chế. Nghiên cứu nhằm đánh giá các đặc điểm nông học và định lượng saponin của 12 mẫu giống mướp đắng rừng thu thập ở khu vực Nam Trung Bộ, Việt Nam. Kết quả nghiên cứu bước đầu cho thấy 12 mẫu giống thu thập được có tỉ lệ nảy mầm thấp (10,34 - 44,19 %), thời gian nảy mầm không đồng đều (2 - 20 ngày), thời gian thu quả đợt đầu từ 37 đến 55 ngày sau trồng. Đặc điểm về lá, quả và hạt khác nhau giữa các mẫu giống: kích thước lá (11,90 - 20,23 cm x 9,95 - 16,4 cm), phiến lá xẻ thùy sâu với 5 hoặc 7 thùy; kích thước quả (4,09 - 8,77 cm x 2,05 - 3,5 cm) và năng suất quả tươi (114,84 - 614,045 g quả tươi/cây); kích thước hạt (9,36 - 13,18 mm x 5,0 - 8,0 mm). Hàm lượng saponin tổng số từ 13,26 đến 18,98 mg/g quả khô. BD4 (614,04 g/cây) và QN2 (613,55 g/cây) có triển vọng trong hướng chọn giống về năng suất. Các mẫu giống có hàm lượng saponin cao là BD3 (18,98 mg/g quả khô) và QN3 (16,77 mg/g quả khô) thích hợp với hướng chọn giống tăng hàm lượng saponin.

Từ khóa: Mướp đắng rừng, Nam Trung Bộ, đặc điểm nông học, hàm lượng saponin.

*Tác giả liên hệ chính.

Email: buihonghai@qnu.edu.vn

Evaluation of the agronomical traits and saponin content of wild bitter gourd accessions (*Momordica charantia* L. var. *abbreviata* Ser.) collected from South Central region of Vietnam

Bui Hong Hai*, Nguyen Thi Y Thanh, Ho Tan

Faculty of Natural Sciences, Quy Nhon University, Vietnam

Received: 30/04/2024; Revised: 08/07/2024

Accepted: 15/07/2024; Published: 28/10/2024

ABSTRACT

Bitter gourds (*Momordica charantia* L.) have been widely grown and used as a vegetable and medicinal herb to treat diabetes, hepatitis, kidney (stone), cancer, anti-oxidation... as they contain active pharmaceutical substances like glycosides, alkaloids, triterpenes, steroids, and especially saponins with a high concentration. Wild bitter gourds (*Momordica charantia* L. var. *abbreviata* Ser.) have higher medicinal properties and resistance than their commercial varieties. However, the number of studies evaluating wild bitter gourd varieties was still very limited. This study aimed to evaluate the agronomical traits and quantify the saponin content of 12 wild bitter gourd accessions collected from the South Central region of Vietnam. The research results showed that 12 wild bitter gourd accessions had a low germination rate (10.34 - 44.19%), various germination time (2 - 20 days), and different first fruit harvest time ranging from 37 to 55 days after planting. The characteristics of leaves, fruits, and seeds differed among collected wild bitter gourd accessions: leaf size (11.90 - 20.23 cm x 9.95 - 16.4 cm), petiole length (3.97 - 8.71 cm), deeply lobed leaf blades with 5 or 7 lobes; fruit size (4.09 - 8.77 cm x 2.05 - 3.5 cm), fruit weight (4.41 - 30.09 g.fruit⁻¹), fresh fruit yield (114.84 - 614.045 g fresh fruit.plant⁻¹); and seed size (9.36 - 13.18 mm x 5.0 - 8.0 mm). The total saponin content ranged from 13.26 to 18.98 mg.g⁻¹ dry fruit. The BD4 and QN2 accessions were found to have a high yield (614.04 g.plant⁻¹ and 613.55 g.plant⁻¹, respectively), which should be used for breeding to enhance their yield. The BD3 and QN3 were found to have a high total saponin content (18.98 mg.g⁻¹ dry fruit and 16.77 mg.g⁻¹ dry fruit, respectively) which should be suitable for breeding research to increase saponin properties.

Keywords: *Wild bitter gourd, South Central region, agronomical traits, saponin content.*

1. INTRODUCTION

Bitter gourds (*Momordica charantia* L.) or bitter melons which belong to Cucurbitaceae, have been widely grown in Vietnam and other countries in Asia, East Africa, Amazon, and the Caribbean countries.^{1,2} Besides being used as a vegetable, bitter gourds are also used as a

herbal medicine to treat diabetes, kidney (stone), eczema, malaria, hepatitis, gout, laxatives, AIDS, cancer...³⁻⁵ Bitter gourds contain active pharmaceutical substances including glycosides, saponins, alkaloids, triterpenes, and steroids.^{4,6} The most important pharmaceutical components are saponin and monordicin.⁴ Saponin is a

*Corresponding authors.

Email: buihonghai@qnu.edu.vn

compound containing charantin, which acts as an insulin having a safe effect on lowering blood sugar. Momordicin is a mixture of charantin, protein, adenine, betanin, vitamins B, C, etc., which can kill bacteria and viruses, effectively inhibiting cancer cells,^{4,5} preventing DNA damage caused by H₂O₂ inhibiting the activity of enzyme α -amylase, fighting inflammatory,⁷ improving the antioxidant activity of enzymes,⁸ and supporting the control of disorders blood lipids and oxidative stress in type 2 diabetic patients.⁹

Related research results showed that wild bitter gourd varieties (*Momordica charantia* L. var. *abbreviata* Ser.) had higher pharmacological activities than their commercial varieties.^{10,11} More especially, these wild varieties also had better resistance to bacteria, to cytotoxic activities,^{12,13} and to fruit flies (*Bactrocer cucurbitae*) than cultivated bitter gourd varieties.¹¹ However, recent research just focused on expanding wild bitter melon production due to market demand and their potential to increase

income for producers. The number of studies on growth and agronomical characteristics of wild bitter gourd varieties was still limited in Vietnam. In recent research, 09 wild bitter melon accessions/varieties were collected in Binh Phuoc province and evaluated, including the AB5 variety with high yield potential and the BT2 variety with high saponin content.¹⁴ The genetic diversity of 12 wild bitter melon varieties in many different regions was also evaluated and analyzed.¹⁵ In the South Central region of Vietnam, there are diverse natural conditions that promise an abundance of wild bitter melon accessions. Therefore, this research aimed to collect, analyze, and evaluate wild bitter melon accession, thereby providing potential data for later breeding research.

2. METHODS

2.1. Materials

Twelve wild bitter gourd accessions collected from the South Central region were presented in Table 1.

Table 1. List of wild bitter gourd accessions.

No.	Accession Code	Place of collection	Number of used seeds
1	QN1	An My ward, Tam Ky city, Quang Nam province	116
2	QN2	Dien Phuong commune, Dien Ban district, Quang Nam province	108
3	QN3	Nui Thanh town, Nui Thanh district, Quang Nam province	86
4	QNg1	Tinh Long commune, Quang Ngai city, Quang Ngai province	108
5	QNg2	Nghia Ha commune, Tu Nghia district, Quang Ngai province	96
6	QNg3	Binh Phuoc commune, Binh Son district, Quang Ngai province	58
7	BD1	Tay Giang commune, Tay Son district, Binh Dinh province	62
8	BD2	Cat Minh commune, Phu Cat district, Binh Dinh province	96
9	BD3	My Chau commune, Phu My district, Binh Dinh province	92
10	BD4	Nhon Phu ward, Quy Nhon city, Binh Dinh province	48
11	PY1	An Nghiep commune, Tuy An district, Phu Yen province	50
12	BT1	Thuan Hoa commune, Ham Thuan Bac district, Binh Thuan province	90

2.2. Methods

2.2.1. Experimental design

The experiment was conducted from September 2023 to February 2024 in the net house at the Applied Research Garden for Agriculture and

Environment, Quy Nhon University, Binh Dinh, Vietnam. Experimental conditions during the study period included average temperature (28.15 °C), relative humidity (80%), average sunshine hours per month (140 hours.month⁻¹), and average monthly rainfall (334 mm.month⁻¹).

The experiment was laid out as a completely randomized design with 12 plots of 5 m² (5 m x 1 m). Each plot planted 10 plants of a wild bitter gourd accession, with two plants 50 cm apart and two rows 80 cm apart. The total experimental area is 60 m² (10 m x 6 m) not including protection paths and isolation distances.

2.2.2. Cultivation process

The experimental procedure was according to QCVN 01-153:2014/BNNPTNT,¹⁶ the wild bitter gourd plants were grown in black nylon bags (20 cm x 40 cm). The planting substrate includes 50% treated coconut fiber, 20% smoked rice husk, 20% soil, and 10% Song Gianh microbial organic fertilizer (v:v). The nutrient solution was mixed from 10 g of GATIT NPK soluble fertilizer 17-6-17+TE (<https://www.gatvn.com/products/gatit-npk-17-6-17-te>) in 30 liters of water and watered by using a drip system. The amount of nutrient solution was adjusted to adapt to each stage of the plant growth.

The seeds were washed, soaked in warm water for 2 hours, and incubated in a damp cloth until germination. Germinated seeds were then planted in a plastic tray containing Tribat clean soil, and kept moist by periodic watering. 13 days after sowing, each seedling with 3 true leaves was transplanted in a black nylon bag.

2.3.3. Indicators and measurement

The germination rate (%) is percentage of the number of germinated seeds compared to the total used seeds which ranged from 48 to 116 seeds (Table 1).

Agronomical traits were evaluated according to QCVN 01-153:2014/BNNPTNT¹⁶ including growth stages (the time point is when 50% of individuals reveal the traits). Accordingly, the leaf length (cm), leaf width (cm), petiole length (cm), fruit length (cm), fruit diameter (cm), fruit weight (g), seed length (mm), seed width (mm), seed thickness (mm), and seed mass (mg) were determined through 15

samples. The number of fruits per plant (fruits) and fresh fruit yield (g.plant⁻¹) were measured in 30 days starting from the first fruit harvest date. The fruits were harvested once they fully matured when the middle fruits and warts were plumped. The total saponin content was determined by using spectrophotometry at 560 nm with UV-Vis 1800i equipment (Shimadzu, Japan).¹⁷ Dry fruit sample was ultrasonically extracted with ethanol 80%. The filtrate was distilled to recover the solvent under reduced pressure to get the extract which was dissolved in distilled water and extracted with n-butanol. The extract mixture was then removed n-butanol to get extract which was dissolved with ethanol 80% and made a color-forming reaction with vanilin 8% and H₂SO₄ 72% at 60°C for 15 minutes. The absorption of the mixture was measured at 560 nm. The reference was momordicoside - G dissolved in ethanol 80% at 35, 75, 150, 300, 600, và 800 µg.ml⁻¹. The dry weight of fruit (%) was determined by weighing and drying at 55 °C for 36 hours.

The research data were analyzed by using MS. Excel 2010 and Statistix (ver. 8.0). The experimental results were statistically analyzed by one-way ANOVA analysis with LSD (least significant difference) test for multiple comparisons with a significance level of 0.05.

3. RESULTS AND DISCUSSION

3.1. Growth stages and yield

The results of the growth stages were shown in Table 2. Research data indicated that the collected wild bitter gourd accessions had a low germination rate which ranged from 10.34% (QN1) to 44.19% (QN3) and the germination time was not consistent among the accessions, ranging from 2 days (QN3 and PY1) to 20 days (QN2). Male flowering time was usually earlier than the female counterpart and this indicator also differed greatly among the studied accessions, ranging from 27 days after planting (DAP) with BT1 to 40 DAP with PY1. Female flowering time also varied, from 20 DAP with

BT1 to 41 DAP with QN2. Interestingly, some studied wild bitter gourd accessions had their male flowers appeared before their female flowers such as BT1 (7 days in advance) and PY1 (2 days in advance), while BD1 had their male and female flowers appeared on the same

day. The first fruit harvest time also witnessed a difference among the accessions, ranging from 37 DAP (BT1) to 55 DAP (QN2), corresponding after the appearance of female flowers from 9 days (BD1, QNg1) to 15 days (QNg3, BD4, and PY1).

Table 2. Growth stages and yield of wild better gourd accessions.

Accession Code	Germination rate (%)	Germination time (day)	Male flowering time (DAP)	Female flowering time (DAP)	First harvest time (DAP)	Num. of fruit per plant (fruit)	Yield (g)
QN1	10.34	18	34	39	49	24.7 ^{de}	261.82 ^d
QN2	11.11	20	35	41	55	35.1 ^b	613.55 ^a
QN3	44.19	2	31	36	49	28.3 ^c	283.00 ^d
QNg1	11.11	13	38	45	54	39.7 ^a	457.74 ^c
QNg2	18.75	13	30	37	51	36.9 ^{ab}	453.87 ^c
QNg3	34.48	3	31	35	50	24.8 ^{de}	248.00 ^{de}
BD1	38.71	6	39	39	48	22.3 ^{ef}	532.97 ^b
BD2	12.50	16	31	36	48	38.6 ^a	582.86 ^a
BD3	13.04	13	24	34	47	38.2 ^{ab}	255.94 ^d
BD4	25.00	6	25	29	44	20.4 ^{fg}	614.04 ^a
PY1	44.00	2	40	38	53	17.8 ^g	206.48 ^e
BT1	37.78	12	27	20	37	26.1 ^{cd}	114.84 ^f
<i>LSD</i> _{0.05}	-	-	-	-	-	3.34	44.94
<i>CV</i> (%)	-	-	-	-	-	12.79	13.15

Note: Mean followed by the same letter is not significantly different within the columns according to an *LSD* test at the significance level of 0.05; *CV* - coefficient variance; *LSD*: least significant difference; *DAP*: number of days after planting.

After 30 days of harvesting, the number of fruits per plant of wild bitter melon accessions ranged from 17 fruits.plant⁻¹ (PY1) to 39.1 fruits.plant⁻¹ (QNg1), and the fresh yield ranged from 114.84 to 614.04 g.plant⁻¹. The accessions with high fresh fruit yield were BD4 (614.04 g.plant⁻¹), QN2 (613.55 g.plant⁻¹), BD2 (582.86 g.plant⁻¹), and BD1 (532.97 g.plant⁻¹). The accessions with low fresh fruit yield were BT1 (114.84 g.plant⁻¹), and PY1 (206.48 g.plant⁻¹). Although the number of fruits per plant was lower, the fresh fruit yield of the wild bitter melon accessions was equivalent to the yield of some varieties

collected in Binh Phuoc during the same harvest time, such as BT4 (500 g.plant⁻¹) and HC2 (610 g.plant⁻¹).¹⁴

3.2. Leaf morphological traits

Leaf is an important plant organ that performs photosynthesis to create organic matter, thereby generating energy for all living activities. Leaf size shows the photosynthetic ability of a plant as well as its yield potential.¹⁸ The characteristics of leaf blade size, petiole length, and leaf descriptions of wild bitter gourd accessions were shown in Table 3 and Figure 1.

Table 3. Leaf morphological traits of wild better gourd accessions.

Accession Code	Leaf length (cm)	Leaf width (cm)	Petiole length (cm)	Leaf descriptions
QN1	12.63 ^{fg}	9.95 ^g	4.37 ^h	Small leaf blade size, light green, deeply lobed leaf blade with 5 lobes, heart-shaped at the base of leaf blade, and slightly pointed at the tip of a leaf lobe.
QN2	14.53 ^e	12.13 ^d	6.03 ^e	Medium leaf blade size, dark green, deeply lobed leaf blade with 7 lobes, heart-shaped at the base of leaf blade, and pointed at the tip of a leaf lobe.
QN3	17.47 ^b	15.30 ^b	4.73 ^g	Medium leaf blade size, light green, deeply lobed leaf blade with 5 lobes, heart-shaped at the base of leaf blade, and slightly pointed at the tip of a leaf lobe.
QNg1	14.87 ^e	10.97 ^{ef}	5.10 ^f	Small leaf blade size, light green, deeply lobed leaf blade with 5 lobes, heart-shaped at the base of leaf blade, and slightly pointed at the tip of leaf lobes.
QNg2	15.97 ^d	14.17 ^c	4.45 ^{gh}	Medium leaf blade size, medium green, deeply lobed leaf blade with 7 lobes, heart-shaped at the base of leaf blade, and pointed at the tip of leaf lobes.
QNg3	13.40 ^f	11.53 ^{de}	5.19 ^f	Medium leaf blade size, medium green, deeply lobed leaf blade with 7 lobes, heart-shaped at the base of leaf blade, and pointed at the tip of leaf lobes.
BD1	17.23 ^b	14.03 ^c	8.02 ^b	Large leaf blade size, medium green, deeply lobed leaf blade with 7 lobes, heart-shaped at the base of leaf blade, and pointed at the tip of leaf lobes.
BD2	17.19 ^{bc}	11.98 ^d	6.98 ^c	Small leaf blade size, light green, deeply lobed leaf blade with 7 lobes, heart-shaped at the base of leaf blade, and pointed at the tip of leaf lobes.
BD3	12.70 ^{fg}	10.97 ^{ef}	6.48 ^d	Small leaf blade size, dark green, deeply lobed leaf blade with 5 lobes, heart-shaped at the base of leaf blade, and slightly obtuse at the tip of leaf lobes.
BD4	20.23 ^a	16.40 ^a	8.71 ^a	Large leaf blade size, light green, deeply lobed leaf blade with 7 lobes, heart-shaped at the base of leaf blade, and pointed at the tip of leaf lobes.
PY1	16.30 ^{cd}	13.53 ^c	8.11 ^b	Medium leaf blade size, light green, deeply lobed leaf blade with 5 lobes, heart-shaped at the base of leaf blade, and slightly pointed at the tip of leaf lobes.
BT1	11.90 ^g	10.40 ^{fg}	3.97 ⁱ	Small leaf blade size, light green, deeply lobed leaf blade with 5 lobes, heart-shaped at the base of leaf blade, and slightly pointed at the tip of leaf lobes.
<i>LSD</i> _{0.05}	0.92	0.81	0.35	-
<i>CV</i> (%)	8.31	8.90	8.11	-

The data in Table 3 showed that the wild bitter gourd accessions had a variety in terms of leaf lengths from 11.90 cm to 20.23 cm and leaf blade widths from 9.95 cm to 16.4 cm. The accessions with small leaf sizes were BT1 (11.9 cm x 10.4 cm), QN1 (12.63 cm x 9.95 cm), and BD3 (12.7 cm x 10.97 cm). The accessions with large leaf sizes are BD4 (20.23 cm x 16.4 cm), QN3 (17.47 cm x 15.3 cm), and BD1 (17.23 cm x 14.03 cm). The petiole length also varied among the accessions. Specifically, BT1 was recorded as the shortest with 3.97 cm and BD4 was the longest with 8.71 cm. Interestingly, the leaf size of the studied wild bitter gourd accessions was equivalent to its counterpart in the study of Bui Thi Xuan (12.1 - 24.2 cm x 9.1 - 16.7cm).¹⁵

The leaf blades of all wild bitter gourd accessions were heart-shaped at the leaf bases. The tip of leaf lobes was slightly pointed, clearly pointed, or slightly obtuse. The leaf blades were deeply lobed with 5 or 7 lobes. The accessions that had the leaf blade with 5 deep lobes were QN1, QN3, QNg1, BD3, PY1, and BT1. The ones that had 7 deep lobes per leaf blade were QN2, QNg2, QNg3, BD1, BD2, and BD4 (Figure 1). The leaf morphological characteristics of the wild bitter gourd accessions were similar to those found in the study of Bui Thi Xuan.¹⁵

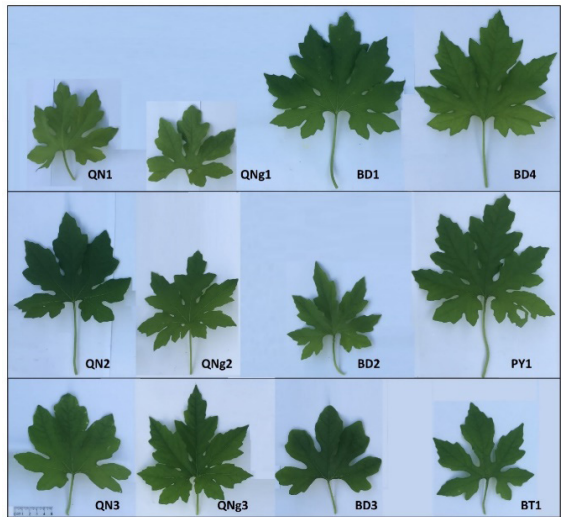


Figure 1. Leaf morphology of the wild bitter gourd accessions.

3.3. Fruit morphological traits

The traits of fruits including fruit length, fruit diameter, cross-section shape, color of skin, and wart size were used to classify group varieties in breeding testing.¹⁶ The fruit traits of studied wild bitter gourd accessions were shown in Table 4 and Figure 2.

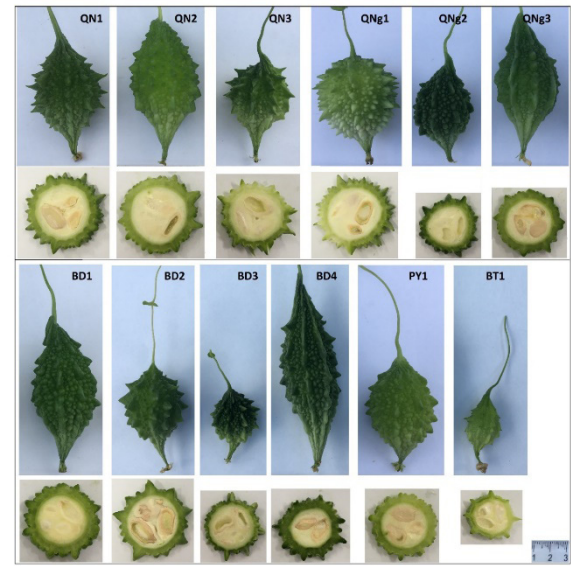


Figure 2. Shape and cross-section of the fruit of wild bitter gourd accessions.

Data in Table 4 showed that the studied fruit length ranged from 4.09 cm to 8.77 cm, the fruit diameter ranged from 2.05 cm to 3.5 cm, and the fruit weight ranged from 4.41 g.fruit⁻¹ to 30.09 g.fruit⁻¹. The wild bitter gourd accessions with large fruit sizes were BD4 (8.77 cm x 3.53 cm, 30.09 g.fruit⁻¹), and BD1 (7.19 cm x 3.24 cm, 23.87 g.fruit⁻¹). The accessions with small fruit sizes were BT1 (4.09 cm x 2.05 cm, 4.41 g.fruit⁻¹), and BD3 (4.15 cm x 2.99 cm, 6.74 g.fruit⁻¹). The fruit size of the wild bitter gourd accessions in this study was smaller than the fruit size of the wild bitter melon accessions studied by Bui Thi Xuan (5.5 - 14.9 cm x 2.9 - 6.7 cm)¹⁵ but it was greater than the wild bitter melon varieties studied by Phan Dang Thai Phuong (2.79 - 4.38 cm x 1.54 - 1.82 cm).¹⁴ This difference might be due to the differences in genotype among accessions collected, as well as the different time and experimental conditions of the studies.

Other fruit morphological traits including longitudinal shape, skin color, and wart size varied among the studied wild bitter gourd accessions. In terms of longitudinal shapes of fruits, some accessions were rhomb (QN1, BD2, and PY1), some were oval (QN2, QNg3, BD1, BD4, and BT1), and the others were ovoid (QN3, QNg1, QNg2, and BD3). Regarding the color of skin, some accessions were dark green (QNg2 and BD4), some were medium green (QN1, BD1, BD2, and BD3), and others were light green (QN2, QN3, QNg1, QNg3, PY1, and BT1). The fruit surfaces also varied among the wild bitter gourd accessions in terms of wart sizes, the shape of top wart, and ridges. Some had acute warts (QN1, QN3, QNg1, BD2, BD3, BT1), obtuse warts (QN2, QNg2, QNg3, BD1, BD4, PY1), and ridges include (QNg3, BD1, and BD4).

Table 4. Fruit morphological traits of wild bitter gourd accessions.

Accession Code	Fruit length (cm)	Fruit diameter (cm)	Fruit mass (g)	Fruit descriptions
QN1	5.32 ^{de}	2.97 ^{cd}	10.62 ^e	Fruits with rhombic shape, acute at base and apex, medium green, surface with medium-large acute warts and many small acute warts.
QN2	6.55 ^c	3.07 ^{bc}	17.48 ^c	Fruits with oval shape, acute at base and apex, white green, surface with many medium acute warts.
QN3	4.88 ^e	2.93 ^{cde}	10.05 ^{ef}	Fruits with avoid shape, acute at base and apex, white green, surface with few large acute warts and many small obtuse warts.
QNg1	5.53 ^d	3.03 ^{bc}	11.53 ^e	Fruits with avoid shape, obtuse at base and acute at apex, white green, surface with few large acute warts and many small acute warts.
QNg2	5.35 ^{de}	3.15 ^{bc}	12.32 ^{de}	Fruits with avoid shape, acute at base and apex, dark green, surface with few large acute warts and many small obtuse warts.
QNg3	4.88 ^e	3.14 ^{bc}	10.05 ^{ef}	Fruits with oblong oval shape, acute at base and apex, white green, surface with many small obtuse warts and long ridges.
BD1	7.19 ^b	3.24 ^{7b}	23.87 ^b	Fruits with oval shape, acute at base and apex, medium green, surface with many small obtuse warts and few short ridges.
BD2	5.57 ^d	3.50 ^a	15.09 ^{cd}	Fruits with rhombic shape, acute at base and apex, medium green, surface with medium-large acute warts and medium-small acute warts.
BD3	4.15 ^f	2.99 ^{ed}	6.74 ^{fg}	Fruits with avoid shape, acute at base and apex, medium green, surface with medium-large acute warts and many small acute warts.
BD4	8.77 ^a	3.53 ^a	30.09 ^a	Fruits with oblong oval shape, acute at base and apex, dark green, surface with many small obtuse warts and long ridges.
PY1	5.33 ^{de}	2.73 ^e	11.65 ^e	Fruits with rhombic shape, acute at base and apex, white green, surface with many small rounded warts and few obtuse warts.
BT1	4.09 ^f	2.05 ^f	4.41 ^g	Fruits with oval shape, acute at base and apex, white green, surface with few large acute warts and many small obtuse warts.
<i>LSD</i> _{0,05}	5.82	2.33	3.33	-
<i>CV</i> (%)	13.96	10.74	32.21	-

3.4. Seed morphological traits

The studied wild bitter gourd accessions also had significant differences in their seed size as shown in Table 4 and Figure 3. In detail, seed length ranged from 9.36 mm (BT1) to 13.18 mm (BD2), seed width ranged from 5.0 mm (QNg2) to 8.0 mm (BD4), grain thickness ranged from 2.61 mm (BD5) to 3.97 mm (BD2), and grain

weight ranged from 71.64 mg.seed⁻¹ (BT1) to 180.64 mg.seed⁻¹ (BD4). The accessions with small seeds were BT1, QN1, QN3, QNg1, and QNg2. The accession with large seeds include QN2, QNg3, BD1, BD2, BD3, BD4, and PY1. The seed size of the accessions in this study was similar to the one in the study of Bui Thi Xuan (9.1 - 15.6 mm x 5.0 - 9.4 mm x 3.0 - 4.8 mm).¹⁵

Table 5. Seed morphological traits of the wild bitter gourd accessions.

Accession Code	Seed length (mm)	Seed width (mm)	Seed thickness (mm)	Seed mass (mg)	Seed descriptions
QN1	11.68 ^{cd}	6.32 ^e	3.21 ^{cd}	99.91 ^f	Oval shape, dark gray, smooth surface with or without black spots in the middle, and small edges.
QN2	12.09 ^{bc}	7.45 ^b	3.79 ^{ab}	140.36 ^{cd}	Oval shape, light brown, lumpy surface with stripes, an indentation in middle, and small edges.
QN3	11.50 ^d	6.09 ^e	3.40 ^c	101.09 ^f	Oval shape, light gray, smooth surface with or without black spots, and very small edges.
QNg1	12.36 ^b	6.73 ^d	2.95 ^e	134.09 ^{de}	Oval shape, dark gray, smooth surface with unclear stripes, and small edges.
QNg2	10.55 ^e	5.00 ^f	3.03 ^{de}	81.73 ^g	Long oval shape, light brown, lumpy surface with clear stripes, and very small edges.
QNg3	12.09 ^{bc}	7.09 ^{bcd}	3.84 ^{ab}	149.73 ^{bc}	Oval shape, dark brown, lumpy surface with clear stripes, and small edges.
BD1	12.09 ^{bc}	7.86 ^a	3.94 ^{ab}	154.36 ^b	Oval shape, dark brown, smooth surface with unclear stripes, and small edges.
BD2	13.18 ^a	7.86 ^a	3.97 ^a	175.09 ^a	Oval shape, light brown, lumpy surface with clear stripes and indentation in middle, and small edges.
BD3	12.23 ^b	7.00 ^{cd}	3.73 ^b	155.36 ^b	Oval shape, light brown, smooth surface with few stripes, and small edges.
BD4	13.00 ^a	8.00 ^a	3.82 ^{ab}	180.64 ^a	Oval shape, dark brown, lumpy surface with clear stripes and dark brown indentation in middle, and small edges.
PY1	11.27 ^d	7.27 ^{bc}	3.09 ^{de}	124.18 ^e	Oval shape, light brown, lumpy surface with clear stripes and dark brown indentation in the middle, and small edges.
BT1	9.36 ^f	5.23 ^f	2.61 ^f	71.64 ^g	Oval shape, light gray, smooth surface with or without dark gray spots, and very small edges.
<i>LSD</i> _{0,05}	0.51	0.38	0.24	13.29	
<i>CV</i> (%)	5.15	6.66	8.36	12.05	

The seed surface and color of testa also varied among accessions. The accessions with gray testa included QN1, QN3, QNg1 and BT1. The accessions with brown testa included QN1, QNg2, QNg3, BD1, BD2, BD3, BD4, and PY1. The seed surface with clear stripes was found in QN2, QNg2, QNg3, BD2, BD4, and PY1 accessions.



Figure 3. Seeds of wild bitter melon accessions.

3.5. Saponin content and dry weight of fruit

Saponin is the most important medicinal substance in bitter gourds which was found to have a great beneficial impact on health. It could lower blood sugar,⁹ kill bacteria and viruses, effectively inhibit cancer cells,^{4,5} limit DNA damage caused by H_2O_2 , fight inflammation,⁷ and resist oxidation.⁸ Quantifying the saponin content was believed to facilitate the evaluation of potential medicinal values of bitter melon accessions in this study, which was shown specifically in Figure 4.

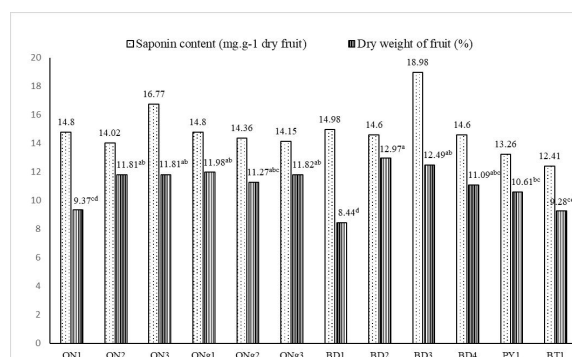


Figure 4. Total saponin content and dry weight of fruit of wild bitter melon accessions.

The total saponin content in dried fruit of wild bitter melon accessions was relatively uniform among accessions, ranging from 13.26 to 18.98 mg.g⁻¹ dry fruit. The two accessions with highest total saponin content were BD3 (18.98 mg.g⁻¹ dry fruit) and QN3 (16.77 mg.g⁻¹ dry fruit).

The dry weight of fruit ranged from 8.44% to 12.97%. The accessions with high fruit dry matter content included BD2 (12.97%) and BD3 (12.49%). The accessions with low fruit dry matter content included BD1 (8.44%), BT1 (9.28%), and QN1 (9.37%).

4. CONCLUSION

The collected 12 wild bitter melon accessions had a low germination rate (from 10.34% to 44.19%), various germination time (from 2 to 20 days), and different first harvest time (from 37 to 55 days after planting). The traits of leaves, fruits, and seeds differed among wild bitter melon accessions. In terms of leaf characteristics, leaf blade size ranged from 11.90 - 20.23 cm x 9.95 - 16.4 cm, petiole length ranged from 3.97 cm to 8.71 cm, number of deep lobes per leaf blade was 5 or 7 lobes; fruit size ranged from 4.09 - 8.77 cm x 2.05 - 3.5 cm, fruit weight ranged from 4.41 to 30.09 g.fruit⁻¹, and fresh fruit yield ranged from 114.84 to 614.045 g.plant⁻¹. Regarding seeds, seed size ranged from 9.36 - 13.18 mm x 5.0 - 8.0 mm. The total saponin content in the studied accessions ranged from 13.26 to 18.98 mg.g⁻¹ dry fruit, and the dry weight of fruit ranged from 8.44 to 12.97%.

Results of the study revealed that the accessions having the highest yield were BD4 (614.04 g.plant⁻¹) and QN2 (613.55 g.plant⁻¹) which should be followed by breeding selection for high yield experiments and studies. The accessions having high saponin content were BD3 (18.98 mg.g⁻¹ dry fruit) and QN3 (16.77 mg.g⁻¹ dry fruit) which should be suitable for breeding to enhance total saponin content.

Acknowledgment

This research is conducted within the framework of a science and technology project at the institutional level of Quy Nhon University under the project code T2023.799.09.

REFERENCES

1. N. Ahmad, N. Hasan, Z. Ahmad, M. Zishan, S. Zohrameena. *Momordica charantia*: for traditional uses and pharmacological actions, *Journal of Drug Delivery and Therapeutics*, **2016**, 6(2), 40-44.
2. R. M. Reyes. *Current status of bitter gourd production and marketing in Asia and market potential of bitter gourd as a functional vegetable*, International Bitter Gourd Conference, BiG 2014, Hyderabad, 2014.
3. S. Y. Arafat, M. Nayeem, S. Jahan, Z. Karim, H. M. Reza, M. H. Hossain, M. Shohel, M. A. Alam. Ellagic acid-rich *Momordica charantia* fruit pulp supplementation prevented oxidative stress, fibrosis and inflammation in the liver of alloxan-induced diabetic rats, *Oriental Pharmacy and Experimental Medicine*, **2016**, 16(4), 267-278.
4. J. K. Grover, S. P. Yadav. Pharmacological action and potential uses of *Momordica charantia*: a review, *Journal of Ethnopharmacology*, **2004**, 93(1), 123-132.
5. W. J. Ru. A triterpenoid from wild bitter gourd inhibits breast cancer cells, *Scientific Reports*, **2016**, 6, 22419.
6. A. Raman, C. Lau. Anti-diabetic properties and phytochemistry of *Momordica charantia* L. (Cucurbitaceae), *Phytomedicine*, **1996**, 2, 349-362.
7. T. M. H. Pham, D. H. Ngo, D. N. Ngo, T. S. Vo. Investigation of biological activities of wild bitter melon (*Momordica charantia* L. var. *abbreviata* Ser.), *Biomolecules*, **2019**, 9(6), 211.
8. A. T. Sagor, M. R. Chowdhury, N. Tabassum, H. Hossain, M. M. Rahman, M. A. Alam. Supplementation of fresh ucche (*Momordica charantia* L. var. *muricata* Willd) prevented oxidative stress, fibrosis and hepatic damage in CCl₄ treated rats, *BMC Complementary and Alternative Medicine*, **2015**, 115, 1-9.
9. S. Poovitha, M. Parani. Protein extracts from *Momordica charantia* var. *charantia* and *M. charantia* var. *muricata* show anti-lipidemic and antioxidant properties in experimental type 2 diabetic rats, *Journal of Food Biochemistry*, **2020**, 44(9), e13370.
10. Y. L. Lu, Y. H. Liu, J. H. Chyuan, K. L. Cheng, W. L. Liang, W. C. Hou. Antioxidant activities of different wild bitter gourd (*Momordica charantia* L. var. *abbreviata* Seringe) cultivars, *Botanical Studies*, **2012**, 53, 207-214.
11. M. K. Dhillon, J. S. Nareesh, R. Singh, N. K. Sharma. Reaction of different bitter gourd (*Momordica charantia* L.) genotypes to melon fruit fly, *Bactrocera cucurbitae* (Coquillett), *Indian Journal of Plant Protection*, **2005**, 33(1), 55-59.
12. Y. L. Lu, Y. H. Liu, W. L. Liang, J. H. Chuang, K. T. Cheng, H. J. Liang, W. C. Hou. Antibacterial and cytotoxic activities of different wild bitter gourd cultivars (*Momordica charantia* L. var. *abbreviata* Seringe), *Botanical Studies*, **2011**, 52, 427-434.
13. N. P. Dam, V. M. Dien, L. H. V. Thanh, P. T. T. Hien, N. T. T. Tram. Investigation of antimicrobial activity and chemical constituents of *Momordica charantia* L. var. *abbreviata* Ser., *Vietnam Journal of Science and Technology*, **2019**, 57(2), 155-161.
14. D. T. P. Phan, T. D. T. Tran, V. B. Vu. Initial assessment of yields and medicinal value of some wild bitter melon varieties in Binh Phuoc (in Vietnamese), *Vietnam Journal of Science and Technology*, **2017**, 13(2), 13-17.
15. T. X. Bui, T. L. Tran, M. T. To, T. K. D. Tran, Q. T. Nguyen, V. T. Nguyen. Evaluation of agromorphological characterisation and genetic diversity of wild bitter gourd (*Momordica charantia* L. var. *abbreviata* Ser.) accessions, *Journal of Agriculture and Rural Development*, **2022**, 1(4), 12-23.
16. Department of Agriculture and Rural Development. *QCVN 01-153:2014/BNNPTNT* - <https://doi.org/10.52111/qnjs.2024.18504>

National technical regulation on testing for distinctness, uniformity and stability of bitter gourd varieties, Hanoi, Vietnam, 2014.

17. S. P. Tan, Q. V. Vuong, C. E. Stathopoulos, S. E. Parks, P. D. Roach. Optimized aqueous extraction of saponin from bitter melon for production of a saponin-enriched bitter melon powder, *Journal of Food Science*, **2014**, 79(7), 1372-1381.
18. T. Lincoln, Z. Eduardo. *Plant physiology and development* (6th edition), Sinauer Associate Inc., Sunderland, Massachusetts, USA, 2015.



© 2024 by the authors. This Open Access Article is licensed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license (<https://creativecommons.org/licenses/by-nc/4.0/>).

Nghiên cứu, chế tạo hệ thống đóng cắt và giám sát thông số hệ thống điện dựa trên xử lý ảnh và IoT

Lương Ngọc Toàn*, Đoàn Đức Tùng, Lê Tuấn Hộ

Khoa Kỹ thuật và Công nghệ, Trường Đại học Quy Nhơn, Việt Nam

Ngày nhận bài: 03/05/2024; Ngày sửa bài: 11/07/2024

Ngày nhận đăng: 15/07/2024; Ngày xuất bản: 28/10/2024

TÓM TẮT

Trong bài báo này chúng tôi đã đi vào nghiên cứu, chế tạo hệ thống đo lường, giám sát các thông số trong hệ thống điện dựa trên công nghệ IoT, và đóng cắt thiết bị trong phạm vi rộng dựa trên công nghệ xử lý ảnh để phát hiện đối tượng. Phần cứng của hệ thống được chế tạo với bộ thiết bị nhận diện đối tượng, UART PZEM-004T, esp32, Aptomat thông minh để thu thập các thông số dòng điện, điện áp, công suất, hệ số công suất, điện năng tiêu thụ và nhận diện đối tượng trong phạm vi giám sát. Các dữ liệu này sẽ được gửi lên server Node-Red chạy ở chế độ cục bộ hoặc Cloud thông qua giao thức Message Queuing Telemetry Transport (MQTT). Sau khi chế tạo hệ thống vận hành ổn định, có khả năng tự động đóng cắt thiết bị khi phát hiện đối tượng, người sử dụng có thể theo dõi trạng thái, điều khiển và cài đặt các thông số thông qua giao diện trên máy tính và điện thoại. Hệ thống còn có chức năng lưu trữ dữ liệu trên GoogleSheet, cảnh báo qua Email, điện thoại. Phần cứng của hệ thống được chế tạo với phần khung vỏ của các thiết bị đang có trên thị trường nên có khả năng ứng dụng cao trong thực tế.

Từ khóa: *IoT, TensorFlow, Node-Red, hệ thống điện.*

**Tác giả liên hệ chính.*

Email: luongngoctoan@qnu.edu.vn

Research and development of a monitoring and switching system of electrical parameters based on image processing and IoT

Luong Ngoc Toan*, Doan Duc Tung, Le Tuan Ho

Faculty of Engineering and Technology, Quy Nhon University, Vietnam

Received: 03/05/2024; Revised: 11/07/2024

Accepted: 15/07/2024; Published: 28/10/2024

ABSTRACTS

This paper studies on designing an Internet of Things (IoT) system for measuring and monitoring some electrical parameters. This system applies image processing technologies and object recognition devices to measure and monitor some electrical parameters such as current, voltage, power, and temperature. The data collected is then sent to the Node-Red server using the Message Queuing Telemetry Transport (MQTT) protocol for further analysis. Once the system is in operation, it can automatically switch off devices upon object detection. Users can monitor the system's status, control, and set parameters through an interface on their computer or phone. The system can also store data on Google and alert users via email or phone in case of any issues or anomalies. According to the paper, the system's hardware has been designed using the shell of existing devices available in the market, making it highly applicable in real-life scenarios.

Keywords: *IoT, TensorFlow, Node-Red, Power-system.*

1. INTRODUCTION

It is fascinating to see how the industry is evolving with the 4.0 revolution, Internet of Things (IoT) technology, and Artificial Intelligence (AI). With the development of such technologies, we are witnessing the creation of more functional devices, particularly in electrical systems and sensors. Moreover, the servers being free for users and the introduction of embedded computers with small sizes and strong configurations like Raspberry Pi and Jesson Nano have made it easier for users to access IoT and AI.

The application of image processing technology in real-life scenarios is becoming increasingly popular, primarily in industry,

environmental monitoring, and traffic surveillance.¹ In the electrical system, this technology is also applied to identify winding faults in transformers² and to determine power quality.²⁻⁴

In Vietnam, the number of offices and public sector facilities is increasing rapidly. However, it can be challenging to save electricity in these areas based on user behavior, and current efforts are often inefficient.⁵ With the rapid development of IoT technology, the application of smart switchgear such as switches and automats has partly solved the limitations in this area.

However, these devices often work with small currents, switching based on the control on

*Corresponding author.

Email: luongngoctoan@qnu.edu.vn

the phone or in combination with sensors such as motion sensors, and infrared sensors. Current devices are designed to either control or monitor electrical equipment, but they lack integration and the ability to detect people in large spaces such as classrooms or halls. Image processing technology has not been widely implemented in these devices to enable this functionality.

Based on the aforementioned issues, the author has conducted research and developed a system that includes functions such as wide-range image recognition for human detection to automatically switch electrical equipment. Additionally, the system can monitor, control, warn, and store data on parameters in the electrical system such as voltage, current, and power.

Although existing energy monitoring systems in previous studies can monitor electrical parameters and store data in the cloud, these systems have not yet integrated human recognition devices to automatically control equipment switching. This not only ensures greater flexibility and security but also improves convenience and efficiency in power system management.^{6,7}

2. RESEARCH AND DEVELOPMENT OF THE SYSTEM

2.1. IoT system

IoT stands for the Internet of Things. A network of sensors, software, and other technologies that connect objects and devices, enabling them to collect and exchange data with each other.^{8,9}

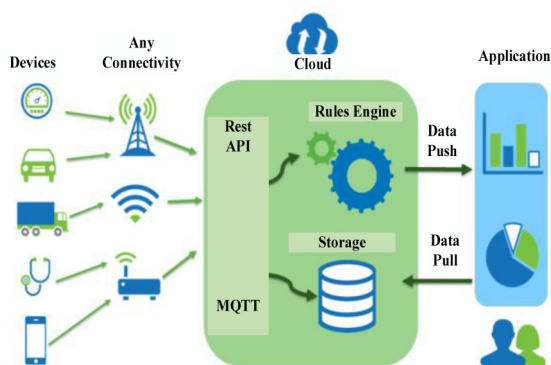


Figure 1. The working principle of an IoT system.

The Internet of Things has vast application potential across all sectors. It is important to note that every complete IoT system encompasses four key steps as shown in Figure 1: data collection, data sharing, data processing, and decision-making.

An IoT system typically comprises four main components: devices (Things), gateways (Gateways), network infrastructure (Network and Cloud), and analysis and data processing (Service and Solution Layers).^{10,11} These components work together to create a network of connected devices that can collect and exchange data with each other, enabling the creation of smart systems and applications.

Sensors within the research system measure electrical parameters, recognize objects, and convert them into data for the Internet. Users can easily process the signals and make necessary changes through phone or computer applications.

2.2. Image processing based on TensorFlow

TensorFlow has indeed become the most popular deep-learning library in recent years. With TensorFlow, users have the flexibility to create various Deep Learning models such as convolutional Neural Networks (CNNs),¹² Recurrent Neural Networks (RNNs), and artificial neural networks. It is a powerful tool that has transformed how we approach machine learning and data analysis.¹³

TensorFlow is a powerful software library that enables numerical computations through data flow graphs. It is highly adaptable and can be used to program and train neural networks in multiple programming languages such as C, Java, and Python. Additionally, it allows for deploying applications on various platforms, including the cloud, machine local server, and mobile devices.

2.2.1. Core Elements in TensorFlow

TensorFlow is a library used for numerical computation and graphical representation of

data that helps in creating and training machine learning models. These graphs consist of nodes and edges that represent calculations and corresponding data in the model. In general, the process of training a machine learning model in TensorFlow follows the steps shown in Figure 2.

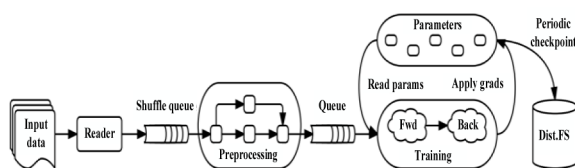


Figure 2. Data flow chart for the training process.

In this training process, the user defines the computational graph structure by declaring variables and operations in the model. The processing is then defined to measure the difference between the model's predicted output and the actual value.

The data obtained after processing will be analyzed to determine the best parameter values that can minimize the loss function. The resulting values are then used to train and update the model. After the training, the model is tested with fresh data to assess its performance. Finally, the trained model is utilized to predict and categorize new data.¹⁴

2.2.2. Application of TensorFlow library on Raspberry Pi

Raspberry is a computer that is designed to be compact and embedded, making it perfect for use in IoT projects. Suppose you want to object detection models. In that case, you can use TensorFlow Lite, which is a TensorFlow library that is specifically designed to work with Raspberry Pi 3 and Raspberry Pi 4 when they are running either Raspbian Buster or Raspbian Stretch. The process for installing and running this program on Raspberry is illustrated in Figure 3.

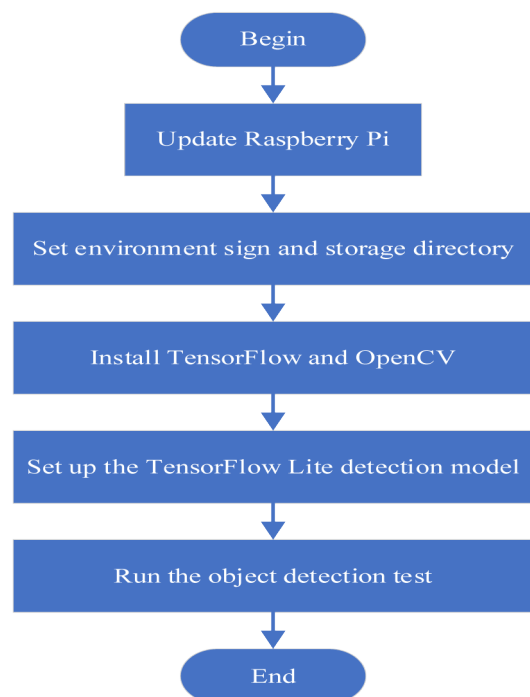


Figure 3. TensorFlow installation procedure.

During the test, the results were displayed as depicted in Figure 4. To capture the camera frame, we used the OpenCV object. After capturing the frame, it was flipped horizontally and vertically for proper alignment and conversion to the required format that the model could process. The output of the model was a list of objects present in the frame, along with their corresponding parameters.

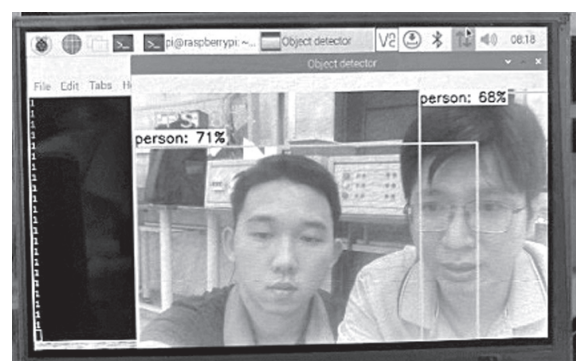


Figure 4. Object detection with TensorFlow Lite model.

The Object Detection model provides information on the location, class, and reliability of objects detected in a frame. To draw bounding boxes around an object, the model uses the coordinates of its upper-left and lower-right points. The layer's name is displayed with the

bounding frame when a person is detected. Additionally, the class name is compared to a predefined label to determine its presence in the frame and the percentage matching the sample.

2.3. System development research

2.3.1. System design

The following study puts forward a system that is capable of monitoring the parameters in the electrical system and can be operated through object recognition within the observation field. It is combined with both local and remote control through the control interface which is depicted in Figure 5 and system device parameters as shown in Table 1.

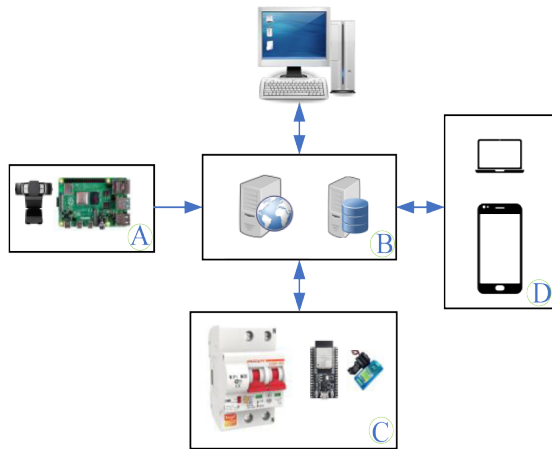


Figure 5. Schematic diagram of online monitoring structure for a distribution power system based on image processing and IoT.

The following are the primary stages of the system.

A: Webcam and Raspberry Pi 4 are combined for image processing to monitor and detect people within its range.

B: The server and database are installed on the computer and run in the cloud.

C: Measurement and switching are carried out using Pzem - 004T and smart automatic TuyA devices. These two measuring and switching devices are connected to Esp32 to transmit data to the server.

D: Monitor and control using your phone and computer.

Table 1. Main Device Specifications in the System.

No.	Device Name	Specifications
1	WebCam-PC-SWC-02	Resolution: 1080P, USB connection, 140-degree scanning angle.
2	Raspberry Pi 4 Model B	RAM: 8GB, GPIO: 40-pin standard, 2 Micro HDMI display outputs, Micro-SD slot for OS and storage.
3	WiFi Tuya Controllable Circuit Breaker	Voltage: 220V ~ 50Hz, Load capacity: 63A (1 phase 2P), Overload and short circuit protection, WiFi: 2.4GHz IEEE 802.11b/g/n.
4	WiFi BLE ESP32 NodeMCU-32S Kit (CH340)	SPI Flash: 32Mbits, Frequency Range: 2400~2483.5MHz, Bluetooth: BLE 4.2 BR/EDR, WiFi: 802.11 b/g/n/e/i.
5	PZEM-004T AC Multi-Functional Meter UART Module	Voltage range: 80 ~ 260VAC / 50 - 60Hz, Accuracy: 0.01, Current range: 0 ~ 100A, UART TTL 5VDC logic level interface, default baud rate: 9600, 8, 1.

The measurement and switching stage (C) involves using the PZEM-004T sensor to measure parameters such as voltage, current, frequency, power, power factor, and power consumption. These measurements are then uploaded to the server via ESP32. At the same time, the identification and image processing (A) system sends a signal to the server to detect the presence or absence of people within the field of view. The server is equipped with a monitoring and control interface and a control program for the data received from (A) and (C).

The control signal is generated based on certain conditions such as the presence or absence of people in the field of view, overcurrent, and overvoltage. Data transmission between the

stages is carried out using the MQTT protocol, and the microcontroller sends an automatic control signal to perform the power supply switching process.

2.3.2. Research and manufacture hardware

The system structure depicted in Figure 4 requires the installation and fabrication of hardware in two stages, namely (A) and (C). In the data recognition and processing stage (A), we make use of WebCam-PC-SWC-02 to connect to Raspberry Pi 4 Mode B. To observe the image processing, we employ a Waveshare 7-inch HDMI screen. Upon linking the devices, we obtain the results as depicted in Figure 6.



Figure 6. Installation of object recognition and handling equipment.

To produce switchgear and measuring equipment, we carry out the following steps:

Step 1. To understand the structure and operating principle of the TuyA smart Aptomat, an experiment was conducted to find the control signal on the device, as shown in Figure 7. As a result, the signal-receiving pin will be detected, then this pin will be connected to the ESP32 Module using an intermediate relay.

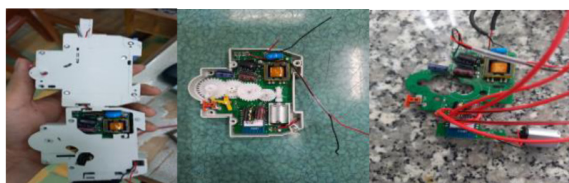


Figure 7. Construction and simulation of control circuit signals for switchgear.

Step 2. To measure all parameters in the power system, we need to install a UART PZEM-004T measurement circuit with ESP32. This circuit is highly versatile and allows us to measure the electrical signal and voltage.

The measurement pins are integrated with the Aptopmat to measure the signal even when the power is turned off as Figure 8.

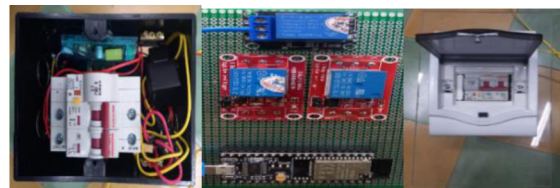


Figure 8. Installation of measuring equipment and switchgear.

Step 3. Fabricate and install a smart switch using an existing mechanical switch, ESP32, and relay to control devices. Following the installation process, we will have a manual manipulator, which is depicted in Figure 9.

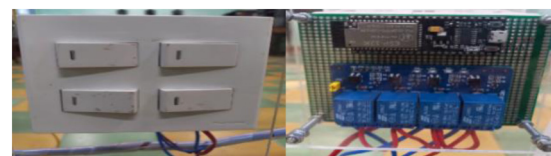


Figure 9. Renovate the mechanical switch.

2.4. Building a server for the system

2.4.1. Server and communication protocols

The system used to build the server involves Node-RED, a programming tool that enables the connection of hardware devices, Application Programming Interfaces (APIs), and online services. It is primarily designed for IoT (Internet of Things) applications, but can also be used to quickly link different service flows for other applications. The construction of the data transmission scheme is illustrated in Figure 10.

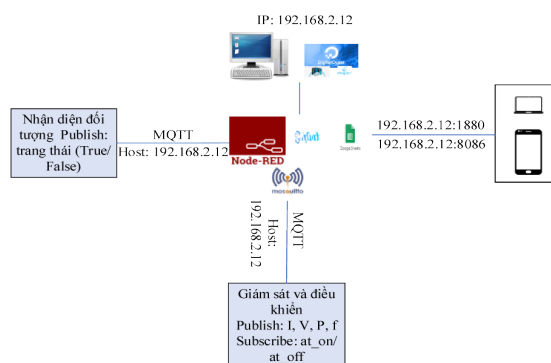


Figure 10. Diagram of data transmission and reception in the system.

The Node-RED tools and software, database, and communication protocol will be installed on the server or cloud server. We will use influxDB and Google sheet to store system data, and all communication within the system will be done using MQTT communication. For this reason, we will install the Mosquitto broker on the server. The server will have an IP address which will serve as the host address to identify the object sending the monitoring state (publish) to the server. The measurement stage will publish the parameters in the power system to the server and, at the same time, receive data from the server through data registration (subscribe).

3. RESEARCH RESULTS

After fabrication, the devices are connected to form a complete system as depicted in Figure 11.

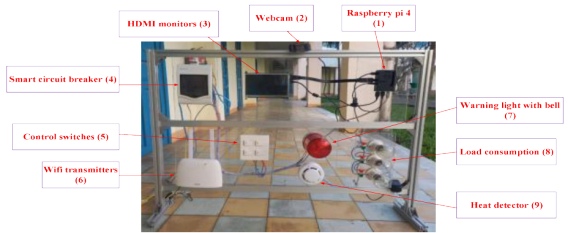


Figure 11. Actual results of the whole system.

After running the test, it was determined that the system operates stably and meets the requirements outlined in Figure 12.



Figure 12. Run a system test.

The system can automatically switch electrical circuits based on signals from image recognition and processing on a Raspberry Pi. It also allows for local switching, remote access through Aptomat, as well as control through computers and phones. Additionally, users can

monitor important electrical parameters such as voltage, current, and power consumption, which can help them evaluate their ability to save electricity while using the product.

The graph interface displays the parameters and stores them on the Google Sheet, as shown in Figure 13 and Figure 14. This feature allows the user to monitor the switchgear status, the status of people being monitored, as well as the real-time values of voltage and current.

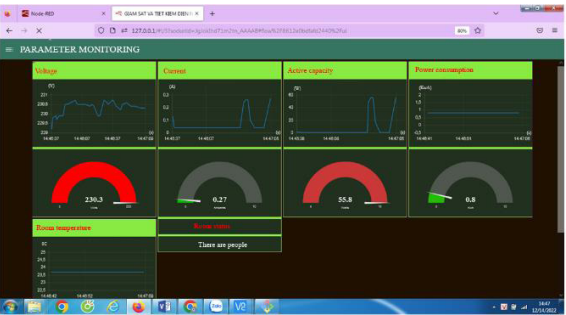


Figure 13. System monitoring interface.

On the control interface, we can set threshold values for alarms and protection against overcurrent, overvoltage, and increased room temperature. These setting values will be based on the process of monitoring parameters at the monitoring interface to ensure system safety. The system has two levels of protection: warning and trouble. Suppose the warning level exceeds the set threshold. In that case, the system will send an email as shown in Figure 14 and the system parameters will be continuously stored on Google Sheets as shown in Figure 15, and display a notification on the phone application phone as shown in Figure 16.

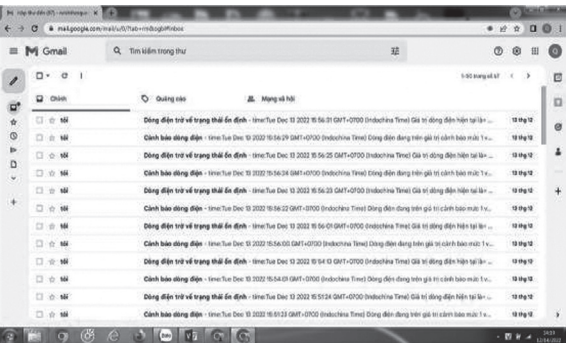


Figure 14. Warning results on Gmail.

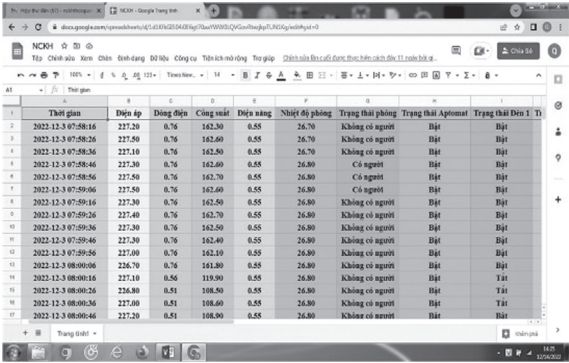


Figure 15. The results of the parameters are stored on Google Sheet.

In addition to computer monitoring and control, the system also allows phone interface and function synchronization as shown in Figure 16.

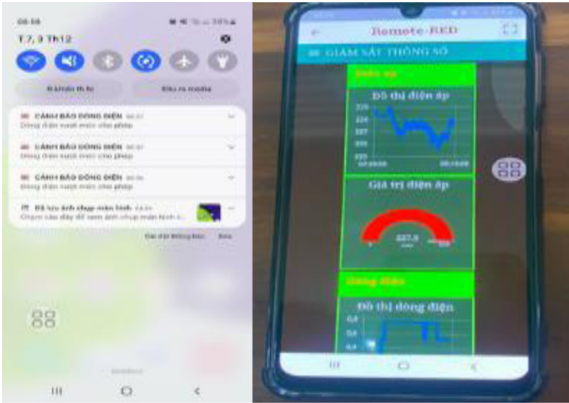


Figure 16. Phone alerts, monitoring, and controls.

4. CONCLUSION

After completing the research, the installation and testing of the system showed stable operation and met requirements. The devices used in the system are built on the frames and housings of devices currently on the market and used by consumers. Preliminary results show that the product has potential for practical applications and can replace existing devices. However, further testing and evaluation is required to ensure product reliability.

The integration of image monitoring units via esp32-Cam image transmission to Raspberry Pi for image processing enhances monitoring scope and device control flexibility.

Acknowledgments

This research is conducted within the framework of a science and technology project at the institutional level of Quy Nhon University under the project code T2023.802.12.

REFERENCES

1. A. S. Putra, H. L. H. S. Warnars. *Intelligent traffic monitoring system (ITMS) for smart cities based on IoT monitoring*, 2018 Indonesian Association for Pattern Recognition International Conference, INAPR 2018, Indonesia, 2018.
2. O. Aljohani, A. A. Siada. Application of digital image processing to detect short-circuit turns in power transformers using frequency response analysis, *IEEE Transactions on Industrial Informatics*, **2016**, 12(6), 2062-2073.
3. H. Shareef, A. Mohamed, A. A. Ibrahim. An image processing-based method for power quality event identification, *International Journal of Electrical Power & Energy Systems*, **2013**, 46, 184-197.
4. M. Abdelsattar, A. Diab, I. Hamdan. *A review on detection of solar PV panels failures using image processing techniques*, The 24th International Conference on Middle East Power System, MEPCON 2023, IEEE, Egypt, 2023.
5. H. V. Hao. Officers' electricity saving in open offices: exploring the role of perceived behavior control and accessibility to control, *VNU Journal of Economics and Business*, **2023**, 3(1), 112-120.
6. P. Karrupusamy. A sensor based IoT monitoring system for electrical devices using Blynk framework, *Journal of Electronics and Informatics*, **2020**, 2(3), 182-187.
7. J. V. Aldás, S. Silva, G. P. Navarro. IoT-based alternating current electrical parameters monitoring system, *Energies*, **2022**, 15(18), 6637.
8. G. Bedi, G. K. Venayagamoorthy, R. Singh. *Navigating the challenges of the Internet of Things (IoT) for power and energy systems*, 2016 Clemson University Power Systems Conference, PSC 2016, USA, 2016.

9. A. A. Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari. Internet of things: a survey on enabling technologies, protocols, and applications, *IEEE Communications Surveys and Tutorials*, **2015**, 17(4), 2347-2376.
10. R. Khan, S. Khan, R. Zaheer, S. Khan. *Future Internet: the Internet of Things architecture, possible applications, and key challenges*, The 10th International Conference on Frontiers of Information Technology, FIT 2012, Pakistan, 2012.
11. Z. Yang, Y. Peng, Y. Yue, X. Wang, Y. Yang, W. Liu. *Study and application of the architecture and key technologies for IoT*, International Conference on Multimedia Technology, ICMT 2011, Pakistan, 2011.
12. Z. Zeng, Q. Gong, J. Zhang. *CNN model design of gesture recognition based on the TensorFlow framework*, IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2019, Chengdu, 2019.
13. F. Ertam, G. Aydin. *Data classification with deep learning using TensorFlow*, International Conference on Computer Science and Engineering, UBMK 2017, Turkiye, 2017.
14. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, *TensorFlow: a system for large-scale machine learning*, The 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 16, USA, 2016.



© 2024 by the authors. This Open Access Article is licensed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license (<https://creativecommons.org/licenses/by-nc/4.0/>).

Nghiên cứu về ảnh hưởng của hàm lượng xi măng đến khả năng chịu lực của cọc xi măng đất để xử lý đất sét yếu ở khu đô thị sông Hà Thanh, thành phố Quy Nhơn, tỉnh Bình Định, Việt Nam

Hoàng Công Duy*, Phạm Văn Trục, Phạm Thị Lan, Nguyễn Thị Khánh Ngân

Khoa Kỹ thuật và Công nghệ, Trường Đại học Quy Nhơn, Việt Nam

Ngày nhận bài: 03/05/2024; Ngày sửa bài: 11/07/2024

Ngày nhận đăng: 15/07/2024; Ngày xuất bản: 28/10/2024

TÓM TẮT

Trong thực tiễn hiện tại của các tòa nhà xây dựng được đặt trên đất mềm ở Việt Nam, công nghệ cọc xi măng được sử dụng ngày càng rộng rãi vì những lợi thế nổi bật của nó. Cọc đất xi măng là cọc được làm từ đất địa phương trộn với một lượng xi măng và phụ gia nhất định tùy thuộc vào đặc tính địa kỹ thuật của đất trong khu vực dự án. Mục đích chính của nghiên cứu này là xác định hàm lượng xi măng tối ưu cho cọc xi măng - đất. Nhiều yếu tố ảnh hưởng đến cường độ của cọc đất xi măng, trong đó hàm lượng xi măng được sử dụng đóng một vai trò quan trọng và cần được nghiên cứu cẩn thận. Nghiên cứu này nhằm mục đích cung cấp thông số hàm lượng cấp phối xi măng phù hợp cho các cọc đất xi măng được áp dụng cho các dự án xây dựng ở khu vực đất yếu của sông Hà Thanh, thành phố Quy Nhơn. Các mẫu được tạo ra bằng cách trộn đất với hàm lượng xi măng thay đổi từ 5% đến 25%. Nhóm tác giả tiến hành xác định ứng suất dọc trục của các mẫu cọc đất xi măng ở thời điểm 7 ngày tuổi và 28 ngày tuổi bằng thí nghiệm nén một trục nở hông. Kết quả cho thấy hàm lượng xi măng tối ưu cho cọc đất xi măng đạt được từ 12% đến 15% đối với đất sét yếu khu đô thị Hà Thanh, thành phố Quy Nhơn, tỉnh Bình Định.

Từ khóa: Cọc xi măng - đất, sức chịu tải của cọc, đất yếu, xử lý nền đất yếu.

*Tác giả liên hệ chính.

Email: hoangcongduy@qnu.edu.vn

Research on influence of cement content on bearing capacity of soil-cement piles to treat soft clay in Ha Thanh river urban area, Quy Nhon city, Binh Dinh province, Vietnam

Hoang Cong Duy*, Pham Van Truc, Pham Thi Lan, Nguyen Thi Khanh Ngan

Faculty of Engineering and Technology, Quy Nhon University, Vietnam

Received: 03/05/2024; Revised: 11/07/2024

Accepted: 15/07/2024; Published: 28/10/2024

ABSTRACT

In the current practice of building foundations placed on soft soil in Vietnam, cement-soil pile technology is increasingly widely used because of its outstanding advantages. Cement-soil pile is made from local soil mixed with a certain amount of cement and additives decided by the geotechnical characterization of the soil in the project area. The main purpose of this study is to propose the optimal cement content for cement-soil piles. Many factors affect the compressive strength of cement-soil piles, in which the cement content used holds an important role and needs to be calculated. This study aims to provide appropriate cement gradation content parameters for cement-soil piles applied to construction projects in the soft soil area of Ha Thanh River, Quy Nhon City. In the paper, the samples were created by mixing soil with cement content varying from 5% to 25%. The authors determined the axial stress of soil-cement pile samples at 7 days old and 28 days old using the unconfined compressive test. The results show that the optimal cement content for cement-soil piles ranges from 12% to 15% for soft clay in the Ha Thanh urban area, Quy Nhon City, Binh Dinh province.

Keywords: *Soil-cement pile, bearing capacity, soft soil, soft soil treatment method.*

1. INTRODUCTION

Due to constant development of the society, the need for land to build housing and infrastructure projects is essential, especially land in the coastal and lagoon areas of Quy Nhon City, Binh Dinh province, and the construction of high-rise apartment buildings, ports, and roads on soft foundations is extremely urgent. In this research article, the soft soil treatment method will encounter more and more complex problems, thus creating opportunities for developing new soft soil foundation treatment technologies.

Therefore, effective treatment of soft soil has become an essential problem in construction designs. Among them, the cement-soil pile method is widely used for industrial and civil buildings, docks, and highways.

In this study, the survey location was located at the An Phu Thinh social housing apartment building project at land lot B1 - 32, An Phu Thinh new urban area, location: Dong Da ward, Quy Nhon City, Binh Dinh province as shown in Figure 1. Soil stratification of the survey location is illustrated in Figure 1.

*Corresponding author:

Email: hoangcongduy@qnu.edu.vn

In the survey area, the soft clay layer has a relatively large thickness. The foundation of the construction is placed on this ground which is required to treat the soft clays or choose the deep foundation. However, choosing a deep foundation for low-rise buildings is not reasonable. Therefore, the authors recommend that the soft soil treatment method is the cement-soil pile, which increases the bearing capacity of the ground and solves the cost of the foundation for the project.

Thus, we conducted experiments to determine the cement content to optimize the grade for soil cement piles. Based on the results of this research, construction engineers can choose the optimal cement content as recommended by the study to manufacture soil-cement piles. Then, the soil-cement piles are statically compressed at the building site to determine the load capacity that is applied in the design ground consolidation.

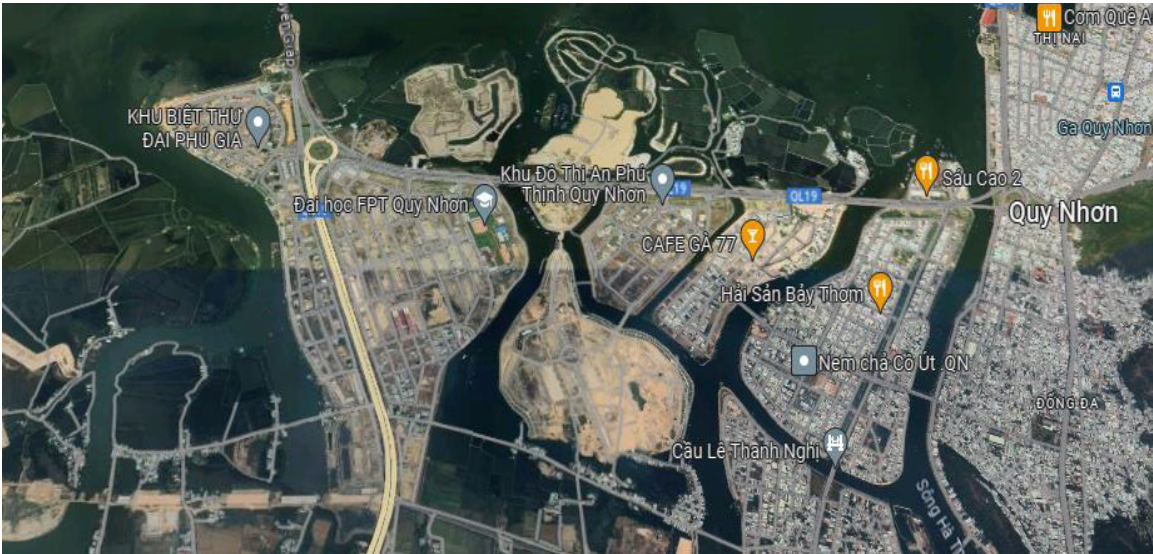


Figure 1. Location of the An Phu Thinh social housing apartment building project.

2. METHODOLOGY

2.1. Geological structural conditions in the survey area

The survey area has relatively flat terrain because it has been partially leveled with river bottom materials. Regarding geomorphology, the area belongs to agglomeration morphology,

the geological formations are river and lake sediments with many types of heterogeneous materials. The soil and water samples were brought in the laboratory to determine physical and mechanical criteria.¹ The geotechnical characteristics of the soft clays are shown in Table 1.

Table 1. Geotechnical parameters of soft clays in the survey area.

Depth(m)	Sample test	Description							SPT (N-Value)
10-52m	Soft clay	Soft, blue-gray, dark gray, organic content of 13.1%, low plasticity							<3
Mechanical consistency of soft clay									
γ	c	ϕ	G _s	W	W _L	W _P	I _p	I _L	e ₀
kN/m³	kN/m²	°		%	%	%			
16.3	27.01	7	2.72	69.9	40.9	22.4	63.3	1.28	1.843

2.2. Stabilization of soft soil - The soil cement column method (TCVN 9403:2012)

In this study, the group of authors used the dry mixing method to cast sample tests. The process includes mechanically loosening the soil in the field and mixing dry cement powder with soil with or without additives. The method of creating soil-cement samples conducted following TCVN 9403:2012. Water and cement in the required amount were mixed manually until there was obtained homogeneous state of

“laitance”. The amount of cement is determined by the weight ratio of dry soil. After that, the cement mortar was mixed with soil which is specific humidity. The obtained mixture was mixed to a homogeneous mass over 5 minutes in cylinders with dimensions $h = 100\text{ mm}$ and $d = 50\text{ mm}$. The samples were pulled from the blocks on the second day and they were retained period till the test when 7 days and 28 days.^{2,3} The parameters of test samples are presented in Table 2.

Table 2. Parameters of test samples in the axial load test.

Number	Sample group	Number of samples	Weight of samples		Size of samples	
			Cement Phuc Son PCB40(g)	Soil (g)	Height (mm)	Diameter (mm)
1	5%	6	74.23	1814.27	100	50
2	8%	6	118.77			
3	10%	6	148.47			
4	12%	6	178.16			
5	15%	6	222.70			
6	18%	6	267.24			
7	20%	6	296.93			
8	25%	6	371.17			

2.3. Axial load test

This test method is used to determine the unconfined compressive strength of cohesive soil in the undisturbed, remolded, or compacted condition by using strain-controlled application of the axial load based on ASTM D2166 standards. The equipment used in the unconfined

compressive test is the Triplex II advanced as shown in Figure 2. In this standard, the unconfined compressive strength (q_u) is extracted as the maximum load attained per unit area or the load per unit area at 20% axial strain.¹

$$q_u = \sigma_1 - \sigma_3$$

(1)

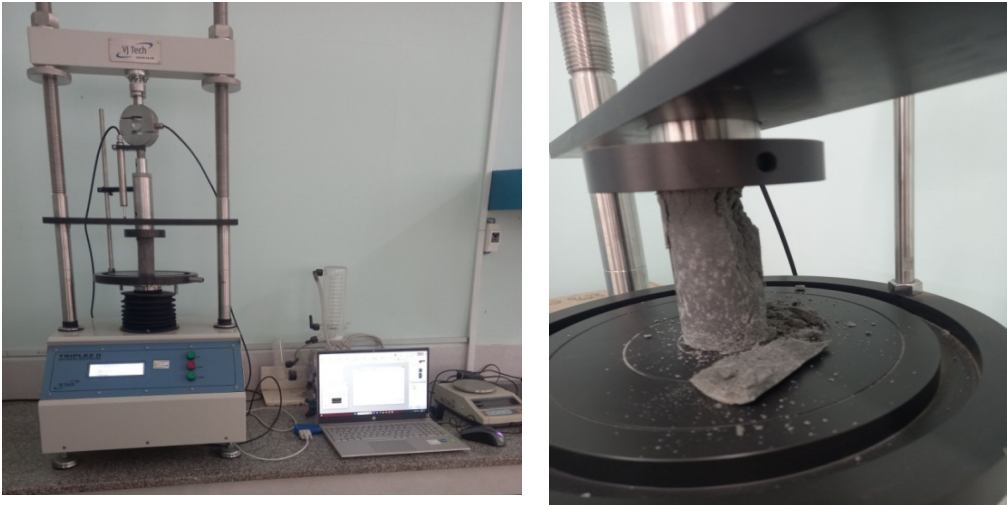


Figure 2. The Triplex II advanced and form of cement-soil piles damage.

3. RESULTS AND DISCUSSION

The cement-soil strength is increasing over time. As the cement content of the cement-soil pile increases, the unconfined compressive strength increases as shown in Figure 3.³⁻⁵

When the cement content increased from 5% to 8%, the compressive strength of samples at 28 days increased from 249.5 Mpa to 842.3 Mpa. The percentage increase of the compressive strength is 238% as shown in Figure 4.

When the cement content increased from 8% to 10%, the compressive strength of samples at 28 days increased from 842.3 Mpa to 961.3 Mpa. The percentage increase of the compressive strength is 14% as shown in Figure 4.

When the cement content increased from 10% to 12%, the compressive strength

of samples at 28 days jumped from 961.3 Mpa to 1632.4 Mpa. The percentage increase of the compressive strength is 14%. Similarly, the percentage increase of the compressive strength is 99% when the cement content increased from 12% to 15% as shown in Figure 4.⁶

When the cement content of the samples increases from 12% to 15%, the compressive strength of cement-soil plies increases from 70% to 99%. However, when the cement content is greater than 15% the compressive strength of samples at 28 days decreased significantly, as presented in Figure 4.

When the cement content of the sample increases from 15% to 20%, the increase in axial compressive strength of the cement-soil pile decreases from 99% to 15% as presented in Figure 4.^{7,8}

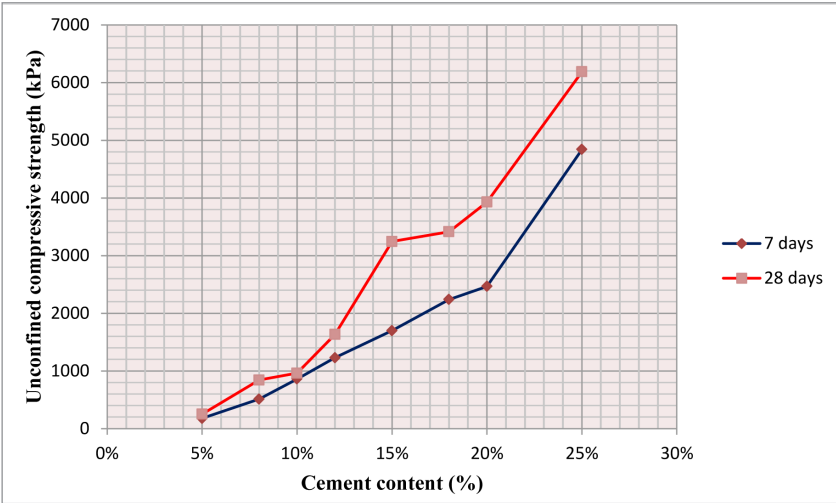


Figure 3. The unconfined compressive strengths of cement-soil piles at 7 days and 28 days.

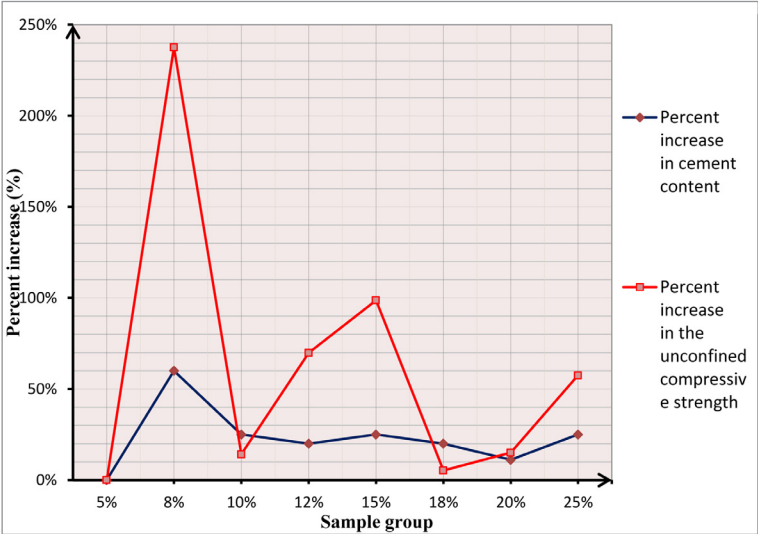


Figure 4. The percent increase in cement content and the percent increase in unconfined compressive strength of the sample groups.

4. CONCLUSIONS

The unconfined compressive strengths of cement-soil piles increase rapidly when cement content increases from 12% to 15% as presented in Figure 4.

The cement content increases or decreases depending on the moisture content of the soil. The authors also proposed the optimal cement content percentages of 10% and 18% for cement-soil piles in the Mekong delta area.

Therefore, the authors suggested reasonable cement content for soft clay in the Ha Thanh River urban area, Quy Nhon City is from 12% to 15%. So the content of cement has a great influence on the unconfined compressive strength of cement-soil piles.

In the geographical area of the study, based on the experimental results regarding the cement content for soil-cement pile specimens as mentioned above, the group of authors proposes an optimal cement content ratio for soil-cement piles ranging from 12% to 15%. The research recommends that design engineers may use a cement content for soil-cement piles from 12% to 15% to manufacture soil-cement piles and conduct static compression tests on-site in the research area for application in design problems.

The cement-soil pile method is proposed to treat soft clay for 4 to 8-storey buildings, located in the Ha Thanh River urban area, Quy Nhon City.

Acknowledgment

This research is conducted within the framework of science and technology projects at institutional level of Quy Nhon University under the project code T2023.805.15.

REFERENCES

1. M. D. Braja. *Advanced soil mechanics (5th edition)*, Taylor and Francis Group, London New York, 2019.
2. N. Zotsenko, Y. Vynnykov, V. Zotsenko. Soil-cement piles by drilling-mixing method, *Construction and Geotechnics*, **2015**, 4, 122-133.
3. X. Jiang, Q. Lu, S. Chen, R. Dai, J. Gao, P. Li. *Research progress of soft soil foundation treatment technology*, The 6th International Conference on Environmental Science and Civil Engineering, China, 2020.
4. U. S. Okyay, D. Dias. Use of lime and cement treated soils as pile supported load transfer platform, *Engineering Geology*, **2010**, 114, 34-44.
5. S. H. Thai, T. M. Tran, V. C. Trinh. Cement content and water-cement ratio or soil stabilization by soil cement deep mixing method in coastal zone of Mekong delta area, *Journal of Water Resources and Environmental Engineering*, **2014**, 44, 58-62
6. S. S. Gue, S. G. Chang. Geotechnical challenges on soft ground, *Journal of Civil Engineering, Science and Technology*, **2022**, 13(2), 84-96.
7. F. Zhang, L. Zhang, L. Yang, T. Feng, L. Chen, X. Zhong. The cement content measurement of cement mixing piles with EDTA titration method, *Geotechnical Engineering*, **2018**, 22, 4306-4315.
8. D. H. Do, T. A. Pham. Investigation of performance of soil-cement pile in support of foundation systems for high-rise buildings, *Civil Engineering Journal*, **2018**, 4, 266-277.



© 2024 by the authors. This Open Access Article is licensed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license (<https://creativecommons.org/licenses/by-nc/4.0/>).

Xây dựng mặt tương tác và đánh giá khả năng chịu lực của cột bê tông cốt thép tiết diện hình chữ nhật

Lê Thị Mai Trang*, Vàng Hiếu Quang

Khoa Kỹ thuật và Công nghệ. Trường Đại học Quy Nhơn, Việt Nam

Ngày nhận bài: 22/04/2024; Ngày sửa bài: 07/06/2024
Ngày nhận đăng: 17/07/2024; Ngày xuất bản: 28/10/2024

TÓM TẮT

Khi thiết kế kết cấu cột bê tông cốt thép chịu nén lệch tâm xiên, việc kiểm tra khả năng chịu lực của cột sau khi đặt cốt thép là cần thiết để xác định tính hợp lý của cốt thép được bố trí. Nghiên cứu này phát triển một chương trình xây dựng các mặt tương tác ba chiều của cột bê tông cốt thép chịu nén lệch tâm xiên có mặt cắt ngang hình chữ nhật bằng ngôn ngữ lập trình MATLAB, theo mô hình biến dạng phi tuyến. Mômen nội lực từ phần mềm như ETABS dùng để kiểm tra khả năng chịu lực của cột được xử lý đưa vào hệ số uốn dọc theo tiêu chuẩn TCVN 5574:2018, sau đó so sánh với mặt tương tác 3D đã thiết lập để đánh giá khả năng chịu lực của cột. Bài báo trình bày hai phương pháp đơn giản để kiểm tra khả năng chịu của cột. Các ví dụ minh họa được trình bày trong bài viết này để xác nhận độ tin cậy và khả năng ứng dụng của chương trình đã đề xuất. So sánh với ETABS và CSI-Col, sai số luôn nhỏ hơn 7,2%. Chương trình có thể được dùng cho các nghiên cứu sâu hơn như nghiên cứu ảnh hưởng của các tham số đầu vào đến khả năng chịu lực của cột, phân tích xác suất, phân tích độ nhạy hoặc tạo dữ liệu cho học máy.

Keywords: Mặt tương tác 3D, cột bê tông cốt thép, khả năng chịu lực, tiết diện chữ nhật, TCVN 5574:2018.

*Tác giả liên hệ chính.

Email: lethimaitrang@qnu.edu.vn

Generating 3D interaction surfaces and assessing load-bearing capacity of rectangular reinforced concrete columns

Le Thi Mai Trang*, Vang Hieu Quang

Faculty of Engineering and Technology, Quy Nhon University, Vietnam

Received: 22/04/2024; Revised: 07/06/2024

Accepted: 17/07/2024; Published: 28/10/2024

ABSTRACT

On designing reinforced concrete (RC) columns subjected to combined axial load and biaxial bending, checking the bearing capacity of the column after placing the reinforcement is necessary to determine the integrity of the placed reinforcement. This study developed a tool to generate the 3D interaction surfaces of rectangular RC columns subjected to biaxial flexural and axial loads using MATLAB software, according to the theory of non-linear material model. The internal moments from the software, such as ETABS, are processed to include the buckling factor according to TCVN 5574:2018, then compared with the established 3D interaction surface to assess the bearing capacity of the designed column. Two methods to check the load-bearing capacity of the columns were introduced. Examples are illustrated in this article to apply and verify the proposed program. Compared with ETABS and CSI-Col, the errors are less than 7.2%. This program can be used for further studies, such as investigating the effect of initial parameters on the column capacity, probabilistic analysis, sensitive analysis, or generating data for machine learning.

Keywords: *3D interaction surfaces, reinforced concrete columns, load-bearing capacity, rectangular cross-section, TCVN 5574:2018.*

1. INTRODUCTION

The required reinforcement for rectangular reinforcement concrete columns subjected to uniaxial flexural and axial loads is calculated based on Vietnamese national design standards TCVN 5574: 2018. However, the method for biaxial flexural and compression forces is sophisticated for practical application. Various approximated approaches were proposed. The superposition method, introduced by Moran, calculates reinforcement separately with (N, M_x) and (N, M_y) and then adds the results.¹

Symmetrically-reinforced rectangular sections can be designed to withstand an increased moment about one axis, as detailed in BS 8110-1:2005 and Nguyen.² Bresler discussed other design criteria for short columns, including three simple methods to generate the failure surfaces.³ From then, several authors have developed approximate methods based on the research of Bresler and ACI standards.⁴⁻⁶ In particular, Mavichak and Furlong conducted experiments on nine rectangular cross-section columns and compared the test data with an analytical model to validate the equation for strength

*Corresponding author.

Email: lethimaitrang@qnu.edu.vn

approximations.⁵ Some authors suggested methods to draw 2D interaction diagrams based on TCVN 5574:2012.⁷⁻⁹ A combination of Bresler's equations and uniaxial interaction diagrams based on TCVN 5574:2012 was proposed by Nguyen to estimate the required reinforcement or check the load-bearing capacity.¹⁰ Nguyen presented the interaction diagram of circular reinforced concrete columns under fire, based on TCVN 5574:2012 and Eurocode 2.¹¹ Hoach and Tran et al. developed software for creating interaction diagrams based on TCVN 5574:2018 for biaxially loaded members.^{12,13} Existing software for generating interaction surfaces based on Vietnamese standards for biaxial flexural columns has its limitations. Most software is written in Excel, with 2D graphs drawn according to the Mx - My or N - M relationship. For the new Vietnamese load and impact standards TCVN 2737:2023, the cases of dangerous internal forces that need to be considered have significantly increased. Checking the column's bearing capacity for each axial force or moment is time-consuming.

Pham proposed approximate approaches, which refer to uniaxial compression, to create the interaction diagram and check the load-bearing capacity for columns under eccentric compression based on TCVN 5574:2012.¹⁴ Hoach and Nguyen explained a method to calculate the coefficient of safety for reinforced concrete columns, considering the slenderness of columns according to TCVN 5574:2018.^{15,16} The proposed method to identify safety factors requires discrete points of the interaction surface to be displayed in a spherical coordinate system.

Even imported software like ETABS and CSI-Col, which integrate the function of drawing interaction surfaces, have drawbacks. They do not support Vietnamese standards, allow users to modify related coefficients, or provide visual load-bearing capacity checks. These limitations underscore the need for our program.

This research presents a novel program, written by MATLAB, for generating 3D interaction surfaces of columns under biaxial flexural and axial forces. The program is based on the limited strain and non-linear material models, considering the buckling factor per Vietnamese standard 5574:2018. It also includes a tool for assessing the load-bearing capacity of the columns. The paper provides a step-by-step procedure for checking the load-bearing capacity corresponding to the internal forces calculated from the load combination. Additionally, the paper includes an illustrative example to demonstrate the program's capabilities further.

2. METHODOLOGY

2.1. Assumptions

The strain distribution is linear across the section of the column (the cross-section of the column is always a plane).

The stress in the steel and concrete is given by the stress-strain graphs as shown in Figure 1 and Figure 2.

The tensile resistance of the concrete is negligible.

The concrete strain at failure is 0.0035.

2.2. Stress-strain model of concrete

According to TCVN 5574:2018, for normal-strength concrete, the stress-strain graph, which includes three or two segments, can be used when calculating reinforced concrete per a non-linear deformation model. In this research, a three-segment stress-strain graph is selected.

$$\text{When } 0 \leq \varepsilon_b \leq \varepsilon_{b1}: \sigma = E_b \varepsilon_b \quad (1)$$

$$\text{When } \varepsilon_{b1} < \varepsilon_b < \varepsilon_{b0}:$$

$$\sigma_b = \left[\left(1 - \frac{\sigma_{b1}}{R_b} \right) \frac{\varepsilon_b - \varepsilon_{b1}}{\varepsilon_{b0} - \varepsilon_{b1}} + \frac{\sigma_{b1}}{R_b} \right] R_b \quad (2)$$

$$\text{When } \varepsilon_{b0} \leq \varepsilon_b \leq \varepsilon_{b2}: \sigma_b = R_b \quad (3)$$

$$\text{Where: } \sigma_{b1} = 0.6R_b; \quad \varepsilon_{b1} = \sigma_{b1} / E_b$$

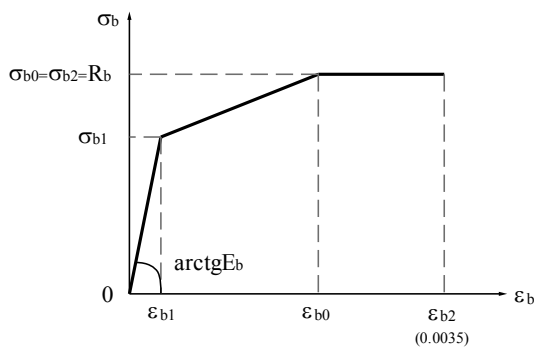


Figure 1. Stress-strain model of concrete.

2.3. Stress-strain model of steel

According to TCVN 5574:2018, it is possible to use a simplified stress-strain chart with two segments for the steel, which has a yield point, as described below.

$$\text{When } 0 \leq \varepsilon_s < \varepsilon_{s0}: \sigma_s = \varepsilon_s E_s \quad (4)$$

$$\text{When } \varepsilon_{s0} \leq \varepsilon_s \leq \varepsilon_{s2}: \sigma_s = R_s \quad (5)$$

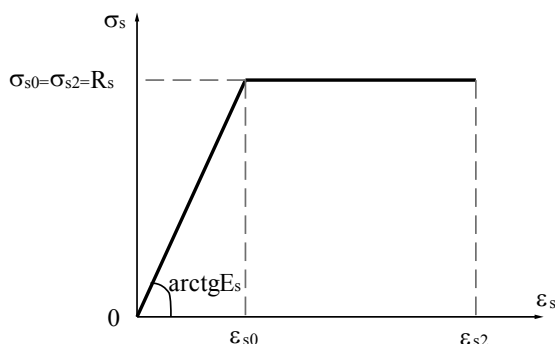


Figure 2. Stress-strain model of steel.

The tensile and compressive stress-strain graphs of steel are identical.

2.4. Generation of 3D interaction surfaces

Considering the section of a reinforced concrete column with pre-arranged reinforcement, the Oxy axis system with the center O coincides with the column section's center of gravity. Each steel bar is modeled by a circle with diameter ϕ and area A_s . The concrete part is discretized into a matrix of equal elements. Each element in the matrix is a square with size du and area du^2 . The size of these elements is relatively small, so it can be assumed that the stress in the elements is considered evenly distributed within

that element. The coordinates of steel bars and concrete elements are (x_{si}, y_{si}) , and (x_{bi}, y_{bi}) , respectively, as shown in Figure 3.

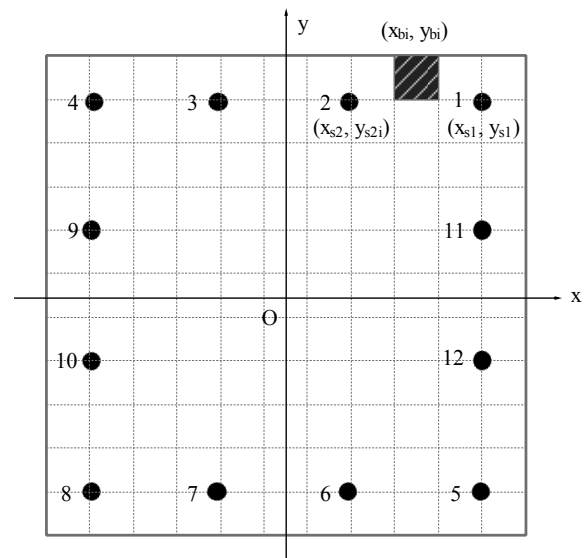


Figure 3. Concrete elements and steel bar coordinates.

Assume the neutral axis occurs in three cases, as shown in Figure 4. Figure 4a shows the neutral axis inclined (biaxial flexural compression), Figure 4b shows the neutral axis parallel to the x-axis (uniaxial flexural compression in the x-direction), and Figure 4c shows the neutral axis parallel to the y-axis (uniaxial flexural compression in the y-direction).

The neutral axis' angle and distance were changed to cover all the possible scenarios for the column. In this research, the interaction surfaces were constructed for the first quadrant, assuming that the maximum compressive strain of concrete is in the top-right corner, equal to 0.0035.

The strain of each steel and concrete element is calculated according to Equation (6).

$$\varepsilon_i = \frac{h_{0i} - x}{x} \varepsilon_{b2} \quad (6)$$

Where:

x is the distance between the neutral axis and the top-right corner.

h_{0i} is the distance between the material element and the straight line paralleled to the neutral axis across the farthest point of the

compression zone, as shown in Figure 5. h_{oi} can be calculated from the coordinates (x_i, y_i) of steel bars or concrete elements and the equation of neutral axis.

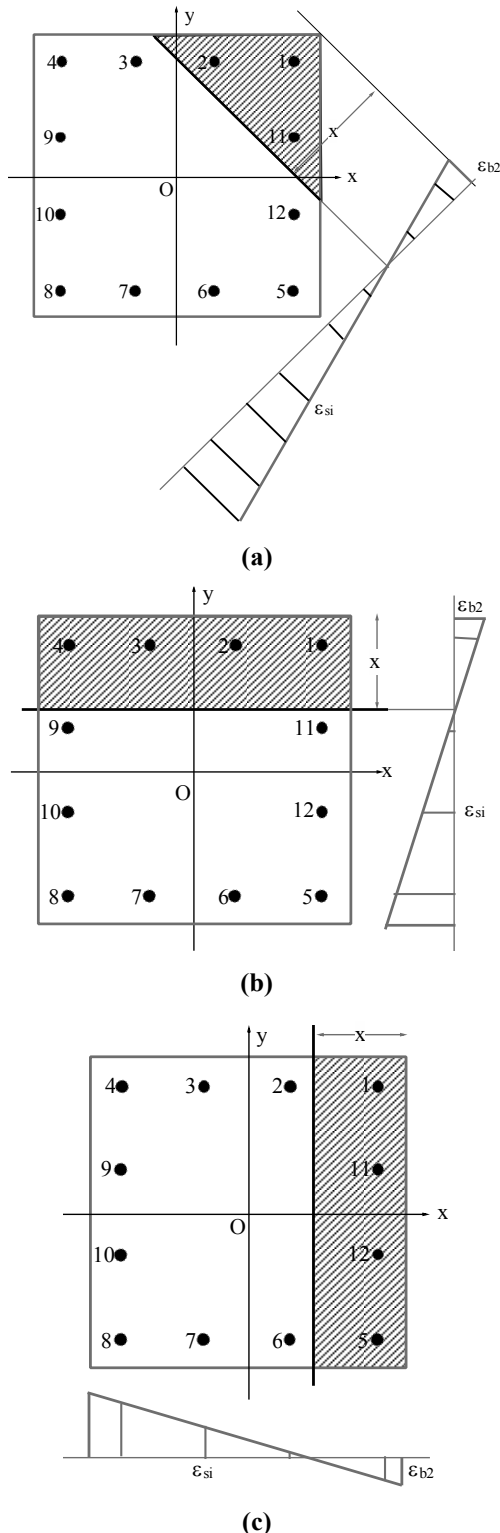


Figure 4. Neutral axis positions and strain distribution: (a) neutral axis inclined, (b) neutral axis parallel to the x-axis, (c) neutral axis parallel to the y-axis.

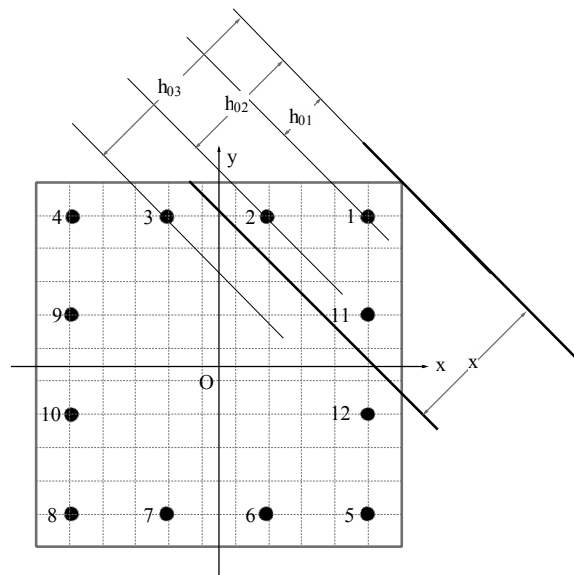


Figure 5. h_{oi} in the general scenario of an inclined neutral axis.

Then, the stress in each steel bar and concrete element is calculated per the stress-strain model in 2.2 and 2.3.

Assume that:

Compression stresses in concrete are positive values;

Compression stresses in steels have negative values, and tensile stresses are positive;

N_{uz} is positive when the column is compressed;

N_{uz} , M_{ux} , M_{uy} on the failure surfaces was calculated as below:

$$N_{uz} = -\sum_{i=1}^n \sigma_{si} A_{si} + \sum_{i=1}^m \sigma_{bi} d_u^2 \quad (7)$$

$$M_{ux} = -\sum_{i=1}^n \sigma_{si} A_{si} y_{si} + \sum_{i=1}^m \sigma_{bi} d_u^2 y_{bi} \quad (8)$$

$$M_{uy} = -\sum_{i=1}^n \sigma_{si} A_{si} x_{si} + \sum_{i=1}^m \sigma_{bi} d_u^2 x_{bi} \quad (9)$$

Where n is the number of steel bars, m is the number of concrete elements.

The maximum compression force columns can withstand is $N_{u' max}$, which occurs in concentric compression. As stated in TCVN 5574:2018, when accounting for the buckling factor, $N_{u' max}$ must be decreased by the coefficient ϕ , whose

values change linearly from 0.9 to 0.85 when L_0/h varies from 10 to 20.

$$N_{u,\max} = \varphi \left(R_s \sum_{i=1}^n A_{si} + R_b b h \right) \quad (10)$$

Where:

b, h are the dimensions of the rectangular cross-section;

L_0 is the effective length of the column.

The interaction surface includes a series of discrete points, as shown in the Figure 6.

The space defined by the interaction surface is called column capacity interaction volume.

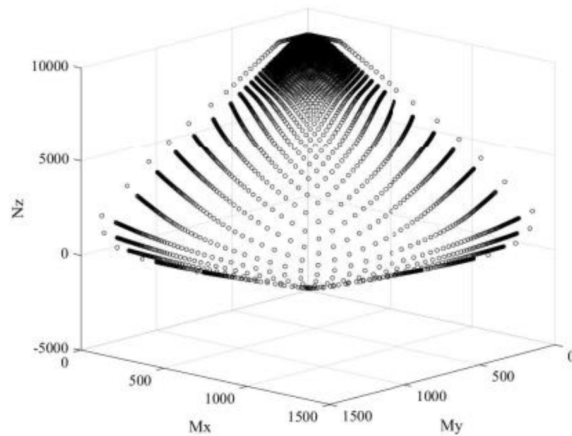


Figure 6. Discrete points of the interaction surface.

2.5. Assessing column capacity

The designer must check the capacity of columns for every load combination at every output station. Based on TCVN 5574:2018, the program follows the flowing steps to check a specific column for a particular load combination at a specific location.

(1) Input the internal moments and forces from the specified load cases and load combination, including N, M_{22} (or M_x), M_{33} (or M_y).

(2) Calculate the eccentricity of axial force.

$$e_{1x} = \frac{M_x}{N}; e_{1y} = \frac{M_y}{N} \quad (11)$$

(3) Define the random eccentricity e_a .

$$\begin{aligned} e_{ax} &= \max (L / 600, b / 30, 10 \text{ mm}) \\ e_{ay} &= \max (L / 600, h / 30, 10 \text{ mm}) \end{aligned} \quad (12)$$

Where L is the length of the column.

(4) Define the initial eccentricity e_0 .

$$e_{0x} = \max (e_{1x}, e_{ax}); e_{0y} = \max (e_{1y}, e_{ay}) \quad (13)$$

(5) Determine the coefficient η_x and η_y due to slender column effect.

$$\eta_x = \frac{1}{1 - \frac{N}{N_{crx}}}; \eta_y = \frac{1}{1 - \frac{N}{N_{cry}}} \quad (14)$$

$$N_{crx} = \frac{\pi^2 D_x}{L_0^2}; N_{cry} = \frac{\pi^2 D_y}{L_0^2} \quad (15)$$

$$\begin{aligned} D_x &= k_{bx} E_b I_x + k_s E_s I_{sx} \\ D_y &= k_{by} E_b I_y + k_s E_s I_{sy} \end{aligned} \quad (16)$$

Where:

I_x, I_y are the inertia moments of the cross-section,

I_{sx}, I_{sy} are the inertia moments of the steel bars, $k_s = 0.7$

$$k_{bx} = \frac{0.15}{\varphi_{Lx} (0.3 + \delta_{ex})}; k_{by} = \frac{0.15}{\varphi_{Ly} (0.3 + \delta_{ey})} \quad (17)$$

$$\varphi_{Lx} = 1 + \frac{M_{L1x}}{M_{Lx}}; \varphi_{Ly} = 1 + \frac{M_{L1y}}{M_{Ly}} \quad (18)$$

$$\delta_{ex} = e_{0x} / b; \delta_{ey} = e_{0y} / h; 0.15 \leq \delta_e \leq 1.5 \quad (19)$$

M_L is the internal moment of total load and M_{L1} is the internal moment of dead load about the axis across the farthest tension steel bar. In the most dangerous cases, ϕ_L is equal to 2.

(6) Determine total design moments by multiplying the coefficient e_x and e_y by the factored axial forces N_z obtained from the analysis.

$$e_x = e_{0x} \eta_x; e_y = e_{0y} \eta_y \quad (20)$$

$$M_x^* = N e_x; M_y^* = N e_y \quad (21)$$

(7) Determine whether the point, defined by the resulting axial load and biaxial moment set (N, M_x^*, M_y^*) lies within the interaction volume.

There are different ways to examine whether the load points lie within the interaction volume. Beside using a 3D rotational interaction surface, this program uses the method mentioned below.

First, calculate the neutral axis angle of each load point.

$$\tan \alpha = \frac{M_y^* I_x}{M_x^* I_y} \quad (22)$$

Second, cut the interaction surface at the above angles to obtain the N - M interaction diagram. Then, determine the resultant moment on the N - M curve corresponding to the axial force N . Finally, if Equation (23) is true, the column can withstand the load and vice versa (Figures 8 and 9).

$$M^* = \sqrt{(M_x^*)^2 + (M_y^*)^2} \leq \text{Resultant } M \quad (23)$$

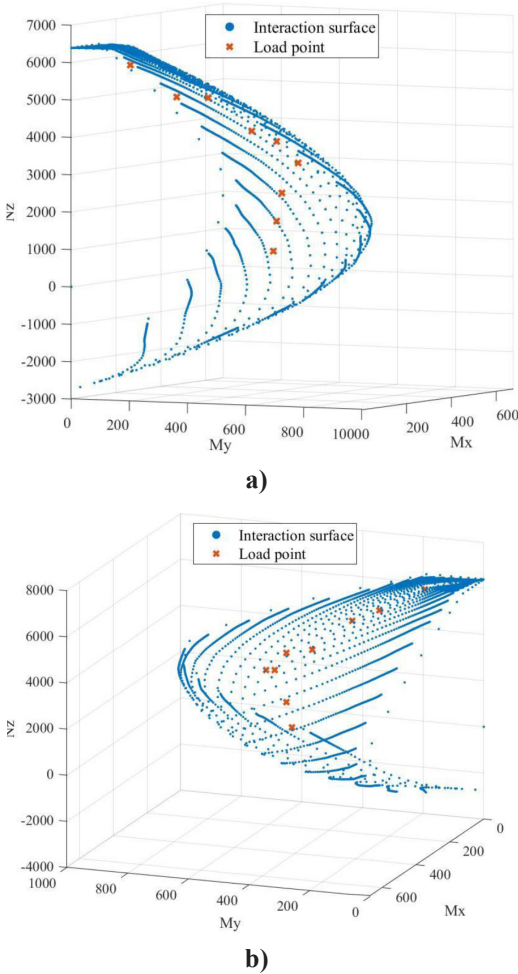


Figure 7. Checking load-bearing capacity using 3D interaction surface: (a) Front view, (b) Rear view.

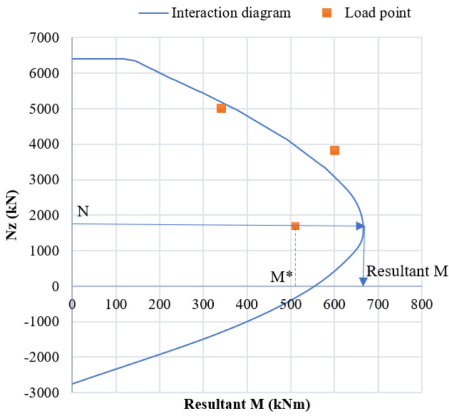


Figure 8. Checking load-bearing capacity using N - M interaction diagram-passed load case.

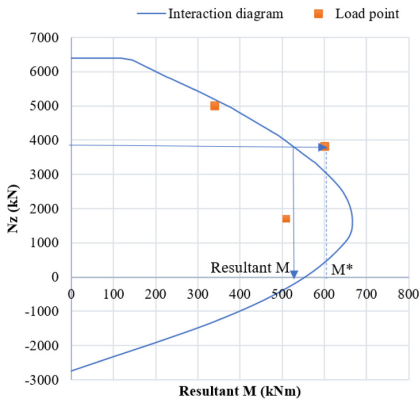


Figure 9. Checking load-bearing capacity using N - M interaction diagram-failed load case.

3. MODEL VERIFICATION AND APPLICATION

The 3D model generated by the proposed program is verified in comparison to the result from ETABS ver18.1.1 and CSI-Col ver10.0.0.

3.1. Verification for a square column

A column has a square cross-section of 70 cm x 70 cm, concrete grade B25, reinforcement CB400-V, 28 Ø 25 mm arranged evenly along the perimeter. The distance between the edge and the central of steel bars $a_0 = 4$ cm. The length of the column is 7 m. Work condition coefficient of concrete $\gamma_b = 0.85$.

The proposed program, ETABS, and CSI-Col give the same results, with an error of less than 7.2%. There are no noticeable differences in the uniaxial scenario. The maximum error occurs at the balance point, the transition between compression-controlled and tension-

controlled points, in case the neutral axis angle is 45° . The discrepancy can happen because of the method for calculating the force and moment due to the concrete compression area, stress-strain relationships, failure criteria of concrete, and work condition coefficients. In the proposed program, the maximum value of N is 9717.18 kN due to the coefficient φ , as mentioned in Equation (10).

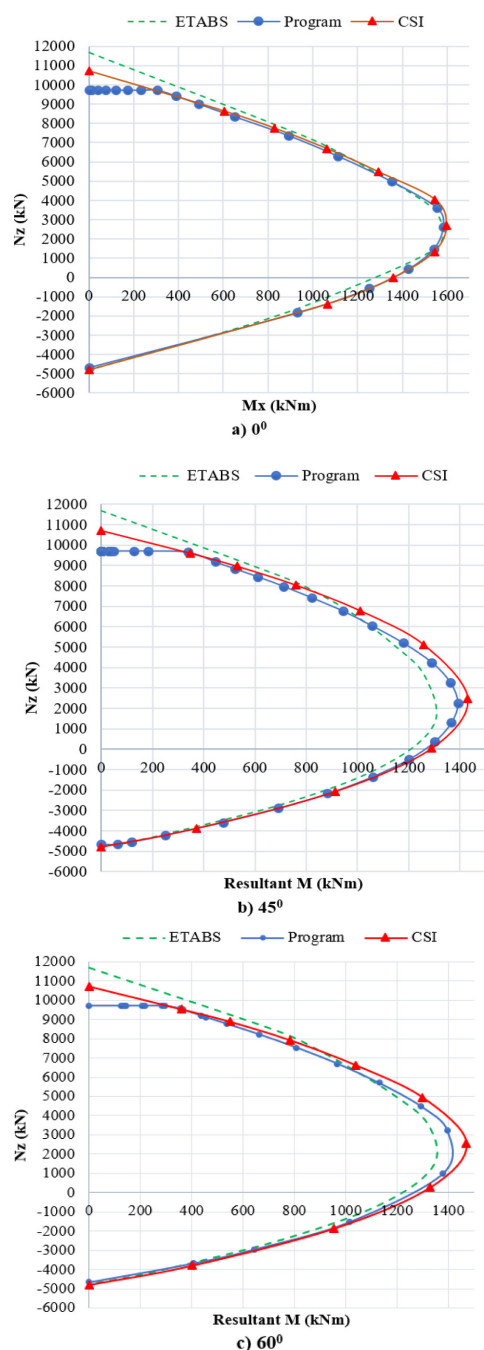


Figure 10. N - M Interaction diagram of square column for variety neutral axis angles: (a) $\alpha = 0^\circ$, (b) $\alpha = 45^\circ$, (c) $\alpha = 60^\circ$.

The proposed program, ETABS, and CSI-Col give the same results, with an error of less than 7.2%. There are no noticeable differences in the uniaxial scenario. The maximum error occurs at the balance point, the transition between compression-controlled and tension-controlled points, in case the neutral axis angle is 45° . The discrepancy can happen because of the method for calculating the force and moment due to the concrete compression area, stress-strain relationships, failure criteria of concrete, and work condition coefficients. In the proposed program, the maximum value of N is 9717.18 kN due to the coefficient φ , as mentioned in Equation (10).

3.2. Verification for a rectangular column

A column has rectangular cross-section 0.7 m x 0.5 m, concrete grade B25, reinforcement CB400-V, 16 $\varnothing$ 25 mm as shown in Figure 11. The distance between the edge and the central of steel bars $a_0 = 4$ cm. The length of the column is 4.5 m. Work condition coefficient of concrete $\gamma_b = 0.85$.

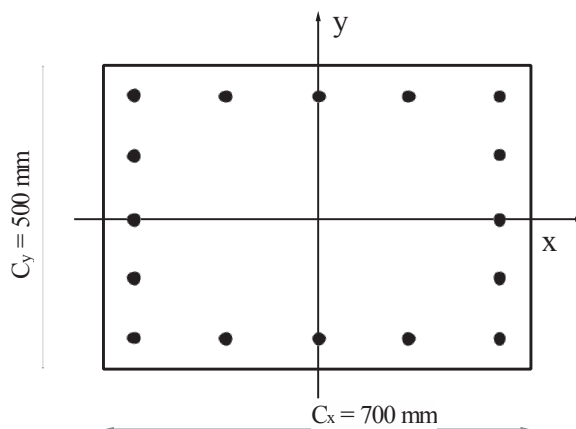


Figure 11. Rectangular column's cross-section.

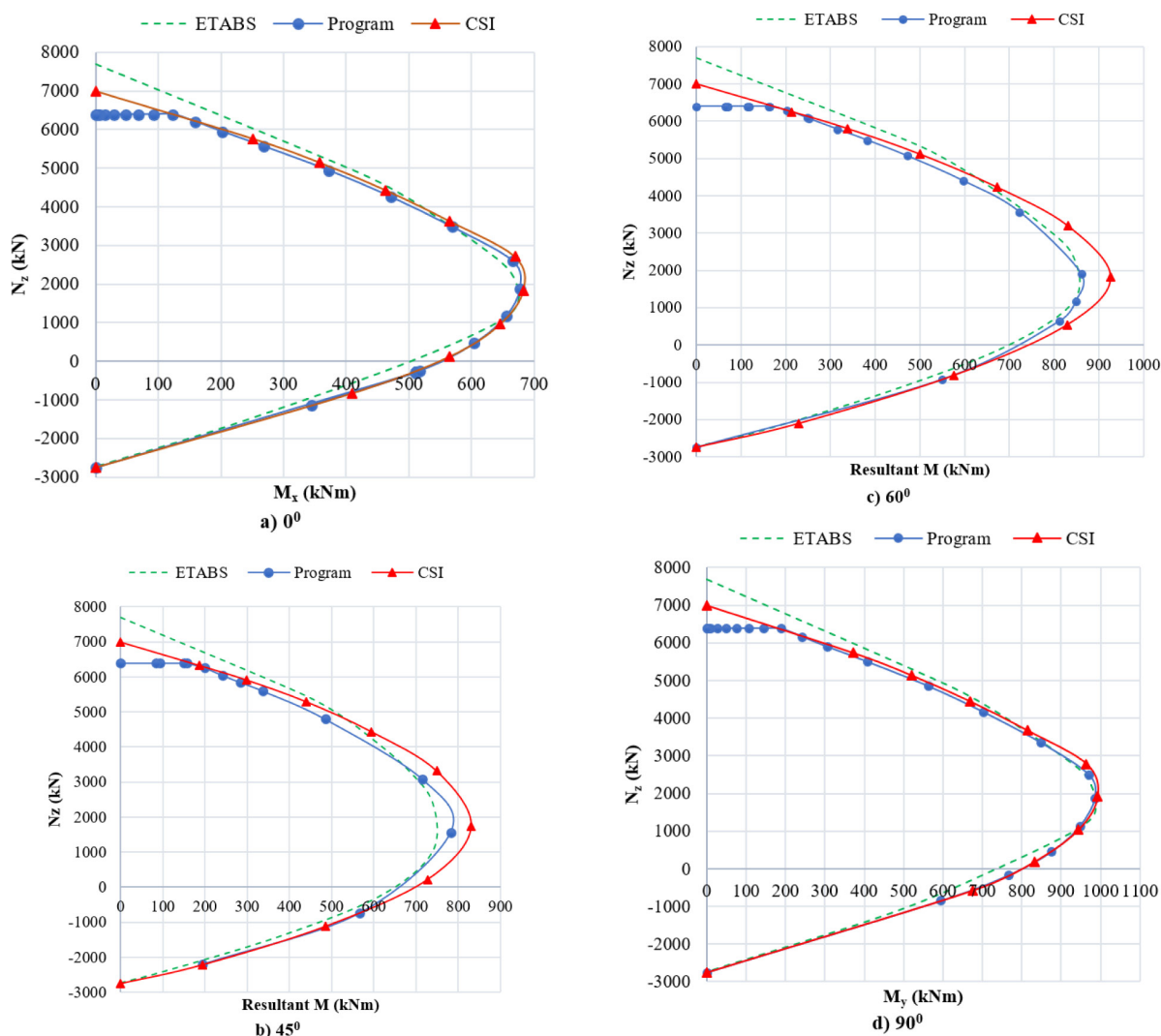


Figure 12. N - M Interaction diagram of rectangular column for variety neutral axis angles: (a) $\alpha = 0^\circ$, (b) $\alpha = 45^\circ$, (c) $\alpha = 60^\circ$, (d) $\alpha = 90^\circ$

Notice: The relationship between the neutral axis angle and the point's coordinates is

$$\tan \alpha = \frac{M_{uy} I_x}{M_{ux} I_y}.$$

As can be seen from Figure 12, in cases where the neutral axis angle is 0° and 90° , the discrepancies between the three software are negligible. The max error occurs at the balance point in case the neutral axis angle is 60° and is less than 7.2%. The maximum axial forces that the column can withstand when considering the buckling factor is 6394.64 kN.

Compared with other research, where the errors are up to 16.8% for square column¹³ and

8.74 for rectangular column,^{13,15} the proposed program errors are less than 7.2% for both types of cross-section. As a result, the proposed program is reliable. It can be used in practical design, further research and teaching.

3.3. Application for checking load-bearing capacity

The column is as shown in the subsection 3.2. The internal moments and axial forces from the load combinations are given in Table 1. Internal moments of dead load are $M_{Llx} = 80$ kNm, $M_{Lly} = 300$ kNm.

Question: Checking the load-bearing capacity of the column.

Table 1. Forces and moments from load combinations.

Combo	N (kN)	M _x (kNm)	M _y (kNm)
1	5400	90	315
2	5310	180	360
3	6300	45	180
4	4320	270	450
5	4050	270	540
6	2700	180	630
7	1800	270	540
8	900	315	495
9	3420	315	585

The load points are determined from Equations (11-23), as shown in Table 2.

Table 2. Load points after considering slender column effect.

Combo	N (kN)	M _x [*] (kNm)	M _y [*] (kNm)	M [*] (kNm)
1	5400	134.7	326.1	352.8
2	5310	190.3	372.2	418.1
3	6300	162.2	188.4	248.6
4	4320	282.3	462.3	541.7
5	4050	281.6	553.8	621.3
6	2700	185.5	641.6	667.9
7	1800	275.7	547.2	612.8
8	900	319.0	498.9	592.2
9	3420	326.6	598.0	681.4

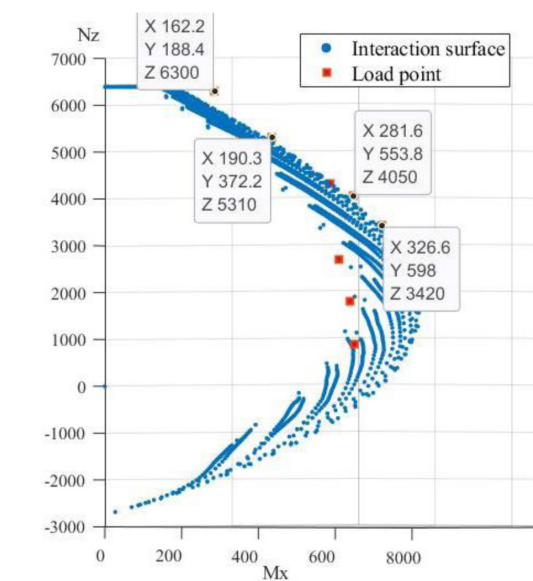


Figure 13. Checking load-bearing capacity using 3D interaction surface.

Using 3D interaction surface, the program points out 4 load points outside of the interaction volume, as shown in Figure 13. They come from combination 2, 3, 5 and 9.

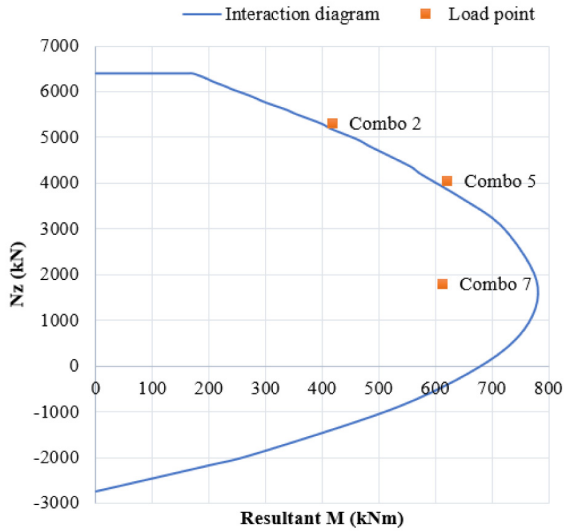


Figure 14. Checking load-bearing capacity using N-M interaction diagram ($\alpha = 45^\circ$).

Notice: The relationship between the neutral axis angle and the point's coordinates is $\tan \alpha = \frac{M_y^* I_x}{M_x^* I_y}$.

The integrated module in the program to checking the load-bearing capacity using N-M curve yields the result in Table 3.

Table 3. Checking load-bearing capacity using N-M interaction diagram.

Com	M [*] (kNm)	Angle (degree)	Resultant M	Check
1	352.8	51	386.1	Pass
2	418.1	45	395.8	Fail
3	248.6	31	174.3	Fail
4	541.7	40	542.4	Pass
5	621.3	45	594.0	Fail
6	667.9	60	815.7	Pass
7	612.8	45	776.6	Pass
8	592.2	39	720.9	Pass
9	681.4	43	664.9	Fail

As can be seen from Figure 14 and Table 3, the two methods of checking load-bearing capacity show identical results.

4. CONCLUSIONS

This study develops a program to generate 3D interaction surfaces of rectangular reinforced concrete columns subjected to axial force and biaxial bending. The method is explained in detail, using non-linear deformation models and considering the buckling factor per Vietnamese standard 5574: 2018. The program also includes a module to evaluate the load-bearing capacity of the columns.

The proposed method and program are highly accurate compared to imported software such as ETABS and CSI-Col. The most discrepancy is less than 7.2% for both square and rectangular columns. In addition, 3D views allow the program to check fast the column capacity subjected to various biaxial bending and axial load cases based on TCVN 5574:2018.

A step-by-step procedure to calculate the coefficient of slenderness effect η_x , η_y was introduced. Two simple methods to check the column capacity using 2D interaction diagrams is presented and incorporated into the proposed program. An example of cross-checking the results from 3D views and 2D diagrams was conducted and showed identical checking results.

The proposed program can be used for further research, including investigating the effects of initial conditions on the column capacity, such as uncertainty of materials properties, effective cover, work condition coefficients, stress-strain relationship, etc. However, there is a requirement for validating the program using experiment results.

Acknowledgment

This research is conducted within the framework of science and technology projects at institutional level of Quy Nhon University under the project code T2023.806.16.

REFERENCES

1. R. Park, T. Paulay. *Reinforced concrete structures*, John Wiley and Son, New York, 1975.
2. C. D. Nguyen. *Calculate the cross-section of reinforced concrete columns*, Construction Publisher, Hanoi, 2006.
3. B. Bresler. Design criteria for reinforced columns under axial load and biaxial bending, *Journal of the American Concrete Institute*, **1960**, 481-490.
4. A. L. Parme, J. M. Nieves, A. Gouwens. Capacity of reinforced rectangular columns subject to biaxial bending, *ACI Journal*, **1966**, 63(9), 911-923.
5. V. Mavichak, R. W. Furlong. *Strength and stiffness of RC columns under biaxial bending*, Texas State 5 Department of Highways and Public Transportation, Transportation Planning Division, USA, 1976.
6. T. Hsu. Analysis and design of square and rectangular columns by equation of failure surface, *ACI Structural Journal*, **1988**, 85(2), 167-179.
7. V. H. Hung, D. X. Pham, H. T. Nguyen. *Method to create interaction diagram and calculate amount of reinforcement for biaxial eccentric compression members*, Conference on the 50th Anniversary of Establishment of Vietnam Institute for Building Science and Technology, Hanoi, 2013.
8. L. T. Pham. Creating interactive diagrams for reinforced concrete columns under eccentric compression in Excel, *Vietnam Journal of Construction*, **2018**, 12, 8-12.
9. H. P. P. Nguyen. *Examining and evaluating practical calculation methods of reinforced concrete column subjected to biaxial eccentric compression*, Master thesis, Ha Noi University of Civil Engineering, 2007.
10. L. T. N. Nguyen. Calculate biaxial eccentric compression column using approximate method, combined with interactive diagram according to TCVN 5574:2012, *Journal of Science and Technology in Civil Engineering*, **2016**, 3, 62-67.
11. D. T. Nguyen, D. T. Nguyen. Interaction diagram of round column under high temperature, *Journal of Structural Engineering and Construction Technology*, **2017**, 24, 67-73.

12. T. S. S. Hoach, Q. B. Le. Calculating load capacity of reinforced concrete biaxially loaded members with rectangular cross-section according to TCVN 5574:2018, *Vietnam Journal of Construction*, **2020**, 65-68.
13. T. V. Tran, T. T. Pham, N. T. Nguyen, V. N. Pham. Programming software for load bearing capacity of reinforced concrete biaxially loaded members with arbitrary cross-sections based on TCVN 5574:2018, *Journal of Science and Technology in Civil Engineering*, **2019**, 13(4V), 47-57.
14. L. T. Pham. Check the bearing capacity of reinforced concrete columns under eccentric compression with interactive diagrams in Excel, *Vietnam Journal of Construction*, **2018**, 12, 127-130.
15. H. S. S. Thach, T. Q. Le, H. T. T. Le. Calculating the load-capacity of rectangular reinforced concrete columns using non-linear models of TCVN 5574:2018, *Journal of Materials and Construction*, **2022**, 12(4V), 52-55.
16. P. V. Nguyen, V. Sykhampha, T. T. Nguyen. Determination of load bearing capacity of reinforced concrete columns using material's non-linear models of TCVN 5574:2018, *Journal of Science and Technology in Civil Engineering*, **2020**, 14(3V), 93-107.



© 2024 by the authors. This Open Access Article is licensed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license (<https://creativecommons.org/licenses/by-nc/4.0/>).

Nghiên cứu ứng dụng phần mềm SCADA Designer thiết kế chương trình quản lý nhu cầu sử dụng điện năng

Trần Tiến Đạt, Nguyễn Vũ Hòa, Ngô Minh Khoa*

Khoa Kỹ thuật và Công nghệ, Trường Đại học Quy Nhơn, Việt Nam

*Ngày nhận bài: 27/03/2024; Ngày sửa bài: 18/07/2024;
Ngày nhận đăng: 08/09/2024; Ngày xuất bản: 28/10/2024*

TÓM TẮT

Quản lý nhu cầu sử dụng điện năng (DSM) là một trong những thách thức chính của lưới điện nói chung và lưới điện thông minh nói riêng. Nó sẽ góp phần hỗ trợ khách hàng và đơn vị sản xuất, kinh doanh sử dụng điện một cách hiệu quả, tiết kiệm thông qua các giải pháp phân bổ lại nhu cầu phụ tải giờ cao điểm, thấp điểm và bình thường. Bài báo này nghiên cứu ứng dụng phần mềm SCADA Designer để thiết kế chương trình điều khiển và giám sát DSM cho một hệ thống điện trong phạm vi phòng thí nghiệm. Mô hình thực nghiệm bao gồm các thiết bị phần cứng như: Động cơ không đồng bộ ba pha rô to lồng sóc, động cơ servo, tải điện trở, đồng hồ đo lường và các thiết bị phụ trợ khác được kết nối giả lập cho mô hình phụ tải hỗn hợp của khách hàng sử dụng điện. Bài báo đã đưa ra lưu đồ thuật toán để thiết kế chương trình phần mềm bằng SCADA Designer. Ngoài ra, đồ thị phụ tải cũng được giả định bởi việc điều khiển động cơ servo nhằm thiết lập các kịch bản thí nghiệm kỹ thuật DSM. Các trường hợp thí nghiệm thực tế dựa trên hệ thống mô hình đã kiểm chứng được hiệu quả làm việc của kỹ thuật DSM trong bài báo này, đồng thời nó còn góp phần giảm công suất cực đại giờ cao điểm của hệ thống điện, loại bỏ các đỉnh phụ tải của hệ thống điện.

Từ khóa: *Lưới điện thông minh, quản lý nhu cầu sử dụng điện, tiết kiệm năng lượng, SCADA Designer.*

*Tác giả liên hệ chính.

Email: ngominhkhoa@qnu.edu.vn

A study of SCADA Designer application for developing demand side management program

Tran Tien Dat, Nguyen Vu Hoa, Ngo Minh Khoa*

Faculty of Engineering and Technology, Quy Nhon University, Vietnam

Received: 27/03/2024; Revised: 18/07/2024;

Accepted: 08/09/2024; Published: 28/10/2024

ABSTRACT

Demand side management (DSM) is one of the main challenges of general power systems and smart grids. It plays a vital role in supporting customers and utilities to use electricity efficiently and economically through techniques that redistribute peak, off-peak and normal load demand. This paper has studied the application of SCADA Designer software to design a DSM controlling and monitoring program for an electrical system in our laboratory. The experimental system includes hardware devices such as a three-phase asynchronous motor with a squirrel cage rotor, a servo motor, a resistive load, a measurement meter and other auxiliary devices that are connected to simulate a mixed load of electricity customers. The paper has developed an algorithm flowchart for designing the software program using SCADA Designer. Additionally, the load profile is controlled by the servo motor to establish various DSM operation scenarios. These scenarios based on the experimental system have verified the working efficiency of the DSM technique in this paper, and it also contributes for reducing the peak load of the power system.

Keywords: *Smart grid, demand side management, energy saving, SCADA Designer.*

1. INTRODUCTION

The economic and environmental challenges posed by the use of fossil fuels have led to a restructuring of the conventional power system. As a result, future grids, known as smart grids, incorporate newer types of digital and high-tech devices that enable two-way communication between the supply and demand sides. These systems offer additional options, such as real-time monitoring, controlling and monitoring between supply and demand, which enhance efficiency, reduce energy consumption and improve system reliability, thereby helping the power system operate safely, flexibly and intelligently. These

benefits are achieved through the application of various demand side management (DSM) techniques.^{1,2}

DSM is a set of technical, technological, economic, social and control solutions to use electricity effectively and economically. DSM is a part of the overall Supply Side Management (SSM) and Demand Side Management (DSM) programs. In previous years, to satisfy the increasing energy demand, people were interested in investing in the exploitation and construction of new power plants. Nowadays, due to the rapid development of electricity demand, the amount of capital invested in the electricity

*Corresponding author.

Email: ngominhkhoa@qnu.edu.vn

industry has become a burden for countries. In addition, the use of coal, oil, gas and other resources in power plants is increasing, leading to increasingly serious environmental pollution. This has resulted in the DSM being recognized as the cheapest and cleanest source of electricity supply process.^{3,4} The DSM helps us to mitigate the investment capital in building new power plants, save resources, and reduce environmental pollution. Besides, thanks to DSM, consumers can also be provided with cheap and higher quality electricity.

In fact, the results of DSM implementation in countries worldwide have concluded that the DSM can reduce more than 10% of energy demand at a cost of only 0.3 to 0.5 of the costs required to build power plants and networks to meet the corresponding amount of electricity. As a result, the DSM brings economic as well as environmental benefits to the country, power sectors and customers. The main objective of the DSM is to reshape the load graph and regulate the maximum and minimum daily demand profile to make the most efficient use of energy sources to relieve the need to build new power plants.⁵ This may result in the direction of energy using during normal hours. Virtually all DSM programs aim to maximize efficiency in order to avoid or slow down the construction of new power plants. In addition, other reasons for implementing the DSM programs are social relationships, environmental reasons and changing customers' electricity habits both during peak and normal hours to save energy.⁶

In order to manage, monitor, collect data and control the electrical system with the application of the DSM techniques, the Supervisory Control and Data Acquisition (SCADA) system is one of the optimal solutions to be applied. The SCADA system helps operators to monitor and control the power system effectively, thereby making accurate and timely control decisions. Besides, the SCADA also contributes to improving the reliability of the power system, optimizing energy use, reducing power loss and operating costs of the power system.^{7,8}

From the above issues, this paper focuses on researching the application of SCADA Designer software to design a DSM program within the laboratory range. The hardware structure is designed based on the laboratory equipment including a three-phase squirrel cage rotor asynchronous motor, a variable resistive load, a servo motor, a power meter, etc. In addition, the SCADA Designer software is applied to design a system monitoring control program based on the DSM technique. The guide user interface (GUI) is designed to make it easy for users to manipulate and monitor the operation of the system in real time.

2. DSM AND SCADA

2.1. DSM programs in Vietnam

The national program on electricity Demand Side Management (DSM) - brings great benefits, of which a basic benefit is that the cost to change, adjust to reduce or save 01 MW of power load capacity during peak hours will be cheaper than the cost to provide 01 MW of additional power capacity by building new power plants or mobilizing high-priced power sources and grid infrastructure systems.⁹

Over the past decade, in the global context, primary energy sources (coal, oil, gas) have been increasingly depleted, in contrast to the growing trend of energy consumption, electricity consumption and climate change. In order to ensure electricity supply security, energy security and sustainable development, in addition to paying attention to supply side development, it is also necessary to pay attention to the demand side. For that reason, Vietnam focuses on building research and promoting the DSM programs. Currently, the DSM programs in Vietnam have gone through many stages and have brought many practical results to the power system.¹⁰ Since the implementation of the DSM stages, the DSM program has mainly implemented programs to change customers' electricity use habits, including programs that reduce electricity use, both during peak and

normal hours, without specifically affecting the quality of service provided to customers. The DSM replaces old equipment with modern equipment to provide services at the same (or higher) level for electricity users (e.g., lighting, heating, cooling, etc.) that consume less electricity.^{11,12}

The electricity using reduction programs are implemented during peak hours in the power system of a power company or a specific area of the power transmission or distribution grid. These programs consist of time-of-use pricing, also known as Time of Use/TOU, and direct load control. The programs that adjust electricity rates, equipment cycles, or power outages in response to changes in energy costs or energy sources can provide flexibility in the shape of the load graph, which can be achieved through instant pricing and pro rata pricing of time spent on electricity. Furthermore, breakable load schedules, live load control, and other load management programs may be included as long as they do not interfere with peak load periods.¹³

The programs that change electricity rates, equipment cycles, or power outages in response to specific changes in energy costs or energy sources can achieve flexibility in the shape of the load graph. These include instant pricing and pro rata pricing of time spent on electricity. They may also include breakable load schedules, live load control, and other load management programs when these activities are not limited by peak load periods.¹⁴

Electricity load building programs are designed to increase the use of electrical equipment or shift electricity consumption from peak hours to normal hours in order to increase total electricity sales. These programs include increasing electricity use during normal and off-peak hours to help stabilize the system.

2.2. DSM techniques

The DSM programs bring many benefits including electricity customers, electricity suppliers, and society, but electricity customers

will be the biggest ones. For electricity suppliers, the benefits are to increase investment efficiency, power supply reliability, power quality, optimize supply and demand balance, reduce investment pressure to upgrade new power sources and expand the grid, and reduce production and business costs.¹⁴ For customers, the benefits are reducing electricity costs, being served with a higher power quality, receiving certain supports (depending on each specific DSM program), increasing electricity efficiency, and improving production and business efficiency. For society, the advantage is to reduce the pressure of increasing electricity prices, which contributes to the development of a sustainable electricity and energy sector, socio-economic stability, and sustainable environmental development. In order to bring about the above benefits, the DSM program has applied the following techniques: peak load reduction, power load adjustment, off-peak load filling, strategic load growth, load chart shifting, and efficient and economical use of electricity.^{1,2,10} These techniques are shown visually in Figure 1.

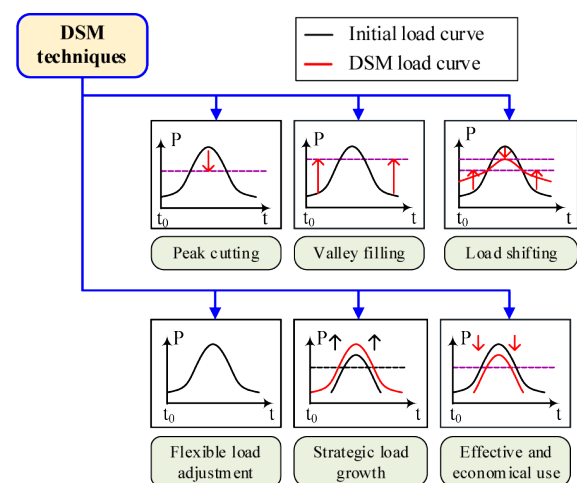


Figure 1. DSM techniques.

Each DSM technique has different advantages and disadvantages. We use the corresponding techniques to bring the best efficiency to customers, depending on the characteristics of the load.^{6,15}

This research model is mainly focused on peak cutting and valley filling techniques as

shown in Figure 2. To analyze the peak cutting method, the initial load factor, the average demand can be calculated using Equations (1) and (2) from the load time curve.

$$LF = \frac{P_{avg}}{P_{max}} \quad (1)$$

$$P_{avg} = \frac{\sum_{i=1}^n P_i}{\sum_{t=1}^{24} T_i} \quad (2)$$

$$C_d = LF * C_e * \sum T \quad (3)$$

where LF is the load factor value; P_{avg} is the average capacity of the load; P_{max} is the maximum capacity of the load; n is the number of loads; P_i is the capacity of all types of loads $i = 1, \dots, n$; T_i is the time period for each type of load $t = 1, \dots, 24$; C_d is the demand cost; C_e is the energy cost.

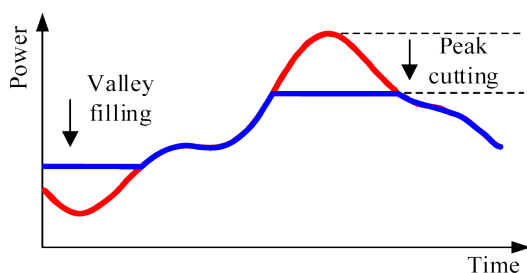


Figure 2. Load curve illustrating DSM program.

2.3. SCADA

SCADA is the process of collecting real-time data from objects to process, perform, store, analyse and potentially manipulate those objects.³⁻⁵ A SCADA system can be divided into two parts: a hardware system for data collection, communication, control and operation, and a software system for data storage, construction, visualization, optimization, alarm management, etc. A SCADA system is shown in Figure 3 and has the following basic components:

- **Remote Terminal Unit (RTU):** the RTU is to perform processing and control tasks in real-time mode (There are two types of real-time systems: hard real-time systems and soft real-time systems). The RTUs range from primitive sensors that collect information

from objects to multiprocessing machines that process information and control it in real-time. The use of a processor-enabled RTU reduces the requirement for the speed of the transmission channel in connection with the control center.

- **Master Terminal Unit (MTU):** the MTU is a dispatch center that performs data processing and control at a high level in real-time mode. One of the basic functions of MTU is to provide an interface between people and observers of the system. The MTU can take different forms, from a single computer with old devices to a mainframe system including servers and clients.

- **Communication System (CS):** the communication channel is required to transmit data from remote locations to the MTU and transmit the control signal to the RTU.

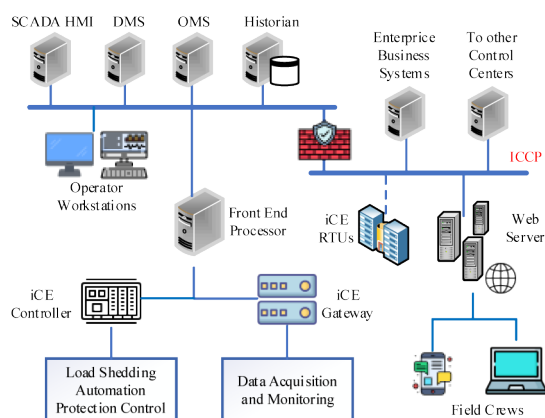


Figure 3. SCADA system model.³⁻⁵

At the first level of the SCADA system, there are sections with monitoring and control functions for each individual device. The most common component in the power system is the protection relays. When the equipment is in trouble, these relays can completely calculate and act on predetermined parameters without communicating with the superior system. In addition to the control function, the elements of this level also have the function of collecting data and parameters from the devices to send to the servers at the station. In modern systems, these elements are collectively referred to as intelligent electronic devices (IDEs) that have

different working principles and functions but have the same communication standards, allowing this IED to exchange information with other IEDs in the same station (peer to peer) and communicate with the server of the station. In principle, failure of an IED in maintenance will not affect other IEDs in the system.¹⁶

The second level of the SCADA system is the Substation Server, with the main function of collecting data from the IEDs managed by it, stored in the database, serving the needs of reading local data via HMIs (Human Machine Interfaces).

The third level is the Control Center of the whole system, which collects data from Substation Servers, performs calculation functions to assess the state of the system, forecast load demand, and performs important control functions such as redistributing power generation between factories and planning the operation of the whole system.

Due to the large scale of the power transmission system, the central control stations can also be divided into levels - central control center (Central Control Center or Central Dispatching Center) and area control stations (Area Control Center). Currently, people often use the SCADA/EMS system.¹⁷

3. EXPERIMENTAL SYSTEM

3.1. Hardware configuration

The experimental system within the laboratory range includes some basic equipment such as a power source, a power meter, a multifunctional relay and loads such as resistive loads and motor loads in addition to some other loads such as capacitive loads, inductive loads, bulb loads. The technical specifications of these devices are shown in Table 1. Here we assume the motor load is the critical load and continuous operation, and the resistive load is the less critical load and can be operated at any time. This model allows us to simulate the typical process of energy management through two techniques of bottom lifting and peak cutting, thereby helping us

understand the process of managing electricity demand. The hardware equipment system and the wiring diagram between the hardware devices are shown in Figure 4 and Figure 5.

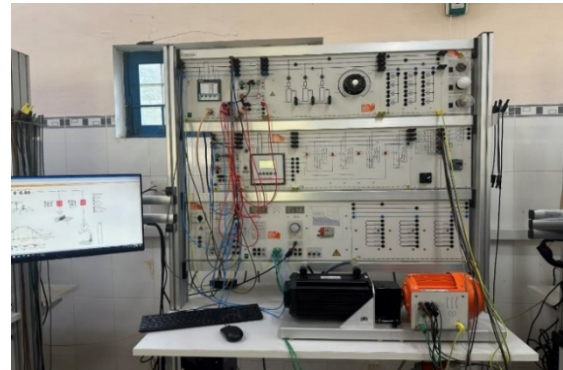


Figure 4. Hardware configuration.

The system consists of two loads: the static load is the resistive load and the dynamic load is the three-phase asynchronous motor of the squirrel cage rotor. In this system, the critical load requires continuous operation and the resistive load is a secondary load and less important can be operated at any time. The system operates through the SCADA Viewer software. Here, we set the maximum total power value and the minimum total power when running in auto DSM mode, and then the system will run according to a pre-programmed cycle. When the maximum power value is set, the system will cut the resistive load out of the system, leading to a decrease in power and when the power is less than the minimum value when set, the system closes the static load into the cycle. The bottom part is the interface of a simple electricity demand management simulation program.

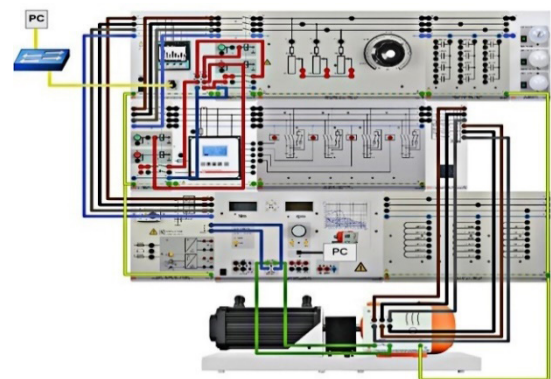


Figure 5. Wiring diagram of test system.

Table 1. The technical specifications.

No.	Device name	Technical specifications
1	Three-phase power meter	Siemens PAC4200
2	Servo power tester	$P_{nominal} = 1.4 \text{ kW}$; $U_{nominal} = 390 \text{ V}$; $I_{nominal} = 3.3 \text{ A}$; $n = 2000 \text{ rpm}$; $\cos\varphi = 0.75$
3	Three-phase asynchronous motor	$P_{nominal} = 1 \text{ kW}$; $U_{nominal} = 390/690 \text{ V}$; $I_{nominal} = 2.1/1.2 \text{ A}$; $n = 2900 \text{ rpm}$; $\cos\varphi = 0.83$
4	Resistive load	$P_{max} = 1 \text{ kW}$

3.2. Software program

The software program is designed based on the algorithm flowchart as shown in Figure 6. This flowchart consists of the following basic steps:

Step 1: Start

Step 2: Initialize the initial signal and set the parameters of the system. In this step, the program will take the input signals such as voltage, frequency, current, and power from the meters and also receive the initial set values such as the total maximum power ΣP_{Max} , the total minimum power ΣP_{Min} , the load profile and other set values from the SCADA control monitor.

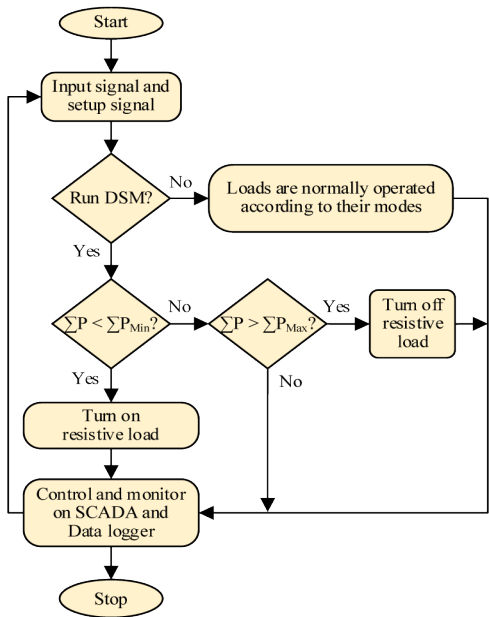


Figure 6. Algorithm flowchart.

Step 3: Run the DSM technique. The step of running the DSM technique is that the program will check whether the DSM mode is active or not. If the DSM mode does not work, the program will return the wrong (NO) conditions, then the loads will operate normally according to the operating mode of each load and return the data to the SCADA control monitor. Conversely, when the DSM mode is activated, the program returns the true (YES) condition and continues to check the next condition.

Step 4: Check conditions in the DSM mode. In this step, the program will check the condition of the total load power (ΣP) with the minimum total load power ΣP_{Min} . If the total load power is less than the total minimum load power ($\Sigma P < \Sigma P_{Min}$) then the program will return the true (YES) condition and turn on the resistive load. In contrast, if the total load power is greater than the total minimum load power ($\Sigma P > \Sigma P_{Min}$), the program will return the wrong (NO) condition and continue to check the condition of the total load power with the total maximum load power (ΣP_{Max}). Under this condition, if the total load power is greater than the total maximum power ($\Sigma P > \Sigma P_{Max}$), the program will return the true (YES) condition and turn off the resistive load. Conversely, if the total load power is less than the total maximum load power ($\Sigma P < \Sigma P_{Max}$), the program will return the wrong (NO) condition and send a signal to the SCADA monitor and controller, then continue to check the original condition. In this step, when turning off or turning on the resistive load, the regulator returns the data to the SCADA control monitor.

Step 5: Stop.

3.3. GUI design

The SCADA Designer software is applied to design the GUI of the SCADA program in this paper. The SCADA Designer is a software with a hardware key that is equipped in the laboratory to assist in the design of controlling and monitoring programs for electrical grids in the laboratory range.¹⁸ The GUI is shown as Figure 7. It consists

of the main parts numbered on the Figure 7 and has the following basic function:

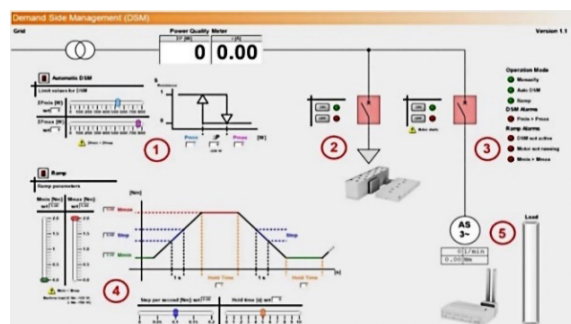


Figure 7. SCADA program GUI.

Block 1 is the load adjustment operation and the configuration of the limits. In this block, the user can mainly set the minimum and maximum load power (ΣP_{Min} and ΣP_{Max}) for the DSM operating mode. With this model, we configure the ΣP_{Max} value range to be 800W and minimum and maximum load power must strictly comply with the following conditions:

$$\Sigma P_{Min} < \Sigma P_{Max} \quad (4)$$

Our set values are visually represented on the register and on a descriptive chart. The graph clearly shows the limit values and also the power difference ΔP .

Block 2 is the manual load control via the load switches. The load blocks are used to describe two typical loads of the model, motor load and resistive load. Motor load is considered a critical load of the system and has a predominant operating time. The impedance load is the secondary load, which is less important and is turned on at any time during the operation of the system. Here the resistive load can be replaced by other types of loads such as lamp loads, capacitive loads, etc. On the GUI, the load block is operated in two automatic and manual modes through emulator buttons on the interface.

Block 3 is to display the mode and the operating status. The status block visually represents the functions in operation.

Block 4 is the ramp gradient configuration for the motor load. The ramp block allows the motor load to operate according to the

pre-setting curve with the established parameters. This includes settings such as the torque, holding time, and step. Thereby it is easy to observe the load demand controlling process.

Block 5 shows the load level of the dynamic load.

The data logger window in Figure 8 allows us to record the operation parameters of the loads in the form of a profile so that we can easily observe the operation in the different modes. From there, it is easy to assess the operating status of the system as well as the power management process.¹⁹

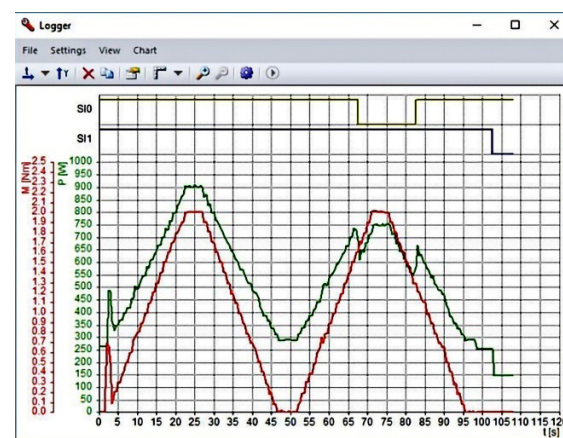


Figure 8. Data logger logging window.

4. RESULTS AND DISCUSSION

The experimental results from the test system are analysed and discussed in this section. From these results, it can be seen how the two techniques of peak cutting and bottom lifting work in the process of power demand management. When the system is operated in the auto mode or DSM mode, it is easy to see that when the total load power (ΔP) reaches the maximum load power (ΣP_{Max}), the resistive load is turned off from the network and when the total load power (ΣP) is reduced to the value lower than the minimum load power (ΣP_{Min}), the resistive load is turned on. Thereby, in order to stabilize the supply capacity and help the loads operate most efficiently and optimally, thereby helping to save economically for both the supply side and the demand side. The two cases studied in this paper are as follows:

Case 1: Without DSM

In Case 1, the DSM technique is not activated, the experimental results are shown in Figure 9. The torque and total load power curves in Figure 9c illustrate the motor torque and the load power profile during three working cycles. From this figure, it can be observed that the total load power is operated according to a given curve, with a peak load power of approximately 900 W and a minimum load power of approximately 400 W. The difference between the peak and valley load power is significant. Consequently, in conjunction with the algorithm flowchart, it can be deduced that when the DSM mode is not activated, these loads operate independently with their respective modes, resulting in a large power difference. Simultaneously, when examining the switching state chart of the loads in Figures 9(a), (b) the SI0 and SI1 switch states remain at the constant value of 1 throughout the operation period.

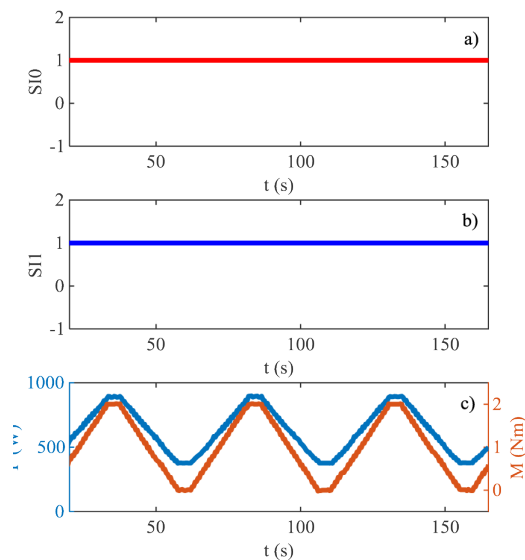


Figure 9. Experimental results without DSM: (a) SI0 contact state, (b) SI1 contact state, (c) Total load power and the torque of the motor.

Case 2: With DSM

In Case 2, the DSM mode is activated, the motor torque and total load power curves in Figure 10c, this figure shows that the difference between peak and valley load power is reduced in this case compared with the first case. This reduction is due to the application of the peak

cutting and valley filling techniques in the DSM program. The program sets the total maximum load power to 760 W and the total minimum load power to 530 W, resulting in a difference of 230 W between these two values. The curve clearly demonstrates this technique, as the load power quickly decreases when the total load power reaches 760 W, and increases again when it falls below 530 W. This corresponds to Step 4 of the algorithm flowchart. In this step, the program checks the conditions of the DSM mode and, upon detecting a decrease in total load power to the minimum load power of 530 W, it will turn on the resistive load. This causes the power curve to increase again. When the graph reaches the maximum total load power value of 760 W, the system will turn off the resistive load, leading to a decrease in the curve. This process is repeated throughout the operation of the system, as shown in Figure 10c. Figures 10(a), (b) depict the switching state of the loads. When the DSM technique is applied, the SI0 contact continuously switches its status between 0 and 1, due to continuous switching of the resistive load during working cycles. The SI1 contact is always maintained at the level 1, because it represents a critical load. Therefore, the DSM technique allows for flexible operations of the load and reduces the large difference between peak and valley load power, resulting in a flatter aggregate load graph.

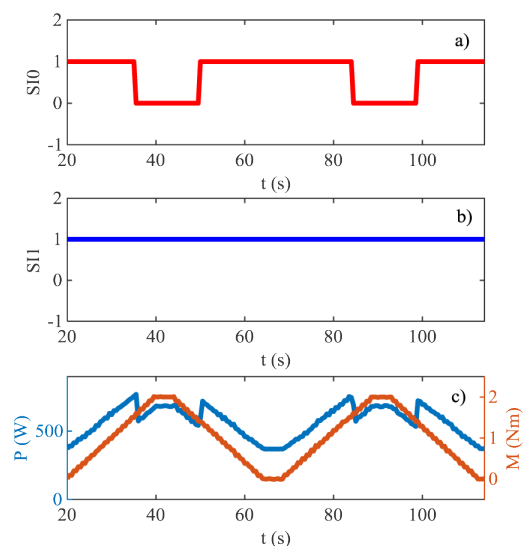


Figure 10. Experimental results with DSM: (a) SI0 contact state, (b) SI1 contact state, (c) Total load power and the torque of the motor.

5. CONCLUSION

This paper applies SCADA Designer software to design a program for demand side management (DSM) technique. The article studies the calculation of energy saving solutions for factories and enterprises to bring economic efficiency to companies and avoid unnecessary loss of electrical energy through the flexible adjustment of load using SCADA. In addition, the research results show that the application of the SCADA program for energy management is an inevitable trend. It helps the operation cycles to communicate with each other, thereby creating an energy ecosystem. Combining SCADA with typical energy management measures such as DSM not only benefits companies and businesses but also reduces capacity demand, saves electricity, and reduces CO₂ emissions. This helps to mitigate the greenhouse effect. Furthermore, it makes it easy to manage special electricity consumption activities, which is very important for energy management for the national power system.

REFERENCES

1. F. N. Budiman, M. A. M. Ramli, H. R. E. H. Boucekara, A. H. Milyani. Optimal scheduling of a microgrid with power quality constraints based on demand side management under grid-connected and islanding operations, *International Journal of Electrical Power and Energy Systems*, **2024**, 155, 109650.
2. R. Elazab, A. T. Abdelnaby, A. A. Ali. Impacts of multiple demand-side management strategies on microgrids planning: a literature survey, *Clean Energy*, **2024**, 8(1), 36-54.
3. N. Ortiz, A. A. Cardenas, A. Wool. SCADA world: An exploration of the diversity in power grid networks, *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, **2024**, 8(1), 1-32.
4. R. Pandit, J. Wang. A comprehensive review on enhancing wind turbine applications with advanced SCADA data analytics and practical insights, *IET Renewable Power Generation*, **2024**, 18(4), 722-742.
5. V. L. Srinivas, J. Wu, B. Singh, S. Mishra. Hybrid state-estimation in combined heat and electric network using SCADA and AMI measurements, *International Journal of Electrical Power and Energy Systems*, **2024**, 156, 109726.
6. H. Sun, X. Cui, H. Latifi. Optimal management of microgrid energy by considering demand side management plan and maintenance cost with developed particle swarm algorithm, *Electric Power Systems Research*, **2024**, 231, 110312.
7. M. Wadinger, M. Kvasnica. Adaptable and interpretable framework for anomaly detection in SCADA-based industrial systems, *Expert Systems With Applications*, **2024**, 246, 123200.
8. M. Wolsink. Conceptualizations of smart grids—anomalous and contradictory expert paradigms in transitions of the electricity system, *Energy Research and Social Science*, **2024**, 109, 103392.
9. M. Yin, H. Cai, A. Gattiglio, F. Khayatian, R. S. Smith, P. Heer. Data-driven predictive control for demand side management: theoretical and experimental results, *Applied Energy*, **2024**, 353, 122101.
10. N. M. Khoa, T. T. Dat, N. V. Hoa. An application of genetic algorithm for optimizing demand side management problem in distribution network systems, *Journal of Applied Science and Engineering*, **2024**, 27(10), 3283-3292.
11. M. S. Bakare, A. Abdulkarim, M. Zeeshan, A. N. Shuaibu. A comprehensive overview on demand side energy management towards smart grids: challenges, solutions, and future direction, *Energy Informatics*, **2023**, 6(4), 1-59.
12. S. Althaher, P. Mancarella, J. Mutale. Automated demand response from home energy management system under dynamic pricing and power and comfort constraints, *IEEE Trans Smart Grid*, **2015**, 6(4), 1874-1883.
13. D. A. Aviles, J. Pascual, F. Guinjoan, L. Marroyo, P. Sanchis, M. P. Marietta. Low complexity energy management strategy for grid profile smoothing of a residential grid-connected microgrid using generation and demand forecasting, *Applied Energy*, **2017**, 205, 69-84.

14. A. Arteconi, N. J. Hewitt, F. Polonara. State of the art of thermal storage for demand-side management, *Applied Energy*, **2012**, 93, 371-389.
15. G. Strbac. Demand side management: benefits and challenges, *Energy Policy*, **2008**, 36(12), 4419-4426.
16. J. Reeser, T. Jankowski, G. M. Kemper. Maintaining HMI and SCADA systems through computer virtualization, *IEEE Transactions on Industry Applications*, **2015**, 51(3), 2558-2564.
17. M. S. Thomas, P. Kumar, V. K. Chandna. Design, development, and commissioning of a supervisory control and data acquisition (SCADA) laboratory for research and training, *IEEE Transactions on Power Systems*, **2004**, 19(3), 1582-1588.
18. T. Asenov. *Improvement of a laboratory micro/nanogrid SCADA system*, The 12th Electrical Engineering Faculty Conference, BulEF2020, Varna, Bulgaria, 2020.
19. N. M. Khoa, L. V. Dai, N. A. Toan, D. D. Tung. A new design of IoT-based network architecture for monitoring and controlling power consumption in distribution grids, *International Journal of Renewable Energy Research*, **2021**, 11(3), 1460-1468.



© 2024 by the authors. This Open Access Article is licensed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license (<https://creativecommons.org/licenses/by-nc/4.0/>).

Ứng dụng học đồ thị trong hệ gợi ý du lịch

Vũ Sơn Lâm^{1,*}, Lê Quang Hùng¹, Trần Đình Luyện¹,
Phan Thị Thùy², Tăng Nhật Hưng¹

¹Khoa Công nghệ thông tin, Trường Đại học Quy Nhơn, Việt Nam

²Khoa Toán và Thống kê, Trường Đại học Quy Nhơn, Việt Nam

Ngày nhận bài: 01/07/2024; Ngày sửa bài: 10/09/2024

Ngày nhận đăng: 12/09/2024; Ngày xuất bản: 28/10/2024

TÓM TẮT

Trong bối cảnh ngành du lịch Việt Nam đang phát triển mạnh mẽ, việc hỗ trợ du khách lựa chọn các điểm tham quan phù hợp với sở thích cá nhân trở nên rất quan trọng. Bài báo này nghiên cứu ứng dụng học đồ thị trong hệ gợi ý, từ đó xây dựng một hệ thống gợi ý du lịch, nhằm đưa ra các gợi ý về điểm tham quan dựa trên các đánh giá của người dùng. Hệ thống gợi ý được xây dựng dựa trên Graph Neural Network, cụ thể là thuật toán LightGCN, một kỹ thuật học máy tiên tiến cho phép học các đặc trưng và mối quan hệ từ dữ liệu có cấu trúc đồ thị. Chúng tôi đã thực nghiệm thuật toán trên ba bộ dữ liệu khác nhau trước khi áp dụng vào hệ thống thực sự. Hệ thống đang được triển khai và thử nghiệm trên mạng Internet.

Từ khóa: Hệ gợi ý, gợi ý, học đồ thị, mạng Nơ-ron đồ thị, du lịch.

*Tác giả liên hệ chính.

Email: vusonlam@qnu.edu.vn

Application of graph learning in travel recommender system

Vu Son Lam^{1,*}, Le Quang Hung¹, Tran Dinh Luyen¹,
Phan Thi Thuy², Tang Nhat Hung¹

¹Department of Information Technology, Quy Nhon University, Vietnam

²Department of Mathematics and Statistics, Quy Nhon University, Vietnam

Received: 01/07/2024; Revised: 10/09/2024

Accepted: 12/09/2024; Published: 28/10/2024

ABSTRACT

In the context of Vietnam's rapidly growing tourism industry, supporting tourists in choice of attractions that suit their personal interests has become very important. This paper studies the application of graph learning in recommender systems, thereby building a travel recommender system, aiming to provide suggestions about attractions based on user ratings. The recommender system is built on Graph Neural Network, specifically the LightGCN algorithm, an advanced machine learning technique that allows learning features and relationships from graph-structured data. We have tested the algorithm on three different datasets before applying it to the real system. The system is being deployed and tested on the Internet.

Keywords: Recommender systems, recommendations, graph learning, graph neural networks, travel.

1. INTRODUCTION

With the advancement of information technology, an increasing number of products and services are being introduced to the digital space. As the volume of available products and services grows, users face significant challenges in selecting the most suitable ones. On the other hand, personalized recommendations, those based on user profile information, can greatly enhance the likelihood of selecting appropriate products. Therefore, the application of recommender systems (RSs) to provide appropriate suggestions is very important.¹⁻³

To build recommender systems, one of the emerging approaches being adopted is graph

learning. Graph learning is a technique that applies machine learning to graph-structured data.⁴ This is an emerging technique in the field of artificial intelligence and has been developing rapidly in recent years. In practice, the data collected and provided to recommender systems can naturally be represented in a graph structure. Specifically, entities within the recommender system, including users, items, attributes, and contexts, are often tightly interconnected, either explicitly or implicitly, forming either homogeneous or heterogeneous graphs. Moreover, graph learning has the capability to learn complex relationships. In fact, many graph learning techniques have been developed to understand relationships modeled by graphs.⁵⁻⁷ Consequently, applying

*Corresponding author.

Email: vusonlam@qnu.edu.vn

graph learning in recommender systems is one of the research directions that is of interest and urgency today.

In Vietnam, the tourism industry has seen significant investment and robust growth in recent years. Vietnam, with its numerous renowned attractions (such as the ancient town of Hoi An, Sapa, Ha Long Bay, Fansipan Mountain, and more), is recognized internationally as an ideal destination. Travel magazines consistently rank Vietnam among the top attractive destinations, showing the appeal of Vietnamese tourism. Alongside the growth of the tourism industry, choosing sightseeing spots to plan a trip has become increasingly challenging. Travelers often spend considerable time and effort reading reviews on various attractions or consulting friends for travel recommendations. This necessitates the development of systems that can recommend attractions to tourists based on their preferences.

Based on these issues, we have conducted research to develop a tourism recommender system using graph learning. The system is designed to provide recommendations for tourist attractions based on user preference profiles. This system is currently being tested online at the following address: <https://travel.fansis.vn>.

2. RELATED WORKS

Recommendation systems are widely applied across various domains, including online shopping, news reading, travel, music, movie, and social networking. These systems act as information filters, suggesting products or services that users might be interested in, based on their behavior history or the matching characteristics between products/services and user profiles.

Among the many approaches to building recommender systems, Graph Learning is currently attracting significant attention. Graph learning applies machine learning techniques to graph-structured data and is rapidly evolving.⁴ In recent years, several graph-based

recommendation models have been proposed and studied.^{5,6} The application of graph learning to recommender systems can be categorized into three main approaches: (i) Random Walk Based Approaches.^{8,9} These methods explore the graph by simulating random walks to generate sequences of nodes, which are then used for making recommendations. While effective in capturing the connectivity within the graph, these approaches are heuristic-based and lack model parameters for optimization. (ii) Graph Embedding Based Approaches.¹⁰⁻¹³ This method is efficient for analyzing complex relationships embedded in the graph. (iii) Graph Neural Network Based Approaches.¹⁴⁻¹⁸ GNNs apply neural network techniques to graph data. They iteratively aggregate information from a node's neighbors to learn representations. GNN-based models have become increasingly popular due to their ability to effectively model the intricate dependencies in graph data, leading to significant advancements in recommendation performance. In Vietnam, a group of authors proposed using GNN to build session-based recommendation models.¹⁹ Their model was evaluated using common metrics in recommender systems, such as Recall@20 and Mean Reciprocal Rank (MRR)@20, demonstrating the effectiveness of GNNs in capturing user preferences and making accurate recommendations.

3. GRAPH LEARNING IN RS

3.1. Problem definition

The recommender system allows making predictions about user preferences for products/services, thereby suggesting items that users might be interested in. With a graph learning approach, the recommender system can be modeled as a bipartite graph.⁶ Here, each vertex represents either a user or an item, meaning there are two types of vertices: user vertices and item vertices. Each edge connecting a user vertex to an item vertex represents the interaction between the user and the corresponding item. These interactions can be views, purchases, ratings, and more.

Formally, let U denote the set of users and I denote the set of items. We construct a graph $G = (V, E)$ where V is the set of vertices and E is the set of edges representing recorded interactions in history. In other words, $E = \{(u, i) \mid u \in U, i \in I, r(u, i) \neq 0\}$ where $r(u, i)$ denotes the interaction of user $u \in U$ with item $i \in I$. If the interaction is a rating given by user u on item i , we can define u interacting with i when u 's rating on i is greater than a certain threshold θ (where $\theta \geq 1$).

The input to the RS is the set of interactions between each user and item. The task of the RS is to predict items that users may be interested in the future. This can be viewed as a link prediction problem in the bipartite graph G (as illustrated in *Figure 1*). In other words, we need to find a function f , such that $f(u, i)$ represents the likelihood of interaction between user u and item i .

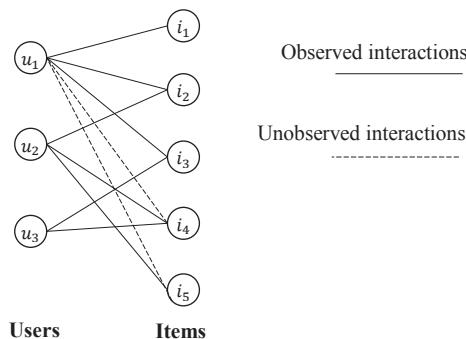


Figure 1. Bipartite graphs represent interactions between users and items.

3.2. The graph neural network approach

The Graph Neural Network (GNN) approach is a method that applies neural networks to data structured as graphs. GNNs leverage this structure to learn features and relationships from the data, offering better efficiency compared to traditional methods. GNNs enable RS to exploit the complex structure of the user-item graph to uncover deeper hidden relationships. For instance, a RS using GNNs can analyze not only the interaction history of users but also the relationships between items, this means capturing collaborative signals from higher-order

neighbors (multi-hop connections), thereby providing more accurate recommendations to each user.⁷

GNNs focus on learning representations of nodes by propagating information from neighboring nodes through multiple iterations on the graph structure and aggregating it. This is similar to the Convolutional Neural Network (CNN) technique, with the key difference being that CNNs operate on grid-like structures where the number and order of neighbors are fixed, whereas in GNNs, the number of neighbors is determined by the historical interaction data.

3.3. Graph neural network architecture

The architecture of the GNN model consists of four components (illustrated in *Figure 2*):^{20,21}

- Embeddings
- Graph Convolution
- Layer Combination
- Prediction

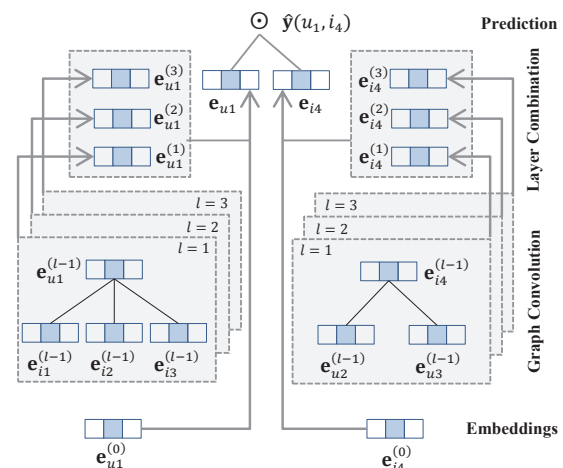


Figure 2. The process of learning representations and predicting links (between node u_1 and i_4) through the GNN model architecture.

First, in the *Embeddings* component, each user and item is represented by an initial embedding vector $\mathbf{e}_u^{(0)}$, $\mathbf{e}_i^{(0)}$, respectively. Here, $\mathbf{e}_u^{(0)} \in \mathbb{R}^d$, $\mathbf{e}_i^{(0)} \in \mathbb{R}^d$, where d is the dimension of the initial embedding vector. These embedding vectors are propagated through the interaction graph to learn and generate the final representation for each corresponding user and item ($\mathbf{e}_u$ and $\mathbf{e}_i$).

Next, in the *Graph Convolution* component, features from neighboring nodes are transmitted along edges to the active node to aggregate and create a new representation for that node. For example, consider the interaction data illustrated in Figure 1, the corresponding graph for the root node u_1 with an interaction depth of 3, as shown in Figure 3. In the first layer, the neighboring nodes of u_1 are i_1, i_2, i_3 , which transmit their information and contributing to the representation of node u_1 ; similarly, this process continues for the second and third layers. Specifically, let $\mathbf{e}_u^{(k+1)}$ be the embedding vector of node u at the $(k + 1)$ -th layer. Then, $\mathbf{e}_u^{(k+1)}$ is constructed based on the following formula:

$$\mathbf{e}_u^{(k+1)} = \text{AGG}(\mathbf{e}_u^{(k)}, \{\mathbf{e}_i^{(k)} : i \in N_u\}) \quad (1)$$

Where,

- AGG is the aggregation function (for instance, using a summation function).
- $\mathbf{e}_u^{(0)}$ and $\mathbf{e}_i^{(0)}$ represent the initial embedding vectors of user u and item i .
- $\mathbf{e}_u^{(k)}$ and $\mathbf{e}_i^{(k)}$ denote the embedding vectors of user u and item i after k layers of propagation.
- N_u represents the set of items interacted with by user u .

Each different specific algorithm in the GNN approach will have a different implementation. For example, in the NGCF algorithm, $\mathbf{e}_u^{(k+1)}$ is specifically calculated as follows:²⁰

$$\mathbf{e}_u^{(k+1)} = \sigma \left(\mathbf{w}_1 \mathbf{e}_u^{(k)} + \sum_{i \in N_u} \frac{1}{\sqrt{|N_u||N_i|}} (\mathbf{w}_1 \mathbf{e}_i^{(k)} + \mathbf{w}_2 (\mathbf{e}_i^{(k)} \odot \mathbf{e}_u^{(k)})) \right) \quad (2)$$

$$\mathbf{e}_i^{(k+1)} = \sigma \left(\mathbf{w}_1 \mathbf{e}_i^{(k)} + \sum_{u \in N_i} \frac{1}{\sqrt{|N_u||N_i|}} (\mathbf{w}_1 \mathbf{e}_u^{(k)} + \mathbf{w}_2 (\mathbf{e}_u^{(k)} \odot \mathbf{e}_i^{(k)})) \right) \quad (3)$$

Where,

- $\mathbf{W}_1, \mathbf{W}_2$: are weight matrices.
- σ : is the nonlinear activation function.
- $\odot$: denotes the element-wise product.

Meanwhile, the LightGCN algorithm simplifies this propagation function by removing the nonlinear activation function σ and weight matrices, $\mathbf{W}_1, \mathbf{W}_2$.²¹ The formula for calculating $\mathbf{e}_u^{(k+1)}$ and $\mathbf{e}_i^{(k+1)}$ in LightGCN is as follows:

$$\mathbf{e}_u^{(k+1)} = \sum_{i \in N_u} \frac{1}{\sqrt{|N_u||N_i|}} \mathbf{e}_i^{(k)} \quad (4)$$

$$\mathbf{e}_i^{(k+1)} = \sum_{u \in N_i} \frac{1}{\sqrt{|N_u||N_i|}} \mathbf{e}_u^{(k)} \quad (5)$$

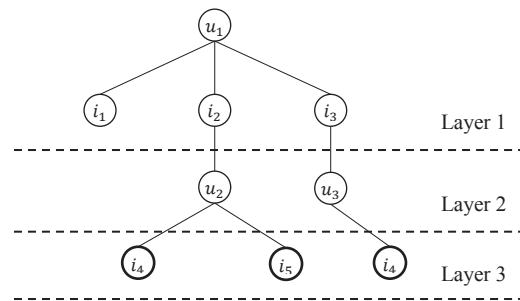


Figure 3. Information propagation from neighboring nodes to node u_1 with 3 layers of connections.

The next part of the GNN is *Layer Integration*. After k layers, we combine the embedding vectors from all layers to form a single representation for the node under consideration. Specifically:

$$\mathbf{e}_u = \text{COMBINE}(\{\mathbf{e}_u^{(k)}, k = [0..K]\}) \quad (6)$$

Here, *COMBINE* can be a concatenation of vectors as used in the NGCF algorithm:

$$\mathbf{e}_u = \mathbf{e}_u^{(0)} \parallel \mathbf{e}_u^{(1)} \parallel \dots \parallel \mathbf{e}_u^{(K)}; \quad \mathbf{e}_i = \mathbf{e}_i^{(0)} \parallel \mathbf{e}_i^{(1)} \parallel \dots \parallel \mathbf{e}_i^{(K)} \quad (7)$$

Or it can be a summation, as in the LightGCN algorithm:

$$\mathbf{e}_u = \sum_{k=0}^K \alpha_k \mathbf{e}_u^{(k)}; \quad \mathbf{e}_i = \sum_{k=0}^K \alpha_k \mathbf{e}_i^{(k)} \quad (8)$$

With $\alpha_k \geq 0$ as a hyperparameter, it represents the importance of the embedding vector at the k -th layer in contributing to the final embedding vector. In the LightGCN algorithm, the authors set $\alpha_k = \frac{1}{K+1}$.

Finally, we have the *Link Prediction* component. Once the embeddings for each

node have been established, the predicted link $\hat{y}_{ui}$ between two nodes u and i , can be simply computed using the inner product:

$$\hat{y}_{ui} = \mathbf{e}_u^T \mathbf{e}_i \quad (9)$$

The higher this result, the stronger the link between u and i . Recommending items to user u is done by calculating the link values between u and various candidate items i . These links are then sorted, and the *top-K* strongest links are selected for recommendation.

3.4. Model training

Using a loss function to optimize parameters in the model. One popular loss function is BPR (*Bayesian Personalized Ranking*) designed to optimize the ranking of recommendations:²²

$$L = - \sum_{(u,i,j) \in D} \ln \sigma(\hat{y}_{ui} - \hat{y}_{uj}) + \lambda \|\Theta\|^2 \quad (10)$$

In which,

- D : is the set of triples (u, i, j) , where (u, i) is the observed interaction of user u with item i , while (u, j) is the interaction of user u with item j that is not observed.

- $\hat{y}_{ui}$ and $\hat{y}_{uj}$: are the predicted values of the strength of the connections between the pairs (user, item): (u, i) and (u, j) , computed by the model.

- σ : is the *sigmoid* function.

- Θ : are the parameters of the model. In the LightGCN algorithm, these parameters are the embedding vectors representing each user and item at layer $k = 0$.

- λ : is the regularization coefficient.

The BPR loss function encourages the model to predict links so that items the users have interacted with (e.g., viewed or rated highly) are ranked higher than items they have not interacted or disliked. Through the above calculation steps, the model will adjust the parameters to minimize the BPR loss function, thereby improving the personalized ranking.

4. TRAVEL RECOMMENDATION APPLICATION

4.1 System architecture

Our system is built using a layered architecture, specifically consisting of four layers: the *Database layer*, the *RS (Recommendation System) layer*, the *Service layer*, and the *Application layer*, as illustrated in Figure 4.

The *Database layer* provides data for the system, comprising two types of data: (i) Rating data for tourist attractions and (ii) Side information, which includes detailed information about the attractions (such as category, address, images). In particular, the Rating data shows the interest and satisfaction level (from 1-5 stars) of tourists with the attractions. This is the basis for building a prediction model, which helps in generating suitable recommendations. These data will be periodically extracted (e.g., every 3 days) from the application database, then they will be stored (using *Google Cloud Storage* service) to serve the recommender system.

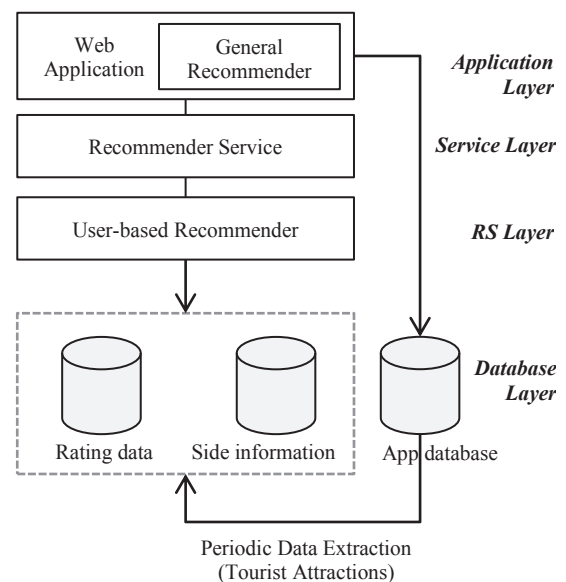


Figure 4. Architecture for the Travel Recommendation Application.

The *RS layer* contains the *Tourist Attraction Recommendation Module*. This layer is responsible for building the prediction model based on the collected rating data and generating

a personalized list of recommendations for each user based on their rating history. We will provide detailed information on the implementation of this module in the following section.

The *Service layer* creates Web APIs to expose the functionalities of the recommender system to various applications, allowing them to call these APIs and present the results. The primary purpose of this layer is to enable different types of applications across various platforms and technologies to utilize the recommender system. We implement this module using *Flask* (a framework for developing web applications on Python) and deploy it on *Google Cloud*.

Finally, the *Application layer* includes applications that interact with end-users and present the results. Since this layer is built upon the *Service layer*, these applications can be deployed on various platforms such as Mobile, Desktop, and Web. In this study, we implement a Web application for testing, using *Wordpress*—a powerful and widely-used Content Management System (CMS). The *General Recommender* module within the application provides recommendations for users who are not logged into the system or do not have profiles on the system. This module recommends highly-rated attractions to users based on simple statistical methods. In the future, this component could be enhanced to offer recommendations based on seasonality, regions, and other factors.

4.2. Recommendation process & experiments

The *User-based Recommender* module in the *RS layer* provides personalized attraction recommendations for a specific user based on their past rating history. In which:

- Input: User ID.
- Output: *Top-K* tourist attractions the user may be interested in.

We propose a general process for generating recommendations, as illustrated in Figure 5. Specifically, it includes main

components such as: *Data*, *Filter*, *Model Training*, *Links Prediction & Ranking*.

The *Data* component includes two types: Rating Data and Side Information Data. The Rating Data includes user ratings for various attractions and is used to train the prediction model. Meanwhile, the Side Information Data is used to filter candidate attractions.

The *Filter* module is designed to select candidate attractions from all available attractions in the system based on user-specific criteria. For example, it can filter attractions that belong to categories such as spiritual tourism or adventure tourism. However, within the scope of this study, the module is implemented in a simplified manner by including all attractions in the system, except for those the user has already interacted with, as input for the recommendation process.

The *Model Training* module is responsible for training the prediction model. The output includes the trained model and the final embeddings for each user and attraction. In this study, we utilize the LightGCN²¹ algorithm for model training and prediction.

Links Prediction & Ranking: This module predicts the links between users and attractions, followed by ranking them. Specifically, for each attraction, the system predicts the strength of the connection between the user and that attraction. The attractions are then sorted in descending order of connection strength, and the *top-K* attractions with the highest connection strengths are returned as recommendations.

One noteworthy point is that we do not store the trained model. Instead, for link prediction, we only save the final embeddings of each user and attraction (Users/Items embeddings) and compute the inner product between these vectors. This approach simplifies server processing since it eliminates the need to reference libraries for loading the model.

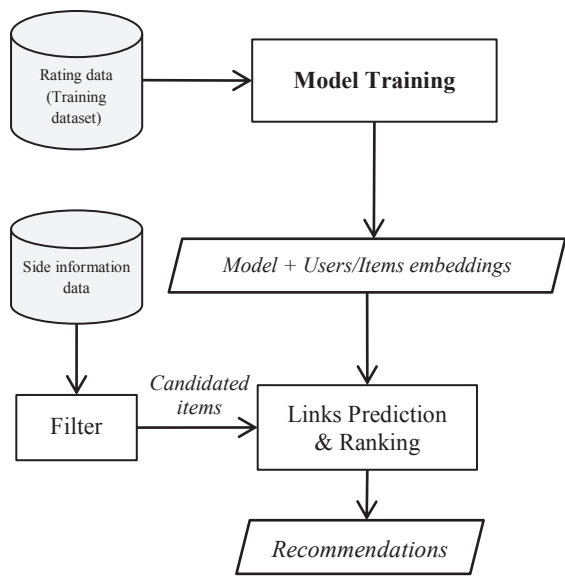


Figure 5. Recommendation process.

To evaluate the effectiveness of the prediction model, we conducted experiments on three datasets: MovieLens, Yelp2018,²¹ and TripAdvisor²³. Both Yelp2018 and TripAdvisor are datasets about the tourism domain. Statistics on the size of these datasets are described in Table 1. Specifically, we used the MovieLens dataset with 100,000 ratings, the Yelp2018 dataset with 500,000 rating (we reduce the data size to fit computational resources), and the TripAdvisor dataset with 13,697 ratings for our experiments. The datasets are divided into 2 parts, specifically, 75% for training and 25% for evaluation. The metrics recall@20, ndcg@20, and precision@20 were used to assess the model's performance.^{24,25} The graph learning algorithm used in these experiments is LightGCN. The initial embedding size is set to 64. The model was optimized using the Adam optimizer with a learning rate of 0.005.²⁶ The number of propagation layers in the graph was set to 3. Each batch size was 1,024, and the model was trained for 100 epochs. We utilized Google Colab (a cloud service provided by Google, allowing writing and execution of Python code through a web browser) to train the model and to learn the embeddings for users and attractions.

Table 1. Experimental datasets.

Datasets	Ratings #	Users #	Items #
MovieLens	100.000	943	1.682
Yelp2018	500.000	8.154	36.837
TripAdvisor	13.697	2.371	2.269

The training process to optimize the model is illustrated through the *training-loss* curve in Figure 6-8, corresponding to the MovieLens, Yelp2018 and TripAdvisor data sets.

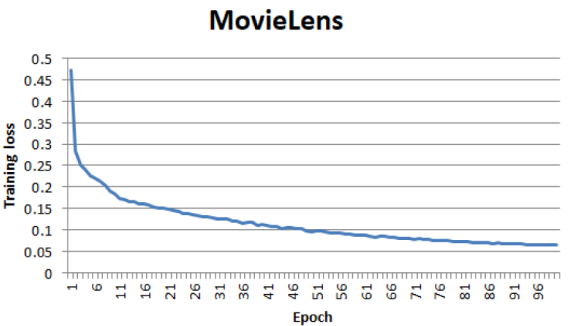


Figure 6. Training-loss curve on the MovieLens dataset.

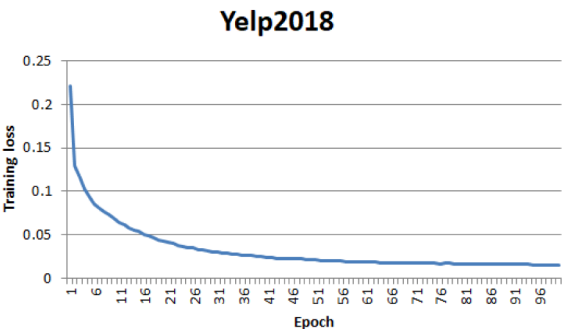


Figure 7. Training-loss curve on the Yelp2018 dataset.

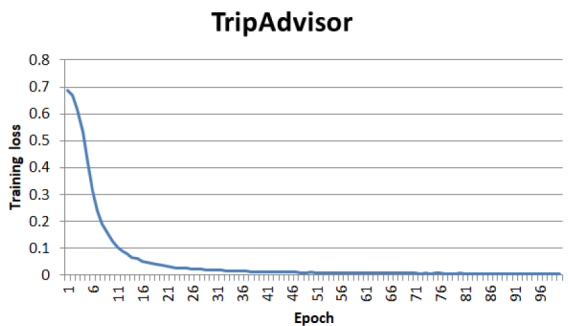


Figure 8. Training-loss curve on the TripAdvisor dataset.

Experimental results of the LightGCN model on the datasets are shown in Table 2.

Table 2. Experimental results of the LightGCN.

Datasets	Metrics		
	Recall	Precision	NDCG
MovieLens	0,212395	0,386744	0,439850
Yelp2018	0,082281	0,052545	0,083219
TripAdvisor	0,088338	0,005332	0,039634

Experimental results show that the LightGCN algorithm works effectively, shown through Recall, Precision, and NDCG metrics. In particular, with the MovieLens data set, the best results were obtained with Recall, Precision and NDCG values of 21.2%, 38.7% and 44%, respectively. For the other two datasets, the results were similar: for the Yelp2018 dataset, the results were 8.2%, 5.3%, and 8.3%; and for the TripAdvisor dataset, the results were 8.8%, 0.5%, and 4%. LightGCN was evaluated to be more effective compared to the NGCF algorithm and other algorithms^{20,21}. In this study, we do not aim to compare the LightGCN algorithm with other algorithms, instead, we test the algorithm with different data sets, to ensure that it performs well before integrating it into the travel recommendation application.

4.3. The web application

The web application is built on the *WordPress* platform, providing a useful and convenient tool for users to explore attractions. The application serves two main types of users: administrators and tourists. In particular, the administrator has main functions such as: Managing attractions (posting, editing articles...) and Managing users; Meanwhile, tourists can search for attractions by province/city, leave reviews about their experiences with attractions, and view recommendations based on their rating history. If the user is not logged in, the system suggests attractions with high average ratings.

As described in the previous section, this Web application will connect to the Web API

to fetch recommended attractions for users. Some images of the application are depicted in Figure 9-11.

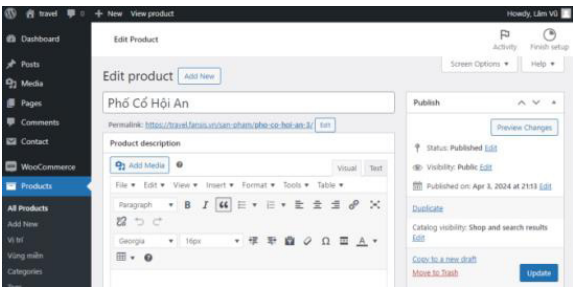


Figure 9. The function of Posting attractions.

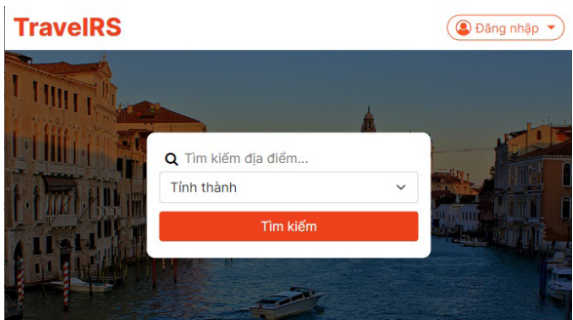


Figure 10. The function of searching attractions.

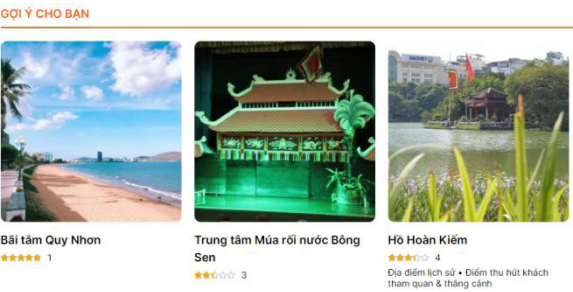


Figure 11. The function of suggesting attractions.

To validate the recommendation functionality of the web application, we simulated creating 3 users: u_1 , u_2 and u_3 . These users rated 5 tourist attractions (encoded as $i_1, i_2 \dots i_5$) out of a total of 50 attractions in the system. To facilitate the creation of visual graphs, we have encoded users and attractions with their IDs. Specific information about the attractions is given in Table 3. The rating data is illustrated in Figure 1. We plotted graphs (Figures 3, 12, and 13) with the root nodes corresponding to each user to find attractions the user might be interested in through the connections in the graph.

Table 3. Information about some attractions.

Attraction ID	Attraction name	Location
i_1	Hồ Hoàn Kiếm	Hà Nội
i_2	Trung tâm Múa rối nước Bông Sen	Hà Nội
i_3	Đền Ngọc Sơn	Hà Nội
i_4	Phố cổ Hội An	Quảng Nam
i_5	Bãi tắm Quy Nhơn	Bình Định

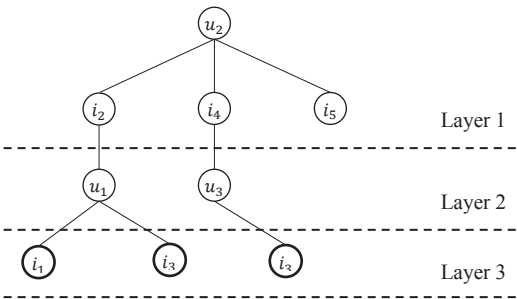


Figure 12. Graph with 3 connection layers for root vertex u_2 .

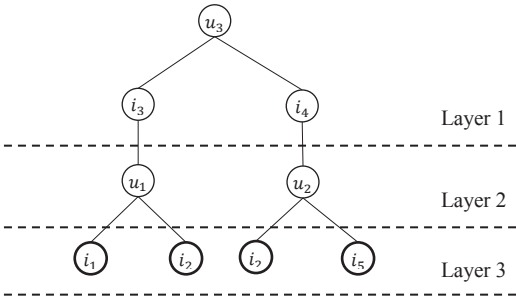


Figure 13. Graph with 3 connection layers for root vertex u_3 .

By visually analyzing these small graphs, the attractions that should be suggested for each user are identified and listed in Table 4 below. These are the leaf nodes on the graph, connected to the root nodes (users) through intermediary nodes.

Table 4. Attractions that should be suggested to each user, derived from the graph analysis.

Users	Attractions should be suggested
u_1	i_4, i_5
u_2	i_3, i_1
u_3	i_2, i_1, i_5

Following the recommendation process described earlier, we learned the representations for each user and attraction to calculate the strength of the connections between them. The predicted connection strengths are provided in Table 5 below. Based on these results, we observe that the constructed model accurately predicts the attractions that users might be interested in. Specifically:

- For u_1 , the two attractions with the highest positive connection strengths are i_4 and i_5 .
- For u_2 , the two attractions with the highest positive connection strengths are i_3 and i_1 .
- For u_3 , the two attractions with the highest positive connection strengths are i_2, i_1 and i_5 .

Table 5. Predicted connection strengths between users and attractions.

User	Attractions	Predicted Links
u_1	i_4	4,399
u_1	i_5	3,109
u_1	$i_{17}, i_{12}, i_7, i_{34}, i_{36}, i_{39}, i_{46}, i_{10}$	-1,567, -1,694, -1,699, -1,705, -1,713, -1,718, -1,725, -1,726
u_2	i_3	4,378
u_2	i_1	3,184
u_2	$i_{17}, i_7, i_{12}, i_{39}, i_{33}, i_{36}, i_{14}, i_{41}$	-1,640, -1,747, -1,752, -1,764, -1,773, -1,776, -1,794, -1,797
u_3	i_2	3,669
u_3	i_1	2,735
u_3	i_5	2,697
u_3	$i_{17}, i_{34}, i_7, i_{40}, i_{41}, i_{46}, i_{11}$	-1,369, -1,415, -1,441, -1,442, -1,459, -1,466, -1,471

Thus, the web application has demonstrated a travel recommender system according to the proposed process described earlier. Accordingly, for users who already have a profile on the system, the application will suggest attractions that they may be interested

in based on their review history. Here, we only created a very limited number of users and attractions to illustrate the system's workflow. Building a larger dataset can be undertaken once the system is deployed in a real-world operational environment.

5. CONCLUSION

In this paper, we have presented the application of graph learning in a travel recommender system. Specifically, we have illustrated the use of the LightGCN algorithm in the Graph Neural Network approach to model and process data in the form of a bipartite graph. This method allows the recommender system to exploit the complex structure and hidden relationships between users and attractions, thereby providing accurate suggestions to users. At the same time, we also tested the algorithm on three different data sets to evaluate their effectiveness. Besides, we have proposed the travel recommender system application architecture. This architecture is general, and based on that, we can build systems with different recommendation algorithms, with different technologies.

In the future, we plan to continue to improve and expand the system. Specifically, some future research directions include: (i) Integrating contextual recommendation. We aim to enhance the recommender system by incorporating contextual information. For example, the system could suggest tourist attractions based on additional context such as the time to visit or the type of companions (family, friends, solo travel). (ii) Exploration of advanced machine learning techniques. We plan to investigate and apply more advanced machine learning techniques to optimize the system. (iii) Enhancement of the web application. We intend to further develop the web application by adding more diverse information and functionalities. This could include features like tour booking, itinerary planning, recommendations for accommodations (hotels, hostels) and dining options (restaurants, cafes).

In short, building a travel recommender system is important and necessary, especially in the context of the rapid growth of the tourism industry in Vietnam. We hope that these research results will be an important stepping stone to developing more intelligent travel recommender systems in the future.

Acknowledgement

This research is conducted within the framework of science and technology projects at institutional level of Quy Nhon University under the project code T2023.807.17.

REFERENCES

1. L. Q. Hung, V. S. Lam, L. A. Cuong. A comparative analysis of various approaches for incorporating contextual information into recommender systems, *Journal of Computer Science*, **2022**, 18(3), 187-203.
2. V. S. Lam, L. Q. Hung. A deep learning based approach for context-aware multi-criteria recommender systems, *Computer Systems Science and Engineering*, **2022**, 44(1), 471-483.
3. L. Q. Hung, V. S. Lam, N. T. K. Phuong, L. T. Xinh. A state-of-the-art survey on context-aware recommender systems and applications, *International Journal of Knowledge and Systems Science*, **2021**, 12(3), 1-20.
4. Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, S. Y. Philip. A comprehensive survey on graph neural networks, *IEEE Transactions on Neural Networks and Learning Systems*, **2020**, 32(1), 4-24.
5. Q. Guo, F. Zhuang, C. Qin, H. Zhu, X. Xie, H. Xiong, Q. He. A survey on knowledge graph-based recommender systems, *IEEE Transactions on Knowledge and Data Engineering*, **2020**, 34(8), 3549-3568.
6. S. Wang, L. Hu, Y. Wang, X. He, Q. Z. Sheng, M. A. Orgun, P. S. Yu. *Graph learning based recommender systems: a review*, Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Montreal, 2021.

7. S. Wu, F. Sun, W. Zhang, X. Xie, B. Cui. Graph neural networks in recommender systems: a survey, *ACM Computing Surveys*, **2022**, 55(5), 1-37.
8. Z. Jiang, H. Liu, B. Fu, Z. Wu, T. Zhang. *Recommendation in heterogeneous information networks based on generalized random walk model and bayesian personalized ranking*, Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM 2018, Marina Del Rey, CA USA, 2018.
9. H. Bagci, P. Karagoz. *Context-aware friend recommendation for location based social networks using random walk*, Proceedings of the 25th International Conference Companion on World Wide Web, WWW 2016, Montréal Québec, Canada, 2016.
10. Z. Wang, H. Liu, Y. Du, Z. Wu, X. Zhang. *Unified embedding model over heterogeneous information network for personalized recommendation*, Proceedings of the 28th International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, 2019.
11. Y. Cen, X. Zou, J. Zhang, H. Yang, J. Zhou, J. Tang. *Representation learning for attributed multiplex heterogeneous network*, Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK USA, 2019.
12. B. Hu, C. Shi, W. X. Zhao, P. S. Yu. *Leveraging meta-path based context for top-n recommendation with a neural co-attention model*, Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2018, London, United Kingdom, 2018.
13. Y. Deng. Recommender stems based on graph embedding techniques: a review, *IEEE Access*, **2022**, 10, 51587-51633.
14. R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, J. Leskovec. *Graph convolutional neural networks for web-scale recommender systems*, Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2018, London, United Kingdom, 2018.
15. W. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. Tang, D. Yin. *Graph neural networks for social recommendation*, The World Wide Web Conference, WWW 2019, San Francisco, CA USA, 2019.
16. Z. Cui, Z. Li, S. Wu, X. Y. Zhang, L. Wang. *Dressing as a whole: outfit compatibility learning based on node-wise graph neural networks*, The World Wide Web Conference, WWW 2019, San Francisco, CA USA, 2019.
17. C. Gao, X. Wang, X. He, Y. Li. *Graph neural networks for recommender system*, Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, WSDM 2022, Virtual Event, AZ USA, 2022.
18. D. H. Tran, Q. Z. Sheng, W. E. Zhang, A. Aljubairy, M. Zaib, S. A. Hamad, N. H. Tran, N. L. D. Khoa. HeteGraph: graph learning in recommender systems via graph convolutional networks, *Neural Comput and Applic*, **2023**, 35, 13047–13063.
19. N. T. Khang, N. T. Anh, M. T. Nga, N. H. An, N. V. Anh. Session-based recommendation using graph neural network (in Vietnamese), *Research and Development on Information and Communication Technology*, **2022**, 2022(2), 73-82.
20. X. Wang, X. He, M. Wang, F. Feng, T. S. Chua. *Neural graph collaborative filtering*, Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, 2019.
21. X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, M. Wang. *LightGCN: simplifying and powering graph convolution network for recommendation*, Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2020, Virtual Event China, 2020.
22. S. Rendle, C. Freudenthaler, Z. Gantner, L. S. Thieme. *BPR: Bayesian personalized ranking from implicit feedback*, Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, UAI 2009, Montreal Quebec, Canada, 2009.

23. Y. Zheng, B. Mobasher, R. Burke. *Context recommendation using multi-label classification*, 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), WI and IAT 2014, Warsaw, Poland, 2014.
24. Y. M. Tamm, R. Damdinov, A. Vasilev. *Quality metrics in recommender systems: do we calculate metrics consistently?*, Proceedings of the 15th ACM Conference on Recommender Systems, RecSys 2021, Amsterdam Netherlands, 2021.
25. V. S. Lam, L. Q. Hung, N. V. Vinh. *Evaluation of recommender systems: survey and experiment (in Vietnamese)*, Proceedings of the 23rd National Conference Some selected issues of Information and Communication Technology, Ha Long, Quang Ninh, 2020.
26. D. P. Kingma, J. Ba. *Adam: a method for stochastic optimization*, International Conference on Learning Representations, ICLR 2015, San Diego, 2015.



© 2024 by the authors. This Open Access Article is licensed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license (<https://creativecommons.org/licenses/by-nc/4.0/>).

Nghiên cứu khả năng hấp phụ và tách lọc carbon dioxide của vật liệu MIL-53(Cr) bằng phương pháp mô phỏng cổ điển

Nguyễn Thị Xuân Huynh^{1,2,*}, Trần Thị Diễm Thanh^{1,2}, Nguyễn Lê Bảo Trân^{1,2},
Nguyễn Ngọc Khoa Trường¹

¹Khoa Khoa học Tự nhiên, Trường Đại học Quy Nhơn, Việt Nam

²Phòng thí nghiệm Hóa tính toán và Mô phỏng, Trường Đại học Quy Nhơn, Việt Nam

Ngày nhận bài: 20/08/2024; Ngày sửa bài: 10/09/2024;

Ngày nhận đăng: 18/09/2024; Ngày xuất bản: 28/10/2024

TÓM TẮT

Nồng độ khí CO₂ đang tăng rất nhanh đe dọa sự nóng lên toàn cầu, cảnh báo đến môi trường và sức khỏe của con người. Do đó, việc làm giảm hoặc chuyển đổi CO₂ thành các hợp chất có giá trị là hết sức cấp thiết. Trong những thập kỷ gần đây, việc ứng dụng các vật liệu có độ xốp cao như vật liệu khung hữu cơ kim loại (MOF) để hấp phụ và tách lọc khí CO₂ đang rất thu hút. Trong nhóm MOF, vật liệu MIL, điển hình là MIL-53(Cr) rất được chú ý với khả năng hấp phụ và tách lọc cao nhờ vào tính ổn định nhiệt, diện tích bề mặt riêng rất lớn và chứa các vị trí kim loại mở, v.v. Do vậy, nghiên cứu này sử dụng chính phương pháp mô phỏng Monte Carlo chính tắc lớn, một phương pháp mô phỏng phân tử cổ điển, để tính lượng khí hấp phụ trong MIL-53(Cr) và làm sáng tỏ khả năng hấp phụ CO₂ và tách lọc CO₂/H₂ ở nhiệt độ phòng và áp suất dưới 50 bar. Kết quả chỉ ra khả năng hấp phụ CO₂ và tách lọc CO₂/H₂ của MIL-53(Cr) được đánh giá cao. Cụ thể, lượng khí hấp phụ CO₂ tinh khiết trong MIL-53(Cr) đạt đến 9,18 mmol/g ở 298 K và 50 bar. Độ tách lọc CO₂/H₂ lớn nhất, $S_{CO_2/H_2(max)}$ của MIL-53(Cr) đạt 116 ở 298 K. Nghiên cứu cũng chỉ ra tỷ lệ mol CO₂:H₂ gần như không làm thay đổi giá trị $S_{CO_2/H_2(max)}$ mà chỉ làm thay đổi áp suất đạt $S_{CO_2/H_2(max)}$ đó. Hơn nữa, việc giảm nhiệt độ cũng làm tăng đáng kể khả năng tách lọc khí CO₂/H₂, cụ thể $S_{CO_2/H_2(max)} = 245$ khi giảm nhiệt độ đến 273 K.

Từ khóa: Hấp phụ CO₂, bắt giữ CO₂, tách lọc CO₂/H₂, vật liệu MIL-53, mô phỏng cổ điển.

*Tác giả liên hệ chính.

Email: nguyenthixuanhuynh@qnu.edu.vn

A classical simulation study for carbon dioxide adsorption and separation capacity of MIL-53(Cr)

Nguyen Thi Xuan Huynh^{1,2,*}, Tran Thi Diem Thanh^{1,2}, Nguyen Le Bao Tran^{1,2},
Nguyen Ngoc Khoa Truong¹

¹Faculty of Natural Sciences, Quy Nhon University, Vietnam

²Lab of Computational Chemistry and Modelling, Quy Nhon University, Vietnam

Received: 20/08/2024; Revised: 10/09/2024;

Accepted: 18/09/2024; Published: 28/10/2024

ABSTRACT

Carbon dioxide (CO₂) concentrations are increasing, threatening global warming limits and leading to an alarming impact on the environment and human health. Therefore, reducing or converting CO₂ into high-value chemicals is necessary. In recent decades, CO₂ capture and separation based on the adsorption of highly porous materials, especially metal-organic framework (MOF) materials have become an issue of great interest. Among MOFs, MIL material, notably MIL-53(Cr), has been a top concern for its noteworthy adsorption and separation capacity due to its thermal stability, ultra-high specific surface, open metal sites, and more. Therefore, this work mainly uses grand canonical Monte Carlo simulation, a classical molecular simulation method, to calculate the amount of gas adsorbed in MIL-53(Cr) and study CO₂ adsorption and CO₂/H₂ separation capacity in the mixture of hydrogen (H₂) and CO₂ at room temperatures and low pressures below 50 bar via adsorption isotherms. This research explored the high CO₂ adsorption capacity and the impressive CO₂/H₂ selectivity of MIL-53(Cr). For pure CO₂, the absolute CO₂ uptake in MIL-53(Cr) is 9.18 mmol/g at 298 K and 50 bar. Besides, the maximum CO₂/H₂ selectivity of MIL-53(Cr) is $S_{\text{CO}_2/\text{H}_2(\text{max})} = 116$ at 298 K. Remarkably, the CO₂/H₂ mole fraction has almost no effect on the value of the maximum SO₂/H₂ separation, $S_{\text{CO}_2/\text{H}_2(\text{max})}$, but leads to the change of pressure reaching $S_{\text{CO}_2/\text{H}_2(\text{max})}$. Furthermore, reducing the temperature significantly increases the separation capacity of CO₂/H₂; specifically, $S_{\text{CO}_2/\text{H}_2(\text{max})} = 245$ when the temperature drops to 273 K.

Keywords: CO₂ adsorption, CO₂ capture, CO₂/H₂ selectivity, MIL-53 material, classical simulations.

1. INTRODUCTION

Fossil fuels are our primary energy source, and their consumption is increasingly uncontrollable. This problem has a significantly negative impact on the environment and human health due to the increasing amount of emissions such as CO₂, CO, NO_x, and more, and causes severe energy shortage issues. Furthermore, CO₂ capture, H₂ purification, and CO₂ separation over H₂ are significant in combustion plants and refineries.¹⁻³

To overcome those problems, the demands for clean energy production, gas separation, and efficient storage have grown dramatically. Tremendous advances have been achieved in both theory and experiments thanks to the outstanding properties of porous materials.⁴⁻⁶

Metal-organic framework (MOF), a new type of inorganic-organic hybrid porous material with a periodic network, appears as a potential technology.⁶⁻⁹ Since the early 1990s, MOFs have

*Corresponding author.

Email: nguyenthixuanhuynh@qnu.edu.vn

gained increasing attention, leading to more MOFs that have been successfully synthesized by experiments and proposed by simulation methods. Researchers have successfully synthesized about 100,000 MOFs through experiments and have predicted over 500,000 structures by simulations or theories, but only nearly 3,500 works related to MOFs have been published.^{10,11}

MILs (Material Institute of Lavoisiers), a class of MOFs, have gained recognition as promising adsorbents for H₂ purification and CO₂ capture thanks to unique properties such as extraordinary chemical and thermal properties, water stability, ultra-large specific surface areas, high porosity, coordinately unsaturated or open metal sites, and more.^{12–15} In 2002, the first MIL, published by Ferey and co-workers, was the MIL-53 with M = chromium(III) (Cr³⁺).^{16,17} MIL-53, obtained by the hydrothermal method, has many potential applications because of its good thermal stability and flexibility compared to other MOFs.^{18,19} Several works have displayed that MIL-53 (Cr or Al) materials have superior CO₂ capture^{20–22} and high selective adsorption for binary mixtures such as CO₂/H₂, CO₂/CH₄, and CO₂/N₂.^{19,23,24}

Due to the potential of MIL-53, in this work, we use the classical grand canonical Monte Carlo (GCMC) simulation (a suitable approach to simulate the equilibrium of the physisorption of small molecules in porous at well-defined pressure and temperature) to study the capacities of CO₂ capture and CO₂/H₂ separation on MIL-53(Cr).

2. COMPUTATIONAL APPROACHES

GCMC simulations are carried out using the RASPA software to calculate the gravimetric amount of gases (CO₂ and H₂) adsorbed in a solid porous material, MIL-53(Cr),²⁵ at room temperature and pressure up to 50 bar. The GCMC simulations run up to 3.0 × 10⁵ cycles after an initial equilibration of 1.5 × 10⁴ steps.

In order to obtain the simulation box for GCMC, a unit cell of MIL-53 (including 20 hydrogen, 32 carbon, 20 oxygen, and 4 chromium) was optimized using the density functional theory (DFT),²⁶ and then repeated the unit cell to 2, 2, and 4 times for *a*, *b*, and *c* lengths (*a* = 16.73 Å, *b* = 13.04 Å, *c* = 6.81 Å), leading to *L_a* = 33.46 Å, *L_b* = 26.08 Å, and *L_c* = 27.24 Å, respectively (Figure 1).

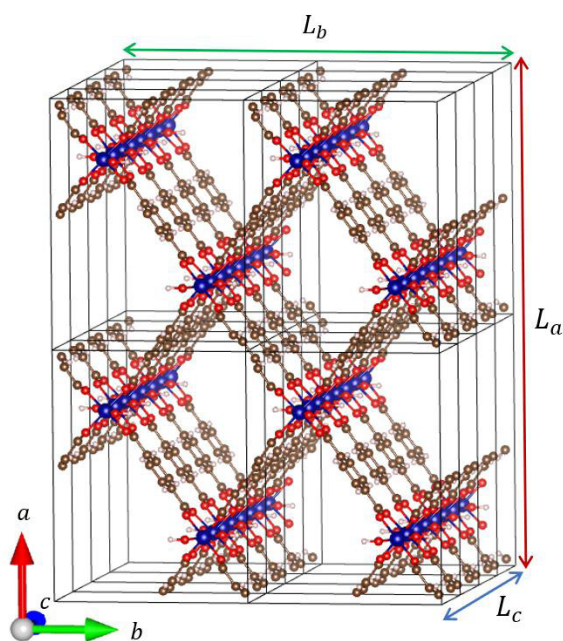


Figure 1. The simulation box of MIL-53(Cr). Each rectangular cube represents a unit cell.

The interactions between gases (CO₂, H₂) and MIL-53(Cr) include (i) van der Waals (vdW) interactions using the Lennard-Jones (LJ) model and (ii) electrostatic interactions.

First, the LJ 12-6 repulsive and attractive interactions between pairs of atoms of gases and MIL-53 are presented by

$$V_{ij}^{LJ} = 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right], \quad (1)$$

herein, ε_{ij} and σ_{ij} are pair LJ parameters, calculated using the pair combining rule of Lorentz and Berthelot,²⁵ in which σ_i and ε_i are taken from the force fields for MOFs with universal force field for metal (Cr) and DREIDING for other elements

of MIL-53 (H, C, O),²⁷ listed in Table 1. The cut-off radius used for the vdW interaction is 16 Å, tested for reliability.

Second, electrostatic or Coulomb interactions between pairs of atoms are as follows:

$$V_{ij}^{elec} = k_c \frac{q_i q_j}{r_{ij}}, \tag{2}$$

where k_c is the Coulomb constant and q_i is the partial point charge of atom i , calculated by the DFT-based density-derived electrostatic and chemical charges (Table 1 with the labeled atoms as in Figure 2).²⁸ Additionally, TrapPE force fields are used for CO₂ and H₂ (Table 1).

The adsorption separation or selectivity of CO₂ over H₂ (CO₂/H₂) in their mixture is calculated via the formula¹³

$$S_{CO_2/H_2} = \frac{q_{CO_2}}{q_{H_2}} \bigg/ \frac{n_{CO_2}}{n_{H_2}}, \tag{3}$$

where q_{CO_2} and q_{H_2} are the amount of adsorbed CO₂ and H₂ (mol/g) in MIL-53, respectively; n_{CO_2} and n_{H_2} are mole fractions of the CO₂ and H₂ components in the mixture, respectively. The investigated molar fractions of CO₂:H₂ includes 1:9, 2:8, 3:7, 4:6, 5:5, 6:4, 7:3, 8:2, 9:1.

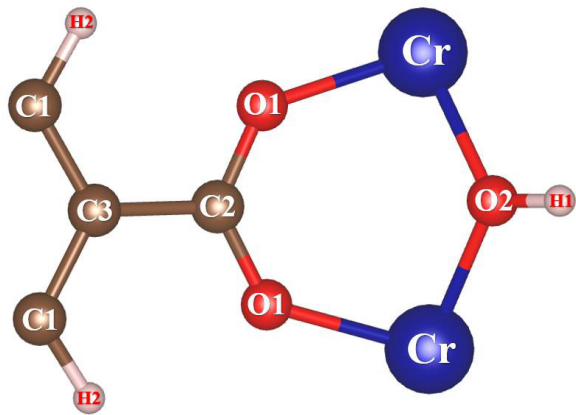


Figure 2. The labeled atoms of MIL-53(Cr), where hydrogen (H), carbon (C), oxygen (O), and chromium (Cr) atoms are displayed by soft pink, brown, red, and blue balls, successively.

Table 1. LJ parameters and partial atomic charges of MIL-53 and gases (H₂ and CO₂).

Atoms	ϵ/k_B (K)	σ (nm)	Partial point charges (e)
Cr	7.548	0.269	+1.581
H1	7.649	0.285	+0.414
H2			+0.100
C1	47.856	0.347	-0.077
C2			+0.636
C3			-0.033
O1	48.158	0.303	-0.589
O2			-0.938
C [CO ₂] ²⁹	27.000	0.280	+0.700
O [CO ₂] ²⁹	79.000	0.305	-0.350
H _{COM} [H ₂] ³⁰	36.700	0.296	-0.936
H [H ₂] ³⁰	-	-	+0.468

3. RESULTS AND DISCUSSION

3.1. CO₂ adsorption of MIL-53(Cr)

Firstly, the absolute CO₂ adsorption capacity or uptake is determined and compared with available experimental data. The results (Figure 3) show that the CO₂ uptake increases rapidly in the pressure region below 5 bar, then increases slightly with increasing pressure. The absolute CO₂ uptake is 9.18 ± 0.02 mmol/g or 403.70 ± 0.88 mg/g at the maximum pressure of 50 bar (blue solid line). However, the maximum excess CO₂ uptake is 8.02 ± 0.02 mmol/g or 387.23 ± 0.93 mg/g at the saturated pressure of 30 bar (red dashed line). Our findings show that the absolute amount of CO₂ adsorption on MIL-53(Cr) is consistent with the available data of Bourrelly et al., about 10 mmol/g at 30 bar and 304 K.²⁰ Moreover, the obtained CO₂ adsorption heat of MIL-53(Cr) is 30.04 kJ/mol, consistent with data of Bourrelly and co-workers (about 32 kJ/mol).²⁰ We also find that, at 298 K, the absolute CO₂ adsorption isotherm obtained from molecular simulations (blue line with solid circles) also agrees well with that of the Schneemann group (back open circles for experimental data) (Figure 3).³¹ These agreements between simulation and experimental measures indicate the reliability of the GCMC simulations.

At room temperature, MIL-53(Cr) is not able to capture CO₂ as strongly as the most highly evaluated MOFs available today, such as MOF-200 (2400 mg/g at 50 bar), MOF-210 (2396 mg/g at 50 bar), MOF-117 (33.5 mmol/g), MIL-100(Cr) (18 mmol/g), and more;¹⁴ however, these results are comparable to the MIL-88 series (4.0 – 12.1 mmol/g), relatively well recorded for CO₂ uptake with the isosteric heat of adsorption of the MIL-88 series (31.8 – 34.9 kJ/mol).¹⁴

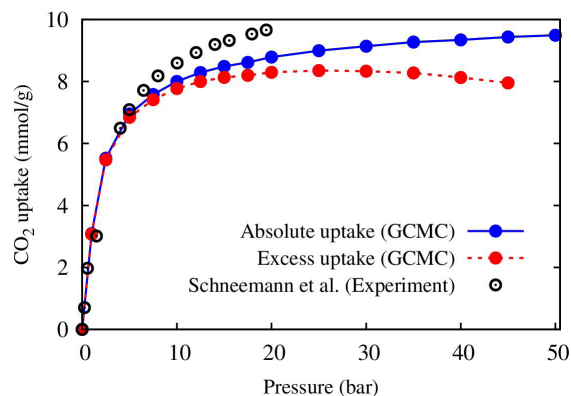


Figure 3. The comparison of the absolute CO₂ adsorption capacity on MIL-53(Cr) at 298 K between GCMC simulation and experimental data.³¹

Next, we study the co-adsorption capacity of MIL-53 for H₂ and CO₂ in the binary mixture at many different mole fractions (CO₂:H₂ = 1:9, 2:8, 3:7, 4:6, 5:5, 6:4, 7:3, 8:2, 9:1) at 298 K and pressures below 50 bar (Figure 4 and Table 2). At 298 K, our simulations also exhibit that the uptake of CO₂ is much higher than that of H₂. This conclusion is reasonable because the average adsorption heat of CO₂ (30.04 kJ/mol) is more than two times higher than that of H₂ (12.87 kJ/mol) in the low-pressure range. Indeed, Figure 4 shows that MIL-53(Cr) adsorbs CO₂ (9.16 → 6.26 mmol/g) superiorly than H₂ (0.02 → 0.55 mmol/g) when the molar ratio of CO₂:H₂ changes from 1:9 → 9:1. The exciting result is that the CO₂ uptake insignificantly changes when the molar ratio of H₂ is less than or equal to 60% in the mixture. For example, the CO₂ uptake decreases from 9.18 mol/g for pure CO₂ adsorption to 8.37 mmol/g for the case of CO₂:H₂ = 4:6 (40% CO₂ and 60% H₂), making an only 8.82% reduction compared to pure CO₂. The presence of a high mole fraction of H₂ in the mixture leads to a significant decrease in CO₂ uptake of MIL-53(Cr). More specifically, compared to pure CO₂ uptake (9.18 mmol/g), the CO₂ uptake decreases by 12.64% to 8.02 mmol/g, 18.52% to 7.48 mmol/g, and 31.81% to 6.26 mmol/g when the molar ratio of H₂ in

the binary mixture occupies 70% ($\text{CO}_2\text{:H}_2 = 3\text{:}7$), 80% ($\text{CO}_2\text{:H}_2 = 2\text{:}8$), and 90% ($\text{CO}_2\text{:H}_2 = 1\text{:}9$) (Table 2), respectively. These impressive results demonstrate that MIL-53(Cr) has stable and good selective adsorption for CO_2 over H_2 in their binary mixture.

Table 2. The absolute CO_2 and H_2 uptakes of MIL-53(Cr) with the different mole fractions of $\text{CO}_2\text{:H}_2$, compared to pure CO_2 uptake at room temperature.

$\text{CO}_2\text{:H}_2$ mole- fraction	CO_2 uptake (50 bar)	Decrease in comparison to pure CO_2 (%)	H_2 uptake (50 bar)
Pure CO_2	9.18	-	-
	10 (30 bar) ²⁰		-
9:1	9.16	0.22	0.02
8:2	9.08	1.09	0.04
7:3	8.95	2.51	0.06
6:4	8.80	4.14	0.09
5:5	8.61	6.21	0.12
4:6	8.37	8.82	0.16
3:7	8.02	12.64	0.22
2:8	7.48	18.52	0.32
1:9	6.26	31.81	0.55

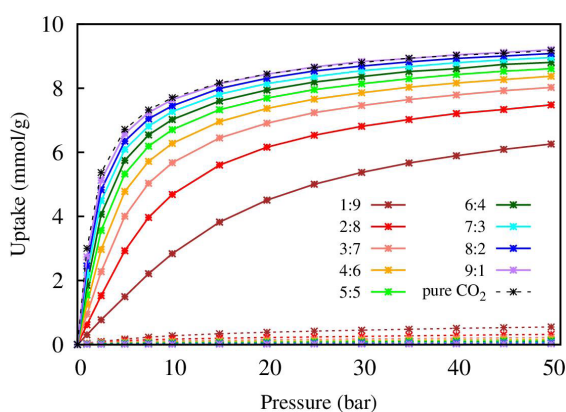


Figure 4. The CO_2 and H_2 co-adsorption isotherms of MIL-53(Cr) at 298 K with different $\text{CO}_2\text{:H}_2$ ratios from 1:9 to 9:1. Solid and dot lines refer to the adsorption amount of CO_2 and H_2 on MIL-53(Cr). The black dashed line represents pure CO_2 adsorption on MIL-53(Cr).

3.2. $\text{CO}_2\text{/H}_2$ separation of MIL-53(Cr)

From Figure 5a, it is evident that at 298 K, the highest selectivity achieved for each $\text{CO}_2\text{:H}_2$ mole-fraction greatly depends on pressure. Specifically, the $\text{CO}_2\text{/H}_2$ selectivity sharply increases at low pressure, peaks at the maximum $\text{CO}_2\text{/H}_2$ selectivity [$S_{\text{CO}_2\text{/H}_2(\text{max})}$], and then gradually decreases with pressure to 50 bar. As a result, the maximum $\text{CO}_2\text{/H}_2$ selectivity varies for $\text{CO}_2\text{:H}_2$ mole ratios, listed in Column 3 of Table 3 (298 K). In addition, the pressure corresponding to the maximum $\text{CO}_2\text{/H}_2$ selectivity decreases with an increase in CO_2 mole fraction and a decrease in H_2 mole fraction. More detail, the maximum selectivity $S_{\text{CO}_2\text{/H}_2(\text{max})} = 107.87$ at 25 bar for $\text{CO}_2\text{:H}_2 = 1\text{:}9$, 114.30 at 15 bar for $\text{CO}_2\text{:H}_2 = 2\text{:}8$, ..., and 118.25 at 5 bar for $\text{CO}_2\text{:H}_2 = 9\text{:}1$ (at 298 K of Table 3). Surprisingly, the maximum $\text{CO}_2\text{/H}_2$ selectivity in MIL-53(Cr) remains nearly consistent regardless of the change in the molar ratio between CO_2 and H_2 , averaging $S_{\text{CO}_2\text{/H}_2(\text{max})} = 116$ at 298 K (pink line in Figure 6). This tendency is similar to the case of MOF-5 and Cu-BTC studied by experiments.^{32,33} Remarkably, our data also shows that the $\text{CO}_2\text{/H}_2$ selectivity, $S_{\text{CO}_2\text{/H}_2}$ of MIL-53(Cr) is higher than that of many previous MOFs, such as microporous silica (3.5), activated carbon (45.0), zeolites Na-4A (70.7), Cu-BTC (150 at 12.5 bar), MOF-5 (36 at 50 bar)³² and IR-MOF- n with $n = 9, 10, \dots, 14$ (≤ 100).³³

Furthermore, our study delves into the influence of temperature on the selectivity of $\text{CO}_2\text{/H}_2$ in MIL-53(Cr). To accomplish this, we perform further investigation at temperatures of 273 K, 323 K, and 348 K surrounding 298 K, visualized in Figures 5 b, c, and d, respectively. Our results show the pressure with the maximal value of $S_{\text{CO}_2\text{/H}_2}$ decreases with the lowering temperature (Figure 5 and Table 3). For instance, at $\text{CO}_2\text{:H}_2 = 5\text{:}5$, the maximum selectivities are $S_{\text{CO}_2\text{/H}_2(\text{max})} = 39.81$ at 15 bar (348 K), 65.17 at 10 bar (323 K), 118.31 at 7.5 bar (298 K), and 246.43 at 2.5 bar (273 K). Figure 5 also shows

that the peaks for CO_2/H_2 separation broaden as the temperature increases (Figure 5). The maximum CO_2/H_2 selectivity and the pressure corresponding to different mole ratios at four temperatures are listed in Table 3.

Similar to the case of 298 K discussed above, the maximum $S_{\text{CO}_2/\text{H}_2}$ values in MIL-53(Cr) also remain constant with changes in the molar ratio between CO_2 and H_2 at 273 K, 323 K, and 348 K. The average maximum selectivities are 245, 116, 64, and 39 at 273 K, 298 K, 323 K, and 348 K, respectively (Figure 6). These results indicate that decreasing temperature significantly increases CO_2/H_2 separation. More importantly, for the maximum $S_{\text{CO}_2/\text{H}_2}$, when the temperature drops from 348 K to 273 K, the variation of CO_2/H_2 separation dramatically increases. Specifically, the maximum CO_2/H_2 selectivity increases 25, 53, and 128 when the temperature drops from 348 K to 323 K, 323 K to 298 K, and 298 K to 273 K, successively. As reported in the literature, the gas adsorption

capacity at different temperatures strongly depends on the gas-MOF interactions.³⁴ At higher temperatures, the reduction in CO_2/H_2 selectivity is attributed to the weaker interaction between CO_2 and MIL-53, causing a remarkable decrease in the CO_2 adsorption of MIL-53. Conversely, the H_2 adsorption in MIL-53 is minimal, so it remains largely unaffected by changes in temperature compared to CO_2 .¹ Therefore, the CO_2/H_2 separation capacity strongly depends on the temperatures. It should also be noted that in order to gain insight into the interactions between H_2 , CO_2 and MIL-53(Cr), first-principles calculations are essential, and it takes a long time to implement; therefore, we will explore this issue further in future studies.

It can be concluded that to significantly boost the CO_2 selective adsorption in the mixture of CO_2 and H_2 , lowering the temperature is able to achieve the desired CO_2/H_2 separation at appropriate pressures based on the determined $\text{CO}_2:\text{H}_2$ mole fraction.

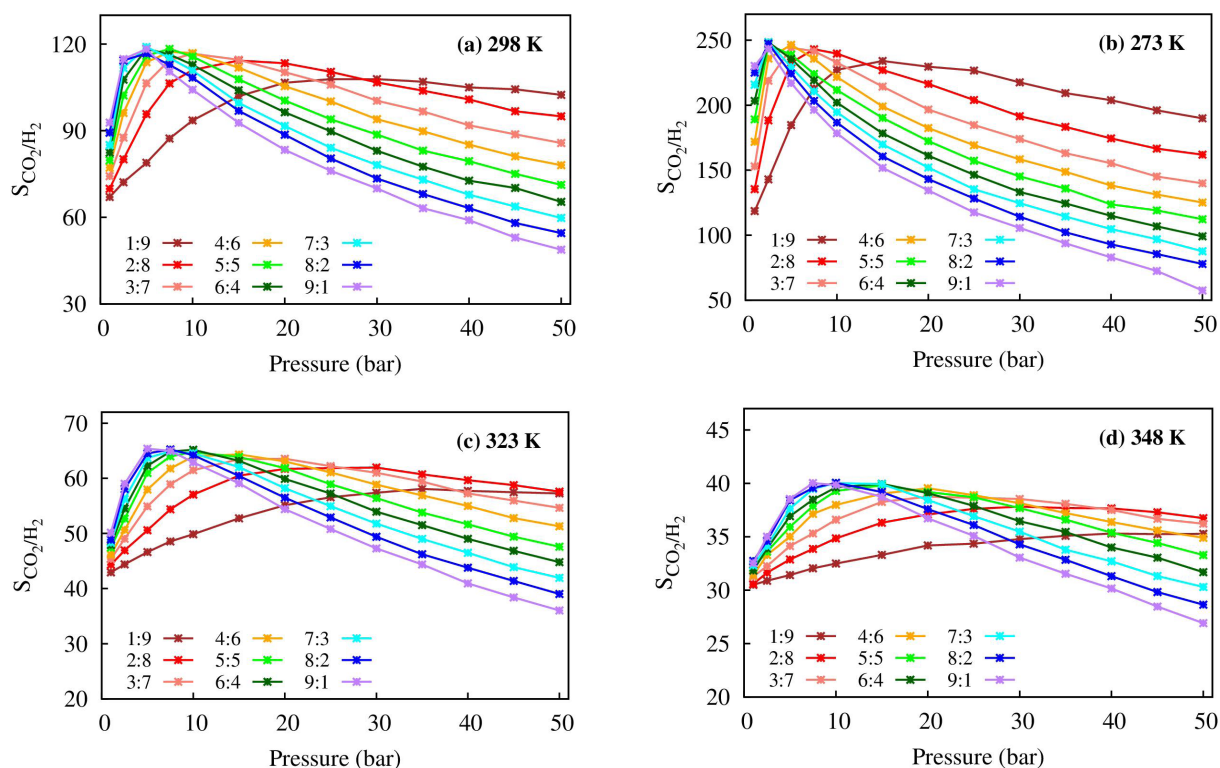


Figure 5. CO_2/H_2 selectivity of MIL-53(Cr) with different CO_2/H_2 mole fractions at temperatures: (a) 298 K, (b) 273 K, (c) 323 K, and (d) 348 K.

Table 3. Maximum CO₂/H₂ selectivity versus mole ratios of CO₂/H₂ at 273 K, 298 K, 323 K, and 348 K.

CO ₂ :H ₂	273 K	298 K	323 K	348 K
1:9	233.97 (15.0 bar)	107.87 (25.0 bar)	57.73 (40.0 bar)	35.27 (50.0 bar)
2:8	243.04 (7.5 bar)	114.30 (15.0 bar)	61.95 (30.0 bar)	37.82 (30.0 bar)
3:7	244.17 (5.0 bar)	116.68 (10.0 bar)	63.55 (20.0 bar)	38.75 (20.0 bar)
4:6	246.44 (5.0 bar)	117.30 (7.5 bar)	64.33 (15.0 bar)	39.54 (20.0 bar)
5:5	246.43 (2.5 bar)	118.31 (7.5 bar)	64.74 (10.0 bar)	39.81 (15.0 bar)
6:4	248.46 (2.5 bar)	118.34 (5.0 bar)	65.17 (10.0 bar)	39.95 (15.0 bar)
7:3	248.44 (2.5 bar)	118.96 (5.0 bar)	64.89 (7.5 bar)	40.02 (10.0 bar)
8:2	247.38 (2.5 bar)	116.78 (5.0 bar)	65.23 (7.5 bar)	40.04 (10.0 bar)
9:1	243.44 (2.5 bar)	118.25 (5.0 bar)	65.37 (5.0 bar)	40.02 (7.5 bar)

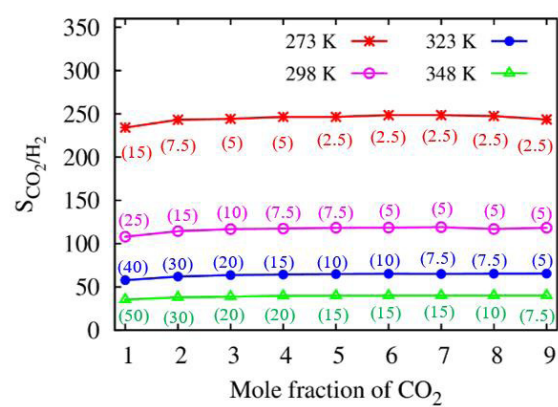


Figure 6. Dependence of maximum CO₂/H₂ selectivity on the mole ratios of CO₂ at 273 K, 298 K, 323 K, and 348 K. The values in parentheses refer to the pressure corresponding to the maximum CO₂/H₂ selectivity for each CO₂:H₂ mole fraction.

4. CONCLUSIONS

This study obtained significant results:

- The CO₂ uptakes in MIL-53(Cr) are high with the absence and presence of H₂. The CO₂ capture capacity insignificantly decreases when the molar fraction of H₂ increases no more than 60%, i.e., decreasing from 9.18 mmol/g (pure CO₂) to 8.37 mmol/g (CO₂:H₂ = 4:6) with 8.82%.
- The high CO₂/H₂ selectivity is also elucidated at various molar fractions of CO₂:H₂ and different temperatures (273 K, 298 K, 323 K, and 348 K). The results exhibit that when the CO₂ mole-fraction in the binary mixture increases, the maximum CO₂/H₂ selectivity remains nearly

constant, but the pressure corresponding to this maximal selective adsorption remarkably decreases. Moreover, the temperature strongly affects the separation capacity of CO₂ over H₂. Specifically, the maximum CO₂/H₂ selectivity decreases from 116 to 245 with the decreasing temperature from 298 K to 273 K. This work evidences that reducing the temperature can increase the selective adsorption efficiency of CO₂/H₂. Additionally, adjusting the mole fraction between hydrogen and carbon dioxide can help achieve the CO₂/H₂ selectivity at low pressure.

Acknowledgments

This research is conducted within the framework of science and technology projects at the institutional level of Quy Nhon University under project code T2023.794.04. We also acknowledge support for the calculations from the Lab of Computational Chemistry and Modeling (LCCM), Quy Nhon University, Vietnam.

REFERENCES

1. G. Avci, I. Erucar, S. Keskin. Do new MOFs perform better for CO₂ capture and H₂ purification? Computational screening of the updated MOF database, *ACS Applied Materials and Interfaces*, **2020**, 12, 41567-41579.

2. H. Demir, G. O. Aksu, H. C. Gulbalkan, S. Keskin. MOF membranes for CO₂ capture: past,

- present and future, *Carbon Capture Science and Technology*, **2022**, 2, 100026.
3. B. Dziejarski, J. Serafin, K. Andersson, R. Krzyżyńska. CO₂ capture materials: a review of current trends and future challenges, *Materials Today Sustainability*, **2023**, 24, 100483.
 4. L. F. Chuah, A. Bokhari, S. Asif, J. J. Klemeš, D. J. Dailin, H. E. Enshasy, A. H. M. Yusof. A review of performance and emission characteristic of engine diesel fuelled by biodiesel, *Chemical Engineering Transactions*, **2022**, 94, 1099-1104.
 5. X. Tan, S. Wang, N. Han. Metal organic frameworks derived functional materials for energy and environment related sustainable applications, *Chemosphere*, **2023**, 313, 137330.
 6. Y. Cao, X. Li, G. Yu, B. Wang. Regulating defective sites for pharmaceuticals selective removal: structure-dependent adsorption over continuously tunable pores, *Journal of Hazardous Materials*, **2023**, 442, 130025.
 7. G. Férey. Hybrid porous solids: past, present, future, *Angewandte Chemie International Edition*, **2008**, 37, 191–214.
 8. H. Furukawa, K. E. Cordova, M. O. Keeffe, O. M. Yaghi. The chemistry and applications of metal-organic frameworks, *Science*, **2013**, 341, 1230444.
 9. I. Hussain, S. Iqbal, C. Lamiel, A. Alfantazi, K. Zhang. Recent advances in oriented metal-organic frameworks for supercapacitive energy storage, *Journal of Materials Chemistry A*, **2022**, 10, 4475-4488.
 10. S. M. Moosavi, A. Nandy, K. M. Jablonka, D. Ongari, J. P. Janet, P. G. Boyd, Y. Lee, B. Smit, H. J. Kulik. Understanding the diversity of the metal-organic framework ecosystem, *Nature Communications*, **2020**, 11, 1-10.
 11. H. Daglar, S. Keskin. Recent advances, opportunities, and challenges in high-throughput computational screening of MOFs for gas separations, *Coordination Chemistry Reviews*, **2020**, 422, 213470.
 12. B. E. Keshta, H. Yu, L. Wang. MIL series-based MOFs as effective adsorbents for removing hazardous organic pollutants from water, *Separation and Purification Technology*, **2023**, 322, 124301.
 13. D. N. Son, N. T. X. Huynh, N. Thoai, P. T. Kien. Highly selective separation of CO₂ and H₂ by MIL-88A metal organic framework, *VNU Journal of Science: Mathematics – Physics*, **2021**, 37, 9-21.
 14. N. T. X. Huynh, O. K. Le, T. P. Dung, V. Chihaia, D. N. Son. Theoretical investigation of CO₂ capture in the MIL-88 series: effects of organic linker modification, *RSC Advances*, **2023**, 13, 15606-15615.
 15. N. T. X. Huynh, V. Chihaia, D. N. Son. Hydrogen storage in MIL-88 series, *Journal of Materials Science*, **2019**, 54, 3994-4010.
 16. S. Tomar, V. K. Singh. Review on synthesis and application of MIL-53, *Materials Today: Proceedings*, **2021**, 43, 3291-3296.
 17. C. Serre, F. Millange, C. Thouvenot, M. Noguès, G. Marsolier, D. Louër, G. Férey. Very large breathing effect in the first nanoporous Chromium(III)-based solids: MIL-53 or Cr^{III}(OH)·{O₂C–C₆H₄–CO₂}·{HO₂C–C₆H₄–CO₂H}_x·H₂O_y, *Journal of the American Chemical Society*, **2002**, 124, 13519-13526.
 18. F. Millange, R. I. Walton. MIL-53 and its isorecticular analogues: a review of the chemistry and structure of a prototypical flexible metal-organic framework, *Israel Journal of Chemistry*, **2018**, 58, 1019-1035.
 19. F. Millange, C. Serre, G. Férey. Synthesis, structure determination and properties of MIL-53as and MIL-53ht: the first Cr^{III} hybrid inorganic-organic microporous solids: Cr^{III}(OH)·{O₂C–C₆H₄–CO₂}·{HO₂C–C₆H₄–CO₂H}_x, *Chemical Communications*, **2002**, 8, 822-823.
 20. S. Bourrelly, P. L. Llewellyn, C. Serre, F. Millange, T. Loiseau, G. Férey. Different adsorption behaviors of methane and carbon dioxide in the isotypic nanoporous metal terephthalates MIL-53 and MIL-47, *Journal of the American Chemical Society*, **2005**, 127, 13519-13521.

21. D. Panda, S. Patra, M. K. Awasthi, S. K. Singh. Lab cooked MOF for CO₂ capture: a sustainable solution to waste management, *Journal of Chemical Education*, **2020**, 97, 1101-1108.
22. J. L. D. Miranda, T. P. D. Abreu, J. M. B. Neto, D. D. P. Souza, I. Coelho, F. Stavale, S. D. S. A. Oliveira, L. C. D. Moura. A case study for an eco-design of aluminum terephthalate metal-organic framework-MIL-53(Al) for CO₂ and methane adsorption, *Sustainable Materials and Technologies*, **2023**, 37, e00689.
23. Y. Jiao, Z. Li, Y. Ma, G. Zhou, S. Wang, G. Lu. The studies on gas adsorption properties of MIL-53 series MOFs materials, *AIP Advances*, **2017**, 7, 085009.
24. P. Rallapalli, K. P. Prasanth, D. Patil, R. S. Somani, R. V. Jasra, H. C. Bajaj. Sorption studies of CO₂, CH₄, N₂, CO, O₂ and Ar on nanoporous aluminum terephthalate [MIL-53(Al)], *Journal of Porous Materials*, **2011**, 18, 205-210.
25. D. Dubbeldama, S. Calerob, D. E. Ellisc, R. Q. Snurr, D. Dubbeldam, S. Calero, D. E. Ellis, R. Q. Snurr. RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials, *Molecular Simulation*, **2016**, 42, 81-101.
26. J. P. Perdew, K. Burke, M. Ernzerhof. Generalized gradient approximation made simple, *Physical Review Letters*, **1996**, 77, 3865-3868.
27. A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard III, W. M. Skiff. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations, *Journal of the American Chemical Society*, **1992**, 114, 10024-10035.
28. N. Q. Vinh, N. T. M. Duyen, N. L. B. Tran, N. Van Nghia, L. T. T. V. Vien, H. T. M. Thanh, N. T. X. Huynh. Computational study on enhancing SO₂ capture capacity of M₂(BDC)₂TED (M = Mg, V, Co, or Ni), *Quy Nhon University Journal of Science*, **2024**, 18, 91-100.
29. J. J. Potoff, J. I. Siepmann. Vapor-liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen, *AIChE Journal*, **2001**, 47, 1676-1682.
30. D. Levesque, A. Gicquel, F. L. Darkrim, S. B. Kayiran. Monte Carlo simulations of hydrogen storage in carbon nanotubes, *Journal of Physics: Condensed Matter*, **2002**, 14, 9285-9293.
31. A. Schneemann, V. Bon, I. Schwedler, I. Senkovska, S. Kaskel, R. A. Fischer. Flexible metal-organic frameworks, *Chemical Society Reviews*, **2014**, 43, 6062-6096.
32. Q. Yang, C. Zhong. Molecular simulation of carbon dioxide/methane/hydrogen mixture adsorption in metal-organic frameworks, *Journal of Physical Chemistry B*, **2006**, 110, 17776-17783.
33. Y. Qingyuan, X. U. Qing, L. I. U. Bei, Z. Chongli. Molecular simulation of CO₂/H₂ mixture separation in metal-organic frameworks: effect of catenation and electrostatic interactions, *Chinese Journal of Chemical Engineering*, **2009**, 17, 781-790.
34. A. Mallick, S. Saha, P. Pachfule, S. Roy, R. Banerjee. Selective CO₂ and H₂ adsorption in a chiral magnesium-based metal organic framework (Mg-MOF) with open metal sites, *Journal of Materials Chemistry*, **2010**, 20, 9073-9080.



© 2024 by the authors. This Open Access Article is licensed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license (<https://creativecommons.org/licenses/by-nc/4.0/>).

MỤC LỤC

1.	Ứng dụng quá trình điểm thời gian trong giám sát quá trình kinh doanh Nguyễn Trương Thành Hưng, Trần Thị Cẩm Thanh, Trương Thị Cẩm Mai	5
2.	Khảo sát tham số ảnh hưởng đến nội lực đài cọc bằng phương pháp phần tử hữu hạn Hoàng Công Vũ, Nguyễn Thị Thảo Nguyên, Trần Bá Cảnh	21
3.	Phương pháp định vị sự cố dựa trên truyền sóng một đầu sử dụng biến đổi wavelet rời rạc Ngô Minh Khoa, Đoàn Đức Tùng	29
4.	Đánh giá đặc điểm nông học và hàm lượng saponin của một số mẫu giống mướp đắng rừng (<i>Momordica charantia</i> L. var. <i>abbreviata</i> Ser.) thu thập ở khu vực Nam Trung Bộ, Việt Nam Bùi Hồng Hải, Nguyễn Thị Y Thanh, Hồ Tân	43
5.	Nghiên cứu, chế tạo hệ thống đóng cắt và giám sát thông số hệ thống điện dựa trên xử lý ảnh và IoT Lương Ngọc Toàn, Đoàn Đức Tùng, Lê Tuấn Hộ	55
6.	Nghiên cứu về ảnh hưởng của hàm lượng xi măng đến khả năng chịu lực của cọc xi măng đất để xử lý đất sét yếu ở khu đô thị sông Hà Thanh, thành phố Quy Nhơn, tỉnh Bình Định, Việt Nam Hoàng Công Duy, Phạm Văn Trực, Phạm Thị Lan, Nguyễn Thị Khánh Ngân	65
7.	Xây dựng mặt tương tác và đánh giá khả năng chịu lực của cột bê tông cốt thép tiết diện hình chữ nhật Lê Thị Mai Trang, Vàng Hiếu Quang	71
8.	Nghiên cứu ứng dụng phần mềm SCADA Designer thiết kế chương trình quản lý nhu cầu sử dụng điện năng Trần Tiến Đạt, Nguyễn Vũ Hòa, Ngô Minh Khoa	83
9.	Ứng dụng học đồ thị trong hệ gợi ý du lịch Vũ Sơn Lâm, Lê Quang Hùng, Trần Đình Luyện, Phan Thị Thùy, Tăng Nhật Hưng	95
10.	Nghiên cứu khả năng hấp phụ và tách lọc carbon dioxide của vật liệu MIL-53(Cr) bằng phương pháp mô phỏng cổ điển Nguyễn Thị Xuân Huynh, Trần Thị Diễm Thanh, Nguyễn Lê Bảo Trân, Nguyễn Ngọc Khoa Trường	109

