TẠP CHÍ KHOA HỌC
JOURNAL OF SCIENCE

**CHUYÊN SAN KHOA HỌC TỰ NHIÊN VÀ KỸ THUẬT
ISSUE: NATURAL SCIENCES AND ENGINEERING**

MỤC LỤC

Phương pháp mới dự báo theo chuỗi thời gian mờ dựa trên Đại số gia tử

Lê Xuân Việt[1,*], Dương Hoàng Huyên[1], Lê Xuân Vinh[2], Lê Thị Thu Vân[3]

[1]Khoa Công nghệ thông tin, Trường Đại học Quy Nhơn, Việt Nam
[2]Phòng Đào tạo Đại học, Trường Đại học Quy Nhơn, Việt Nam
[3]Tổ Tin học, Trường THPT chuyên Lê Khiết, Việt Nam

TÓM TẮT

Cho đến nay đã có nhiều tác giả đề xuất các phương pháp dự báo theo chuỗi thời gian mờ. Trong bài báo này, chúng tôi đề xuất một phương pháp mới dự báo chuỗi thời gian mờ dựa trên đại số gia tử. Để kiểm chứng tính hiệu quả của phương pháp, chúng tôi dự báo dựa trên dữ liệu về lượng khách du lịch quốc tế đến Việt Nam và so sánh với phương pháp ARIMA. Kết quả cho thấy phương pháp mới đề xuất cho sai số nhỏ hơn. Hơn nữa, với bộ dữ liệu là số lượng sinh viên nhập học trường Đại học Alabama, phương pháp vừa đề xuất cũng cho kết quả dự báo tốt hơn so với các phương pháp của các tác giả khác.

Từ khóa: *Dự báo, chuỗi thời gian mờ, đại số gia tử, hạng từ ngôn ngữ.*

---

*Tác giả liên hệ chính.
Email: lexuanviet@qnu.edu.vn

A Novel Fuzzy Time Series Method for Forecasting Based on Hedge Algebras

Le Xuan Viet[1,*], Duong Hoang Huyen[1], Le Xuan Vinh[2], Le Thi Thu Van[3]

[1]Department of Information Technology, Quy Nhon University, Vietnam

[2]Undergraduate Training Office, Quy Nhon University, Vietnam

[3] Department of Information Technology, Le Khiet high school for gifted student, Vietnam

ABSTRACT

So far, many methods have been proposed to deal with forecasting problems using fuzzy time series. In this paper, we have proposed a new fuzzy time series forecasting method based on hedge algebras. To examine efficiency of our method, we employed the data of the Vietnam's outbound tourists in order to predict and compare with ARIMA. It is clear that our method has smaller error than ARIMA method. Furthermore, experimenting our method with the data of the enrollment of Alabama University, our forecasting results are better than ones presented in other studies.

Keywords: *Forecasting, fuzzy time series, hedge algebras, linguistic term.*

1. INTRODUCTION

The forecasting activities play an important role in our daily life. Accurate forecasting will be used to help people making more suitable decisions. The classical time series methods can not deal with forecasting problems in which the values of time series are linguistic terms represented by fuzzy sets. So, Song and Chissom[1] presented the theory of fuzzy time series to overcome the drawback of the classical time series methods. Over the years, fuzzy time series has been widely used for forecasting data. A lot of studies have been discussed for forecasting used fuzzy time series such as enrollment,[2-9] the stock index,[10] foreign tourists[11] and financial forecasting,[12] etc.

As we know, when solving problems using fuzzy sets, many authors used hedge algebra instead.[13-17] The advantage of using the hedge algebra is that the linguistic values are arranged in semantic order. Moreover, we can easily calculate on these linguistic values because they are quantified into real values by using semantic quantitative function. In this paper, we shall propose a new method based on hedge algebra to solve the forecasting problem. The advantage of this method is that from the fuzziness of linguistic terms of hedge algebras we can determine the corresponding intervals of the discourse of linguistic variable. Moreover, thanks to the parameters of hedge algebras we can adjust the lengths of the intervals, thereby helping the forecasting process to obtain better results. This paper is organized by 5 sections. The first is the introduce. In Section 2, the basic concepts of fuzzy sets and fuzzy time series are

described. In Section 3, we have proposed a new method to solve the forecasting problem based on hedge algebras and given some examples to handle. The details are presented in Section 4. The verification and comparison of our method are done with some models.[18-23] Finally, the conclusions and comments are mentioned in Section 5.

2. FUZZY TIME SERIES

2.1. Basic concepts of fuzzy time series

Fuzzy time series model was firstly given by Q. Song and B.S Chissom.[1,7,8] Then, it is improved by S.M Chen[4] to process some arithmetic calculations. From that points, they can get more exactly forecasting results. In this session, we briefly review the concepts of fuzzy time series as in Chen.[19]

Let U be the universe of discourse, where $U = \{u_1, u_2,..., u_n\}$. A fuzzy set defined in the universe of discourse U can be represented as follows:

$$A = f_A(u_1)/u_1 + f_A(u_2)/u_2 + \cdots + f_A(u_n)/u_n,$$

where f_A denotes the membership function of the fuzzy set A, $f_A : U \rightarrow [0, 1]$, and $f_A(u_i)$ denotes the degree of membership of u_i belonging to the fuzzy set A, and $f_A(u_i) \in [0, 1]$, and $1 \leq i \leq n$.

Definition 2.1. Let $Y(t)$ $(t = ...,0,1,2,...)$ be the universe of discourse and be a subset of R. Assume $f_i(t)$ $(i = 1,2,...)$ are defined on $Y(t)$, and assume that $F(t)$ is a collection of $f_1(t), f_2(t), ...,$ then $F(t)$ is called a fuzzy time series definition $Y(t)$ $(t = ...,0,1,2,...)$.

Definition 2.2. Assume that $F(t)$ is caused by $F(t-1)$ only, denoted as $F(t-1) \rightarrow F(t)$, then this relationship can be expressed as $F(t) = F(t-1) \circ R(t, t-1)$, where $F(t) = F(t-1) \circ R(t, t-1)$ is called the first-order model of $F(t)$, $R(t, t-1)$ is the fuzzy relationship between $F(t-1)$ and $F(t)$, and "$\circ$" is the Max-Min composition operator.

Definition 2.3. Let $R(t, t-1)$ be a first-order model of $F(t)$. If for any t, $R(t, t-1) = R(t-1, t-2)$, then $F(t)$ is called a time-invariant

fuzzy time series. Otherwise, it is called a time-variant fuzzy time series.

Definition 2.4. Assume that the fuzzified input data of the ith year/month is A_j and the fuzzified input data of the $i+1^{th}$ year/month is A_k, where A_j and A_k are two fuzzy sets defined in the universe of discourse U, then the fuzzy logical relationship can be represented by $A_j \rightarrow A_k$, where A_j is called the current state of the fuzzy logical relationship.

If we have $A_i \rightarrow A_{j1}, A_i \rightarrow A_{j2}, ..., A_i \rightarrow A_{jk}$ then we can write $A_i \rightarrow A_{j1}, A_{j2}, ..., A_{jk}$.

2.2. Rules for fuzzy time series forecasting

Assume that A_j is the value of $F(t-1)$, the forecasted output $F(t)$ be defined with some rules[2]:

i) If there exist a relation 1-1 within group of the relations where A_j on the left of rule, suppose that $A_j \rightarrow A_k$, and the maximum membership value of A_k occurs at interval u_k, then the output of $F(t)$ is middle point of u_k.

ii) If $A_k = \varnothing$, that mean $A_j \rightarrow \varnothing$ and the maximum membership value of A_j occurs at interval u_j, then the output of $F(t)$ is middle point of u_j.

iii) If we have $A_j \rightarrow A_1, A_2, ..., A_n$, and the maximum membership values of $A_1, A_2, ..., A_n$ occur at intervals $u_1, u_2, ..., u_n$ respectively, then the output of $F(t)$ is average of the middle points $m_1, m_2, ..., m_n$ of $u_1, u_2, ..., u_n$, that is $(m_1 + m_2 + ... + m_n)/n$.

3. THE MODEL OF FORECASTING TIME SERIES BASED ON HEDGE ALGEBRAS

In this section, we shall give a short overview on the algebraic approach to the semantics of vague words in natural languages investigated in papers[13-17] and construct a new method to forecast rely on hedge algebras theory.

3.1. Hedge Algebras: A Short Overview

Given linguistic variable X, every word-domain $\mathcal{X}$ of a variable X can be considered as an abstract algebra AX = $(\mathcal{X}, \mathbb{G}, \mathbb{C}, \mathcal{H}, \leq)$, where

● $\mathcal{H}$ is the set of linguistic hedges or modifiers considered as *1-ary operations* of the algebra AX;

● $\mathbb{C} = \{0, W, 1\}$ is a set of *special words* which are, respectively, the least, the medium and the greatest elements of $\mathcal{X}$ and regarded as *constants* of AX since they are fixed points;

● $\mathbb{G} = \{c^-, c^+\}$ is a set of the primary or atomic words of the variable X, the first one is called the *negative* word, say "*young*" of AGE, and the second, the *positive* one, say "*old*".

● $\mathbb{G} \cup \mathbb{C}$ is the set of the generators of the algebra AX that is $\mathcal{H}(\mathbb{G} \cup \mathbb{C}) = \mathcal{X} = \mathbb{C} \cup \mathcal{H}(\mathbb{G})$, the underlying set of AX where for a subset Z of $\mathcal{X}$, the set $\mathcal{H}(Z)$ denotes the set of all elements freely generated from the words in Z. I.e. $\mathcal{H}(Z) = \{\sigma x : x \in Z \text{ and } \sigma \in \mathcal{H}^*\}$, where $\mathcal{H}^*$ is the set of all strings of hedges in $\mathcal{H}$, including the empty string ε. Note that for $\sigma = \varepsilon$, $\varepsilon x = x$ and, hence, $Z \subseteq \mathcal{H}(Z)$. In the case $Z = \{x\}$ we shall write $\mathcal{H}(x)$ instead of $\mathcal{H}(\{x\})$.

● $\leq$ is a Semantical Order Relation (SOR) upon $\mathcal{X}$.

If a word $x \in \mathcal{X}$ is generated from a given element u by means of hedges h_i, $i = 1 \ldots, n$, then we will write $x = h_n \ldots h_1 u$ and it is called a string representation of x with respect to (w.r.t.) u.

For illustration, let us consider an HA of the variable AGE, AX = $(\mathcal{X}, \mathbb{G}, \mathbb{C}, \mathcal{H}, \leq)$, where

$\mathbb{G} = \{c^-, c^+\}$, $\mathcal{H}^- = \{R, L\}$ and $\mathcal{H}^+ = \{V, E\}$, where c^- and c^+ stand for "young" and "old" and R, L, V and E for "Rather", "Little", "Very" and "Extremely", respectively. The order and the generality-specificity relation and the set-containing relation of the fuzziness models of the words can be represented by a labeled graph given in Figure 1, in which every *arrow* means that the word at its root generates the word at its peak and the relationships between the words and between the fuzziness models at the respective positions of the arrow are specified by an inequality and a set-containing relation associated with the arrow.

Consider an HA AX = $(\mathcal{X}, \mathbb{G}, \mathbb{C}, \mathcal{H}, \leq)$ of an attribute X with numeric reference interval domain U normalized to be $[0,1]$, for convenience in a unified presentation of the quantification of the hedge algebras. Formally, the numeric semantics of the words of X can be determined by a so-called Semantically Quantifying Mapping (SQM), $f : \mathbb{X} \to [0, 1]$, defined as follows.

Definition 3.1. A mapping $f: \mathcal{X} \to [0, 1]$ is said to be an SQM of AX, if we have:

(SQM1) f is an *order isomorphism*, i.e. it is one-to-one and for $\forall x, y \in \mathcal{X}, x \leq y \Rightarrow f(x) \leq f(y)$.

(SQM2) The image of $\mathcal{X}$ under f, $f(\mathcal{X})$, is topologically dense in the universe $[0, 1]$.
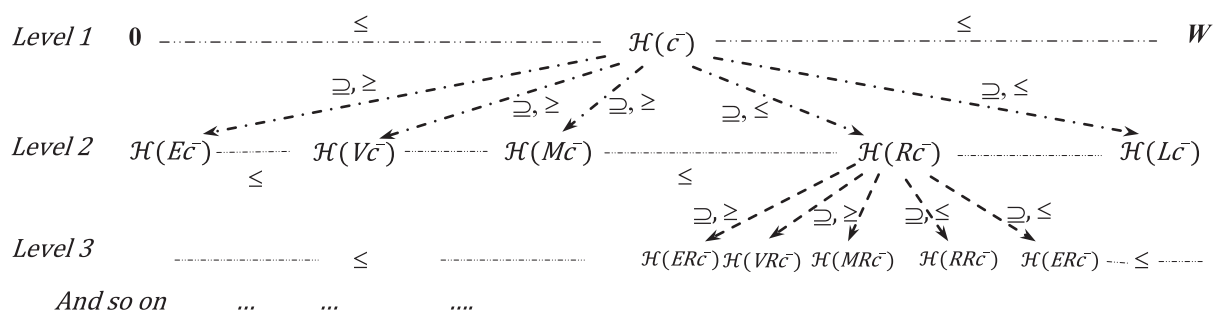

Figure 1. A piece of the graph representation of the semantic structure of the words of X generated from the atomic word c^- and their fuzziness models, which involves the order-based relationships and the set-containing relationships between the fuzziness models of words and the word generality-specificity relationships

Definition 3.2. A function $fm: \mathcal{X} \rightarrow [0, 1]$ is said to be a fuzziness measure of the HA AX associated with the given variable X, if it satisfies the following axioms, for any $x \in \mathcal{X}$ and $h \in \mathcal{H}$:

(fm1) $fm(c^-) + fm(c^+) = 1$.

(fm2) $\sum_{-q \leq j \leq p, j \neq 0} fm(h_j x) = fm(x)$.

(fm3) $fm(hx) = \mu(h)fm(x)$, where $\mu(h)$ is called for convenience the fuzziness measure of h as well.

(fm4) For $x = h_n h_{n-1} \ldots h_1 c$, $fm(x) = fm(h_n h_{n-1} \ldots h_1 c) = \mu(h_n)\mu(h_{n-1}) \ldots \mu(h_1)fm(c)$, $c \in \mathbb{G} = \{ c^-, c^+ \}$.

(fm5) Setting $\sum_{-q \leq j \leq -1} \mu(h_j) = \alpha$ & $\sum_{1 \leq j \leq p} \mu(h_j) = \beta$, we have $\alpha + \beta = \sum_{-q \leq j \leq p, j \neq 0} \mu(h_j) = 1$.

In the general case, for given values of the fuzziness parameters of X we can stablish a recursive expression to compute the SQM v_{fm}, called the SQM induced by the given fm, as follows:

• $v_{fm}(W) = \kappa = fm(c^-)$, $v_{fm}(c^-) = \kappa - \alpha fm(c^-) = \beta fm(c^-)$, $v_{fm}(c^+) = \kappa + \alpha fm(c^+)$;

• $v_{fm}(h_j x) = v_{fm}(x) + sign(h_j x) \times$

$$\left(\sum_{i = sign(j)}^{j-1} \mu_i(h_i) + (1 - \omega(h_j x)) \mu(h_j) \right) fm(x)$$

where

$$\omega(h_j x) = \frac{1}{2}[1 + sign(h_j x) sign(h_p h_j x)(\beta - \alpha)] \in \{\alpha, \beta\}$$

for all $j \in [-q \ldots p], j \neq 0$, and $sign()$ function is defined as in Ho N.C.[16]

3.2. Semantization and desemantization

To convert the values from the reference domain to semantic domain of a variable X and vice versa, we synthesize some transformations as below. Assume that $[a, b]$ is a reference domain of the variable X, and $[a_s, b_s]$ is semantic domain. The linear conversion from $[a, b]$ to $[a_s, b_s]$ is called *linear semantization* and the conversion from $[a_s, b_s]$ to $[a, b]$ is called *linear*

desemantization. The semantic domain by using hedge algebras is usually $[0, 1]$, thus the linear semantization is named *normalization* and linear desemantization is named *denormalization*. We have some functions as below:

LinearSemantization $(x) =$
$\quad x_s = a_s + (b_s - a_s)(x - a)/(b - a)$
LinearDesemantization $(x_s) =$
$\quad x = a + (b - a)(x_s - a_s)/(b_s - a_s)$
Normalization $(x) = x_s = (x - a)/(b - a)$
Denormalization $(x_s) = x = a + (b - a)x_s$

For flexibility in semantization or desemantization, we can expand from linear to nonlinear by adding some parameters $sp, dp \in [-1, 1]$, for example:

NonlinearSemantization $(x) = f(x_s, sp)$, satisfy the conditions $0 \leq f(x_s, sp) \leq 1$, $f(x_s = 0, sp) = 0$, $f(x_s = 1, sp) = 1$.
NonlinearDesemantization $(x_s) = g(x, dp)$, satisfy conditions $a \leq g(x, dp) \leq b$, $g(x = a, dp) = a$, $g(x = b, dp) = b$.

In this paper, we shall use the functions:

NonlinearNormalization $(x) = f(x_s, sp) = sp \times x_s(1-x_s) + x_s$
NonlinearDenormalization $(x_s) = dp \times$ (Denormalization($f(x_s, sp) - a) \times (b -$ Denormalization($f(x_s, sp)))/(b - a) +$ Denormalization $(f(x_s, sp))$, where Denormalization $(f(x_s, sp)) = (sp \times x(1 - x) + x) \times (b - a) + a$.

The *NonlinearDenormalization* be denoted *ND* for short.

3.3. A new method on forecasting

Inputs: n values of data $\{y(t_1), y(t_2),\ldots, y(t_n)\}$ with $t_1, t_2, \ldots, t_n$ are point times.

Output: $y(t_{n+1})$ at the time t_{n+1}.

Step 1. Define the discourse U

Put $U = [D_{min}, D_{max}]$ where $D_{min} = \min\{y(t_1), y(t_2),\ldots, y(t_n)\}$ and $D_{max} = \max\{y(t_1), y(t_2),\ldots, y(t_n)\}$.

Step 2. Building the intervals upon U by using fuzziness model of hedge algebras

We assume that U is a discourse of an abstract variable $\mathcal{X}$ and the linguistic domain of $\mathcal{X}$ to be considered as an algebra AX = $(\mathcal{X}, \mathbb{G}, \mathbb{C}, \mathcal{H}, \leq)$. Then, by chosen k, the discourse U is divided into k intevals $u_1, u_2, ..., u_k$ w.r.t level k as in Figure 1. The inteval u_i is labeled A_i, $i = 1, 2,..., k$ satifying $A_1 < A_2 < ... < A_k$. We caculate the length of intervals of u_i denoted f_{ui}, $f_{ui} = fm(A_i) \times (D_{max} - D_{min})$, $i = 1, 2, ..., k$. So we have $u_1 = [u_{1d}, u_{1c}] = [D_{min}, D_{min} + f_{u1}]$, $u_2 = [u_{2d}, u_{2c}] = [u_{1c}+1, u_{2d}+f_{u2}]$, ..., $u_k = [u_{kd}, u_{kc}] = [u_{(k-1)c}+1, u_{kd}+f_{uk}]$.

Step 3. Quantifying semantic of the linguistic values $A_1, A_2, ..., A_k$.

To quantify the semantic of $A_1, A_2, ..., A_k$, we use SQM v_{fm} as $SA_1 = v_{fm}(A_1)$, $SA_2 = v_{fm}(A_2)$, ..., $SA_k = v_{fm}(A_k)$. By properties of hedge algebras, it is clear that $SA_1 < SA_2 < ... < SA_k$.

Step 4. Constructing the relationships

Suppose that, $F(t-1)$ is A_i, $F(t)$ is A_j, and $F(t)$ is caused by $F(t-1)$. Clearly, we have relation between A_i and A_j, denoted $A_i \rightarrow A_j$.

Step 5. Grouping relationship

If $A_i \rightarrow A_{j1}, A_i \ A_{j2},..., A_i \rightarrow A_{jm}$, then we estalish the relation by grouping all of them to unique relation $A_i \rightarrow A_{j1}, A_{j2}, ..., A_{jm}$.

Step 6. Caculating output value

From group of the relations in Step 5, applying the rules similarity in the way of Section 2.2 we get the results of $F(t)$, scilicet:

• If there is a relation $A_i \rightarrow A_j$, then $F(t) = ND(SA_j)$ upon $u_j = [u_{jd}, u_{jc}]$.

• If $A_i \rightarrow \varnothing$ then $F(t) = ND(\varnothing)$ upon $u_i = [u_{id}, u_{ic}]$.

• If $A_i \rightarrow A_{j1}, A_{j2}, ..., A_{jk}$ then $F(t) = ND(W_{i,j1} \times SA_{j1} + W_{i,j2} \times SA_{j2} + ... + W_{i,jk} \times SA_{jk})$ upon $[\min\{u_{j1d}, u_{j2d}, ..., u_{jkd}\}, \max\{u_{j1c}, u_{j2c}, ..., u_{jkc}\}]$ where $W_{i,j}$ is the weights measured in the ratio number of times of real data in the interval u_i to sum of number of times of real data in the intervals $u_{j1}, u_{j2}, ..., u_{jk}$.

4. EXPERIMENTAL RESULTS

Now, we would like to forecast the number of tourists going to Vietnam on Jan 2012 based on the real data from January 2010 to December 2011 (the data are given in column "Real Data" of Table 1). This data are also on website of Vietnam National Administration of Tourism (*http://vietnamtourism.gov.vn*).

For illustration, we shall apply the algorithm step by step to solve this problem.

• The discourse U = $[D_{min}, D_{max}]$ = [286618, 611864].

• Chosen a hedge algebra AX = $(\mathcal{X}, \mathbb{G}, \mathbb{C}, \mathcal{H}, \leq)$, where $\mathbb{G}$ = {*Small, Large*}, $\mathbb{C}$ = {**0, W, 1**}, $\mathcal{H}$ = {*Little, Very*}, $\mu(Little)$ = 0.5, $\mu(Very)$ = 0.5, κ = 0.5. In this case, level k be defined is 3, so we clearly get 8 linguistic values $A_1 < A_2 < A_3 < A_4 < A_5 < A_6 < A_7 < A_8$ with respect to *VeryVerySmall < LittleVerySmall < LittleLittleSmall < VeryLittleSmall < VeryLittleLarge < LittleLittleLarge < LittleVeryLarge < VeryVeryLarge*. Obviously, $fm(A_1) = fm(A_2) = ... = fm(A_8) = 0.125$. Consequently, $f_{u1} = f_{u2} = ... = f_{u8}$ = 40656. Finally, the discouse U be divided into 8 intervals such as u_1 = [286618, 327274], u_2 = [327275, 367931], u_3 = [367932, 408588], u_4 = [408589, 449245], u_5 = [449246, 489902], u_6 = [489903, 530559], u_7 = [530560, 571216], u_8 = [571217, 611864].

• Quantifying semantic of the linguistic values $A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8$ by using SQM v_{fm} of AX, we have SA_1 = 0.0625, SA_2 = 0.1875, SA_3 = 0.3125, SA_4 = 0.4375, SA_5 = 0.5625, SA_6 = 0.6875, SA_7 = 0.8125, SA_8 = 0.9375.

• Constructing the relations from data

Table 1. The relations from data

Order	Time	Real data	Label	The relations
1	1/2010	416249	A_4	
2	2/2010	446323	A_4	$A_4 \rightarrow A_4$
3	3/2010	473509	A_5	$A_4 \rightarrow A_5$
4	4/2010	432608	A_4	$A_5 \rightarrow A_4$

5	5/2010	350982	A_2	$A_4 \to A_2$
6	6/2010	375707	A_3	$A_2 \to A_3$
7	7/2010	410000	A_4	$A_3 \to A_4$
8	8/2010	427935	A_4	$A_4 \to A_4$
9	9/2010	383463	A_3	$A_4 \to A_3$
10	10/2010	440071	A_4	$A_3 \to A_4$
11	11/2010	428295	A_4	$A_4 \to A_4$
12	12/2010	449570	A_5	$A_4 \to A_5$
13	1/2011	506424	A_6	$A_5 \to A_6$
14	2/2011	542671	A_7	$A_6 \to A_7$
15	3/2011	475733	A_5	$A_7 \to A_5$
16	4/2011	460000	A_5	$A_5 \to A_5$
17	5/2011	480886	A_5	$A_5 \to A_5$
18	6/2011	446966	A_4	$A_5 \to A_4$
19	7/2011	460000	A_5	$A_4 \to A_5$
20	8/2011	490000	A_6	$A_5 \to A_6$
21	9/2011	286618	A_1	$A_6 \to A_1$
22	10/2011	518477	A_6	$A_1 \to A_6$
23	11/2011	611864	A_8	$A_6 \to A_8$
24	12/2011	593408	A_8	$A_8 \to A_8$

- Group of relations from the Table 1

$A_1 \to (A_6)$

$A_2 (A_3)$

$A_3 (A_4, A_4)$

$A_4 (A_2, A_3, A_4, A_4, A_4, A_5, A_5, A_5)$

$A_5 (A_4, A_4, A_5, A_5, A_6, A_6)$

$A_6 (A_1, A_7, A_8)$

$A_7 (A_5)$

$A_8 (A_8)$

Table 2. The weights of group relations

Intv.	Month/ Year	Amount of tourists	Group relations and weights $W_{i,j}$	Num. of values in u_i
u_1	9/2011	286618	$A_1 \to (A_6)$ $W_{1,6} = 3/3 = 1$	1
u_2	5/2010	350982	$A_2 \to (A_3)$ $W_{2,3} = 2/2 = 1$	1

u_3	6/2010	375707	$A_3 \to (A_4, A_4)$	2
	9/2010	383463	$W_{3,4} = 8/(2\times8) = 1/2$	
u_4	1/2010	416249	$A_4 \to (A_2, A_3, A_4, A_4, A_4,$ $A_5, A_5, A_5)$	8
	2/2010	446323		
	4/2010	432608	$W_{4,2} = 1/(1+2+8\times3+6\times3)$ $= 1/45$	
	7/2010	410000	$W_{4,3} = 2/(1+2+8\times3+6\times3)$ $= 2/45$	
	8/2010	427935		
	10/2010	440071	$W_{4,4} = 8/(1+2+8\times3+6\times3)$ $= 8/45$	
	11/2010	428295	$W_{4,5} = 6/(1+2+8\times3+6\times3)$ $= 6/45$	
	6/2011	446966		
u_5	3/2010	473509	$A_5 \to (A_4, A_4, A_5, A_5,$ $A_6, A_6)$	6
	12/2010	449570	$W_{5,4} = 8/(8\times2+6\times2+3\times2)$ $= 8/34$	
	3/2011	475733		
	4/2011	460000	$W_{5,5} = 6/(8\times2+6\times2+3\times2)$ $= 6/34$	
	5/2011	480886	$W_{5,6} = 3/(8\times2+6\times2+3\times2)$ $= 3/34$	
	7/2011	460000		
u_6	1/2011	506424	$A_6 \to (A_1, A_7, A_8)$ $W_{6,1} = 1/(1+1+2) = 1/4$	3
	8/2011	490000	$W_{6,7} = 1/(1+1+2) = 1/4$	
	10/2011	518477	$W_{6,8} = 2/(1+1+2) = 2/4$ $= 1/2$	
u_7	2/2011	542671	$A_7 \to (A_5)$ $W_{7,5} = 6/6 = 1$	1
u_8	11/2011	611864	$A_8 \to (A_8)$	2
	12/2011	593408	$W_{8,8} = 2/2 = 1$	

For example, to calculate number of tourists at the time February 2010, because the current state is A_4, we consider the relation $A_4 \to (A_2, A_3, A_4, A_4, A_4, A_5, A_5, A_5)$. According to Table 2, the data in this group belong to intervals u_2, u_3, u_4, u_5 where u_2 contains 1 value, u_3 contains 2 values, u_4 contains 8 values, but A_4 appears three times, thus the number of values of u_4 is (3×8). Similarly, u_5 contains (3×6) values. So, sum of them equal to $(1+2+3\times8+3\times6 = 45)$ and quantified semantic x_s is 0.47639.

With the manual chosen $sp = 0.6$ and $dp = -0.2$, $x = Denormalization(x_s) = f(0.47639, 0.6) = (0.6 \times 0.47639 \times (1 - 0.47639) + 0.47639) \times (489902 - 327275) + 327275 = 429089$ and

$ND(x)$ = $g(429089,-0.2)$ = $-0.2 \times (429089-327275) \times (489902-429089)/(489902-327275) + 429089 = 421474$.

We find out the forecasted value on February 2010 is 421474. Completely similar as above, all forecated values are represented in Figure 2.
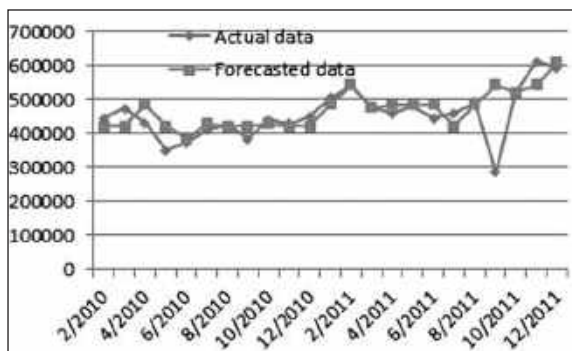

Figure 2. Curves of forecasted data and actual data

In the following, we use the mean square error (MSE) to compare the forecasting results of different forecasting methods, where the mean square error is calculated as follows:

$$MSE = \frac{\sum\limits_{i=1}^{n} \left(Actual\_Data_i - Forecasted\_Data_i \right)^2}{n}$$

where *Actual_Data i* denotes the actual data of month *i*, and *Forecasted_Data i* denotes the forecasted data of month *i*. In Table 3, we compare the forecasting results of the proposed method with the one of the existing method[4] through five month from 1/2011 to 5/2011.

Table 3. A comparison results between our method and ARIMA method[18]

Month/Year	Real data	ARIMA method[18]	The proposed method
1/2011	506424	443853	485611
2/2011	542671	426238	542401
3/2011	475733	611361	476444
4/2011	460000	558587	485611
5/2011	480886	609786	485611
MSE		**249402**	**33347**

To further affirm the effectiveness of our method, we experiment with data of enrollment students of Alabama University from 1971 to 1992.

With chosen hedge algebra AX = ($\mathcal{X}$, $\mathbb{G}$, $\mathbb{C}$, $\mathcal{H}$, $\leq$), $\mathbb{G}$ = *{Small, Large}*, $\mathbb{C}$ = {**0, W, 1**}, $\mathcal{H}$ = *{Little, Very}*, where the parameters are set $\mu(Little)$ = 0.44, $\mu(Very)$ = 0.56, κ = 0.4 and sp = 0.2, dp = 0.0, the calculation is completely similar to the one above. We get the result as shown in Table 4.

Table 4. MSE of some methods with admission data of Alabama University

Year	Num. of students	Chen Method[19]	Lee Method[20]	Manikandan Method[21]	Qiu Method[22]	Huarng Method[23]	The proposed Method
1971	13055						
1972	13563	14000	13833	13739	14195	14000	13233
1973	13867	14000	13833	14639	14424	14000	13233
1974	14696	14000	13833	15539	14593	14000	14608
1975	15460	15500	15500	15557	15589	15500	15192
1976	15311	16000	15722	15539	15645	15500	15611
1977	15603	16000	15722	15557	15634	16000	15611
1978	15861	16000	15722	16739	16100	16000	16159
1979	16807	16000	15722	17639	16188	16000	16159
1980	16919	16833	16750	16439	17077	17500	17155
1981	16388	16833	16750	15539	17105	16000	17155
1982	15433	16833	16750	15557	16369	16000	16159
1983	15497	16000	15722	15557	15643	16000	15611
1984	15145	16000	15722	15539	15648	15500	15611
1985	15163	16000	15722	15539	15622	16000	15611
1986	15984	16000	15722	16739	15623	16000	15611
1987	16859	16000	15722	17639	16231	16000	16159
1988	18150	16833	16750	19139	17090	17500	17155
1989	18970	19000	19000	19439	18325	19000	19235
1990	19328	19000	19000	19289	19000	19000	19235
1991	19337	19000	19000	19289	19000	19500	19235
1992	18876	19000	19000	19439	19000	19000	19235
MSE		407507	397537	325158	261473	226611	220401

5. CONCLUSIONS

In this paper, we have proposed a fuzzy time series method for forcasting based on hedge algebras. The proposed method is both simple and effective. Especially, the intervals of the discourse are made upon the k-level of hedge algebras. We have experimented our method with the data of tourists going to Vietnam and data of enrollment student of University of Alabama. The forecasted results obtained by our method are better than ones of the other methods. It can be proven with MSE as in Table 3&4.

The errors of the forecasted results of our method are directly influenced by the parameters of hedge algebras, but these were intuitively chosen. In the future, we shall put forward a way to set up these parameters automatically, for instance using particle swarm optimization to reach a higher forecasting accuracy rate.

REFERENCES

1. Song Q, Chissom B.S. Fuzzy time series and its models. *Fuzzy Sets and Syst.*, **1993**, *54*, 269 - 277.

   2. Chen, S. M. Forecasting Enrollments Based on Fuzzy Time Series. *Fuzzy Sets and Syst.* **1996**, *81*, 311 - 319.

3. Chen S. M. Forecasting Enrollments based on High Order Fuzzy Time Series. *Cybernetics and Systems An International Journal*, **2002**, *33*, 1 - 16.

4. Chen S. M and Chung N. Y. Forecasting enrollments using high-order fuzzy time series and genetic algorithms, *Int. Journal of Intelligent Systems*, **2006**, *21*, 485 - 501.

5. Lee M. H, Efendi R, Ismad Z. Modified Weighted for Enrollments Forecasting Based on Fuzzy Time Series. *MATEMATIKA*, **2009**, *25* (1), 67 - 78.

6. Lin, C.J.; Chen, H.F.; Lee, T.S. Forecasting tourism demand using time series, artificial neural networks and multivariate adaptive regression splines. *Evidence from Taiwan. Int. J. Bus. Adm.*, **2011**, *2*, 14 - 24.

7. Song Q, Chissom B.S. Forecasting enrollments with fuzzy time series – part 1. *Fuzzy Sets and Syst.*, **1993**, *54*, 1 - 9.

8. Song Q, Chissom B.S. Forecasting enrollments with fuzzy time series - part 2. *Fuzzy Sets and Syst.*, **1994**, *62*, 1 - 8.

9. Huarng, K. Heuristic Models of Fuzzy Time Series for Forecasting, *Fuzzy Sets and Syst.* **2001**, *123*, 369 - 386.

10. C.H. Cheng, T.L. Chen & C.H. Chiang. Trend-weighted fuzzy time series model for TAIEX forecasting, *ICONIP, Part III, LNNC*, **2006**, *4234*, 469 - 477.

11. Hu. Predicting Foreign Tourists for the Tourism Industry Using Soft Computing-Based Grey–Markov Models, *Sustainability*, **2017**, *9*(7), 1-12.

12. C.H.L.Lee, A.Liu & W.S.Chen. Pattern Discovery of Fuzzy time series for financial prediction, *IEEE Transactions on Knowledge and data Engineering*, **2006**, *18*, 613 - 625.

13. N.C. Ho, W. Wechler. Hedge algebras: An algebraic approach to structures of sets of linguistic domains of linguistic truth variable, *Fuzzy Sets and Systems*, **1990**, *35* (3), 281 - 293.

14. N. Cat Ho and W. Wechler, Extended hedge algebras and their application to Fuzzy logic, *Fuzzy Sets and Systems*, **1992**, *52*, 259 - 281.

15. N.C. Ho H.V. Nam. Towards an Algebraic Foundation for a Zadeh Fuzzy Logic, *Fuzzy Sets and System*, **2002**, *129*, 229 - 254.

16. N.C. Ho. A Topological Completion of Refined Hedge Algebras and a Model of Fuzziness of Linguistic words, *Fuzzy Sets and Systems*, **2007**, *158* (4), 436 - 451.

17. N.C Ho, V.N. Lan, L.X Viet. Optimal hedge-algebras-based controller: Design and Application, *Fuzzy Sets and Systems*, **2008**, *159*, 968 - 989.

18. D.Q Giam, V.T Han, L.T.L. Phuong, N.T. Thuy. Building ARIMA model to forecast Vietnam outbound tourists, *Journal of Science and Development Hanoi Agricultural University*, **2012**, *10* (2), 364 - 370.

19. Chen, S.M. Forecasting Enrollments Based on Fuzzy Time Series, *Fuzzy Sets and Syst*, **1996**, *81*, 311 - 319.

20. Lee M H, Efendi R, Ismad Z. Modified Weighted for Enrollments Forecasting Based on Fuzzy Time Series. *MATEMATIKA*, **2009**, *25*(1), 67 - 78.

21. Manikandan. M, Dr. Senthamarai Kannan. K, Deneshkumar. V. Computational Method Based on Distribution in Fuzzy Time Series Forecasting, **2013**, *2* (8).

22. Qiu W, Liu X, Li H. Generalized Method for Forecasting Based on Fuzzy Time Series, *Expert Systems with Applications*, **2011**, *38*, 10446 - 10453.

23. Huarng, K. Heuristic Models of Fuzzy Time Series for Forecasting, *Fuzzy Sets and Syst*, **2001**, *123*, 369 - 386.

Dự báo lượng mưa hàng tháng của Quy Nhơn bằng mô hình Sarima

Nguyễn Quốc Dương[1,*], Lê Phương Thảo[1], Đinh Thị Quỳnh Như[1],
Cao Thị Ái Loan[2], Phùng Thị Hồng Diễm[2], Lê Thanh Bính[2]

[1]Khoa Sư phạm, Trường Đại học Quy Nhơn, Việt Nam
[2]Khoa Toán và Thống kê, Trường Đại học Quy Nhơn, Việt Nam

TÓM TẮT

Mô hình SARIMA được sử dụng rộng rãi để phân tích dữ liệu chuỗi thời gian. Trong bài báo này, chúng tôi sẽ áp dụng phương pháp này với sự trợ giúp của phần mềm thống kê R để dự báo lượng mưa hàng tháng tại thành phố Quy Nhơn, tỉnh Bình Định, Việt Nam. Dữ liệu được thu thập từ tháng 1 năm 2000 đến tháng 12 năm 2018 để thiết lập mô hình và dự báo.

Keywords: *ARIMA, phương pháp Box-Jenkins, SARIMA, chuỗi thời gian, lượng mưa Quy Nhơn.*

---

[*]Tác giả liên hệ chính.
Email: nguyenquocduongqnu1999@gmail.com

Monthly rainfall forecast of Quy Nhon using SARIMA model

Nguyen Quoc Duong[1,*], Le Phuong Thao[1], Dinh Thi Quynh Nhu[1],
Cao Thi Ai Loan[2], Phung Thi Hong Diem[2], Le Thanh Binh[2]

[1]Faculty of Education, Quy Nhon University, Viet Nam
[2]Faculty of Mathematics and Statistics, Quy Nhon University, Viet Nam

ABSTRACT

The SARIMA model is widely used to analyze time series data. In this paper, we will apply this method with the help of R statistical software to forecast monthly rainfall in Quy Nhon city, Binh Dinh Province, Viet Nam. Mean monthly rainfall from 2000 to 2018 were used for modelling and forecasting.

Keywords: *ARIMA, Box-Jenkins methodology, SARIMA, time series, Quy Nhon rainfall.*

1. INTRODUCTION

Quy Nhon is a coastal city in Binh Dinh Province in central Vietnam. Here, weather is characterized by tropical monsoon climate with two distinct seasons, rainy season and dry season. Rainy season lasts from September to December, while dry season runs from January to August.

Changes in rainfall precipitation will be one of the most critical factors determining the overall impact of climate change. Therefore, its model-ling and forecasting are needed for water resources management, irrigation scheduling, agricultural management and reservoir opera-tion, tourism. Prediction of rainfall is tough due to its non linear pattern and a large variation in intensity. Till today, numerous techniques have been used to forecast rainfall. Among them, Autoregressive Integrated Moving Average (ARIMA) modeling, introduced by Box and Jenkins is an effective method.[1] The Box-Jenkins Seasonal ARIMA (SARIMA) model has several advantages over other models, parti-cularly over exponential smoothing and neural network, due to its forecasting capability and richer information on time related changes.[2] ARIMA model considers the serial correlation, which is the most important characteristic of time series data, and also provides a systematic option to identify a better model. Another advantage of ARIMA model is that the model uses less parameter to describe a time series. Therefore, we use the SARIMA model to predict rainfall in Quy Nhon.

Besides mathematical, software tools today also play an important role in forecasting. There are many software tools for highly effective data analysis such as SPSS, Eviews, Python, etc. In this study, we use R statistical software to analyze the rainfall data. The advantages of R programming are open source programming language, providing exemplary support for data organization, package arrays, quality plotting and graphing, highly compatible, platform independent reporting and machine learning activity.

Corresponding author.
Email: nguyenquocduongqnu1999@gmail.com

The rainfall data in Quy Nhon are obtained from General Statistics Office of Viet Nam (https://www.gso.gov.vn), and cover monthly observations from 2000 to 2018. We will proceed to build the appropriate forecasting model and compare evaluation between actual data and forecast data.

2. METHODOLOGY

2.1. The seasonal ARIMA model

Seasonal Autoregressive Integrated Moving Average, SARIMA or Seasonal ARIMA, is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component. SARIMA$(p, d, q)(P, D, Q)_s$ model can be most succinctly expressed using the backward shift operator:

$$\Phi_P(B^s)\phi_P(B)(1 - B^s)^D(1 - B)^d y_t = c + \Theta_Q(B^s)\theta_q(B)\varepsilon_t,$$

where Φ_p, ϕ_p, Θ_Q and θq are polynomials of orders P, p, Q, and q, respectively. The parameter p and q define the order of the AR and MA processes with its non-seasonal lags, respectively, and d defines the degree of differencing of the series with its non-seasonal lags. Likewise, the P and Q parameters represent the corresponding order of the seasonal AR and MA processes of the series with its seasonal lags, and D defines the degree of differencing of the series with its non-seasonal lags. In general, the model is non-stationary, although if $D = d = 0$ and the roots of the characteristic equation all exceed unity in absolute value, the resulting model would be stationary.

2.2. Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF)

ACF is an (complete) auto-correlation function which gives us values of auto-correlation of any series with its lagged values. We plot these values along with the confidence band and tada! We have an ACF plot. In simple terms, it describes how well the present value of the series is related with its past values. A time series can have components like trend, seasonality, cyclic and residual. ACF considers all these com-ponents while finding correlations hence it is a 'complete auto-correlation plot'.[3]

PACF is a partial auto-correlation function. Basically instead of finding correlations of present with lags like ACF, it finds correlation of the residuals (which remains after removing the effects which are already explained by the earlier lag(s)) with the next lag value hence 'partial' and not 'complete' as we remove already found variations before we find the next correlation. So if there is any hidden information in the residual which can be modeled by the next lag, we might get a good correlation and we will keep that next lag as a feature while modeling. Remember while modeling we do not want to keep too many features which are correlated as that can create multicollinearity issues. Hence we need to retain only the relevant features.[3]

2.3. Modelling procedure

When fitting an SARIMA model to a set of time series data, the following procedure provides a useful general approach.

❖ Step 1: Data preparation: Here, we prepare the data for the training and testing process of the model. This step includes splitting the series into training (in-sample) and testing (out-sample) partitions, creating new features (when applicable), and applying a transformation if needed (for example, log transformation, scaling, and so on).

❖ Step 2: Train the model: Here, we used the training partition to train a statistical model. The main goal of this step is to utilize the training set to train, tune, and estimate the model coefficients that minimize the selected error criteria. The fitted values and the model estimation of the training partition observations will be used later on to evaluate the overall performance of the model.

❖ Step 3: Test the model: Here, we utilize the trained model to forecast the corresponding observations of the testing partition. The main

idea here is to evaluate the performance of the model with a new dataset (that the model did not see during the training process).

❖ Step 4: Model evaluation: Last but not least, after the model was trained and tested, it is time to evaluate the overall performance of the model on both the training and testing partitions.[4]

2.4. Tuning the SARIMA model

The tuning process of the SARIMA model follows the same logic as one of the ARIMA models. However, the complexity of the model increases as there are now six parameters to tune, that is, p, d, q, P, D and Q, as opposed to three with the ARIMA model. Luckily, the tuning of the P, D, and Q seasonal parameters follows the same logic as the ones of p, d, q, respectively, with the use of the ACF and PACF plots. The main difference between the tuning of these two groups of parameters (non-seasonal and seasonal) is that the non seasonal parameters are tuned with the non-seasonal lags, as we saw previously with the ARIMA model. On the other hand, the tuning of the seasonal parameters are tuned with the seasonal lags (for example, for monthly series with lags 12, 24, 36, and so on).[4,5]

2.4.1. Tuning the non-seasonal parameters

Applying the same logic that we used with the ARIMA model, tuning the non-seasonal parameters of the SARIMA model is based on the ACF and PACF plots:

❖ An AR(p) process should be used if the non-seasonal lags of the ACF plot are tailing off, while the corresponding lags of the PACF plots are cutting off on the p lag.

❖ Similarly, an MA(q) process should be used if the non-seasonal lags of the ACF plot are cutting off on the q lag and the corre-sponding lags of the PACF plots are tailing off.

❖ When both the ACF and PACF non-seasonal lags are tailing off, an ARMA model should be used.

❖ Differencing the series with the non-seasonal lags should be applied when the non-

seasonal lags of the ACF plot are decaying in a linear manner.[4]

2.4.2. Tuning the seasonal parameters

Tuning the seasonal parameters of the SARIMA model with ACF and PACF follows the same guidelines as the ones we used for selecting the ARIMA parameters:

❖ We will use a seasonal autoregressive process with an order of P, or SAR(P), if the seasonal lags of the ACF plot are tailing off and the seasonal lags of the PACF plot are cutting off by the P seasonal lag.

❖ Similarly, we will apply a seasonal moving average process with an order of Q, or SMA(Q), if the seasonal lags of the ACF plot are cutting off by the Q seasonal lag and the seasonal lags of the PACF plot are tailing off.

❖ An ARMA model should be used whenever the seasonal lags of both the ACF and PACF plots are tailing off.

❖ Seasonal differencing should be applied if the correlation of the seasonal lags are decaying in a linear manner.[6]

2.5. A step-wise procedure for traversing the model space

Suppose that we have seasonal data, we consider ARIMA$(p, d, q)(P, D, Q)_s$ models, where p and q can take values from 0 to 3, and P and Q can take values from 0 to 1. When $c = 0$ there is a total of 288 possible models, and when $c \neq 0$ there is a total of 192 possible models, giving 480 models altogether. If the values of p, d, q, P, D and Q are allowed to range more widely, the number of possible models increases rapidly. If d and D are known, we can select the orders p, q, P and Q via an information criterion such as the AIC:

$$AIC := -2\log(L) + 2(p + q + P + Q + k),$$

where $k = 1$ if $C \neq 0$ and 0 otherwise, and L is the maximized likelihood of the model fitted to the differenced data $(1 - B^s)^D (1 - B)^d y_t$. The likelihood of the full model for y_t is not actually

defined and so the value of the AIC for different levels of differencing are not comparable.[5]

Consequently, it is often not feasible to simply fit every potential model and choose the one with the lowest AIC. Instead, we need a way of traversing the space of models efficiently in order to arrive at the model with the lowest AIC value. We propose a step-wise algorithm as follows.

Step 1: We try four possible models to start with.

❖ ARIMA(2, d, 2) if $s = 1$ and ARIMA(2, d, 2)(1, D, 1) if $s > 1$;

❖ ARIMA(0, d, 0) if $s = 1$ and ARIMA(0, d, 0)(0, D, 0) if $s > 1$;

❖ ARIMA(1, d, 0) if $s = 1$ and ARIMA(1, d, 0)(1, D, 0) if $s > 1$;

❖ ARIMA(0, d, 1) if $s = 1$ and ARIMA(0, d, 1)(0, D, 1) if $s > 1$.

If $d + D \leq 1$, these models are fitted with $C \neq 0$. Otherwise, we set $c = 0$. Of these four models, we select the one with the smallest AIC value. This is called the 'current' model and is denoted by ARIMA-(p, d, q) if $s = 1$ or ARIMA(p, d, q)(P, D, Q)$_s$ if $s > 1$.

Step 2: We consider up to 13 variations on the current model:

❖ where one of p, q, P and Q is allowed to vary by ±1 from the current model;

❖ where p and q both vary by ±1 from the current model;

❖ where P and Q both vary by ±1 from the current model;

❖ where the constant c is included if the current model has $c = 0$ or excluded if the current model has $C \neq 0$.

Whenever a model with lower AIC is found, it becomes the new 'current' model and the procedure is repeated. This process finishes when we cannot find a model close to the current model with lower AIC.

There are several constraints on the fitted models to avoid problems with convergence or near unit roots. The constraints are outlined below:

❖ The values of p and q are not allowed to exceed specified upper bounds (with default values of 5 in each case).

❖ The values of P and Q are not allowed to exceed specified upper bounds (with default values of 2 in each case).

❖ We reject any model which is 'close' to non-invertible or non-causal. Specifi-cally, we compute the roots of $\phi(B)\Phi(B)$ and $\theta(B)\Theta(B)$. If either have a root that is smaller than 1.001 in absolute value, the model is rejected.

❖ If there are any errors arising in the non-linear optimization routine used for estimation, the model is rejected. The rationale here is that any model that is difficult to fit is probably not a good model for the data.

The algorithm is guaranteed to return a valid model because the model space is finite and at least one of the starting models will be accepted (the model with no AR or MA parameters). The selected model is used to produce forecasts.[5,6]

2.6. Forecast evaluation methods

Once you finalize the model tuning, it is time to test the ability of the model to predict observations that the model did not see before (as opposed to the fitted values that the model saw throughout the training process). The most common method for evaluating the forecast's success is to predict the actual values with the use of an error metric to quantify the forecast's overall accuracy. The selection of a specific error metric depends on the forecast accuracy's goals. This study only considers common error metric is as follow:

Root Mean Squared Error (RMSE): This is the root of the average squared distance of the actual and forecasted values:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2},$$

where Y_t and $\hat{Y}_t$ are the actual value of the original series and predicted value from the proposed hybrid model, respectively. The smallest value of RMSE indicates the best model.[4]

3. RESULTS AND DISCUSSION

The data has been collected at the Quy Nhon station from 2000 to 2018. We have a monthly series with 228 observations (19 years) and the goal is to forecast the next two years (24 months). The corresponding command packages and libraries for model prediction are *forecast, readxl, tseries, TSstudio*.[6-8] Let us load the rainfall series from file *datats.xlsx*.

Let us plot the series with the ts_plot function and review the main characteristics of the series by the code below:

```
#plot time series data with
ts_plot function
ts_plot(datats, title=
"Monthly Rainfall in Quy
Nhon", Ytitle="Rainfall(mm)",
Xtitle="Year")
```

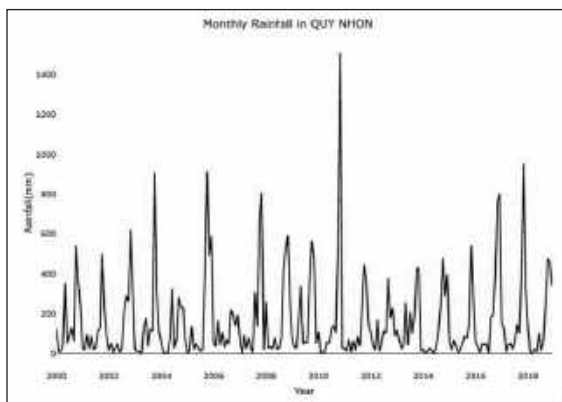*We attain the output as shown in Figure 1:*

Figure 1. Time series plot of observed mean monthly rainfall from 2000 to 2018.

From Figure 1, the *datats* series has a strong seasonal pattern, therefore the SARIMA model is the most appropriate one for our data. In addition, the series is trending up, so we can already conclude that the series is not stationary and some differencing of the series is required. We would use the first 216 observations for training and test the performance using the last 12 observations. Creating partitions in R can be done manually with the *ts_split* function from the stats package. For instance, let is split the *mydata* series into partitions, leaving the last 12 observations of the series as the testing partition and the rest as training:

```
QN_rainfall<-ts_split(datats,
sample.out = 12)
train <- QN_rainfall$train
test <- QN_rainfall$test
```

In statistics, an *augmented Dickey–Fuller (ADF) test* the null hypothesis that a unit root is present in a time series sample. The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend stationarity. We obtain the following output:

```
adf.test(train, alternative =
"stationary")
##Dickey-Fuller = -8.0188,
Lag order = 5, p-value = 0.01
##alternative hypothesis:
stationary
```

Before we start the training process of the SARIMA model, we will conduct diagnostics in regards to the series correlation with the ACF and PACF functions. Since we are interested in viewing the relationship of the series with its seasonal lags, we will increase the number of lags to calculate.

```
par(mfrow=c(1,2))
acf(ts(train),main="ACF For
Rainfall", col="blue",lwd = 4)
pacf(ts(train),main="PACF For
Rainfall",col="coral",lwd = 4)
```
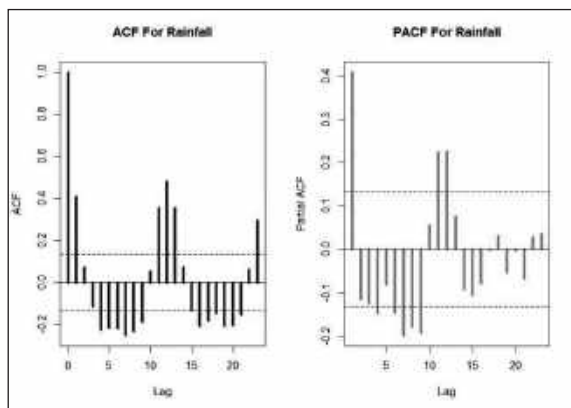
Figure 2. ACF and PACF plot of rainfall for Quy Nhon station.

The preceding ACF plot indicates that the series has a strong correlation with both the seasonal and non-seasonal lags. Furthermore, the linear decay of the seasonal lags indicates that the series is not stationary and that seasonal differencing is required. We will start with a seasonal differencing of the series and plot the output to identify whether the series is in a stationary state. The R commans and output are as follows:

```
ndiffs(train) #to determine d
(the number of seasonal
differences to use)
##0
nsdiffs(train)#to determine D
(the number of ordinary
differences to use)
##1
QN_rainfall_12 <- diff(train,
lag = 12, differences = 1)
ts_plot(QN_rainfall_12, title =
"Quy Nhon Monthly Rainfall -
First Seasonal
Difference",Ytitle = Rainfall
(First Difference)",Xtitle =
"Year")
```
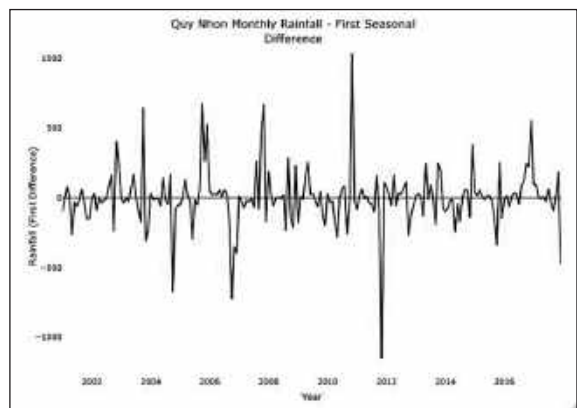
By the *ts_plot* function, we get the output as Figure 3:

Figure 3. Plot of seasonal differencing of the series.

After taking the first order differencing, along with the first order seasonal differencing, the series seems to stabilize around the zero x axis line (or fairly close to being stable). After transforming the series into a stationary state, we can review the ACF and PACF functions again to identify the required process:

```
par(mfrow=c(1,2))
acf(ts(QN_rainfall_12), main=
"ACF For First Seasonal
Difference", col="blue", lwd = 4)
pacf(ts(QN_rainfall_12), main=
"ACF For First Seasonal
Difference", col="coral", lwd = 4)
```

It should note that this step is very important. We need to the observe from the ACF and PACF plots in order to choose simple models.

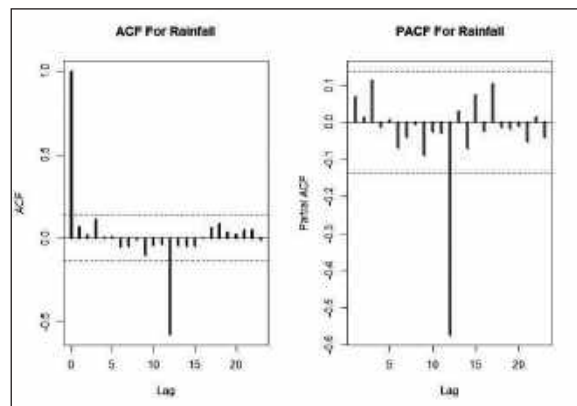The output of the commans above is shown in Figure 4:

Figure 4. ACF and PACF plot of rainfall for Quy Nhon station.

The main observation from the preceding ACF and PACF plots is that both the nonseasonal and seasonal lags (in both plots) are tailing off. Hence, we can conclude that after we difference the series and transform them into a stationary state, we should apply an ARMA process for both the seasonal and non-seasonal components of the SARIMA model. Our aim now is to find an appropriate ARIMA model based on the ACF and PACF plot. The significant spike at lag 1 in the ACF suggests a non-seasonal MA(1) component, and the signsifycant spike at lag 12 in the ACF suggests a seasonal MA(1) component. Consequently, we begin with an $SARIMA(1,0,0)(1,1,0)_{12}$ model, indicating seasonal difference, and non-seasonal and seasonal MA(1) components. Based on the PACF plot, we can choose another simple model $SARIMA(0,0,0)(0,1,0)_{12}$. Of these two models, the best is the SARIMA model (i.e., it has the smallest AIC value).

```
md1<- Arima(train,
order=c(1,0,0),
seasonal=c(1,1,0))
summary(md1)
##AIC: 2683.2
md2<- Arima(train,
order=c(0,0,0),
seasonal=c(0,1,0))
summary(md2)
##AIC: 2762.83
```

Consequently, this initial analysis suggests that a possible model for these data is an $SARIMA(1,0,0)(1,1,0)_{12}$. We fit this model, along with some variations on it, compute the AIC values and test set evaluation shown as in Table 1.

Table 1. AIC and RMSE values for various SARIMA models applied to the monthly rainfall data.

Models	AIC	RMSE
$SARIMA(1,0,0)(1,1,0)_{12}$	2683.20	164.4443
$SARIMA(0,0,0)(1,1,0)_{12}$	2682.27	164.8799
$SARIMA(2,0,0)(1,1,0)_{12}$	2684.69	164.2065
$SARIMA(1,0,1)(1,1,0)_{12}$	2681.33	162.7866
$SARIMA(1,0,0)(0,1,0)_{12}$	2763.84	203.7255
$SARIMA(1,0,0)(2,1,0)_{12}$	2658.01	152.5503
$SARIMA(1,0,0)(1,1,1)_{12}$	2650.34	146.6646
$SARIMA(2,0,1)(1,1,0)_{12}$	2683.32	162.7785
$SARIMA(1,0,0)(2,1,1)_{12}$	2649.60	145.6458

Of these models, the best is the SARIMA-$(1,0,0)(2,1,1)_{12}$ model (which has the lowest RMSE value on the training set, and the best AIC value amongst models with only seasonal differencing). Before we finalize the forecast, let's evaluate the selected model's performance on the testing set. We will retrain the model using the settings of the selected model:

```
QNrainfall_best_md <-
Arima(train, order = c(1,0,0),
seasonal =list(order=
c(2,1,1)))
```

We then get the output as described in Table 2:

Table 2. Summary of $SARIMA(1,0,0)(2,1,1)_{12}$.

SARIMA(1,0,0)(2,1,1)$_{12}$				
SARIMA	*ar1*	*sar2*	*sar2*	*sma1*
Coefficients	0.067	-0.052	0.1	-0.881
s.e.	0.068	0.092	0.087	0.093
AIC = 2650.34				

Let us use the QNrainfall_best_md trained model to forecast the corresponding observations of the testing set:

```
QNrainfall_test_fc <- forecast(QNrainfall_
best_md, h = 12)
```

We will assess the model's performance by put

$$H = \frac{Actual\ Value - Forecast\ Value}{Actual\ Value}$$

and these values are calculated as in Table 3:

Table 3. Assess performances of SARIMA(1,0,0) - $(2,1,1)_{12}$ model.

Month	Forecast Value	Actual Value	H
January	81.81485	128.6	0.36
February	32.32366	2.8	-10.54
March	34.07819	1.6	-20.29
April	46.95980	20.0	-1.34
May	82.40952	9.4	-7.76
June	44.63459	103.7	0.56
July	57.38414	14.0	-3.09
August	122.92861	51.1	-1.40
September	209.18078	235.5	0.11
October	420.49951	476.7	0.11
November	574.92124	462.0	-0.24
December	321.89014	337.9	0.04

From Table 3, we see the predicted model has two values big deviations in February and March. This is unavoidable because climate change is very complex. However, the forecast value and actual value of the rainy season from September to December for very low error. Moreover, these predicted values are very suitable for the climate characteristics of Quy Nhon city.

Now, we will use the test_forecast function to get a more intuitive view of the model is performance on the training and testing partitions:

```
test_forecast(datats,forecast.
obj = QNrainfall_test_fc,test
= test)
```
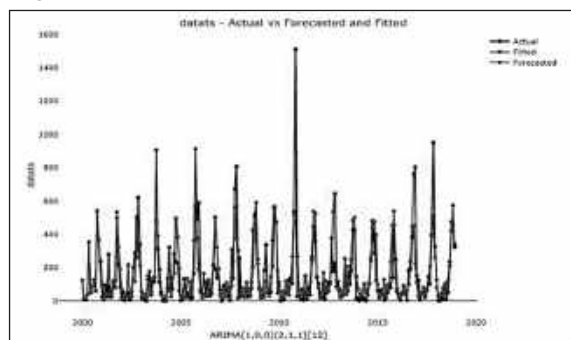
We then have the output as shown in Figure 5:

Figure 5. *Plot of datats – actual & forecasted an fited.*

Now that we have satisfied the preceding conditions, we can move on to the last step of the forecasting process and generate the final forecast with the selected model. We will start by retraining the selected model on all the series:

```
final_md <- Arima(datats,
order = c(1,0,0), seasonal =
list(order=c(2,1,1)))
```

Before we forecast the next 24 months, let is verify that the residuals of the model satisfy the model condition. We execute the code as follows:

```
checkresiduals(final_md)
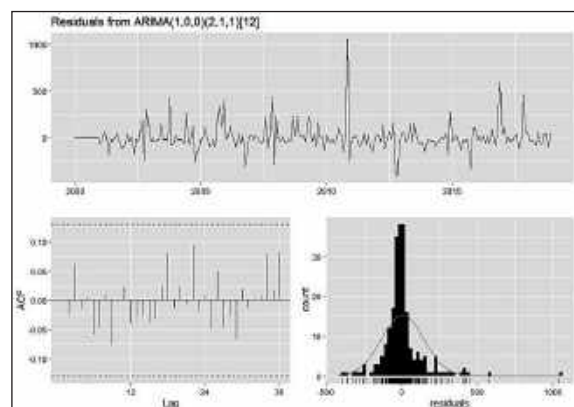```

The output is as Figure 6:



**Figure 6.** Residuals from the SARIMA(1,0,0)-$(2,1,1)_{12}$ model applied to monthly rainfall data.

The output of the Ljung-Box test suggested that the residuals of the model are white noise:

```
## Ljung-Box test
## data: Residuals from
ARIMA(1,0,0)(2,1,1)[12]
## Q* = 9.7273,
df = 20,
p-value= 0.9729
## Model df:4.
## Total lags used: 24
```

By looking at the preceding residuals plot, you can see that the residuals are white noise and normally distributed. Furthermore, the Ljung-Box test confirms that there is no autocorrelation left on the residuals with a p-value of 0.9729, we cannot reject the null hypothesis that the

residuals are white noise. Thus, we now have a seasonal ARIMA model that passes the required checks and is ready for forecasting.

The main goal of the forecasting process is to minimize the level of uncertainty around the future values of the series. Although we cannot completely eliminate this uncertainty, we can quantify it and provide some range around the point estimate of the forecast. The confidence interval is a statistical approximation method that's used to express the range of possible values that contain the true value with some degree of confidence (or probability). We now use the

forecast function to obtain the predicted values for the next 24 months of the data series.

```
QNrainfall_fc <-
forecast(final_md, h = 24)
QNrainfall_fc
```

Forecast package is written by Rob J Hyndman and is available from CRAN here. The R package forecast provides methods and tools for displaying and analysing univariate time series forecasts including exponential smoothing via state space models and automatic ARIMA modelling.

We obtain the result as described in Table 4:

Table 4. Forecasts of the mothly rainfall data using the SARIMA$(1,0,0)(2,1,1)_{12}$ model with 80% and 95% confidence intervals.

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2019	91.76578	-96.60397	280.1355	-196.32090	379.8525
Feb 2019	42.27684	-146.48860	231.0423	-246.41498	330.9687
Mar 2019	33.84519	-154.92190	222.6123	-254.84917	322.5396
Apr 2019	43.94335	-144.82376	232.7105	-244.75103	332.6377
May 2019	83.59559	-105.17151	272.3627	-205.09878	372.2900
Jun 2019	46.60968	-142.15743	235.3768	-242.08470	335.3041
Jul 2019	61.58513	-127.18198	250.3522	-227.10925	350.2795
Aug 2019	116.15631	-72.61080	304.9234	-172.53807	404.8507
Sep 2019	198.88576	10.11865	387.6529	-89.80862	487.5801
Oct 2019	434.81604	246.04893	623.5831	146.12166	723.5104
Nov 2019	592.95521	404.18811	781.7223	304.26084	881.6496
Dec 2019	260.07814	71.31136	448.8449	-28.61573	548.7720
Jan 2020	89.25500	-99.61077	278.1208	-199.59027	378.1003
Feb 2020	27.27827	-161.58825	216.1448	-261.56814	316.1247
Mar 2020	31.56637	-157.30015	220.4329	-257.28005	320.4128
Apr 2020	40.21528	-148.65124	229.0818	-248.63113	329.0617
May 2020	75.65715	-113.20936	264.5237	-213.18926	364.5036
Jun 2020	58.20946	-130.65706	247.0760	-230.63695	347.0559
Jul 2020	53.28597	-135.58055	242.1525	-235.56044	342.1324
Aug 2020	102.82699	-86.03953	291.6935	-186.01943	391.6734
Sep 2020	214.99832	26.13180	403.8648	-73.84809	503.8447
Oct 2020	445.11325	256.24673	633.9798	156.26684	733.9597
Nov 2020	534.54059	345.67407	723.4071	245.69418	823.3870
Dec 2020	265.01471	76.14849	453.8809	-23.83125	553.8607

We can plot historical data with forecasts and confidence intervals by the *plot_forecast* function:

```
plot_forecast(QNrainfall_fc,title = "Quy Nhon Rainfall Forecast",
Ytitle = "Rainfall(mm)",Xtitle = "Year")
```
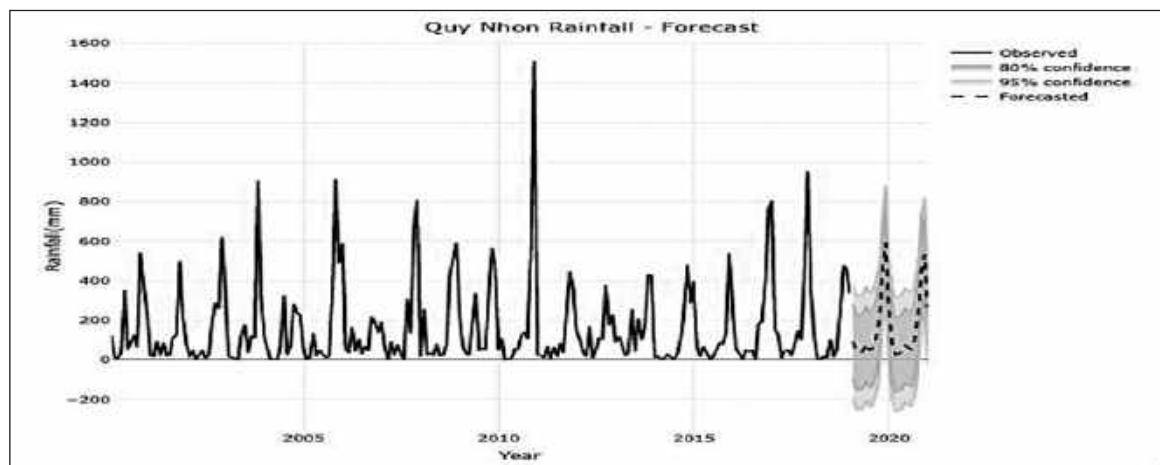
We then obtain the output as Figure 7:

Figure 7. Forecasts of the mothly rainfall data using the ARIMA$(1,0,0)(2,1,1)_{12}$ model with 80% and 95% confidence intervals.

Note that the lower bounds are somewhat funny for rainfall. Is the earth going to sprinkle water back into the clouds? We explain this as follows: the simulation of the forecasts generating a family of forecasts for each period can simply be truncated at 0.0, which is meant to decouple the variance of the errors from the expected value of the model when one can safely ignore values lower than 0.0.

4. CONCLUSIONS

In this study, we used the SARIMA model for forecasting monthly rainfall data of Quy Nhon city. Based on seasonally differenced correlogram characteristics, different SARIMA models were evaluated; their parameters were optimized, and diagnostic check up of forecasts was made by using white noise and heter-oscedasticity tests. The best SARIMA model (corresponding to our data) was chosen based on smallest value of AIC and RMSE. A validation check was performed on residual series. Residuals was found white noise for SARIMA-$(1,0,0)(2,1,1)_{12}$ model. The predicted values from the model were compared with the actual values to determine prediction precision. We found that selected model predicted monthly rainfall with a reasonable accuracy. Therefore, year-long rainfall can be forecasted using these models. Moreover, this model can be applied in the study of the time series in similar fields at Quy Nhon or other cities.

Acknowledgement

REFERENCES

1. G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control,* San Fran-cisco: Holden-Day, 1976.

2. A. K. Mishra and V. R. Desai. Drought fore-casting using stochastic models, *Stochastic Environmental Research and Risk Assessment,* **2005**, *19*, 326-339.

3. J. Salvi. Significance of ACF and PACF Plots In Time Series Analysis, <https://towardsdata-science.com/significance-of-acf-and-pacf-plots-in-time-series-analysis-2fa11a5d10a8>, accessed 20/10/2019.

4. R. Krispin. *Hands-On Time Series Analysis with R,* Packt Publisher, 2019.

5. R. J. Hyndman and G. Athanasopoulos. *Fore-casting: Principles and Practice, 2ⁿᵈ edition,* OTexts Publisher, 2018.

6. R. J. Hyndman and Y. Khandakar. Automatic Time Series Forecasting: The forecast Package for R, *Journal of Statistical Software*, **2008**, *27*(3), 1-22.

7. C. Chatfield. *The Analysis Of Time Series-An Introduction, 5th edition*, Chapman and Hall, 1996.

8. G. James, D. Witten, T. Hastie and R. Tibshirani. *An Introduction to Statistical Learning with Applications in R, 1st edition*, Springer Texts in Statistics, 2007.

Mô hình tiên đoán bệnh tim mạch vành
sử dụng hồi quy logistic dựa vào tập dữ liệu Evans

Lê Thanh Bính[*]

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

TÓM TẮT

Bệnh tim mạch, bao gồm cả bệnh tim mạch vành (CHD), nằm trong số những bệnh phổ biến ở cả những nước phát triển và đang phát triển và được xem là nguyên nhân chủ yếu gây tử vong trên toàn thế giới. Chỉ riêng bệnh tim mạch vành, bệnh này tiếp tục là nguyên nhân hàng đầu gây nên bệnh tật và tử vong ở người trưởng thành tại châu Âu và Bắc Mỹ. Sự tiên lượng sớm bệnh tim mạch vành có thể giúp đưa ra các quyết định thay đổi lối sống ở những bệnh nhân có nguy cơ cao và từ đó làm giảm các biến chứng của bệnh. Vì vậy, sử dụng các thuật toán khai phá dữ liệu có thể hữu ích trong dự đoán bệnh tim mạch vành. Nghiên cứu này nhằm xây dựng một mô hình dự đoán bệnh tim mạch vành sử dụng hồi quy logistic, với sự trợ giúp của phần mềm thống kê R, dựa vào tập dữ liệu Evans về bệnh tim.

Từ khóa: *Bệnh tim mạch, bệnh tim mạch vành (CHD), mô hình hồi quy logistic, tập dữ liệu Evans.*

*Tác giả liên hệ chính.
Email: lethanhbinh@qnu.edu.vn

A coronary heart disease prediction model using logistic regression based on Evans dataset

Le Thanh Binh[1,*]

[1]*Faculty of Mathematics and Statistics, Quy Nhon University, Vietnam*

ABSTRACT

Cardiovascular diseases, including coronary heart disease (CHD), are among the common diseases in both developed and developing countries and regarded as the main cause of death throughout the world. Coronary heart disease itself has been still being the leading cause of morbidity and mortality among adults in Europe and North America. The early prognosis of coronary heart disease can help making decisions in changing lifestyle of high-risk patients, thereby reducing complications of the disease. Therefore, the use of data mining algorithms could be useful in predicting coronary heart disease. This study aimed to create a coronary heart disease prediction model using logistic regression, with the help of statistical software R, based on the Evans heart disease dataset.

Keywords: *Cardiovascular diseases, coronary heart disease (CHD), logistic regression model, Evans dataset.*

1. INTRODUCTION

Cardiovascular diseases (CVDs) are caused by disorders of the heart and bood vessels. CVDs include *coronary heart disease* (CHD, heart attacks), *cerebrovascular disease* (stroke)*, raised blood pressure (hypertension), peripheral artery disease, rheumatic heart disease, congenital heart disease and heart failure.* CVDs are the leading cause of morbidity and mortality worldwide, with 80% of total deaths occurring in developing countries.[1] Based on the report by WHO, in 2017, more than half (54%) of the dealths around the world were caused 10 leading causes, and CVDs which led to 15 million dealths in 2015 constituted the largest group of fatal diseases.[2] CVDs kill millions of people annually and this value may be increased up to 24.8 million by 2020 if preventive measures are not taken.[3]

In Vietnam, CVDs was among the top 10 leading causes of death in 2006, 2007 and 2009.[4,5] Around 32% of deaths from non-communicable diseases in rural areas are caused by CVDs.[6,7] For just coronary heart disease (CHD), according to the WHO data published in 2017 CHD deaths in Vietnam reached 58452 or 11.58% of total deaths. Evidently, the burden of CHD will continue to rise unless effective interventions for addressing its underlying risk factors are put in place.

Early detection of complications helps to treat CHD patients in a comprehensive way. Therefore, medical communities attempt to find a way for the accurate and timely prediction of CHD by using new statistical techniques, such as data mining techniques.[8] These techniques can help to recognize the patterns and factors influencing diseases.

The novel science of data mining is among the 10 developing sciences which have made

Corresponding author.
Email: lethanhbinh@qnu.edu.vn

the next decade face enormous technological evolutions. Using specialized knowledge, it will have extensive applications in the domain of medicine.[9,10] Predictive modeling of the risk of CVDs can render valuable information for planning of health care interventions. Over the last decade, several models have been developed and validated. The first well-known was developed in the United States using data from the Framingham study. This model, however, might be less reliable in other populations and has overestimated or underestimated the CVDs risks in specific settings.[11,12] Accordingly, other models were developed, such as the Systematic Coronary Risk Evaluation model in Europe[13]; a risk model based on QRESEARCH database in the United Kingdom[14]; the Prospective Cardiovascular Munster study in Germany[15] and a risk model based on database of CUORE Cohorts Project in Italy.[16]

In Asia, models to predict CVDs risk have been developed in Thailand, China, Japan, Malaysia, and Singapore. For all Asian populations, a tool was developed on the basis of six cohorts in this region (Asian model).[11,17-22] In Vietnam, with its own environment of biological, behavioral, and social characteristics, there is not yet a specific model to predict CVDs. The largest survey on risk factors for CVDs done in Vietnam applied the Framingham model only.[23]

The literature review showed that different algorithms such as clustering, logistic regression, decision trees, Bayesian network, neural network, scaled conjugate gradient (SCG) and support vector machine (SVM) have been used for predicting CVDs.[24-28] Among these algorithms, logistic regression has some advantages, such as high speed, simplicity. Logistic regression belongs to a family, named *Generalized Linear Model* (GLM), developed for extending the linear regression model to other situations. It is a widely used technique because it is very efficient and does not require too many computational resources.[29] Logistic regression model output is very easy to interpret

compared to other classification methods. In addition, because of its simplicity it is less prone to overfitting than flexible methods such as decision trees. For this reason, we would like to use logistic regression in the present article in order to create a prediction model of coronary heart disease based on the Evans heart disease dataset available in R (*version 3.6.1*).

2. METHODS AND DATA

2.1. Logistic regression model

The logistic regression model is a type of predictive model that can be used when the response variable is binary, meaning that there are only two possible outcomes such as *live/ die, disease/no disease, purchase/no purchase, and win/lose.*[29,30] In short, we want to model the probability of getting a certain outcome, in effect modeling the mean of the variable (which is the same as the probability in the case of binary variables). A logistic regression model can be applied to response variables with more than two categories; however, those cases are less common.

As the responses are not on a continuous measure and as such is not continuous, the use of logistic regression differs somewhat from the well-known linear regression, because, while in both cases we are modeling the mean, the mean in linear regression lies anywhere between $(-\infty, +\infty)$ whereas the mean (or the probability) in logistic regression lies between $[0, 1]$. Thus, we are predicting the probability that Y is equal to 1 (rather than 0) given certain cases of the predictors $X_1,..., X_p$.[30] It is important to make the distinction between these and linear regression models so we can think about how the observed data may be 0 or 1, but the predicted value may lie between $[0, 1]$. For example, we might try to predict the probability of whether a patient will live or die based on the patient's age as well as the number of years of experience his or her operating physician has.

The general form of the the logistic regression model is[29,30]

$$\log\left(\frac{p_1}{1-p_1}\right) = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n.$$

where p_1 is the probability that $Y = 1$ (the event), given $X_1,\ldots, X_n$ are the predictors (covariates), and β_i, $i = 1, 2, \ldots, n$ are known as the regression coefficients, which have to be estimated from the data. One can see that logistic regression model forms a linear combination of the explanatory variables to impact the logit, which is *log{probability of event/ probability of nonevent}*.

2.1.1. Probability

Let probability p_1 denote success and $(1 - p_1)$ denote failure with the results constrained to lie between 0 and 1. On the probability scale, we define[29,30]

$$p_1 = \frac{exp[\beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n]}{1 + exp[\beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n]}$$

The constraints of $0 \leq p_1 \leq 1$ make it impossible to construct a linear equation for predicting probabilities.

2.1.2. Odds

Odd is the ratio of the probability of an event to the probability of a nonevent. For example, flipping a coin and getting a head as an event versus getting tail as the nonevent. On the odds scale, we define[29, 30]

$$odds = \frac{p_1}{1 - p_1}$$
$$= exp[\beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n].$$

They are constrained by $0 \leq \frac{p_1}{1 - p_1} < \infty$, with 1 as the point for which both outcomes are equally likely.

2.1.3. Logits

Logit is the natural logarithm of the odds. On the logit scale, we define[29,30]

$$logit(p_1) = log\ odds = \log\left(\frac{p_1}{1 - p_1}\right)$$
$$= \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n.$$

On this scale, we have linearity. The logits are symmetric. They lie in the range $-\infty$ to $+\infty$.

The value that is equally likely for both outcomes is 0. If the identification of the two outcomes are switched, the log odds are multiplied by -1, since $log(a/b) = - log(b/a)$. The log odds of an event relays equally the same message as the probability of the event, so if a certain predictor has a positive impact on the logit then it has the same directional effect on the odds. When the log odds take on any value between $-\infty$ and $+\infty$, the coefficients form a logistic regression equation that can be interpreted in the usual way, meaning that they represent the change in log odds of the response per unit change in the predictor.[30]

2.2. Description of dataset

Data on coronary heart disease risk factors in Vietnam are currently limited, whereas the default installation of R comes with several relevant built-in datasets. For this reason, to illustrate the application of logistic regression analysis in the current study, we'll be working on the *Evans dataset* in the *lbreg package*. The data are from a cohort study in which 609 while males were followed for 7 years, with coronary heart disease as the outcome of interest. The variables are defined as in Table 1.[29]

Table 1. Description of Evans dataset.

CHD	A dichotomous outcome variable indicating the presence (coded 1) or absence (coded 0) of coronary heart disesase.
CAT	A dichotomous predictor variable indicating high (coded 1) or normal (coded 0) catecholamine level.
AGE	A continuous predictor variable for age.
CHL	A continuous predictor variable for cholesterol, mg/dl.
SMK	A dichotomous predictor variable indicating whether the subject has ever smoked (coded 1) or never smoked (coded 0)
ECG	A dichotomous predictor variable indicating the presence (coded 1) or absence (coded 0) of electrocardiogram abnormality
DBP	A continuous variable for diastolic blood pressure, mmHg

SBP	A continuous variable for systolic boold pressure, mmHg
HPT	A dichotomous variable indicating the presence (coded 1) or absence (coded 0) of high blood pressure. HPT is coded 1 if the diastolic blood pressure is greater than or equal to 160 or the systolic boold pressure is greater than or equal to 95.

Furthermore, the readers should notice that there is a slight difference about the notations. Namely, the outcome variable (indicating coronary heart disease) in the Evans dataset in R is denoted by "CDH" instead of "CHD" as described before. Accordingly, when executing the R codes which will appear in this article, we have to use "CDH" for the outcome variable.

3. RESULTS AND DISCUSSION

3.1. Data preparation

When working with a real dataset we need to take into account the fact that some data might be missing or corrupted, therefore we need to prepare the dataset for our analysis. As a first step we load the data.

```
# Load the required R package
library(lbreg)
# Load the dataset
```

```
data(Evans)
# Give the dimensions of the dataset
dim(Evans)
# Give the names of the variables
names(Evans)
```

Then we get the following:

```
[1]   609   9

[1] "CDH" "CAT" "AGE" "CHL" "SMK"
"ECG" "DBP" "SBP" "HPT"
```

We introduce the *is.na( )*[31,32] function as a tool for finding missing values. By *is.na(Evans),* we can verify that the Evans dataset has no missing values. Moreover, a visual take on the missing values might be helpful: the *Amelia package* has a special plotting function *missmap()*[31,32] that will plot dataset and highlight missing values. With the following R code,

```
# Load the required R package
library(Amelia)
# Draw a map of the missingness in the dataset
missmap(Evans, main="", col=c(4,6))
```

we have the output as Figure 1. The *"col"* parameter is to choose the colors we want.
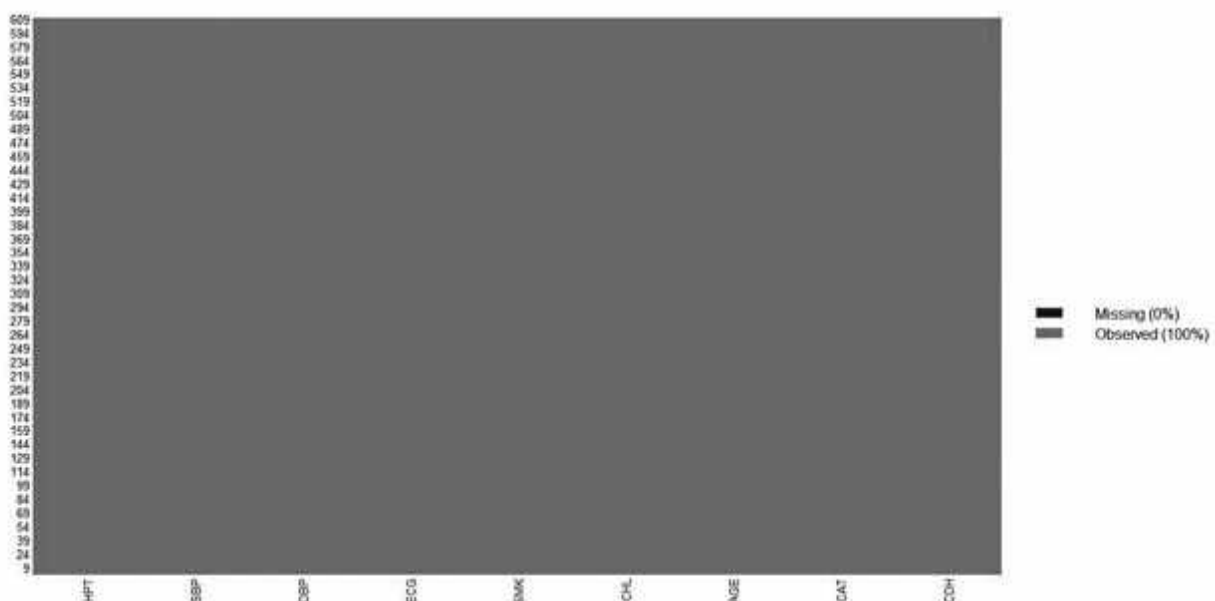

Figure 1. Missing values and observed values.

3.2. Model fitting

R programing language makes it very easy to fit a logistic regression model with the *glm( )* function and the fitting process is not so different from the one used in linear regression. In this section, we would like to fit a binary multiple logistic regression model corresponding to the Evans dataset and explain each step. We'll now split the data into two parts: *training set* for building a predictive model and *testing set* for evaluating the model.

```
# Split the data into two parts
train.data <- Evans[1:500,]
test.data <- Evans[501:609,]
```

The function *glm()*[31,32], for generalized linear model, is used to compute logistic regression. We need to specify the option *"family = binomial"*, which tells to R that we want to fit logistic regression.

The following R code is to build a model to predict the probability of being coronary heart disease-positive (CHD-positive) based on all the predictor variables available in the dataset. This is done using "~.". Using the *summary()*[31,32] function, we obtain the results of our model as described in Figure 2.

```
# Fit the model
model <- glm(CDH ~ ., data = train.data,
family = binomial)
summary(model) # Summarize the model
```

```
Call:
glm(formula = CDH ~ ., family = binomial, data = train.data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1270  -0.5229  -0.4076  -0.3011   2.6619

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -7.167040   1.832461  -3.911 9.19e-05 ***
CAT          0.651742   0.422820   1.541   0.1232
AGE          0.039513   0.017821   2.217   0.0266 *
CHL          0.009155   0.003634   2.519   0.0118 *
SMK          0.749937   0.331131   2.265   0.0235 *
ECG          0.360553   0.323574   1.114   0.2652
DBP          0.024388   0.017117   1.425   0.1542
SBP         -0.015100   0.009362  -1.613   0.1068
HPT          0.412969   0.442455   0.933   0.3506
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 358.88  on 499  degrees of freedom
Residual deviance: 329.12  on 491  degrees of freedom
AIC: 347.12

Number of Fisher Scoring iterations: 5
```

Figure 2. The results of the logistic regression.

From the output above, the coefficients table shows the beta coefficient estimates and their significance levels. Columns are:[31,32]

• *Estimate*: the intercept (β_0) and the beta coefficient estimates associated to each predictor variable.

• *Std. Error*: the standard error of the coefficient estimates. This presents the accuracy of the coefficients. The larger the standard error, the less confident we are about the estimate.

• *z value*: the z-statistic, which is the coefficient estimate (column 2) divided by the standard error of the estimate (column 3).

• *Pr(> |z |):* The *p-value* corresponding to the z-statistic. The smaller the p-value, the more significant the estimate is.

3.3. Interpretation

Now we can analyze the fitting and interpret what the model is telling us. First of all, it can be seen that only 3 out of 8 predictors are significantly associated to the outcome. These includes: AGE, CHL and SMK. As for the statistically significant predictors, CHL has the lowest p-value suggesting a strong association of the cholesterol concentration of the individuals with the probability of being coronary heart disease-positive (CHD-positive).

The coefficient estimate of the variable AGE is 0.039513, which is positive. This means that an increase in AGE is associated with increase in the probability of CHD-positive. Similarly for the variables CHL, SMK. We also see that the coefficient estimate of the variable SBP is negative (note that SBP is not statistically significant), therefore an increase in SBP will be associated with a decreased probability of being CHD-positive.

On the other hand, recall that in the logit model the response variable is given as

$$log\ odds = \beta_0 + \beta_1X_1 + ... + \beta_nX_n$$

The coefficients for CHL and SMK are 0.009155 and 0.749937, respectively. Since SMK

is a dummy variable, being having ever smoked increases the log odds by 0.749937, while one unit increases in the cholesterol concentration increases the log odds by 0.009155. As for the predictor AGE, one unit increases in AGE will increase the log odds by 0.039513.

From the logistic regression results, one should notice that some variables (including CAT, ECG, DBP, SBP and HPT) are not statistically significant. Keeping them in the model may contribute to overfitting. Consequently, they should be eliminated to obtain an optimal model with a reduced set of variables, without compromising the model accuracy. Here, we select manually the most significant variables:

```
# Fit the reduced model with only three
predictor variables
model1 <- glm(CDH ~ AGE + CHL +
SMK, data = train.data, family = binomial)
# Summarize the reduced model
summary(model1)
```

The reduced model is named *model1* and the results of this model, by *summary(model1),* is given as Figure 3:

Figure 3. The results of the reduced model.

3.4. Making predictions

The reduced logistic model can be written as

$$p_1 = \exp(-6.9607 + 0.0472 \times AGE + 0.0086 \times CHL + 0.753 \times SMK)/[1 + \exp(-6.9607 + 0.0472 \times AGE + 0.0086 \times CHL + 0.753 \times SMK)],$$

where p_1 is *the probability of being CHD-positive.* Using this formula, we'll make predictions using the testing data in order to evaluate the

performance of our logistic regression model. The proceduce is twofold:

• Predict the class membership probabilities of observations based on predictor variables;

• Assign the observations to the class with highest probability score (i.e. above 0.5).

The function *predict()*[31,32] can be used for predicting the probability of being CHD-positive, given the values of the predictors. Using the option *type = "response"* to directly get the probabilities.

Predict the probabilities of being CHD-positive:

```
# For easy data manipulation and visualization
library(tidyverse)
#Predict the probability of being CHD-positive
probabilities <- model1 %>% predict(test.
data, type = "response")
# See some first probabilities
head(probabilities)
```

The output is as follows:

501	502	503
0.12714064	0.10954876	0.13142426
504	505	506
0.07268374	0.14418855	0.36544538

Which classes do these probabilities refer to? In our case, the output is the probability that the coronary heart disease test will be positive.

Predict the class of individual: The following R code categorizes individuals into two groups based on their predicted probabilities of being coronary heart disease-positive. Individuals with probability above 0.5 (random guessing) are considered as CHD-positive.

```
# Predict the class of individual
predicted.classes <- ifelse(probabilities > 0.5,
"pos", "neg")
head(predicted.classes)
```

The output is as follows:

501	502	503	504	505	506
"neg"	"neg"	"neg"	"neg"	"neg"	"neg"

3.5. Assessing model accuracy

The model accuracy is measured as *the proportion of observations that have been correctly classified.*[29,30] Inversely, the classification error is defined as the proportion of observations that have been misclassified. We now use the following R code to compute the proportion of correctly classified observations:[30,31]

```
# Classify the individuals in the testing set.
test.data_CDH <- if.else(test.data$CDH ==
1, "pos", "neg")
# Compute and display the model accuracy
accuracy <- mean(predicted.classes ==
test.data_CDH)
print(paste('Accuracy:', accuracy))
```

Then the output is as follows:

[1] "Accuracy: 0.880733944954128"

From above, the *0.88 accuracy* on the testing set is quite a good result. However, you should keep in mind that this result is somewhat dependent on the manual split of the data that we made earlier.

In conclusion, with the ROCR package[30,31], we are going to plot the ROC curve and calculate the AUC (area under the ROC curve) which are typical performance measurements for a binary classifier. The ROC[29,30] is a curve generated by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings while the AUC[29,30] is the area under ROC curve. As a rule of thumb, a model with good predictive ability should have an AUC closer to 1 than to 0.5.

```
# Load the ROCR package
library(ROCR)
#Predict the probability of being CHD-positive
probabilities <- model1 %>% predict(test.
data, type = "response")

pr <- prediction(probabilities, test.data$CDH)
perf <- performance(pr, measure = "tpr",
x.measure = "fpr")

# Plot the ROC curve
plot(perf, xlab = "False positive rate", ylab
= "True positive rate", col = "blue")

# Calculate and display the AUC
auc <- performance(pr, measure = "auc")
auc <- auc@y.values[[1]]
print(paste('AUC:', auc))
```

When excuting the above R code, we will get the ROC plot as shown in Figure 4 and the AUC:
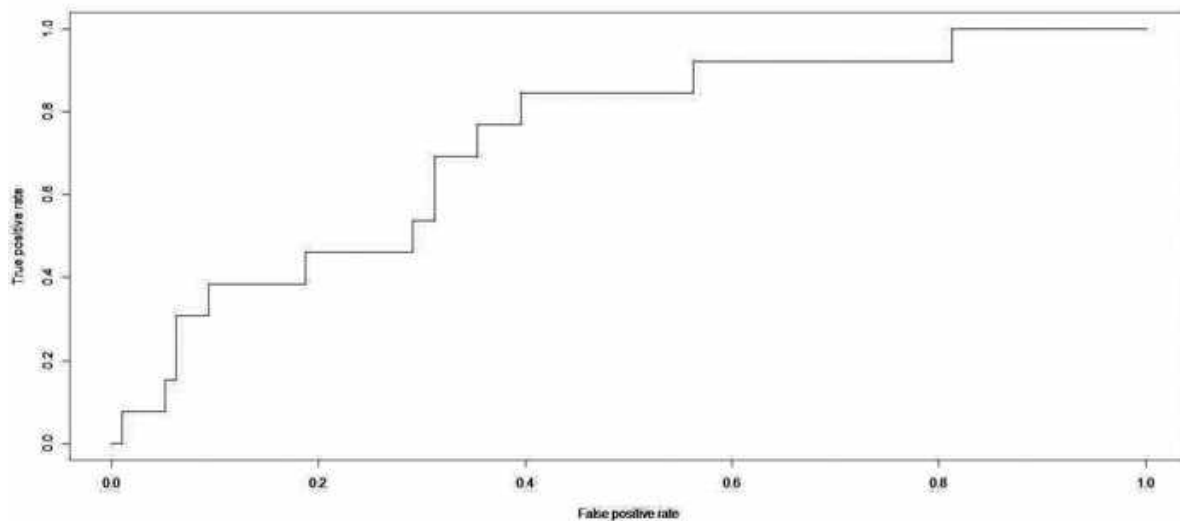
[1] "AUC: 0.729967948717949"

Figure 4. The ROC plot.

4. CONCLUSIONS

The process of disease prediction in medical sciences is as a very important process for decision-making and physicians need to know the risk factors for different diseases. This process can be facilitated by using statistical methods and data mining algorithms, such as logistic regression. In this study, we have described how logistic regression works and provided R codes to compute logistic regression. Moreover, we demonstrated how to make predictions and to assess the model accuracy. The results of the current study showed that the use of logistic regression can be very useful in predicting coronary heart disease and a simple coronary heart disease prediction model was proposed, having the 0.88 accuracy on the testing set, which allows physicians to predict CHD risk in patients. As described before, several CVDs (including CHD) prediction models are available around the world in general and in Vietnam in particular. These models vary in different aspects such as the time horizon used, characteristics of study population included, predictor variables, and outcome, they may all produce different results. Nevertheless, due to its advantages, logistic regression is still a powerful tool, especially in epidemiologic studies, allowing multiple predictor variables being analyzed simultaneously, meanwhile reducing the effect of confounding factors. With the gradual rise in CHD events among Vietnamese, the prevention and treatment of CHD risk factors are important public health concerns. Vietnam should of course ideally have its own cohort to feed models to predict CHD risk and validate models for the national setting. Even building an own Vietnamese model could be considered, may be an adaptation of one of the existing models. Because of the long time needed for that, and because of resource constraints, the suggestion that for Vietnamese populations, our logistic regression model can provide valid and reliable results should however be investigated in further analyses using real-life data for potential confirmation. Also, the current results of the study may be useful to health planners and will provide a basis for a further study on cost-effectiveness modeling of CHD management. In conclusion, we have to emphasize that further research is needed to compare the performance of different models, thereby identifying which models are best for Vietnamese populations.

REFERENCES

1. WHO. 2002. The World Health Report: Reducing Risks, Promoting Healthy Life, <https://www.who.int/whr/2002/en/>, accessed 05/02/2020.

2. WHO. 2017. The top ten causes of death, < https://www.who.int/news-room/fact-sheets/detail-/the-top-10-causes-of-death>, accessed 05/02/2020.

3. WHO. 2018. Physical activity and older adults, <https://www.who.int/dietphysicalactivity/factsheet_olderadults/en/>, accessed 07/02/2020.

4. N. P. Hoa, C. Rao, D. Hoy, N. D. Hinh, N. T. K. Chuc, N. D. Anh. Mortality measures from sample-based surveillance: evidence of the epidemiological transition in Vietnam, *Bulletin of the World Health Organization,* **2012**, *90*(10), 764-772.

5. Ngo D. A, C. Rao, N. P. Hoa, T. Adair, N. T. K. Chuc. Mortality patterns in Vietnam, 2006: findings from a national verbal autopsy survey, *BMC Research Notes,* **2010**, *3*(1).

6. H. V. Minh, D. L. Huong, S. Wall, N. T. K. Chuc. Cardiovascular disease mortality and its association with socioeconomic status: findings from a populationbased cohort study in rural Vietnam, 1999 - 2003, *Preventing Chronic Disease,* **2006**, *3*(3), 1-11.

7. D. L. Huong, H. V. Minh, U. Janlert, D. D. Van. Socio-economic status inequality and major causes of death in adults: a 5-year follow-up study in rural Vietnam, *Public Health,* **2006**, *120*(6), 497-504.

8. P. Rezaei, M. Ahmadi, S. Alizadeh, F. Sadoughi. Use of data mining techniques to determine and predict length of stay of cardiac patients, *Healthc. Inform. Res.,* **2013**, *19*(2), 121-129.

9. R. Bellazzi, F. Ferrazzi, L. Sacchi. Predictive data mining in clinical medicine: A focus on selected methods and applications, *Wiley interdisciplinary reviews: WIREs data mining and knowledge discovery*, **2011**, *1*(5), 416-430.

10. F. Amato, A. López, E. Pena-Méndez, P. Vanhara, A. Hampl, J. Havel. Artificial neural networks in medical diagnosis, *J. Appl. Biomed.,* **2013**, *11*, 47-58.

11. J. Liu, et al. Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the Chinese Multi-Provincial Cohort Study, *JAMA*, **2004**, *291*, 2591-9.

12. S. Kanjilal, *et al.* Application of cardiovascular disease risk prediction models and the relevance of novel biomarkers to risk stratification in Asian Indians, *Vasc. Health. Risk. Manag.,* **2008**, *4*, 199-211.

13. RM. Conroy, *et al.* Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project, *Eur. Heart. J.,* **2003**, *24*, 987-1003.

14. J. Hippisley-Cox, et al. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study, *BMJ*, **2007**, 335-136.

15. G. Assmann, P. Cullen, H. Schulte. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the Prospective Cardiovascular Munster (PROCAM) study. *Circulation*, **2002**, *105*, 310-5.

16. M. Ferrario, et al. Prediction of coronary events in a low incidence population: assessing accuracy of the CUORE Cohort Study prediction equation, *Int. J. Epidemiol.,* **2005**, *34*, 413-21.

17. P. Sritara, *et al.* Twelve-year changes in vascular risk factors and their associations with mortality in a cohort of 3499 Thais: the Electricity Generating Authority of Thailand Study, *Int. J. Epidemiol.*, **2003**, *32*(3), 461-8.

18. Y. Wu, *et al.* Estimation of 10-year risk of fatal and nonfatal ischemic cardiovascular diseases in Chinese adults, *Circulation*, **2006**, *114*(21), 2217-25.

19. I. Saito, *et al.* A low level of C-reactive protein in Japanese adults and its association with cardiovascular risk factors: the Japan NCVC-Collaborative Inflammation Cohort (JNIC)

study, *Atherosclerosis*, **2007**, *194*(1), 238-44.

20. H. Arima, et al. Development and validation of a cardiovascular risk prediction model for Japanese: the Hisayama study, *Hypertens. Res.,* **2009**, *32*(12), 1119-22.

21. C. Purwanto Eswaran, *et al.* Prediction models for early risk detection of cardiovascular event, *J. Med. Syst.,* **2012**, *36*(2), 521-31.

22. I. Baik, *et al.* Dietary information improves cardiovascular disease risk prediction models, *Eur. J. Clin. Nutr.,* **2013**, *67*(1), 25-30.

23. N. Q. Ngoc, *et al.* Cardiovascular disease risk factor patterns and their implications for intervention strategies in Vietnam, *Int. J. Hypertens.,* **2012**.

24. S. Desai, S. Giraddi, P. Narayankar, S. Sulegaon, N. Pudakalakatti. Back-propagation neural network versus logistic regression in heart disease classification, *J. Adv. Comput. Commun. Technol.,* **2019**.

25. N. Kausar, *et al.* Ensemble clustering algorithm with supervised classification of clinical data for early diagnosis of coronary artery disease, *J. Med. Imaging Health Inform.,* **2016**, *6*(1),78-87.

26. L. Guner, N. Karabacak, O. Akdemir, P. Karagoz, S. Kocaman, A. Cengel, M. Unlu. An open-source framework of neural networks for diagnosis of coronary artery disease from myocardial perfusion SPECT, *J. Nucl. Cardiol.,* **2010**, *17*(3), 405-13.

27. K. Orphanou, A. Stassopoulou, E. Keravnou. DBN-extended: a dynamic Bayesian network model extended with temporal abstractions for coronary heartdisease prognosis, *IEEE J. Biomed. Health Inform.,* **2016**, *20*(3), 944-52.

28. J. Kim, J. Lee, Y. Lee. Data-mining-based coronary heart disease risk prediction model using fuzzy logic and decision tree, *Healthc. Inform. Res.,* **2015**, *21*(3), 167-74.

29. D. Kleinbaum, M. Klein. *Logistic Regression: A Self-Learning Text*, Springer, 2010.

30. J. Wilson, K. Lorenz. *Modeling Binary Correlated Responses using SAS, SPSS and R,* Springer, 2015.

31. J. Ledolter. *Data Mining and Bussiness analytics with R,* John Wiley & Sons, 2013.

32. Y. Li, J. Baron. *Behavioral Research Data Analysis with R,* Springer, 2012.

Tính chính quy Milyutin của ánh xạ đa trị

Nguyễn Hữu Trọn[1,*], Đào Ngọc Hân[2]

[1]*Khoa Toán và Thống kê, Trường Đại học Quy Nhơn, Việt Nam.*
[2]*Khoa Giáo dục Tiểu học và Mầm non, Trường Đại học Quy Nhơn, Việt Nam.*

TÓM TẮT

Mục đích của bài báo là nghiên cứu những đặc trưng cho tính chính quy Milyutin của ánh xạ đa trị thông qua độ dốc địa phương và độ dốc toàn cục của hàm bao nửa liên tục dưới của hàm khoảng cách liên kết với ánh xạ đa trị đã cho. Bằng cách sử dụng các đặc trưng này, chúng tôi thu được tính ổn định nhiễu của chính quy Milyutin.

Từ khóa: *Tính chính quy mêtric, tính chính quy Milyutin, tính ổn định nhiễu, độ dốc.*

---

*Tác giả liên hệ chính.
Email: nguyenhuutron@qnu.edu.vn*

On the Milyutin regularity of set-valued mappings

Nguyen Huu Tron[1,*], Dao Ngoc Han[2]

[1]*Faculty of Mathematics and Statistics, Quy Nhon University, Vietnam.*
[2]*Faculty of Preschool and Primary Education, Quy Nhon University, Vietnam.*

ABSTRACT

The aim of this paper is to study characterizations of Milyutin regularity of a set-valued mapping via the local and non-local slope of the lower semicontinuous envelope of the distance function associated with this set-valued mapping. By using of these characterizations, we get the stability under perturbation of the Milyutin regularity.

Keywords: *Metric regularity, Milyutin regularity, perturbation stability, slope.*

1. INTRODUCTION

The emergence of metric regularity is increasingly clear during last decades and considered one of the important concepts of variational analysis by its extensively applications in a large amount of mathematical areas. This property is studied by experts which obtained valuable results such as implicit and inverse function theorem and stability under small variations,.. It is also the basis for qualification conditions in various calculus rules and optimally criteria, etc. The reader is referred to many theoretical results on the metric regularity as well as its applications in works [1-21], and the references given therein.

It is also known that metric regularity is one of powerful tools to examine the solution existence of equations. For equations of the form $f(x) = y$, where $f : X \to Y$ is a single-valued function from a metric space X to metric space Y, the condition ensuring the existence of solutions is the surjectivity of f. As in nonlinear analysis, regularity of a strictly differentiable mapping at some point $\bar{x}$ is equivalent to its derivative at the point is onto.

However, variational analysis, especially optimization theory, appears the objects may lack of smoothness: non-differentiable functions at point of interest, set-valued mappings, etc. Thus, the condition on the surjectivity of the derivative mapping at the point may be failed. One way to overcome this problem is to give an upper estimation for the distance from a point x near a given solution $\bar{x}$ of the generalized equation $y \in F(x)$ to the solution set $F^{-1}(y)$ in terms of the residual $d(y, F(x))$. In applications, the residual is able to calculate or estimate easily, meanwhile the finding the exact solution set might be considerably more complicated. The map $F : X \rightrightarrows Y$ satisfying the above estimation is said to be local metric regularity around $(\bar{x}, \bar{y})$, it means that there exist some positive

Corresponding author.
Email: Nguyenhuutron@qnu.edu.vn

numbers τ, δ, ρ such that

$$d(x, F^{-1}(y)) \leq \tau d(y, F(x))$$

for all $x \in B(\bar{x}, \delta)$ and $y \in B(\bar{y}, \rho)$.

Due to the crucial role of metric regularity theory in the areas of applied mathematics such as optimization, fixed point theory, convergence analysis of algorithms, economics, equilibrium, control theory, so on, many authors extended this property to non-local version. It found that the non-local regularity can be started from well-known Banach's contraction map principle. Extension of this principle on closed ball in a complete metric space established a connection between non-local regularity and fixed point of maps. This was first observed by Arutynov [22], Ioffe [7,15] and some years before Dmitruk, Milyutin, Osmolovskii [3] in connecting to the extremal problems. The reader is referred to the works [10,16,19,23−25] for these developments. Because of many applications in such practical problems, in recent papers [7,15], Ioffe presented a complete model of non-local regularity and its important applications. One of the most important properties in this type is Milyutin regularity. This type of regularity is associated with a regularity horizon function that is convenient to establish the criterion of regularity. In case of Milyutin regularity, there is almost no gap between necessary and sufficient conditons in regularity criteria, but with any regularity horizon function, this gap appears. The fact is that if the considered set is an open set, then the regularity horizon function is positive on it. Thus, dealing with Milyutin regularity, we do not need to be interseted in points outside the set. That is reason why the Milyutin regularity becomes important in non-local context.

In this paper, we will characterize Milyutin regularity via non-local and local slope as well as its applications in establishing the stability of the Milyutin regularity under pertubation.

The organization of paper is as follows. In Section 2, we give some useful notations and definitions such as openness, pseudo-Lipschitz, metric regularity, Milyutin metric regularity and their equivalence. We establish the non-local and local slope characterizations for the Milyutin regularity on fixed sets in Section 3. Section 4 is dedicated for the stability of the Milyutin regularity under suitable pertubation.

2. PRELIMINARIES

In this section, we present some essential definitions and properties that will be used throughout this paper. Let X and Y be metric spaces endowed with metrics both denoted by $d(\cdot, \cdot)$. Let $F : X \rightrightarrows Y$ is a set-valued mapping. We use the notations $\text{gph} F := \{(x, y) \in X \times Y \; : \; y \in F(x)\}$ for the graph of F, $\text{dom} F := \{x \in X \; : \; F(x) \neq \emptyset\}$ for the domain of F and $F^{-1} : Y \rightrightarrows X$ for the inverse of F. This inverse is defined by $F^{-1}(y) := \{x \in X \; : \; y \in F(x)\}$, $y \in Y$ and satisfies

$$(x, y) \in \text{gph} F \iff (y, x) \in \text{gph} F^{-1}.$$

Given $\bar{x} \in X$, $r > 0$, we denote by $B(\bar{x}, r), \overline{B}(\bar{x}, r)$, the open and closed balls with center $\bar{x}$ and radius $r > 0$, respectively.

In recent works [7,8], Ioffe studied a nonlinear non-local regularity model of set-valued mapping on a box $U \times V$ of $X \times Y$. In this paper, we suggest a new version of this property which is slightly different from the mentionned one but on a set of $W \subset X \times Y$. Let W be a subset of $X \times Y$ and a function $\gamma : X \rightarrow \mathbb{R}_+$ which is positive on $P_X W$ and a function $\mu : \mathbb{R}_+ \rightarrow \mathbb{R}_+$.

Definition 2.1. Let X, Y be metric spaces, W be a subset of $X \times Y$ and let $F : X \rightrightarrows Y$ be a set-valued mapping. F is said to be (μ, γ)-*metrically regular* on W if there exists a positive real numbers r, κ such that

$$d(x, F^{-1}(y)) \leq \kappa \mu(d(y, F(x))) \qquad (1)$$

for all $(x, y) \in W$ with $0 < r\mu(d(y, F(x))) < \gamma(x)$, where $\gamma(x) > 0$, for all $x \in P_X W$.

Next, we introduce equivalent versions of the regularity such as (γ, μ)-Hölder property and (γ, μ)-openness of set-valued mappings.

Definition 2.2. Let X, Y be metric spaces, W be a subset of $X \times Y$ and let $F : X \rightrightarrows Y$ be a set-valued mapping. F is (γ, μ)-Hölder on W if there are $r, \kappa > 0$ such that

$$d(y, F(x)) \leq \kappa\mu(d(x, u))$$

holds for all $(x, y) \in W, y \in F(u)$ and $0 < r\mu(d(x, u)) < \gamma(x)$.

Definition 2.3. Let X, Y be metric spaces, W be a subset of $X \times Y$ and let $F : X \rightrightarrows Y$ be a set-valued mapping. F is (γ, μ)-open on W if there are $r, \kappa > 0$ such that the conclusion

$$B(F(x), \mu(rt)) \cap W_x \subset F(B(x, \kappa rt))$$

fullfils for all $x \in P_X W$ and $0 < t < \gamma(x)$.

In the case W is a box $U \times V$, one gets simpler versions of the regularity above as the defintions below.

Definition 2.4. F is said to be γ-metrically regular on (U, V) if there is a $\tau, \kappa > 0$ such that

$$d(x, F^{-1}(y)) \leq \tau d(y, F(x))$$

provided $x \in U, y \in V$ and $\kappa d(y, F(x)) < \gamma(x)$. In the case of $r \equiv \kappa$ we denote by $\text{reg}_\gamma F(U|V)$ the lowest bound of such τ. If no such τ exists, set $\text{reg}_\gamma F(U|V) = \infty$. We shall call $\text{reg}_\gamma F(U|V)$ the modulus (of rate) of γ-metric regularity of F on (U, V).

Definition 2.5. F is said to have δ-pseudo-Lipschitz property on (U, V) if there is a λ such that

$$d(y, F(x)) \leq \lambda d(x, u)$$

if $x \in U, y \in V, \lambda d(x, u) < \delta(y)$ and $y \in F(u)$. Denote by $\text{lip}_\delta F(U|V)$ the lower bound of such λ. If no such λ exists, set $\text{lip}_\delta F(U|V) = \infty$. We shall call $\text{lip}_\delta F$ the δ-Lipschitz modulus of F such on (U, V).

Definition 2.6. F is said to be γ-open (or γ-covering) at a linear rate on (U, V) if there is a $r > 0$ such that

$$B(F(x), rt) \cap V \subset F(B(x, t)),$$

if $x \in U$ and $t < \gamma(x)$. Denote by $\text{sur}_\gamma F(U|V)$ the upper bound of such r. If no such r exists, set $\text{sur}_\gamma F(U|V) = 0$. We shall call $\text{sur}_\gamma F$ the modulus (or rate) of γ-openness of F on (U, V).

Proposition 2.7. *(see [15]) Let $F : X \rightrightarrows Y$ be a set-valued mapping defined between metric spaces X and Y, U, V be subsets of X and Y respectively, and $\gamma : X \to \overline{\mathbb{R}}$ be a extended-valued function which is positive on U. Then, the following three properties are equivalent*

(i) F is γ-open at a linear rate on (U, V);

(ii) F is γ-metrically regular on (U, V);

(iii) F^{-1} has γ-pseudo-Lipschitz property on (V, U).

Moreover, under the convention that $0 \cdot \infty = 1$, one has

$$sur_\gamma F(U|V) \cdot reg_\gamma F(U|V) = 1,$$

$$reg_\gamma F(U|V) = lip_\gamma F^{-1}(V|U).$$

By choosing the gauge function $\gamma(x)$ is $m(x) = d(x, X \setminus U)$ with U be an open subset of X, we get the Milyutin regularity of F on (U, V).

Definition 2.8. F is said to be Milyutin metrically regular on (U, V) if there is a $\tau, \kappa > 0$ such that

$$d(x, F^{-1}(y)) \leq \tau d(y, F(x))$$

provided $x \in U, y \in V$ and $\kappa d(y, F(x)) < m(x)$.

A special case of this property is $U := B(\bar{x}, r), m(x) := r - d(x, \bar{x})$. When $V = Y$, we say that F is Milyutin regular on U. Then, we write $\text{reg}_m F(U)$ rather $\text{reg}_m F(U|Y)$. And, we say that F is globally regular if it is regular on $\text{Dom} F \times Y$ with $\gamma \equiv \infty$.

Remark 2.9. Unlike the local regularity, the domains U, V are essential elements of the definitions above. In the non-local case, the regular domain cannot be freely changed.

Example 2.10. Let $X = Y = \mathbb{R}$, $F(x) = \{x, 3\}$, $U = (0, 1), V = (0, 2)$. Then, F is 1-regular on (U, V) with modulus 1 but F is not γ-regular on (U, V') with $V' = (0, 3)$ for any constant γ.

Indeed, one has that for $x \in U$, $y \in V$ such that $d(y, F(x)) < \gamma(x)$. So, $d(y, F(x)) < 1$. Thus, $d(y, F(x)) = |x - y|$. Therefore, $d(x, F^{-1}(y)) = |x - y| = d(y, F(x))$. It means that F is 1-metrically regular on $U \times V'$. However, F is not γ-regular on $U = (0, 1), V' = (0, 3)$ for any γ because for $x \in U$, there exists $t > 0$ such that $0 < x + t < 1$, then $(3 - t, 3) \subset B(F(x), t) \cap V'$ but $(3 - t, 3) \nsubseteq F(B(x, t)) = (x - t, x + t) \cup \{3\}$.

3. CHARACTERIZATIONS FOR MILYUTIN REGULARITY

Let X be a metric space and let $f : X \to \overline{\mathbb{R}}$ be a given function. As usual, $\text{Dom } f := \{x \in X : f(x) < +\infty\}$ denotes the domain of f. The symbol $[f(x)]_+$ stands for $\max(f(x), 0)$. Recall that the local slope $|\nabla f|(x)$ of a lower semicontinuous function f at $x \in \text{dom} f$ is the quantity defined by $|\nabla f|(x) = 0$ if x is a local minimum of f; otherwise

$$|\nabla f|(x) = \limsup_{y \to x, y \neq x} \frac{f(x) - f(y)}{d(x, y)}.$$

For $x \notin \text{dom} f$, we set $|\nabla f|(x) = +\infty$. The non-local slope of f is defined by

$$|\Gamma f|(x) := \sup_{y \neq x} \frac{[f(x) - f(y)]_+}{d(x, y)}.$$

For $x \notin \text{dom} f$, we set $|\Gamma f|(x) = +\infty$.

It is easy to see that if X be a normed space and f is Fréchet differentiable at x then $|\nabla f|(x) = \|f'(x)\|$ and $|\nabla f|(x) \leq |\Gamma f|(x)$ for all $x \in X$.

Example 3.1. Let $f : \mathbb{R} \to \mathbb{R}$ be given as

$$f(x) := \begin{cases} x^2, & \text{if } x \geq 0, \\ -x, & \text{if } x < 0. \end{cases}$$

Since f attains the (global) minimum at $x = 0$, $|\nabla f|(0) = |\Gamma|(0) = 0$. For $x \neq 0$, f is differentiable, so we have $|\nabla f|(x) = 2x$ if $x > 0$ and $|\nabla f|(x) = 1$ if $x < 0$.

We notice that if F is any set-valued mapping, the distance function $d(\cdot, F(\cdot))$ is not generally a lower semicontinuous function. However, the tools of variational analysis often require the considered function to be lower semicontinuous. Therefore, instead of using the distance function, we often use the lower semicontinuous envelope $(x, y) \to \varphi_y(x)$ (which is always lower semicontinous) of the function $(x, y) \to d(y, F(x))$ defined by, for $(x, y) \in X \times Y$,

$$\varphi_y(x) = \liminf_{u \to x} d(y, F(u)).$$

We need two lemmas in the sequel.

Lemma 3.2. *Let $F : X \rightrightarrows Y$ be a closed multifunction, i.e., its graph is a closed set in $X \times Y$. Then, for each $y \in Y$,*

$$F^{-1}(y) = \{x \in X : \varphi_y(x) = 0\}.$$

Proof. Indeed, if $x \in F^{-1}(y)$, then $0 \leq \varphi_y(x) \leq d(y, F(x)) = 0$, so $\varphi_y(x) = 0$. Conversely, suppose $\varphi_y(x) = 0$. There exists a

sequence $\{x_n\}_{n \in \mathbb{N}}$ converging to x such that $d(y, F(x_n))$ converges to 0. Then, we can find a sequence $\{v_n\}_{n \in \mathbb{N}}$ such that $v_n \in F(x_n)$ and $d(y, v_n) \to 0$. Since the graph of F is closed, then $(x, y) \in \mathrm{gph} F$, i.e. $x \in F^{-1}(y)$.

Lemma 3.3. *Let X, Y be metric spaces, W be a subset of $X \times Y$, let $F : X \rightrightarrows Y$ be a set-valued mapping. If there are positive reals r, κ such that*

$$d(x, F^{-1}(y)) \leq \kappa \varphi_y(x) \qquad (2)$$

for all $(x, y) \in W$ with $0 < r\varphi_y(x) < \gamma(x)$ then F is γ-metrically regular on W. Conversely, if F is γ-metrically regular on the open set W of $X \times Y$ then (2) holds.

Proof. Because of $\varphi_y(x) \leq d(y, F(x))$ for all $(x, y) \in W$, if (2) satisfies then F is γ-metrically regular on W. Conversely, suppose that F is γ-metrically regular on open subset W of $X \times Y$. Let now $(x, y) \in W$ such that $0 < r\varphi_y(x) < \gamma(x)$, there exists a sequence $x_n \in X$ converging to x such that $d(y, F(x_n)) \to \varphi_y(x)$ when n tends to ∞. Because of the openness of $P_X W$ and $x \in P_X W$, $x_n \in P_X W$ for sufficiently large n. Thus, due to the regularity of F, and the continuity of γ, one has for sufficiently large n, $0 < rd(y, F(x_n)) < \gamma(x_n)$. It follows that

$$d(x_n, F^{-1}(y)) \leq \kappa d(y, F(x_n)).$$

Let n tends to ∞ in this inequality, one gets

$$d(x, F^{-1}(y)) \leq \kappa \varphi_y(x).$$

The proof is complete.

The following result establishes the necessary and sufficient condition for Milyutin regularity through the nonlocal slope of the lower envelope of distance function $(x, y) \to d(y, F(x))$.

Theorem 3.4. *Let X be a complete metric space and Y be a metric space, $U \subset X, V \subset Y$ be open subsets of X and Y, respectively and* let $F : X \rightrightarrows Y$ be a closed set-valued mapping. Then, if any $x \in U$ and $y \in V$ with $0 < \tau \varphi_y(x) < m(x)$,

$$|\Gamma \varphi_y|(x) > \tau^{-1},$$

then F is Milyutin regular on $U \times V$ with modulus τ. Conversely, if F is Milyutin regular on $U \times V$ with modulus τ, for $x \in U$ and $y \in V$ with $0 < \tau \varphi_y(x) < m(x)$,

$$|\Gamma \varphi_y|(x) \geq \tau^{-1}.$$

Proof. For the sufficient condition, take $x \in U$ and $y \in V$ with $0 < \varphi_y(x) < \tau m(x)$ such that

$$|\Gamma \varphi_y|(x) > \tau^{-1}. \qquad (3)$$

We shall prove that F is Milyutin regular on $U \times V$ with modulus τ. Indeed, take $x \in U$ and $y \in V$ with $0 < \tau d(y, F(x)) < m(x)$. If follows that $0 < \varphi_y(x) < \tau^{-1} m(x)$. Fixing $y \in V$ and applying now the Ekeland's variational principle for the function $u \to f(u) := \varphi_y(u)$, one can find a point $z \in X$ satisfying the folowing conditions

(i) $d(z, x) \leq m(x)$;

(ii) $f(z) + \tau^{-1} d(x, z) \leq f(x)$;

(iii) $f(u) + \tau^{-1} d(u, z) > f(z), \forall u \neq z$.

We shall prove $f(z) = 0$. Suppose the contradiction that $f(z) > 0$. By (i), $z \in U$. By (ii),

$$f(z) \leq f(x) - \tau^{-1} d(x, z) <$$
$$< \tau^{-1} m(x) - \tau^{-1} d(x, z) \leq \tau^{-1} m(z).$$

By (iii),

$$|\Gamma f|(z) \leq \tau^{-1}.$$

This contradicts to (3). So, $f(z) = 0$. It implies that $y \in F(z)$ From (ii), one obtains that

$$\tau^{-1} d(x, z) \leq f(x).$$

Thus,

$$\tau^{-1} d(x, F^{-1}(y)) \leq \tau^{-1} d(x, z) \leq$$
$$\leq f(x) \leq d(y, F(x)).$$

Consequently,

$$d(x, F^{-1}(y)) \leq \tau d(y, F(x)).$$

One finishes the sufficient condition.

Conversely, suppose that F is Milyutin regular on $U \times V$ with modulus τ. Take now $x \in U, y \in V$ with $0 < \tau\varphi_y(x) < m(x)$. Subsequently, taking into account Lemma 3.3,

$$d(x, F^{-1}(y)) \leq \tau\varphi_y(x).$$

Then, for every $\varepsilon > 0$, there is $u \in F^{-1}(y)$ such that

$$d(x, u) < (\tau+\varepsilon)\varphi_y(x) = (\tau+\varepsilon)\varphi_y(x) - (\tau+\varepsilon)\varphi_y(u).$$

Note that $u \neq x$ and one gets

$$|\Gamma\varphi_y|(x) > (\tau + \varepsilon)^{-1}.$$

Let ε tend to 0, we obtain that

$$|\Gamma\varphi_y|(x) \geq \tau^{-1},$$

and we finish the necessary condition. The theorem is proved.

Using the local slope of the function mentioned above, one only gets the sufficient condition for the Milyutin regularity.

Theorem 3.5. *Let X be a complete metric space and Y be a metric space, $U \subset X, V \subset Y$ be open subsets of X and Y, respectively and let $F : X \rightrightarrows Y$ be a closed set-valued mapping. If for any $x \in U$ and $y \in V$ with $0 < \tau\varphi_y(x) < m(x)$,*

$$|\nabla\varphi_y|(x) > \tau^{-1}$$

then F is Milyutin regular on $U \times V$ with modulus τ.

Proof. The proof of this theorem is very simple with noting that $|\nabla f|(x) \leq |\Gamma f|(x)$ for all $x \in X$. Therefore, if $|\nabla\varphi_y|(x) > \tau^{-1}$ then $|\Gamma\varphi_y|(x) > \tau^{-1}$. So, by Theorem 3.4, one obtains the result.

However, the converse of this result is not true in general and the results above can be extended to more general case when one considers the regularity on an open subset W of $X \times Y$

Theorem 3.6. *Let X be a complete metric space and Y be a metric space, $W \subset X \times Y$ be an open subset of $X \times Y$ and let $F : X \rightrightarrows Y$ be a closed set-valued mapping. If for any $x \in P_X W$ and $y \in P_Y W$ with $0 < \tau\varphi_y(x) < m(x) = d(x, X \setminus P_X W)$,*

$$|\Gamma\varphi_y|(x) > \tau^{-1}$$

then, F is Milyutin regular on W with modulus τ. Conversely, if F is Milyutin regular on W with modulus τ for any $x \in P_X W$ and $y \in P_Y W$ with $0 < \tau\varphi_y(x) < m(x) = d(x, X \setminus P_X W)$,

$$|\Gamma\varphi_y|(x) \geq \tau^{-1}.$$

When F is Milyutin regular on W we get the following characterization.

Theorem 3.7. *Let X be a complete metric space and Y be a metric space, let $W \subset X \times Y$ be an open set. Let $F : X \rightrightarrows Y$ be a closed set-valued mapping. $m : P_X W \rightarrow \mathbb{R}_+$ is Lipschitz function on $P_X W$ with constant 1, $\mu : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a gauge function. Then F is Milyutin regular on W if and only if*

$$\liminf_{\delta \downarrow 0} \left\{ |\Gamma\varphi_y|(x) : \begin{array}{c} \frac{d(x, W_y)}{m(x)} < \delta, \\ y \in P_Y W, \\ 0 < \frac{\varphi_y(x)}{m(x)} < \delta \end{array} \right\} > 0.$$

4. MILYUTIN REGULARITY UNDER PERTURBATION

Using Theorem 3.7, we can establish the stability of the Milyutin regularity under suitable perturbation.

Theorem 4.1. *Let X be a Banach space and Y be a normed space, U, V be open sets in X*

and Y, respectively. Let $F : X \rightrightarrows Y$ be a closed set-valued mapping. If F is Milyutin regular on $U \times V$ with modulus τ, $g : X \to Y$ is Lipschitz on U with constant $\lambda > 0$ satisfying $\tau\lambda < 1$ then $F + g$ is Milyutin regular on W, where $W :=$ $\{(x, y) \in X \times Y : x \in U, B(y-g(x), m(x)) \subset V\}$.

Proof. For in detail, see Ngai, Tron, Han[26].

5. CONCLUSIONS

In this paper, we established the characterizations for the Milyutin regularity of closed set-valued mappings defined on the complete metric spaces throughout the slopes of the lower semicontinuous envelope assiciated to this map. Then, by using obtained results in previous sections, we give the stability of the Milyutin regularity under perturbation.

REFERENCES

1. D. Azé. A unified theory for metric regularity of multifunctions, *J. Convex Anal.*, **2006**, *13* (2), 225-252.

2. J. M. Borwein, D. M. Zhuang. Verifiable necessary and sufficient conditions for openness and regularity for set-valued and single-valued mapps, *J. Math. Anal. Appl.*, **1988**, *134*, 441–459.

3. A. V. Dmitruk, A. A. Milyutin, N. P. Osmolovskii. Lyusternik's theorem and the theory of extrema, *Russian Math. Surveys*, **1980**, *35* (6), 11–51.

4. A. L. Dontchev, R. T. Rockafellar. *Implicit Functions and Solution Mappings: A View from Variational Analysis*, 2rd edition, Springer Series in Operations Research and Financial Engineering, Springer New York, 2016.

5. M. Durea, R. Strugariu. Chain rules for linear openness in metric spaces. Applications to parametric variational systems, *Mathematical Programming*, **2014**, Series A, *143* (1), 147-176.

6. A. D. Ioffe. On regularity concepts in variational analysis, *J. Fixed Point Theory Appl*, **2010**, *8* (2), 339–363.

7. A. D. Ioffe. Regularity on fixed sets, *SIAM J. Optim.*, **2011**, *21* (4), 1345-1370.

8. A. D. Ioffe. Nonlinear regularity models, *Math. Program.*, **2013**, *139* (1), 223-242.

9. A. D. Ioffe. Metric regularity-a survey. Part II. Applications, *J. Aust. Math. Soc.*, **2016**, *101*, 376-417.

10. A. D. Ioffe. Metric regularity and subdifferential calculus, *Russian Mathematical Surveys*, **2000**, *55* (3), 501-558.

11. B. S. Mordukhovich. *Variational analysis and generalized differentiation. I. Basic theory*, Springer-Verlag, Berlin, 2006.

12. H. V. Ngai, N. H. Tron, M. Théra. Directional Holder metric regularity, *J. Optim. Theory Appl.*, **2016**, *171*, 785-819.

13. H. V. Ngai, N. H. Tron, M. Théra. Metric regularity of the sum of multifunctions and applications, *Journal of Optimization: Theory and Applications*, **2014**, *160* (2), 355–390.

14. J. -P. Penot. Metric regularity, openness and Lipschitzian behavior of multifunctions, *Nonlinear Anal.*, **1989**, *13* (6), 629–643.

15. A. D. Ioffe. *Variational Analysis of Regular Mappings: Theory and Applications*, Springer Monographs in Mathematics, Springer, 2017.

16. A. D. Ioffe. Towards variational analysis in metric spaces: Metric regularity and fixed points, *Math. Program.*, **2010**, *123*, 241-252.

17. M. Ivanov, N. Zlateva. On Characterizations of Metric Regularity of Multi-valued Maps, <https://arxiv.org/abs/1811.08243>, reprint, 2020.

18. A. Jourani, L. Thibault. Metric regularity and subdifferential calculus in Banach spaces, *Set Valued Anal.*, **1995**, *3* (1), 87–100.

19. P. Q. Khanh. An induction theorem and general open mapping theorem, *J. Math. Anal. Appl.*, **1986**, *118*, 519-534.

20. S. M. Robinson. Generalized equations and their solutions, part I: Basic theory, *Point-to-Set Maps and Mathematical Programming*, **1979**, *10*, 128–141.

21. S. M. Robinson. Generalized equations and their solutions, part II: Applications to nonlinear programming, *Optimality and Stability in Mathematical Programming*, **1982**, *19*, 200–221.

22. A. V. Arutyunov, E. R. Avakov, S. E. Zhukovskiy. Stability Theorems for Estimating the Distance to a Set of Coincidence Points, *SIAM J. Optim.*, **2015**, *25* (2), 807–828.

23. A. V. Arutyunov, E. R. Avakov, B. Gelman, A. Dmitruk, V. Obukhovski. Locally covering maps in metric spaces and coincidence points, *J. Fixed Point Theory Appl.*, **2009**, *5*, 106-127.

24. A. V. Arutyunov. Covering mappings in metric spaces and fixed points, *Dokl. Math.*, **2007**, *76*, 665-668.

25. A. L. Dontchev, H. Frankowska. Lusternik-Graves theorem and fixed points, *Proc. Amer. Math. Soc.*, **2011**, *139*, 521- 534.

26. H. V. Ngai, N. H. Tron, D. N. Han. Nonlinear regularity on fixed sets, **2020**, *In revision.*

Một bất đẳng thức về độ chính xác của lượng tử

Vương Trung Dũng[1,*], Võ Thị Bích Khuê[2]

[1] *Trường Phổ thông Năng khiếu - Đại học Quốc gia, Thành phố Hồ Chí Minh, Việt Nam*
[2] *Khoa Toán, Trường Đại học Tài chính-Marketing, Thành phố Hồ Chí Minh, Việt Nam*

TÓM TẮT

Độ chính xác của lượng tử là một vấn đề quan trọng trong lý thuyết thông tin lượng tử và lý thuyết hỗn độn lượng tử. Đại lượng này đo khoảng cách giữa các ma trận mật độ hay còn được hiểu là các trạng thái của lượng tử. Mặc dù đại lượng này không phải là một metric tuy nhiên nó có nhiều tính chất hữu dụng giúp xác định một metric trong không gian các ma trận mật độ. Trong bài báo này chúng tôi đưa ra và chứng minh một bất đẳng thức có tham số về độ chính xác của lượng tử. Hệ quả là với hai trạng thái lượng tử A và B sao cho $\frac{64}{81} \leq ||A - B||_1 \leq 16$, bất đẳng thức mà chúng tôi đưa ra là một trường hợp làm chặt hơn cho bất đẳng thức Fuchs-van de Graaf.

Từ khóa: *Thông tin lượng tử, hàm khoảng cách, trung bình nhân, độ chính xác.*

*Tác giả liên hệ chính..
Email: vtdnk2012@gmail.com

An inequality for quantum fidelity

Vuong Trung Dung[1,*], Vo Thi Bich Khue[2]

[1]*High School for the Gifted, Vietnam National University, Ho Chi Minh City,Viet Nam*
[2]*Department of Mathematics, Ho Chi Minh City University of Finance and Marketing, Ho Chi Minh City, Vietnam*

ABSTRACT

Quantum fidelity is an important quantity in quantum information theory and quantum chaos theory. It is a distance measure between density matrices which are considered as quantum states. Although it is not a metric, it has many useful properties that can be used to define a metric on the space of density matrices. In this article, we prove a parameterized inequality for quantum fidelity. As a consequence, for quantum states A and B such that $\frac{64}{81} \leq ||A - B||_1 \leq 16$, our result is a refinement of the well-known Fuchs-van de Graaf 's inequality.

Keywords: *Quantum information, Function distances, Geometric mean, fidelity.*

1. INTRODUCTION

Let $\mathcal{H}$ be the $n-$dimensional Hilbert space $\mathbb{C}^n$. The inner product between two vectors x and y is written as $\langle x, y \rangle$ or as x^*y. We denote by $\mathcal{L}(\mathcal{H})$ the space of all linear operators on $\mathcal{H}$, and by $\mathbb{M}_n(\mathbb{C})$ (or simply $\mathbb{M}_n$) the algebra of $n \times n$ matrices over $\mathbb{C}$. Denote by I the identity matrix of $\mathbb{M}_n$.

Every element A of $\mathcal{L}(\mathcal{H})$ can be identified with its matrix with respect to the standard basis $\{e_j\}$ of $\mathbb{C}^n$. We use the symbol A for this matrix as well. We say A is *positive semidefinite*[1] if

$$\langle x, Ax \rangle \geq 0, \text{ for all } x \in \mathcal{H},$$

and *positive definite* if, in addition,

$$\langle x, Ax \rangle > 0, \text{ for all } x \neq 0.$$

It is clear that a positive semidefinite matrix is a positive definite matrix if only if it is invertible. For convenience, we use the term *positive matrix* for a positive semidefinite, or a positive definite, matrix. Sometimes, if we want to emphasize that the matrix is positive definite, we say that it is *strictly positive*. We use the notation $A \geq 0$ to mean that A is positive, and $A > 0$

to mean it is strictly positive. We denote by A^T the *transpose* of matrix A and the *adjoint matrix* A^* as the complex conjugate of the transpose A^T. If $A = A^*$ then we call A is Hermitian, we also denote $\mathbb{H}_n$ as the real subspace of $\mathbb{M}_n$ consisting of Hermitian matrices. For each $A \in \mathbb{M}_n$, we have $A^*A \geq 0$. Therefore, we can define the matrix $|A| = (A^*A)^{1/2}$ which is called the *absolute value* of A. This matrix can be also defined using functional calculus. We have a result[1], A is positive if only if $A = B^2$ for some positive matrix B. Such B is unique. We write $B = A^{1/2}$ or $B = \sqrt{A}$ and call it the (positive) square root of A. Evidently, A is strictly positive if only if B is strictly positive.

The eigenvalues $\sigma_i(A)$ of $|A|$ are called *singular values* of A. If $A \in \mathbb{M}_n$, then the usual notation is

$$\sigma(A) = \Big(\sigma_1(A), ..., \sigma_n(A) \Big),$$

where

$$\sigma_1(A) \geq \sigma_2(A) \geq ... \geq \sigma_n(A).$$

Let denote by $\mathcal{D}_n$ the cone of positive definite matrices in $\mathbb{M}_n$. The space of density matrices is

defined as

$$\mathcal{D}_n^1 = \{A \in \mathcal{D}_n : Tr A = 1\}.$$

Definition 1.1. Let $A, B \in \mathcal{D}_n$ be positive semidefinite matrices. The *fidelity*[6,7] between two elements A and B is defined as

$$F(A, B) = ||\sqrt{A}\sqrt{B}||_1, \qquad (1.1)$$

where $||.||_1$ is Schatten 1-norm (trace norm),

$$||A||_1 = Tr|A| = Tr\sqrt{AA^*}.$$

Alternatively, the trace norm of an operator (or a matrix) A can be expressed as the sum of its singular values, $||A||_1 = \sum_{i=1}^{n} \sigma_i(A)$.

In quantum theory, quantum fidelity is defined for density matrices, and it can be generalized to the set of positive semidefinite matrices. By (1.1), we have

$$F(A, B) = Tr\left(A^{1/2}BA^{1/2}\right)^{1/2}.$$

Many researchers have paid attentions on different distance functions on $\mathcal{D}_n$ in the past few years. One of the important distance functions is the Bures distance[12]

$$d_b(A, B) = \left(Tr(A + B) - 2Tr((A^{1/2}BA^{1/2})^{1/2})\right)^{1/2}$$

or

$$d_b(A, B) = (Tr(A + B) - 2F(A, B))^{1/2}.$$

When $A, B \in \mathcal{D}_n^1$, quantum fidelity have several important properties[3,5,7], which can be proved in the sense of unital C^*-algebras

(1) Bounds: $0 \leq F(A, B) \leq 1$. Furthermore $F(A, B) = 1$ iff $A = B$, while $F(A, B) = 0$ iff $supp(A) \perp supp(B)$.

(2) Symmetry: $F(A, B) = F(B, A)$.

(3) Unitary Invariance: $F(A, B) = F(UAU^*, UBU^*)$, for any unitary matrix U.

(4) Concavity: $F(A, tB + (1 - t)C) \geq tF(A, B) + (1 - t)F(A, C)$, for $t \in [0, 1]$ and $A, B, C \in \mathcal{D}_n^1$.

(5) Multiplicativity: $F(A \otimes B, C \otimes D) = F(A, C) \cdot F(B, D)$, for A, B, C, and $D \in \mathcal{D}_n^1$.

(6) Joint concavity: $F(tA + (1 - t)B, tc + (1 - t)D) \geq tF(A, C) + (1 - t)F(B, D)$, for $t \in [0, 1]$ and A, B, C, and $D \in \mathcal{D}_n^1$.

In[12], the authors considered the function $f(X) = Tr(AX + BX^{-1})$ on $\mathcal{D}_n$. Using the Frechet derivative of the function $f(X)$ they showed that the geometric mean $X_0 = A^{-1}\sharp B = A^{-1/2}(A^{1/2}BA^{1/2})^{1/2}A^{-1/2}$ is the only critical point of $f(X)$. Hence, $f(X)$ attains minimum at X_0:

$$\min_{X>0} f(X) = f(X_0) =$$
$$Tr\left((A(A^{-1}\sharp B) + B(A\sharp B^{-1}))\right)$$
$$= 2Tr(A^{1/2}BA^{1/2})^{1/2} = 2F(A, B).$$

They also use the block matrix techniques to show the following: For positive definite matrices A and B,

(1) $F(A, B) = \min_{X>0} \sqrt{Tr(AX)Tr(BX^{-1})}$.

(2) $F(A, B) = \max_{X>0}\{|TrX| : A \geq XB^{-1}X^*\}$.

One of the most important inequalities of quantum fidelity is Fuchs de Graaf's inequality[6,11].

Theorem 1.1. *(Fuchs-van de Graaf's inequality) For two density matrices A and $B \in \mathcal{D}_n^1$, we have*

$$1 - \frac{1}{2}||A - B||_1 \leq F(A, B) \leq \sqrt{1 - \frac{1}{4}||A - B||_1^2}. \qquad (1.2)$$

Equivalently,

$$2 - 2F(A, B) \leq ||A - B||_1 \leq 2\sqrt{1 - F(A, B)^2}. \qquad (1.3)$$

The above inequality provides an upper bound and lower bound of quantum fidelity. It is also a tight relationship between different distances between A and B.

The proof of the right inequality of (1.3) is based on Uhlmann's theorem[4] while the proof of the left inequality of (1.3) based on the following result[6].

Lemma 1.1. *Let $A, B \in \mathcal{D}_n$ be positive semidefinite matrices. It holds that*

$$||A - B||_1 \geq ||\sqrt{A} - \sqrt{B}||_2^2,$$

where $||.||_2$ is the Shcatten 2-norm,

$$||A||_2 = \Big(\sum_{i=1}^n \sigma_i^2(A) \Big)^{1/2}.$$

It is worth mentioning that it is difficult to improve the Fuchs-van de Graaf inequality. In[11], the authors established a lower bound for $F(A, B)$ as follows.

Let $\lambda_0 = \lambda_{\max}(B^{-1/2}AB^{-1/2})$, where $\lambda_{\max}(X)$ is used to denote the maximum eigenvalue of the matrix X. Then

$$F(A, B) \geq 1 - \frac{1}{2} \frac{\sqrt{\lambda_0}}{\sqrt{\lambda_0} + 1} ||A - B||_1.$$

2. MAIN RESULTS

In this section we establish an estimate for the trace-norm of the difference for two density matrices A and B and the fidelity of A and the convex combination $tA + (1-t)B, t \in [0,1]$ of A and B.

Before presenting the main result, let's recall the following well-known inequality[8,9,10]

$$d_b(A, B) \leq d_1^{1/2}(A, B),$$

where $A, B \in \mathcal{D}_n$. This inequality was first proved in C^*-algebra setting by Araki in[10]. However, we can prove this inequality by another way as follows.

By Lemma (1.1), we have

$$
\begin{aligned}
||A - B||_1 &\geq ||\sqrt{A} - \sqrt{B}||_2^2 \\
&= Tr(\sqrt{A} - \sqrt{B})^2 \\
&= Tr(A + B - 2\sqrt{A}\sqrt{B}) \\
&\geq TrA + TrB - 2F(A, B) \\
&= d_b^2(A, B),
\end{aligned}
$$

where the last inequality follows from the fact that

$$F(A, B) = Tr(A^{1/2}BA^{1/2})^{1/2} \geq Tr(A^{1/2}B^{1/2}),$$

which is the consequence of the famous Araki-Lieb-Thirring inequality[15].

Theorem 2.1. *Let $A, B \in \mathcal{D}_n^1$ and $t \in [0,1]$. Then*

$$\sqrt{F(A, tAa + (1-t)B)} \geq 1 - \frac{1}{4}(1 - \sqrt{t})||A - B||_1^{1/2}.$$

Proof. Firstly, let us recall the Jensen inequality for trace. Let f be a continuous and concave function on an interval J and m be a natural number. Then for self-adjoint matrices $X_1, \cdots, X_m$ with spectra in J,

$$Tr\Big(f\Big(\sum_{i=1}^m A_i^* X_i A_i \Big) \Big) \geq Tr\Big(\sum_{i=1}^m A_i^* f(X_i) A_i \Big),$$

where $A_1, \cdots, A_m$ satisfy $\sum_{i=1}^m A_i^* A_i = I$.

We have

$$
\begin{aligned}
F(A, tA + (1-t)B) & \\
&= Tr[A^{1/2}(tA + (1-t)B)A^{1/2}]^{1/2} \\
&= Tr[tA^2 + (1-t)A^{1/2}BA^{1/2}]^{1/2} \\
&\geq Tr[tA + (1-t)(A^{1/2}BA^{1/2})^{1/2}] \\
&= t + (1-t)F(A, B),
\end{aligned}
$$

where the inequality is valid according to the concavity of the function $x \mapsto x^{1/2}$ and Jensen's trace inequality.

From $d_b(A, B) \leq d_1^{1/2}(A, B) = ||A - B||_1^{1/2}$, we have

$$
\begin{aligned}
& 1 - \frac{1}{4}(1 - \sqrt{t})||A - B||_1^{1/2} \\
\leq\ & 1 - \frac{1}{4}(1 - \sqrt{t})d_b(A, B)
\end{aligned}
$$

$$= 1 - \frac{1}{4}(1 - \sqrt{t})\sqrt{(2 - 2Tr(A^{1/2}BA^{1/2})^{1/2})}$$
$$= 1 - (1 - \sqrt{t})\sqrt{(1 - F(A, B))}.$$

Thus, it is necessary to prove

$$\sqrt{t + (1-t)F(A,B)} \geq 1 - (1-\sqrt{t})\sqrt{(1 - F(A, B))}.$$

Since $0 \leq t \leq 1$, and $0 \leq F(A, B) \leq 1$, squaring both sides of this inequality, we have

$$t + F - tF \geq 1 - 2(1 - \sqrt{t})\sqrt{1 - F}$$
$$+ (1 - \sqrt{t})^2(1 - F)$$
$$\Leftrightarrow (t-1)(1-F) + 2(1 - \sqrt{t})\sqrt{1-F}$$
$$- (1 - \sqrt{t})^2(1 - F) \geq 0$$
$$\Leftrightarrow (1-F)[(t-1) - (1-\sqrt{t})^2]$$
$$+ 2(1 - \sqrt{t})\sqrt{1-F} \geq 0$$
$$\Leftrightarrow 2(1 - \sqrt{t})[\sqrt{1-F} - (1-F)] \geq 0.$$

In the above transformations, F is used to denote for $F(A, B)$. The last inequality is evident because $0 \leq \sqrt{t} \leq 1$, and $0 \leq 1 - F(A, B) \leq 1$. □

Remark 2.1. For $t = 0$, with $||A - B||_1 \leq 16$ and from the theorem we have

$$F(A, B) \geq (1 - \frac{1}{4}||A - B||_1^{1/2})^2.$$

Let's compare the value $(1 - \frac{1}{4}||A - B||_1^{1/2})^2$ and the value $1 - \frac{1}{2}||A - B||_1$ on the left-hand-side part in the Fuchs-van de Graaf inequality. By a simple computation, if $||A - B||_1 \geq 64/81$ then we have

$$F(A, B) \geq (1 - \frac{1}{4}||A - B||_1^{1/2})^2 \geq 1 - \frac{1}{2}||A - B||_1.$$

Indeed, from the last inequality we have

$$(1 - \frac{1}{4}||A - B||_1^{1/2})^2 \geq 1 - \frac{1}{2}||A - B||_1$$
$$\Leftrightarrow 1 - \frac{1}{2}||A - B||_1^{1/2} + \frac{1}{16}||A - B||_1$$
$$\geq 1 - \frac{1}{2}||A - B||_1$$

$$\Leftrightarrow \frac{||A - B||_1^{1/2}}{2}\Big(\frac{9}{8}||A - B||_1^{1/2} - 1\Big) \geq 0,$$

which is equivalent to that

$$||A - B||_1 \geq 64/81.$$

Therefore, the main result is a refinement of the Fuchs-van de Graaf inequality for a big set of quantum states A and B.

REFERENCES

1. R.Bhatia. *Positive Definite matrices (Princeton Series in Applied Mathematics)*, Princeton University Press, 2007.

2. R.Bhatia. *Matrix Analysis*, Springer, 1997.

3. A.Uhlmann. The transition probability in the state space of a*-algebra, *Mathematical Phys*, **1976**, *9*(2), 273-279.

4. M.A.Nielsen, L.L.Chuang. *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, 2000.

5. Richard Jozsa. Fidelity for Mixed Quantum States, *Journal of Modern Optics*, **1994**, *41(12)*, 2315-2323.

6. J. Watrous. *The Theory of quantum information*, Cambridge University Press, 2008.

7. J.A. Miszczak, Z.Puchala, P.Horodecki, A.Uhlmanx, K.Zyczkowski. Sub-and super-fidelity as bounds for quantum fidelity, *Quantum Information and Computation*, **2009**, *9*(12),103-130.

8. D.Spehner, M.Orszag. Geometric quantum discord with Bures distance, *New journal of physics*, **2013**, *15*.

9. D.Spehner, F.Illuminati, M.Orszag, W.Roga. Geometric Measures of Quantum Correlations with Bures and Hellinger Distances, *Lectures on General Quantum Correlations and their Applications*, **2017**, 105-157.

10. H.Araki. A remark on Bures distance function for normal states, *RIMS Kyoto University,* **1970-1971**, *6*, 477-482.

11. Zhang, K.Bu, J.Wu. A lower bound on the fidelity between two states in terms of their trace-distance and max-relative entropy, *Linear and Multilinear Algebra,* **2016**, *64*(5),801-806.

12. R. Bhatia, T. Jain, Y. Lim. On the Bures-Wasserstein distance between positive definite matrices, *Expositiones Mathematicae,* **2019**, *37*, 165-191.

13. R. Bhatia, T. Jain, Y. Lim. Inequalities for the Wasserstein mean of positive definite matrices, *Linear Algebra and its Applications,* **2019**, *576*, 108-123.

14. R.Bhatia, T. Jain, Y.Lim. Strong convexity of sandwiched entropies and related optimization problems, *Reviews in Mathematical Physics,* **2018**, *30*(9) .

15. K.M.R. Audenaert. On the Araki-Lieb-Thirring inequality, *International Journal of Information and Systems Sciences,* **2008**, *4*(1), 78-83.

Một đặc trưng cho bậc của đa tạp Fano

Đặng Tuấn Hiệp[1], Nguyễn Thị Mai Vân[2,*]

[1]*Khoa Toán - Tin học, Trường Đại học Đà Lạt, Việt Nam*

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

TÓM TẮT

Sử dụng một kết quả gần đây về số giao trên đa tạp Grassmann, chúng tôi đưa ra một đặc trưng cho bậc của đa tạp Fano của các không gian con tuyến tính trên giao đầy đủ trong một không gian xạ ảnh phức dưới dạng hệ số của một đa thức đối xứng.

Từ khóa: *Đa tạp Fano, công thức bậc.*

*Tác giả liên hệ chính.
Email: ntmvan18@gmail.com*

A characterization for the degree of Fano varieties

Dang Tuan Hiep[1], **Nguyen Thi Mai Van**[2,*]

[1]*Faculty of Mathematics and Computer Science, Da Lat University, Vietnam*
[2]*Faculty of Mathematics and Statistics, Quy Nhon University, Vietnam*

ABSTRACT

Using a recent result on intersection numbers over Grassmannians, we propose a characterization for the degree of Fano varieties of linear subspaces on complete intersections in a complex projective space in terms of the coefficient of a symmetric polynomial.

Keywords: *Fano variety, degree formula.*

1. INTRODUCTION

Let X be a general complete intersection of type $\underline{d} = (d_1, \ldots, d_r)$ in the projective space $\mathbb{P}^n$ over the complex field $\mathbb{C}$, provided that $n, d_1, \ldots, d_r$ are natural numbers with $n \geq 4, d_i \geq 2$ for all i. Recall that the Fano variety $F_k(X)$ parametrizing linear subspaces of dimension k contained in X is a smooth subvariety of the Grassmannian $G(k+1, n+1)$ of linear subspaces of dimension k in $\mathbb{P}^n$, provided that

$$(k+1)(n-k) \geq \sum_{i=1}^{r} \binom{d_i + k}{k}$$

and X is not a quadric, in which last case $n \geq 2k + r$ is required[1-2]. The degree of $F_k(X)$ under the Plücker embedding were formulated by Debarre-Manivel (1998)[2] and Hiep (2016)[6].

In this paper, we show that the degree of $F_k(X)$ can be expressed in terms of the coefficient of a symmetric polynomial. For convenience, we set

$$\delta(n, \underline{d}, k) = (k+1)(n-k) - \sum_{i=1}^{r} \binom{d_i + k}{k},$$

which is the expected dimension of $F_k(X)$. Our main result is the following:

Theorem 1. *With the notations as above, if $\delta(n, \underline{d}, k) \geq 0$, then the degree of $F_k(X)$ under the Plücker embedding is given by*

$$\deg(F_k(X)) = \frac{c(n, \underline{d}, k)}{(k+1)!},$$

where $c(n, \underline{d}, k)$ is the coefficient of $x_0^n \cdots x_k^n$ in the polynomial

$$P_{(n,\underline{d},k)}(x_0, \ldots, x_n) \prod_{i \neq j}(x_i - x_j),$$

where

$$P_{(n,\underline{d},k)}(x_0, \ldots, x_n) = \prod_{i=1}^{r} \prod_{a_0 + \cdots + a_k = d_i, a_i \in \mathbb{N}}$$

$$(a_0 x_0 + \cdots + a_k x_k)(x_0 + \cdots + x_k)^{\delta(n,\underline{d},k)}.$$

The statement of Theorem 1 seems to be similar to that of Debarre-Manivel (1998)[2]. However, we here consider the coefficient of the monomial $x_0^n \cdots x_k^n$ in the product of the polynomial $P_{(n,\underline{d},k)}(x_0, \ldots, x_n)$ by the discriminant

$$\Delta = \prod_{i \neq j}(x_i - x_j)$$

instead of that of the monomial $x_0^n x_1^{n-1} \cdots x_k^{n-k}$ in the product of the same polynomial by the Vandermonde determinant

$$V = \prod_{i < j}(x_i - x_j).$$

---
*Corresponding author.
Email: ntmvan18@gmail.com*

In particular, our approach is completely different from that of Debarre-Manivel (1998)[2]. We apply an integral formula for Grassmannians which has been recently explored by Hiep (2019)[8] (see Theorem 3 below).

The rest of the paper is organized as follows: Section 2 presents preliminary results. Section 3 presents the proof of the main theorem.

2. PRELIMINARY RESULTS

In this section, we review the basic notions and results which are known.

2.1. Grassmannians and their Schubert classes.

Let $G(k,n)$ be the *Grassmannian* of k-dimensional linear subspaces in a vector space V of dimension n. The *tautological subbundle S* on $G(k,n)$ is the vector bundle of rank k whose fiber at $W \in G(k,n)$ is the vector subspace $W \subset V$ itself. The *tautological quotient bundle Q* on $G(k,n)$ is the vector bundle of rank $n-k$ whose fiber at $W \in G(k,n)$ is the quotient vector space V/W. The *tangent bundle T* on $G(k,n)$ is isomorphic to $\text{Hom}(S,Q) \cong S^\vee \otimes Q$.

Let $\mathcal{V}$ be a flag in V, that is, a strictly increasing sequence of linear subspaces

$$0 \subset V_1 \subset \cdots \subset V_{n-1} \subset V_n = V,$$

where $\dim V_i = i$.

For any sequence $a = (a_1, \ldots, a_k)$ of integers with

$$n - k \geq a_1 \geq a_2 \geq \cdots \geq a_k \geq 0,$$

we define the *Schubert cycle* by $\Sigma_a(\mathcal{V})$, where

$$\Sigma_a(\mathcal{V}) = \{W \in G(k,n) :$$

$$\dim(V_{n-k+i-a_i} \cap W) \geq i, i = 1, \ldots, k\}.$$

One can show that this is an irreducible subvariety of $G(k,n)$ of codimension

$$|a| = \sum_{i=1}^{k} a_k,$$

and its cycle class $[\Sigma_a(\mathcal{V})]$ does not depend on the choice of flag. We then define the *Schubert class* to be the cycle class $\sigma_a := [\Sigma_a(\mathcal{V})]$.

To shorten the notation, we write Σ_a in place of $\Sigma_a(\mathcal{V})$, write $\Sigma_{a_1,\ldots,a_s}, \sigma_{a_1,\ldots,a_s}$ whenever $a = (a_1, \ldots, a_s, 0, \ldots, 0)$ and $\Sigma_{p^i}, \sigma_{p^i}$ whenever $a = (p, \ldots, p, 0, \ldots, 0)$ with i the first components equal to p. Then the cycle classes $\sigma_i, i = 1, \ldots, n-k$ and $\sigma_{1^i}, i = 1, \ldots, k$ are called *special Schubert classes*.

The special Schubert classes are intimately connected with the tautological bundles on $G(k,n)$, and both $\{\sigma_1, \sigma_2, \ldots, \sigma_{n-k}\}$ and $\{\sigma_1, \sigma_{1^2}, \ldots, \sigma_{1^k}\}$ are minimal generating sets for the Chow ring of $G(k,n)$. More precisely, we have the following statements.

Proposition 1 (Manivel (2001)[10] and Eisenbud-Harris (2016)[3]). *The Chern classes of S and Q are as follows:*

$$c_i(S) = (-1)^i \sigma_{1^i}, \quad i = 1, \ldots, k$$

and

$$c_i(Q) = \sigma_i, \quad i = 1, \ldots, n-k.$$

By Corollary 3.5 in Eisenbud-Harris (2016)[3], the Schubert classes form a free $\mathbb{Z}$-basis for $A(G(k,n))$. The multiplication is determined by the following formulas.

Proposition 2 (Duality formula). *(See Corollary 3.2 and Proposition 3.4 in Eisenbud-Harris (2016)[3]) If $|a| + |b| = k(n-k)$, we have*

$$\sigma_a \cdot \sigma_b = \begin{cases} \sigma_{(n-k)^k} & \text{if } a_i + b_{k-i} = n-k \text{ for all } i, \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, both $(\sigma_{n-k})^k$ and $(\sigma_{1^k})^{n-k}$ are equal to the class of a point in the Chow ring of $G(k,n)$.

Proposition 3 (Pieri formula). *(See Proposition 3.7 in Eisenbud-Harris (2016)[3]) For any Schubert class $\sigma_a \in A^*(G(k,n))$ and any integer i with $0 \leq i \leq n-k$, we have*

$$\sigma_a \cdot \sigma_i = \sum_c \sigma_c,$$

where the sum is over all c with $n - k \geq c_1 \geq a_1 \geq c_2 \geq \cdots \geq c_k \geq a_k \geq 0$, and $|c| = |a| + i$.

Proposition 4 (Giambelli formula). *(see Section 1.5 in Griffiths-Harris (1978)[5]) For any $a = (a_1, \ldots, a_k)$ with $n - k \geq a_1 \geq a_2 \geq \cdots \geq a_k \geq 0$, we have*

$$\sigma_a = \det(\sigma_{a_i + j - i})_{1 \leq i, j \leq k},$$

where $\sigma_0 = 1$ and $\sigma_m = 0$ whenever $m < 0$ or $m > n - k$.

Pieri's formula shows how to determine the product of an arbitrary Schubert class and a special Schubert class. Giambelli's formula shows how to express an arbitrary Schubert class in terms of special ones. Therefore, both formulas give us an effective way to determine the product of two arbitrary Schubert classes.

2.2. Splitting Principle.

The splitting principle is a useful technique for reducing questions concerning vector bundles to questions concerning line bundles.

Let E be a vector bundle of rank r on a variety X. The splitting principle says that we can regard the Chern classes of E as the elementary symmetric polynomials in r variables α_i for all $i = 1, \ldots, r$, which are called the *Chern roots* of E. More precisely, we have

$$c_0(E) = 1,$$
$$c_1(E) = \sum_{1 \leq i \leq r} \alpha_i,$$
$$c_2(E) = \sum_{1 \leq i < j \leq r} \alpha_i \alpha_j,$$
$$\vdots$$
$$c_r(E) = \alpha_1 \alpha_2 \ldots \alpha_r.$$

By the splitting principle and the Chern roots, one has the following statements.

Proposition 5. *(See Remark 3.2.3 and Example 3.2.6 in Fulton (1997)[4]) Let E and F be two vector bundles with Chern roots $(\alpha_i)_i$ and $(\beta_j)_j$, respectively. Then we have the following statements:*

(i) $E^\vee$ has the Chern roots $(-\alpha_i)_i$. Hence $c_k(E^\vee) = (-1)^k c_k(E)$ for all k.

(ii) $E \otimes F$ has the Chern roots
$$(\alpha_i + \beta_j)_{i,j}.$$

(iii) $\mathrm{Sym}^d E$ has the Chern roots
$$(\alpha_{i_1} + \cdots + \alpha_{i_d})_{i_1 \leq \cdots \leq i_d}.$$

(iv) $\wedge^d E$ has the Chern roots
$$(\alpha_{i_1} + \cdots + \alpha_{i_d})_{i_1 < \cdots < i_d}.$$

Here we denote by $\mathrm{Sym}^d E$ the d-th symmetric power of E, and $\wedge^d E$ the d-th extorior power of E.

Example 1. Let E be a vector bundle of rank 2 on a variety X of dimension 4. We want to compute the Chern classes of $\mathrm{Sym}^3 E$ in terms of the Chern classes of E. If α_1 and α_2 are the Chern roots of E, then $\mathrm{Sym}^3 E$ has the Chern roots $3\alpha_1, 2\alpha_1 + \alpha_2, \alpha_1 + 2\alpha_2, 3\alpha_2$. Thus we have

$$c_1(\mathrm{Sym}^3 E) = 3\alpha_1 + 2\alpha_1 + \alpha_2 + \alpha_1 + 2\alpha_2 + 3\alpha_2$$
$$= 6(\alpha_1 + \alpha_2)$$
$$= 6c_1(E).$$

Similarly, we have

$$c_2(\mathrm{Sym}^3 E) = 11c_1(E)^2 + 10c_2(E),$$
$$c_3(\mathrm{Sym}^3 E) = 6c_1(E)^3 + 30c_1(E)c_2(E),$$
$$c_4(\mathrm{Sym}^3 E) = 18c_1(E)^2 c_2(E) + 9c_2(E)^2.$$

2.3. Fano varieties: the hypersurface case.

Let X be a general hypersurface of degree d in the projective space $\mathbb{P}^n$ over the complex field $\mathbb{C}$. The *Fano variety* $F_k(X)$ is defined to be the set of k-dimensional subspaces of $\mathbb{P}^n$ which are contained in X. This is a subvariety of the Grassmannian $G(k+1, n+1)$. For convenience, we set

$$\delta = (k+1)(n-k) - \binom{d+k}{k}.$$

Suppose that $d \neq 2$ (or $n \geq 2k + r$) and $\delta \geq 0$. Langer (1996)[9] showed that $F_k(X)$ is smooth of expected dimension δ. By the language of Schubert calculus, Debarre-Manivel (1998)[2] showed that the degree of $F_k(X)$ is equal to a certain coefficient of an explicit polynomial, gives as the product of linear forms. Hiep (2016)[6] proposed

and proved an explicit formula for computing the degree of $F_k(X)$ via equivariant intersection theory.

Consider the diagonal action of $T = (\mathbb{C}^*)^{n+1}$ on $\mathbb{P}^n$ given in coordinates by

$$(t_0, \ldots, t_n)(x_0 : \ldots : x_n) = (t_0 x_0 : \ldots : t_n x_n)$$

This induces an action of T on the Grassmannian $G(k+1, n+1)$ with $\binom{n+1}{k+1}$ isolated fixed points L_I corresponding to the coordinate k-subspaces in $\mathbb{P}^n$, which are indexed by the subsets I of size $k+1$ of the set $\{1, \ldots, n+1\}$. Let $\mathcal{I}$ denote the set of all these subsets. Then the degree of $F_k(X)$ can be expressed as a sum of rational polynomials, where the sum ranges over all elements I of $\mathcal{I}$.

Theorem 2. *(See Theorem 1.1 in Hiep (2016)[6]) Let $X \subset \mathbb{P}^n$ be a general hypersurface of degree d and $k < n$ be a positive integer. Then the degree of the Fano variety $F_k(X)$ can be computed by the following formula*

$$\deg(F_k(X)) = (-1)^\delta \sum_{I \in \mathcal{I}} \frac{S_I Q_I^\delta}{T_I},$$

where

$$S_I = \prod_{v_i \in \mathbb{N}, \sum_{i \in I} v_i = d} \left(\sum_{i \in I} v_i h_i \right),$$

$$Q_I = \sum_{j \notin I} h_j \ \text{and} \ T_I = \prod_{i \in I} \prod_{j \notin I} (h_i - h_j)$$

are polynomials in $\mathbb{C}[h_1, \ldots, h_{n+1}]$.

Remark 1. The right-hand-side of the formula in Theorem 2 is the sum of rational polynomials, and the above theorem claims in other words that it is in fact a constant function, moreover it is an integer. Namely, for any numbers h_i such that $h_i \neq h_j$ for $i \neq j$, the right-hand-side of the formula becomes the same integer.

Example 2. Let $k = 1$ and $X \subset \mathbb{P}^3$ be a general cubic surface. In this case, the Fano variety $F_1(X)$ has the expected dimension $\delta = 0$. The degree of $F_1(X)$ can be computed as follows:

$$\deg(F_1(X)) = \sum_{\{i_1, i_2\} \subset \{1,2,3,4\}}$$

$$\frac{3h_{i_1}(2h_{i_1} + h_{i_2})(h_{i_1} + 2h_{i_2})3h_{i_2}}{(h_{i_1} - h_{j_1})(h_{i_1} - h_{j_2})(h_{i_2} - h_{j_1})(h_{i_2} - h_{j_2})},$$

where $\{j_1, j_2\}$ is the complement of the subset $\{i_1, i_2\}$ in the set $\{1, 2, 3, 4\}$. After simplifying, we obtain the degree of $F_1(X)$ is 27 as desired.

2.4. Intersection numbers on Grassmannians.

Consider the following intersection number on the Grassmannian $G(k, n)$

$$\int_{G(k,n)} \Phi(\mathcal{S}),$$

where $\Phi(\mathcal{S})$ are respectively characteristic classes of the tautological sub-bundle $\mathcal{S}$.

Using an identity involving symmetric polynomials, Hiep (2019)[8] expressed the intersection number in terms of a coefficient of a certain monomial in the expansion of a symmetric polynomial.

Theorem 3. *(See Corollary 1 in Hiep (2019)[8]) Suppose that $\Phi(\mathcal{S})$ is represented by a symmetric polynomial $P(x_1, \ldots, x_k)$ of degree not greater than $k(n-k)$ in k variables $x_1, \ldots, x_k$ which are the Chern roots of $\mathcal{S}$. Then we have the following formula:*

$$\int_{G(k,n)} \Phi(\mathcal{S}) = (-1)^{k(n-k)} \frac{c(k,n)}{k!},$$

where $c(k, n)$ is the coefficient of $x_1^{n-1} \ldots x_k^{n-1}$ in the polynomial

$$P(x_1, \ldots, x_k) \prod_{j \neq i} (x_i - x_j).$$

2.5. Debarre-Manivel's result.

As mentioned in the introduction, our main result seems to be similar to that of Theorem 4.3 in Debarre-Manivel (1998)[2]. Let us recall the result of Debarre-Manivel for comparison.

Theorem 4. *(See Theorem 4.3 in Debarre-Manivel (1998)[2]) With the notations as mentioned in the introduction and* $\delta(n, \underline{d}, k) \geq 0$. *Then the degree of* $F_k(X)$ *under the Plücker embedding is given by*

$$\deg(F_k(X)) = e(n, \underline{d}, k),$$

where $e(n, \underline{d}, k)$ *is the coefficient of* $x_0^n x_1^{n-1} \cdots x_k^{n-k}$ *in the polynomial*

$$P_{(n, \underline{d}, k)}(x_0, \ldots, x_n) \prod_{i<j}(x_i - x_j).$$

Example 3. Let us come back Example 2 above. By the statement of Theorem 4, the degree of $F_1(X)$ can be computed as follows:

$$\deg(F_1(X)) = e(3, (3), 1),$$

where $e(3, (3), 1)$ is the coefficient of $x_0^3 x_1^2$ in the expansion of the polynomial

$$3x_0(2x_0 + x_1)(x_0 + 2x_1)3x_1(x_0 - x_1).$$

After expanding, we obtain $e(3, (3), 1) = 27$, then the degree is 27 as desired.

3. PROOF OF THEOREM 1

We first prove the following lemma.

Lemma 1. *(see Proposition 6.4 in Eisenbud-Harris (2016)[3] and Lemma 3 in Hiep-Tu-Van (2019)[7]) Let* $X \subset \mathbb{P}^n$ *be a general complete intersection of type* $(d_1, \ldots, d_r)$. *The variety* $F = F_k(X)$ *is the zero locus of a global section of the vector bundle*

$$\mathcal{F} = \bigoplus_{i=1}^r \operatorname{Sym}^{d_i} S^\vee.$$

Proof of Lemma 1. Assume that X is the intersection of r hypersurfaces $X_1, \ldots, X_r$ with $\deg(X_i) = d_i$ for all i. Each $F_k(X_i)$ is the zero locus of a global section s_i of $\operatorname{Sym}^{d_i} S^\vee$. Thus the variety F, which is the intersection of the $F_k(X_i)$, is the zero locus of a global section $s = (s_1, \ldots, s_r)$ of the vector bundle $\mathcal{F}$.

We now prove Theorem 1. By the Gauss-Bonnet formula (see, for example, Section 3.5.3 in Manivel (2001)[10]), the class of $F_k(X)$ is the

top Chern class of the vector bundle $\mathcal{F}$. If $\delta(n, \underline{d}, k) \geq 0$, then the degree of $F_k(X)$ can be expressed as follows:

$$\deg(F_k(X)) =$$

$$\int_{G(k+1, n+1)} \prod_{i=1}^r c_{\operatorname{top}}(\operatorname{Sym}^{d_i} S^\vee) \cdot c_1(S^\vee)^{\delta(n, \underline{d}, k)},$$

where $c_{\operatorname{top}}(E)$ is the top Chern class of the vector bundle E. By the splitting principle, the characteristic class

$$\prod_{i=1}^r c_{\operatorname{top}}(\operatorname{Sym}^{d_i} S^\vee) \cdot c_1(S^\vee)^{\delta(n, \underline{d}, k)}$$

is represented by the symmetric polynomial

$$(-1)^{(k+1)(n-k)} P_{(n, \underline{d}, k)}(x_0, \ldots, x_n).$$

Note that $x_0, \ldots, x_k$ are the Chern roots of the tautological sub-bundle S on the Grassmannian $G(k+1, n+1)$. By Theorem 3, Theorem 1 follows.

Example 4. Let us come back Example 2 above. By the statement of Theorem 1, the degree of $F_1(X)$ can be computed as follows:

$$\deg(F_1(X)) = \frac{c(3, (3), 1)}{2!},$$

where $c(3, (3), 1)$ is the coefficient of $x_0^3 x_1^3$ in the expansion of the polynomial

$$3x_0(2x_0 + x_1)(x_0 + 2x_1)3x_1(x_0 - x_1)(x_1 - x_0).$$

After expanding, we obtain $c(3, (3), 1) = 54$, then the degree is 27 as desired.

REFERENCES

1. C. Borcea. Deforming varieties of k-planes of projective complete intersections, *J. Math.143*, **1990**, *143*, 25–36.

2. O. Debarre, L. Manivel. Sur la variété des espaces linéaires contenus dans une intersection complète, *Math. Ann.*, **1998**, *312*, 549–574.

3. D. Eisenbud, J. Harris. 3264 & *All that: A second course in algebraic geometry*, Cambridge University Press, 2016.

4. W. Fulton. *Intersection theory*, Springer-Verlag, 1997.

5. P. Griffiths, J. Harris. *Principles of Algebraic Geometry*, John Wiley & Sons, 1978.

6. D. T. Hiep. On the degree of Fano schemes of linear subspaces on hypersurfaces, *Kodai Math. J.*, **2016**, *39*, 110–118.

7. D. T. Hiep, N. C. Tu, N.T.M Van. A genus-degree formula for Fano varieties of linear subspaces on complete intersections, *Journal of Science: Quy Nhon University*, **2019**, *13*, 91–97.

8. D. T. Hiep. Identities involving (doubly) symmetric polynomials and integrals over Grassmannians, *Fundamenta Mathematicae*, **2019**, *246*, 181–191.

9. A. Langer. Fano schemes of linear spaces on hypersurfaces, *Manuscripta Math*, **1997**, *93*, 21–28.

10. L. Manivel, *Symmetric functions, Schubert polynomials and degeneracy loci*, American Mathematical Society, 2001.

Ảnh hưởng của tỷ lệ mol giữa ion kim loại/axit citric và nhiệt độ nung đến cấu trúc và hoạt tính quang xúc tác của YFeO$_3$

Dương Đức Lương, Nguyễn Văn Thắng*

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

TÓM TẮT

Vật liệu FeYO$_3$ tổng hợp thành công bằng phương pháp sol-gel đã được khảo sát đặc trưng bằng phương pháp nhiễu xạ tia X, kính hiển vi điện tử quét, phương pháp phổ hồng ngoại, phương pháp phổ phản xạ khuếch tán tử ngoại khả kiến. Ảnh hưởng của điều kiện tổng hợp vật liệu như tỷ lệ mol giữa kim loại/axit citric, nhiệt độ nung đến cấu trúc và hoạt tính quang xúc tác của FeYO$_3$ đã được khảo sát một cách có hệ thống. Kết quả thực nghiệm cho thấy, tất cả các vật liệu YFeO$_3$ thu được có hoạt tính quang xúc tác trong vùng ánh sáng khả kiến thông qua sự phân hủy xanh metylen trong dung dịch nước. Cụ thể, vật liệu YFeO$_3$ được thiêu kết ở 900 °C ứng với tỷ lệ mol 1:2 của kim loại và axit citric cho hiệu suất quang xúc tác tốt nhất so với các vật liệu YFeO$_3$ được tổng hợp trong cùng điều kiện nhưng được nung ở nhiệt độ khác.

Từ khóa: *YFeO$_3$, phương pháp sol-gel, quang xúc tác, ánh sáng nhìn thấy và xanh methylen.*

Tác giả liên hệ chính.
Email: nguyenvanthang@qnu.edu.vn

Influence of molar ratio of metal ions to citric acid and annealing temperature on structure and photocatalytic properties of YFeO$_3$ photocatalysts synthesized by sol-gel method

Duong Duc Luong, Nguyen Van Thang[*]

Faculty of Natural Sciences, Quy Nhon University, Vietnam

ABSTRACT

YFeO$_3$ materials, which were successfully synthesized by the sol-gel method, were characterized by X-ray diffraction (XRD), Scanning Electron Microscopy (SEM), Fourier Transform Infrared Spectroscopy (FT-IR), and Ultraviolet-Visible Diffuse Reflectance Spectroscopy (UV-Vis DRS). The influence of the synthesizing conditions such as the molar ratio of metal ions to citric acid, and the annealing temperature on the structure and photocatalytic properties of YFeO$_3$ samples has been systematically investigated. The experimental results show that all obtained YFeO$_3$ samples exhibit photocatalytic activities under the visible light region through the photodegradation of methylene blue (MB) in an aqueous solution. Specifically, the YFeO$_3$ sample, which was synthesized with the molar ratio of 1: 2 metal ions/citric acid and annealed at 900 °C, shows the highest photocatalytic degradation efficiency compared to YFeO$_3$ samples synthesized under the same conditions but annealed at other temperatures.

Keywords: *YFeO$_3$ photocatalysts, sol-gel method, photocatalysis, visible-light irradiation, and methylene blue degradation.*

1. INTRODUCTION

Environmental concerns, especially water pollution, have been a leading challenge for humanity over the years due to the population and economic growth. Many different methods of treating contaminated water have been reported.[1] However, these methods have proved ineffective because they do not completely remove pollutants or are time-consuming and costly in operation. Therefore, the search for new and more effective water treatment methods is essential.

Recently, a method that has been researched and applied by scientists in the treatment of polluted water is the use of photocatalytic materials in order to decompose toxic organic compounds in water.[2] Among various photocatalysts, TiO$_2$-based catalysts have been extensively studied owing to its high physical and chemical stability, low production cost and high photocatalytic activity under irradiation of UV light.[3–7] Due to the wide band gap (Eg) ($\approx$ 3.2 eV), however, TiO$_2$ is limited in its application under visible light.[8] Hence, it is less effective when used as photocatalysts in the visible light region. Visible light is a source of 43% of the sun's energy spectrum, a source of clean energy and is considered endless. Thus, it is highly desirable to look for photocatalysts with a narrow band gap so that they have high photocatalytic activity under the visible light irradiation.

[*]*Corresponding author.*
Email: nguyenvanthang@qnu.edu.vn

In the search for new materials over the years, iron-based (III) semiconductors such as Fe_2O_3, $FeBiO_3$ and especially $YFeO_3$ become more and more attractive since they have a narrow band gap and thus effectively operating under the irradiation of visible light.[8,9] However, $YFeO_3$ is thermodynamically unstable and can be converted to a more stable form Fe_3O_4 or $Y_3Fe_5O_{12}$ at high temperatures, so it is difficult to synthesize pure $YFeO_3$.[10] On the other hand, most current methods for synthesizing $YFeO_3$ require high-temperature annealing, resulting in a high-energy consumption and the creation of large particle size materials. Therefore, it is highly desirable to optimize the synthesizing conditions for $YFeO_3$.

This work focuses on the fabrication of $YFeO_3$ photocatalysts by the sol-gel method. The influence of both the molar ratio of metal ions to citric acid and the annealing temperature on the crystal structure, surface morphology, and methylene blue (MB) photodegradation efficiency of obtained $YFeO_3$ materials was systematically investigated.

2. EXPERIMENT

2.1. Chemicals

The chemicals used in our experiments were yttrium oxide (Y_2O_3, 99.99% purity), citric acid ($C_6H_8O_7$, 98% purity), iron(III) nitrate nonahydrate ($Fe(NO_3)_3.9H_2O$, 98%), methylene blue ($C_{16}H_{18}N_3SCl$) and nitric acid (HNO_3, 65%).

2.2. Synthesis of $YFeO_3$ by the sol-gel method

In order to prepare $YFeO_3$ samples by the sol-gel method, 3.513 g of Y_2O_3 and 12.560g of $Fe(NO_3)_3.9H_2O$ were dissolved in 100 mL of HNO_3 1 M in a 250 mL glass beaker by magnetically stirring constantly at room temperature. When the mixture became transparent, different amounts of citric acid was added and continuously stirred for 24 hours. After that, the temperature was increased to 80°C and constantly stirred for 2 hours to create the sol. In order to obtain the gel, the temperature was raised to 100 °C to remove water molecules

from the sol. When the water evaporates, the sol changes into a highly viscous gel. The obtained gel was dried in an oven at 200 °C for a period of 2 hours. At this temperature, the gel burns and releases NO_2 gas from the gel layer, causing a significant increase in the gel volume. As a result, a yellow-brown powder called a dry gel was obtained. The powder was then annealed at various temperatures between 700 to 1000 °C for 5 hours in the air to obtain the final products. The resultant samples were denoted as $YFeO_3$.850. (1:n) (n = 1, 2, 3 and 4) and $YFeO_3$.T(1:2) (T = 700, 800, 900 and 1000) for samples annealed at 850 °C with different molar ratios of metal ions to citric acid and samples annealed at different temperatures, respectively.

2.3. Material characterizations

XRD patterns of obtained samples were collected at room temperature by means of a D8 Advance Brucker diffractometer operated at 40 kV and 100 mA with Cu-Kα radiation ($\lambda_{K\alpha}$ = 1,5406 Å) in a 2θ range between 20° and 80° at the step of 0.03°. FT-IR spectra of obtained samples were recorded by means of an IRAffinity-1S in the spectra range of 4000 to 400 cm^{-1}. The optical properties of obtained $YFeO_3$ photocatalysts were investigated by UV-Vis DRS using a UV-Vis spectrometer Jasco-V670. The particle size and surface morphology of the samples were investigated by SEM carried out on Nano SEM-450.

2.4. Photocatalytic activities of $YFeO_3$

Photocatalytic activities of the obtained $YFeO_3$ samples were evaluated through the photodegradation of MB in an aqueous solution under the visible light irradiation. Typically, 0.1 g of the photocatalyst was added to 80 mL of MB solution (concentration of 10 mg/L) in a 250 mL glass beaker. After that, the mixture was continuously stirred in the dark for 2 hours in order to allow the adsorption and desorption processes of MB molecules on the surface of the photocatalyst to reach the equilibrium before investigating its photocatalytic activities under the irradiation of a tungsten filament

lamp 60 W – 220 V in a period of 6 hours. At certain irradiation intervals, 7 mL of the mixed suspension was taken out and centrifuged to remove the photocatalyst from the mixture (6000 rpm, 20-minute centrifugation time). The residual MB concentration as a function of the irradiation time was analyzed by collecting variation of the absorbance at 663 nm using UV-Vis CE-2011 spectrometer. MB photodegradation efficiency is calculated by the following formula:

$$H = \frac{C_o - C}{C_o} \cdot 100$$

In which, C_o is the initial concentration of MB and C is the remaining concentration of MB after each corresponding irradiation time.

3. RESULTS AND DISCUSSION

3.1. Effect of the molar ratio of metal ions to citric acid on the crystal structure, surface morphology, and methylene blue decomposition efficiency

With the aim to optimize the lattice crystal structure of $YFeO_3$ materials, different molar ratios of metal ions to citric acid (1:1, 1:2, 1:3 and 1:4) were employed during the sample synthesis.

3.1.1. Effect of the molar ratio of metal ions to citric acid on the crystal structure

The structure of the obtained $YFeO_3.850(1:n)$ samples was investigated by both XRD and FT-IR. Figure 3.1 shows the XRD patterns collected at room temperature for the $YFeO_3.850(1:n)$ samples.
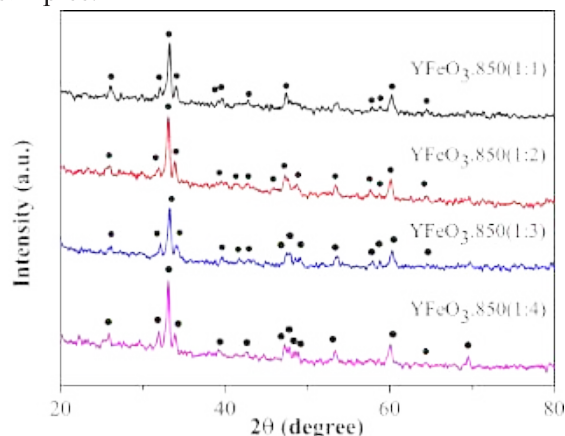
Figure 3.1. XRD patterns for $YFeO_3.850(1:n)$ samples.

As can be clearly seen in Figure 3.1, the diffraction peaks of the $YFeO_3.850(1:n)$ samples are completely consistent with the JCPDS Card No.39-1489. This indicates that all $YFeO_3.850(1:n)$ samples crystallize in the orthorhombic structure (space group Pnma(62)) with deformed perovskite structure [11] and have the lattice parameters $a = 0.5588$ nm, $b = 0.7595$ nm, $c = 0.5274$ nm. The main diffraction peaks are at $2\theta = 25.9, 31.9, 33.1, 33.9, 47.3, 53.4$ and $60.2°$ which correspond to the crystal planes of (111), (200), (121), (002), (040), (311) and (042), respectively. The intensity of diffraction peaks is relatively high, indicating that the obtained materials have high crystallinity. The position and intensity of all diffraction peaks do not change much, suggesting that the molar ratio of metal ions to citric acid hardly changes the crystal structure of the synthetic materials.

Table 3.1. Variation of average crystallite size for $YFeO_3.850(1:n)$ samples

Sample	$YFeO_3.850(1:n)$			
	1:1	1:2	1:3	1:4
Average crystallite size (nm)	19.322	17.058	16.657	17.148

The average crystallite size of $YFeO_3$ samples was calculated by applying the Debye - Scherrer formula to the max intensity peak (121), which was fitted with the OriginPro 8.5.1 software to identify the peak position and the full-width-at-half-maximum. The results obtained in Table 3.1 show that all samples have small crystallite size (< 20 nm), in which the sample corresponding to a molar ratio of 1:1 between the number of metal ions and citric acid has the largest crystallite size.

The FT-IR spectra of $YFeO_3$ samples after annealing at 850 °C for 5 hours shown in Figure 3.2 consist of four main absorption regions: the first one corresponds to a broad absorption band at 3700 - 3200 cm^{-1} due to the O-H stretching of the free moisture in the air [12]; the second region corresponds to peaks at 2353 cm^{-1} due to the oscillation of the C = O group of CO_2 present in

the air [13]; the third region corresponds to a wide absorption band at 1572 - 1385 cm^{-1} which may be attributed to the stretching of C-O bonds [14]; the fourth region corresponds to the characteristic absorption band at 568 cm^{-1} due to the Fe-O stretching mode of the ortho-ferrite systems [14,15]. When changing the molar ratio of metal ions to citric acid, there are no additional peaks on the FT-IR spectra, indicating that the change in the molar ratio of metal ions and citric acid hardly affects the lattice crystal structure of obtained YFeO$_3$ samples. This is in a good agreement with the XRD analysis.
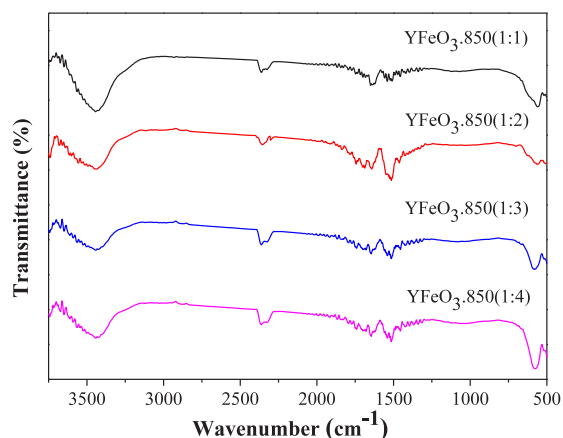
Figure 3.2. FT-IR spectra of YFeO$_3$.850(1:n) samples

3.1.2. Effect of the molar ratio of metal ions to citric acid on the surface morphology

SEM has been used to investigate the effect of the molar ratio of metal ions to citric acid on the surface morphology of YFeO$_3$ samples. The SEM images of the YFeO$_3$ samples synthesized by the sol-gel method and annealed at 850 ^{o}C for 5 hours with different molar ratios of metal ions to citric acid in Figure 3.3 show that the particle size of all samples was at the nanoscale. Moreover, for the samples with the molar ratios of 1:3 and 1:4 between metal ions and citric acid, the particles are more uniform and have less cohesion than samples with the molar ratios of 1:1 and 1:2 between metal ions and citric acid. It is worth noting that the crystallite size calculated by XRD is the smallest – mostly like a single crystal in powder form. Since particle may be present as a single crystal or an agglomeration of several crystals, the particle size determined by SEM is always larger than the crystallite size.
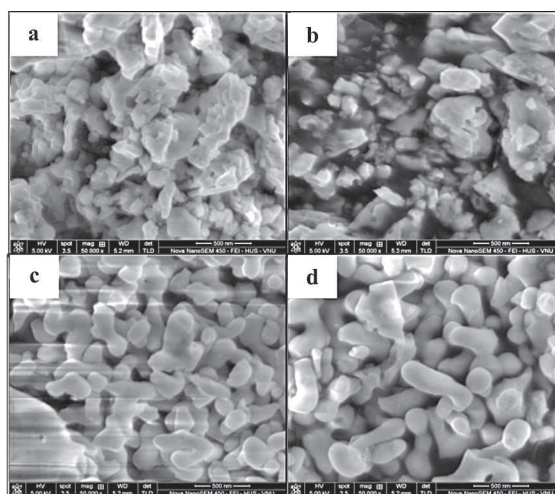
Figure 3.3. SEM images of YFeO$_3$.850(1:1) (a); YFeO$_3$.850(1:2) (b); YFeO$_3$.850(1:3) (c); YFeO$_3$.850(1:4) (d)

3.1.3. Effect of the molar ratio of metal ions to citric acid on the methylene blue photodegradation efficiency

In order to study the effect of the molar ratio of metal ions to citric acid on the MB photodegradation efficiency, the C/C$_o$ versus the irradiation time is plotted and shown in Figure 3.4. In which, C$_o$ is the concentration of MB solution after stirring 2 hours in the dark and C is the concentration of MB solution after irradiation time t (h). The results show that the sample corresponding to the molar ratio of 1:2 between metal ions and citric acid had a faster decrease in the concentration of MB than other samples synthesized under the same conditions but with different molar ratios of the metal ions to citric acid.
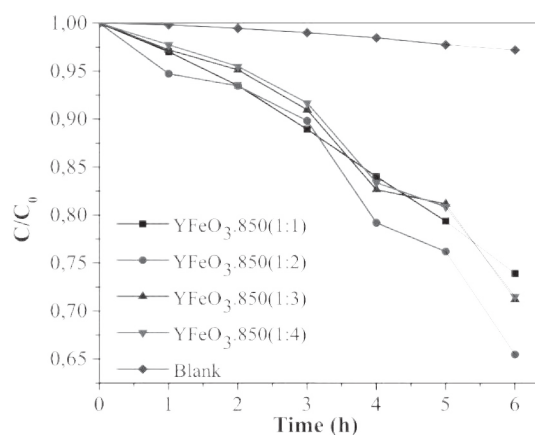

Figure 3.4. The C/C$_0$ of MB as a function of the irradiation time for the blank and YFeO$_3$.850(1:n) samples

Table 3.2 shows the MB photodegradation efficiency with and without the presence of $YFeO_3.850(1:n)$ samples after 6 hours of irradiation. In general, all samples have relatively low photodegradation efficiency. After 6 hours of irradiation, $YFeO_3.850(1:2)$ sample has the highest decomposition efficiency (34.53%). Meanwhile, MB decomposition efficiency for $YFeO_3.850(1:1)$, $YFeO_3.850(1:3)$ and $YFeO_3.850(1:4)$ samples are 26.09, 28.81 and 28.50%, respectively. Besides, for the blank sample, MB photodegradation efficiency is only 2.79%, suggesting that without the presence of $YFeO_3$ MB was hardly decomposed under irradiation conditions. Therefore, it can be concluded that the presence of $YFeO_3$ photocatalysts plays a vital role in the photodegradation of MB.

Table 3.2. MB photodegradation efficiency (%) for blank and $YFeO_3.850(1:n)$ after 6 hours of irradiation

Sample	Photodegradation efficiency (%)
$YFeO_3.850(1:1)$	26.09
$YFeO_3.850(1:2)$	34.53
$YFeO_3.850(1:3)$	28.81
$YFeO_3.850(1:4)$	28.50
Blank	2.79

3.2. Effect of the annealing temperature on the lattice crystal structure, band gap energy and methylene blue photodegradation efficiency of $YFeO_3$

3.2.1. Effect of the annealing temperature on the lattice crystal structure of $YFeO_3$
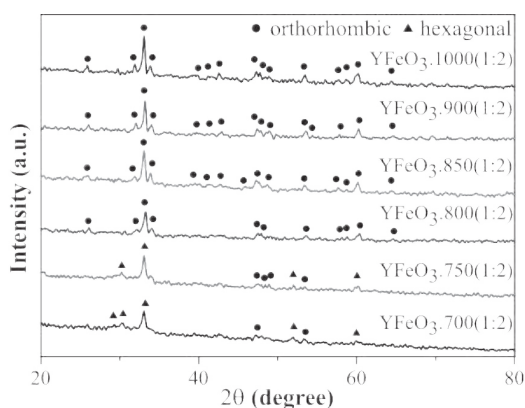
Figure 3.4. XRD patterns collected at room temperature for $YFeO_3.T(1:2)$ samples

The XRD patterns measured at room temperature for $YFeO_3.T(1:2)$ samples annealed at different annealing temperatures (700 - 1000°C) are shown in Figure 3.4. The experimental results show that samples annealed below 800 °C have the presence of both orthorhombic and hexagonal phases. However, when sintered at higher temperatures (800 - 1000 °C), the samples obtained have an orthorhombic single-phase structure. This can be explained by the fact that at high temperature the hexagonal structure phase (space group (P63/mmc)) will change to the orthorhombic phase that is thermodynamically stable.[16] Moreover, the diffraction peak corresponding to the highest intensity at $2\theta \approx 33.1°$ becomes sharper and has a higher intensity for increasing annealing temperature, indicating that the obtained samples have a better crystallinity and a larger crystallite size.

Table 3.3 presents the average crystallite size of $YFeO_3$ samples annealed at different temperatures, which was calculated by applying the Debye - Scherrer formula to the max intensity peak (121) at $2\theta \approx 33$. In general, the average crystallite size of all samples increases with increasing annealing temperature, especially for the sample annealed at 1000 °C.

Table 3.3. Variation of average crystallite size for $YFeO_3.T(1:2)$ samples

Sample	Average crystallite size (nm)
$YFeO_3.700(1:2)$	12.850
$YFeO_3.750(1:2)$	16.524
$YFeO_3.800(1:2)$	15.436
$YFeO_3.850(1:2)$	17.058
$YFeO_3.900(1:2)$	16.627
$YFeO_3.1000(1:2)$	24.128

FT-IR spectra of $YFeO_3.T(1:2)$ samples shown in Figure 3.5 include 4 main absorption regions like $YFeO_3$ samples sintered at 850°C as described in 3.1.1. However, the peak intensity absorbed at 568 cm^{-1} due to the asymmetric elongation of the Fe-O-Fe bond in the FeO_6 octahedron increases for increasing

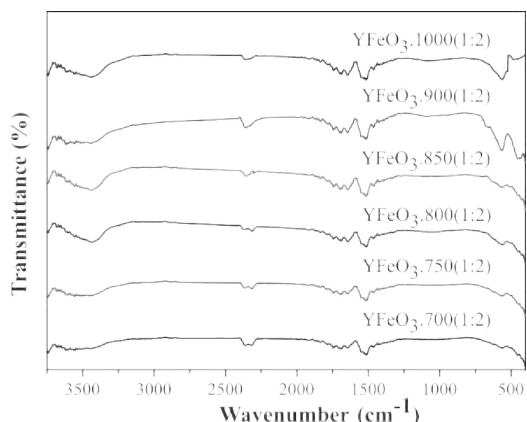the annealing temperature. This proves that the purity of the YFeO₃.T(1:2) samples is enhanced as the annealing temperature increases.

Figure 3.5. FT-IR spectra of $YFeO_3.T(1:2)$ annealed at different temperatures

3.2.2. Effect of the annealing temperature on the band gap energy

UV-Vis DRS was used to investigate the effect of the annealing temperature on the band gap energy of obtained $YFeO_3.T(1:2)$ samples. UV-Vis DRS spectra of these samples are shown in Figure 3.6.

The experimental results show that $YFeO_3.T(1:2)$ samples annealed at different temperatures are active in the visible light region and the absorption band extends from the ultraviolet to the visible. On the other hand, when the annealing temperature increases, the light absorption region of the material shifts to the shorter wavelength region. This is in good agreement with the change in the color of obtained samples as a function of the annealing temperature.
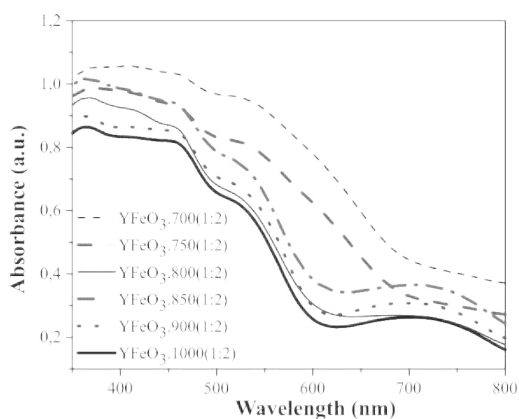
Figure 3.6. UV-Vis diffuse absorption spectra of the $YFeO_3.T(1:2)$ samples

Since $YFeO_3$ is an indirect band gap semiconductor, the Kubelka-Munk function $[F(R)h\nu]^2$ versus the phonon energy was plotted to estimate the band gap energy of all samples. The results presented in Figure 3.7 and Table 3.4 show that the band gap energy increases in the annealing temperature range of 700 - 800 ºC and hardly changes as further increasing the annealing temperature to 1000 ºC.

Table 3.4. Variation in the band gap energy for $YFeO_3.T(1:2)$ samples

Sample	E_g (eV)
YFeO₃.700(1:2)	2.25
YFeO₃.750(1:2)	2.36
YFeO₃.800(1:2)	2.48
YFeO₃.850(1:2)	2.43
YFeO₃.900(1:2)	2.43
YFeO₃.1000(1:2)	2.45
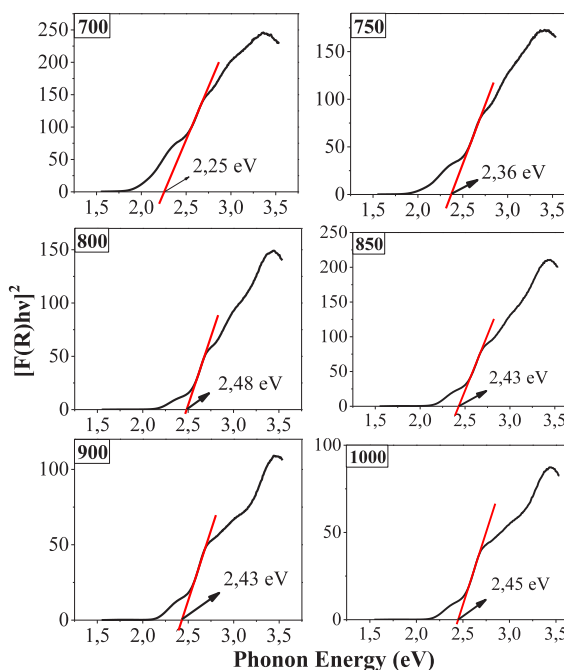
Figure 3.7. Plots of $[F(R)h\nu]^2$ versus phonon energy for $YFeO_3.T(1:2)$ samples

3.2.3. Effect of the annealing temperature on the methylene blue photodegradation efficiency

Table 3.5. MB photodegradation efficiency (%) of YFeO$_3$ samples after a period of six-hour irradiation

Sample	Photodegradation efficiency (%)
YFeO$_3$.700(1:2)	36.94
YFeO$_3$.750(1:2)	27.40
YFeO$_3$.800(1:2)	28.50
YFeO$_3$.850(1:2)	34.53
YFeO$_3$.900(1:2)	48.99
YFeO$_3$.1000(1:2)	35.33
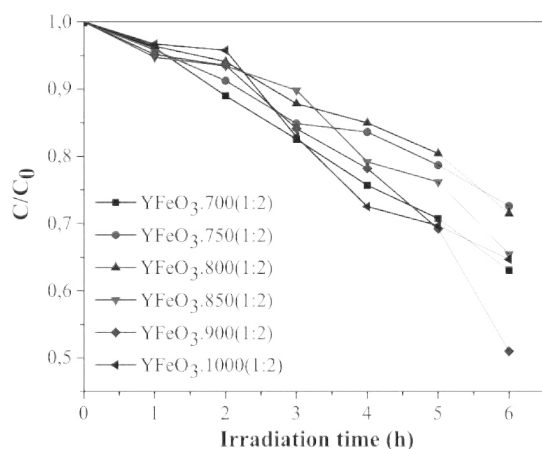
Figure 3.8. Kinetic of the photocatalytic degradation of MB for YFeO$_3$.T(1:2) samples under visible light irradiation

The photocatalytic activity survey results presented in Table 3.5 indicate that the YFeO$_3$ sample synthesized with a molar ratio of 1:2 metal ions/citric acid and annealed at 900 ºC has the highest MB photodegradation efficiency (48.99% after 6 hours of irradiation). Meanwhile, MB decomposition efficiency for YFeO$_3$ samples synthesized under the same conditions but annealed at 700, 750, 800, 850 and 1000ºC was 36.94, 27.40, 28.50, 34.53 and 35.33%, respectively. This may be because the sample annealed at 900 ºC has higher crystallinity than those annealed at 700 - 850 ºC and has a smaller particle size than the samples annealed at 1000ºC.

3.3. Kinetic of photocatalytic reactions for YFeO$_3$

Table 3.6. Data obtained from the YFeO$_3$ samples when evaluating the kinetics of the photocatalytic reactions by employing the Langmuir-Hinshelwood model

| Sample	Kinetic equation	K$_{app}$ (h^{-1})	R^2
YFeO$_3$.850(1:1)	y = 0.05055x − 0.01988	0.05055	0.97979
YFeO$_3$.850(1:2)	y = 0.06685x − 0.03514	0.06685	0.91017
YFeO$_3$.850(1:3)	y = 0.05432x − 0.03275	0.05432	0.91637
YFeO$_3$.850(1:4)	y = 0.05434x − 0.03652	0.05434	0.91582
YFeO$_3$.700(1:2)	y = 0.0771x − 0.01988	0.0771	0.98619
YFeO$_3$.750(1:2)	y = 0.05149x − 0.00636	0.05149	0.98106
YFeO$_3$.800(1:2)	y = 0.05252x − 0.02283	0.05252	0.94071
YFeO$_3$.900(1:2)	y = 0.10124x − 0.07855	0.10124	0.84127
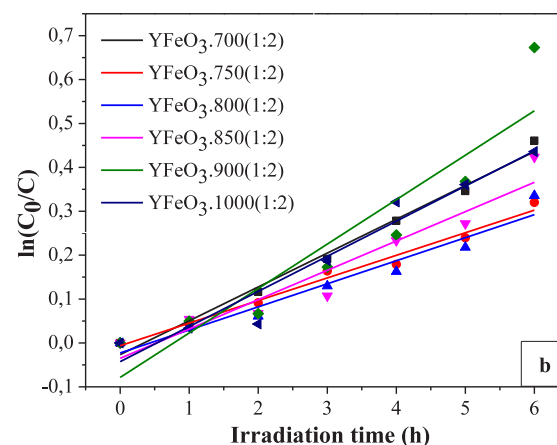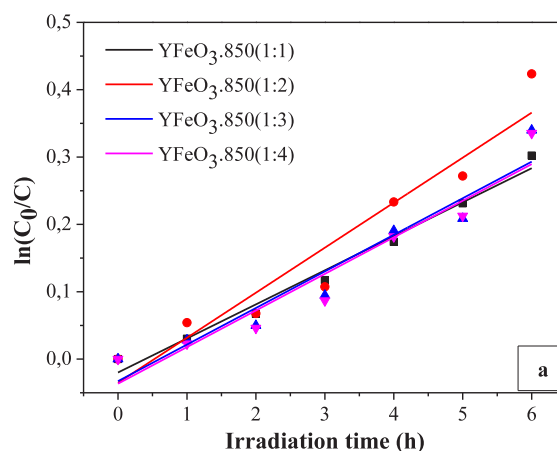
Figure 3.9. The plots of ln(C/C$_o$) versus irradiation time for the photocatalytic degradation of MB of YFeO$_3$.850(1:n) and YFeO$_3$.T(1:2) samples

The plots of $\ln(C_0/C)$ versus the irradiation time shown in Figure 3.9 is linear, revealing that the photocatalytic reactions of $YFeO_3$ materials for MB photodegradation follow the Langmuir-Hinhshelwood kinematic model with the high coefficient of determination.

The analysis results of MB photodegradation kinetics for the $YFeO_3$ samples presented in Table 3.9 show that the $YFeO_3$.900(1:2) sample shows the highest photodegradation rate (k = 0.10124 h^{-1}), which is 1.5 times higher than that of $YFeO_3$.850(1:2) sample.

4. CONCLUSION

In this paper, we systematically investigated the effect of both the molar ratio of metal ions to citric acid and the annealing temperature on structure and photocatalytic properties of $YFeO_3$ samples that were fabricated by the sol-gel method. XRD analysis indicates that all $YFeO_3$ samples obtained mainly crystallize in the orthorhombic structure (space group Pnma (62)) with deformed perovskite structure. All $YFeO_3$ photocatalysts obtained were found to exhibit photocatalytic activities in the visible light region. In particular, the $YFeO_3$ sample annealed at 900 °C with a molar ratio of 1: 2 between metal ions and citric acid has the highest MB catalytic decomposition efficiency compared to $YFeO_3$ samples fabricated under the same condition but annealed at other temperatures.

ACKNOWLEDGEMENTS

REFERENCES

1. P. Rajasulochana and V. Preethy. Comparison on efficiency of various techniques in treatment of waste and sewage water – A comprehensive review, *Resource - Efficient Technologies,* **2016,** *2,* 175–184.

2. Suresh C. Ameta. Advanced Oxidation Processes for Wastewater Treatment, *Advanced Oxidation Processes for Wastewater Treatment: Emerging Green Chemical Technology*, **2018,** 1–12.

3. J. Zhang, Q. Xu, Z. Feng, M. Li, and C. Li. Importance of the Relationship between Surface Phases and Photocatalytic Activity of TiO_2, *Angewandte Chemie - International Edition*, **2008,** *47*(9), 1766–1769.

4. Q. Guo, C. Zhou, Z. Ma, and X. Yang. Fundamentals of TiO_2 Photocatalysis: Concepts, Mechanisms, and Challenges, *Advanced Materials,* **2019,** *31,* 1–26.

5. M. Pawar, S. T. Sendoğdular, P. Gouma. A brief overview of TiO_2 photocatalyst for organic dye remediation: Case study of reaction mechanisms involved in Ce-TiO_2 photocatalysts system, *Journal of Nanomaterials,* **2018,** *2018,* 1–13.

6. L. Zhao, X. Chen, X. Wang, Y. Zhang, W. Wei, Y. Sun, M. Antonietti, and M. M. Titirici.One-step solvothermal synthesis of a carbon@TiO_2 dyade structure effectively promoting visible-light photocatalysis, *Advanced Materials,* **2010,** *22*(30), 3317–3321.

7. S. C. Li and U. Diebold. Reactivity of TiO_2 rutile and anatase surfaces toward nitroaromatics, *Journal of the American Chemical Society,* **2010,** *132*(1), 64–66.

8. S. Sun, W. Wang, H. Xu, L. Zhou, M. Shang, and L. Zhang. $Bi_5FeTi_3O_{15}$ Hierarchical Microflowers: Hydrothermal Synthesis, Growth Mechanism, and Associated Visible-Light-Driven Photocatalysis, *Journal of Physical Chemistry C*, **2008,** *112,* 17835–17843.

9. X. Lu, J. Xie, H. Shua, J. Liu, C. Yin, J. Lin. Microwave-assisted synthesis of nanocrystalline $YFeO_3$ and study of its photoactivity, *Materials Science and Engineering B: Solid-State Materials for Advanced Technology*, **2017,** *138*(3), 289–292.

10. S. Mathur, M. Veith, R. Rapalaviciute, H. Shen, G. F. Goya, W. L. M. Filho, and T.S. Berquo. Molecule Derived Synthesis of Nanocrystalline

YFeO$_3$ and Investigations on Its Weak Ferromagnetic Behavior, *Nanoengineered Nanofibrous Materials,* **2004**, *169*(16), 425–442.

11. D. du Boulay, E. N. Maslen, V. A. Streltsov & N. Ishizawa. A synchrotron X-ray study of the electron density in YFeO$_3$, *Acta Crystallographica Section B Structural Science,* **1995**, *51*(6), 921–929.

12. M. C. Navarroa, E. V. Pannunzio-Minerb, S. Pagolab, M. InesGomeza, R. E. Carboniob. Structural refinement of Nd[Fe(CN)$_6$].4H$_2$O and study of NdFeO$_3$ obtained by its oxidative thermal decomposition at very low temperatures, *Journal Solid State Chem,* **2005**, *178*(3), 847–854.

13. A. Gatelytė, D. Jasaitis, A. Beganskienė, A. Kareiva. Sol-gel synthesis and characterization of selected transition metal nano-ferrites, *Medziagotyra,* **2011**, *17*(3), 302–307.

14. N. T. Thuy, D. L Minh. Size Effect on the Structural and Magnetic Properties of Nanosized Perovskite LaFeO$_3$ Prepared by Different Methods, *Advances in Materials Science and Engineering,* **2012**, *2012*(3), 1–6.

15. Q. Lin, J. Xu, F. Yang, X. Yang and Y. He. The influence of Ca substitution on LaFeO$_3$ nanoparticles in terms of structural and magnetic properties, *Journal of Applied Biomaterials and Functional Materials,* **2018**, *16*(1S), 17–25.

16. J. Li, U. G. Singh, T. D. Schladt, J. K. Stalick, S. L. Scott, and R. Seshadri. Hexagonal YFe$_{1-x}$Pd$_x$O$_{3-\delta}$: Nonperovskite host compounds for Pd^{2+} and their catalytic activity for CO oxidation, *Chemistry of Materials*, **2008**, *20*(20), 6567–6576.

Nghiên cứu lý thuyết về cấu trúc, độ bền và tương tác của các phân tử ampicillin và amoxicillin trên bề mặt anatase-TiO$_2$ (101)

Nguyễn Ngọc Trí, Nguyễn Tiến Trung*

Phòng thí nghiệm Hóa học tính toán và Mô phỏng,
Khoa Khoa học tự nhiên, Trường Đại học Quy Nhơn, Việt Nam

TÓM TẮT

Sử dụng các tính toán phiếm hàm mật độ, chúng tôi đã khảo sát sự hấp phụ các phân tử ampicillin và amoxicillin trên bề mặt anatase-TiO$_2$ (101). Mô hình tuần hoàn và sóng phẳng đã được áp dụng trong tất cả các tính toán. Kết quả chỉ ra rằng các quá trình này được đánh giá là các quá trình hấp phụ hóa học với năng lượng khoảng -31 kcal.mol^{-1}. Độ bền của các cấu hình được đóng góp đáng kể bởi các tương tác hút tĩnh điện Ti···O và các liên kết hydro kiểu O/N/C-H···O. Sự tồn tại và vai trò của các tương tác bề mặt trong việc làm bền phức được phân tích chi tiết dựa vào cách tiếp cận hóa học lượng tử. Kết quả cho thấy khả năng hấp phụ của ampicillin trên bề mặt anatase mạnh hơn so với amoxicillin. Đáng chú ý, sự xắp xếp theo chiều ngang với diện tích bề mặt tiếp xúc lớn được diễn ra ưu tiên trong việc kết dính các phân tử kháng sinh này lên bề mặt anatase.

Từ khóa: *Anatase, kháng sinh, tính toán DFT, bề mặt vật liệu.*

* *Tác giả liên hệ chính.*
Email: *nguyentientrung@qnu.edu.vn*

Theoretical study of geometry, stability and interaction in configurations of ampicillin and amoxicillin molecules on the surface of anatase-TiO$_2$ (101)

Nguyen Ngoc Tri, Nguyen Tien Trung[*]

Laboratory of Computational Chemistry and Modelling, Faculty of Natural Sciences,
Quy Nhon University, Vietnam

ABSTRACT

By using density functional theory calculations, we examined the adsorption of ampicillin (AP) and amoxicillin (AX) molecules on anatase-TiO$_2$ (101) surface (a-TiO$_2$). The periodic slab model and plane wave method were applied in all calculations. Results indicate that these processes are evaluated as chemical adsorptions with associated energies of -31 kcal.mol^{-1}. Stability of configurations is significantly contributed by Ti···O electrostatic interactions and O/N/C-H ··· O hydrogen bonds. Existence and role of surface interactions in complex stabilization are analyzed in detail based on quantum chemical approaches. As a consequence, the adsorption capacity of AP on the a-TiO$_2$ is slightly stronger than that of AX. Remarkably, the horizontal arrangement along with large surface area occurred favorably in adhesion of these antibiotics on the a-TiO$_2$.

Keywords: *Anatase, antibiotics, DFT calculations, material surface.*

1. INTRODUCTION

In many decades, TiO$_2$ has been of the important semiconductor materials applied in fields such as energy, health, food technologies.[1,2] TiO$_2$ was used extensively in photocatalytic, adsorption, decomposition of pollutants due to its considerable surface properties. The investigations on surface properties of TiO$_2$ using experimental technologies and theoretical simulations were interested by scientists.[3] TiO$_2$ existed in three common phases including anatase, rutile and brookite.[2-4] Among phases of TiO$_2$, the anatase has high photocatalytic activity and is observed widely. Furthermore, the (101) surface of anatase is the most stable one and frequently form.[4] The adsorption of organic molecules on TiO$_2$ surfaces has been previously reported.[3-7] As a consequence, TiO$_2$ can be used as potential material for efficient adsorption of organic compounds.

In addition, the nature of intermolecular interactions on TiO$_2$ surfaces and their role to the stability of complexes have not been paid much attention by scientists.[8,9] This understanding is of importance for further evaluation of molecules adsorption, especially for photocalytic reactions occurred onto TiO$_2$ surfaces in order to form simple products such as CO$_2$, H$_2$O.[2,3,5,7] Besides, interactions between antibiotics used widely such as ampicillin (AP), amoxicillin (AX) and material surfaces have not been much paid attention. Our previous investigations indicated that the adsorption of these molecules on vermiculite or adsorption of enrofloxacine molecule on rutile-

TiO$_2$ are strong chemisorptions.[10,11] However, the insight into adsorptive interactions of antibiotics on anatase-TiO$_2$ surfaces has been still not mentioned fully yet. Noticeably, quantum chemistry computations are emerging as useful tools in understanding surface phenomena in nature.[12] Hence, in the present work, we use DFT caculations to investigate the adsorptive interaction of AP, AX molecules on anatase-TiO$_2$ (101) surface. Moreover, this work is performed with an aim to have a convenient evaluation on adsorption capacity of AP, AX on an anatase surface and to also serve in finding efficient material for removal of pollutants.

2. COMPUTATIONAL METHODS

The structures of anatase surface and molecules and configurations are optimized by using the VASP program.[13] The (101) surface of anatase is designed with 4 x 5 x 2 supercell of primitive cell which has been characterized by following dimension parameters: a = 15.22 Å; b = 20.87 Å; c = 30.01 Å. The cutoff energy and grid k-points are considered at 500 eV and 4 x 3 x 1 Gamma points, respectively. The atomic positions are optimized until a force value is smaller than 0.01 eV. The energy convergence is set up at 1.10^{-6} eV. The Perdew-Burke-Ernzernhof (PBE) functional in conjugation with generalized gradient approximation (GGA) is used in all computations.[14] The adsorption energy (E$_{ads}$) is calculated by expression: E$_{ads}$ = E$_{comp}$ - E$_{surf}$ - E$_{mole}$ where E$_{comp}$, E$_{surf}$ and E$_{mole}$ are energy values of optimized configurations, surface and molecules, respectively.

In order to have an insight into the existence and role of the intermolecular interactions upon the adsorption process, quantum chemistry analyses based on Atoms In Molecular (AIM) and Natural Bonds Orbitals (NBO) approaches are carried out at the B3LYP/6-31+G(d,p) level for first-layered structures of the most configurations. The transfers of total electron density (EDT) following interactions formation are also considered for these structures. These calculations are performed by using Gaussian 09, AIM2000 and NBO 5.G programs.[15-17]

3. RESULTS AND DISCUSSION

3.1. Stable structures

The optimized geometries of anatase-TiO$_2$ (101) surface (denoted by a-TiO$_2$) and AP, AX molecules are displayed in Figure 1. Some characterized parameters including bond length, angle are compared to experimental data and theoretical results in previous studies.[3-7] As a matter of fact, the calculated parameters in this work are in a good agreement with previous reports in which the changes of bond lengths and angles are *ca.* 0.01-0.05 Å and 1-3°, respectively. The ground states of configurations for adsorption of AP and AX molecules onto a-TiO$_2$ are illustrated in Figure 2. These structures are denoted by **APi** and **AXi** (i = 1-4), following different arrangements of AP, AX molecules on surface to have a convenient look. The **AX4** is found out only for AX system due to interaction of –OH group attached to benzene ring as compared to AP system. As shown in Figure 2, the stable configurations obtained for two systems have similar geometrical structures, particularly for **AP1** and **AX1**, **AP2** and **AX2**, **AP3** and **AX3**. Besides, the distances of intermolecular contacts and changes of bond lengths, angles are given in Table 1 upon complexation.
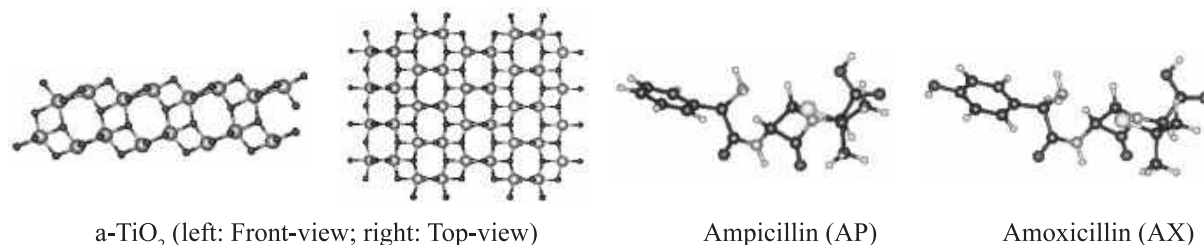

a-TiO$_2$ (left: Front-view; right: Top-view)   Ampicillin (AP)   Amoxicillin (AX)

Figure 1. Optimized structures of a-TiO$_2$ and AP, AX molecules

AP-1   AP-2   AP-3
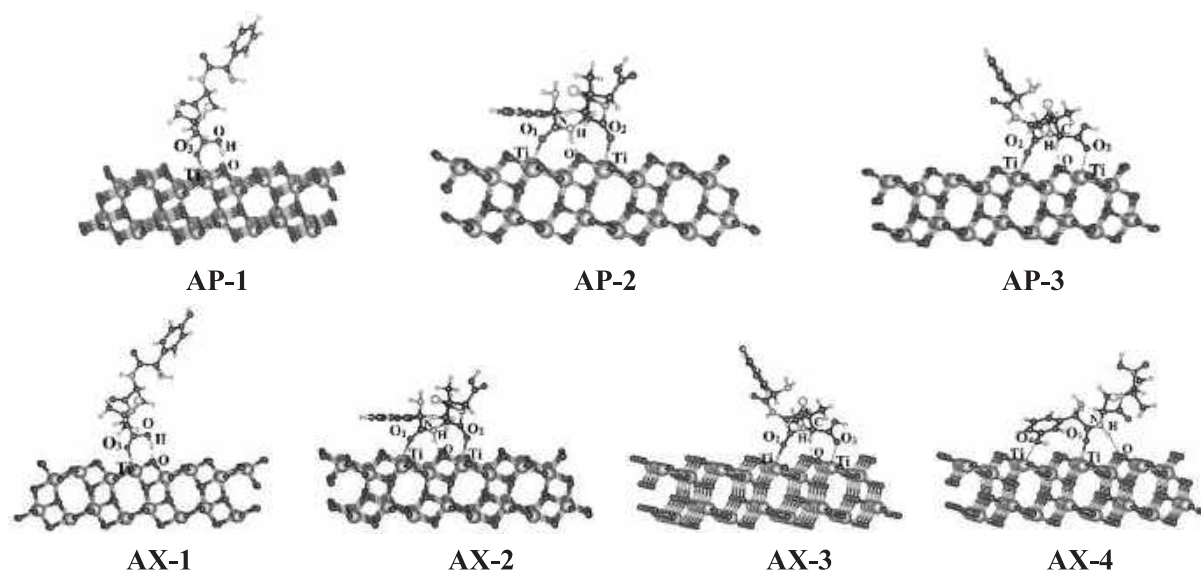
AX-1   AX-2   AX-3   AX-4

Figure 2. The optimized configurations of AP, AX adsorbed on a-TiO_2 (101) *(C, H, O, N, S, Ti atoms are displayed by grey, white, red, small cyan, yellow, cyan colors, respectively)*

Table 1. Some characteristic parameters of the stable configurations

	d(Ti···O)	d(H···O)	a(CO···Ti)	a(O/N/C-H···O)	Δr(C=O)	Δr(X-H)	Δr(Ti-O)**
AP1	2.16³	1.49	131.7	164.1	0.04	0.07	0.02 – 0.06
AP2	2.13¹/2.28²	1.64	153.3¹/135.0	161.9	0.04/0.02	0.03	0.01 – 0.03
AP3	2.22²/2.39³	1.86	163.2/130.8³	143.8	0.01/0.01	0.00	-0.02 – 0.01
AX1	2.19³	1.52	131.9	164.2	0.04	0.06	0.00 – 0.07
AX2	2.11¹/2.30²	1.66	153.6¹/135.5	161.7	0.03/0.02	0.03	-0.00 – 0.04
AX3	2.26²/2.42³	1.89	163.8/130.3³	142.1	0.01/0.01	0.00	-0.03 – 0.02
AX4	2.16¹/2.43⁴	2.28/2.26*	151.1/125.9⁴	160.0/160.2*	0.02/0.02⁴	0.01/0.01*	0.02 – 0.03

1,2,3,4-for O atoms as shown in Fig.2; *-for O-H···O; **-for O atoms at bridging sites; a-angle (°); d-distance (Å)

As presented in Table 1, the distances of Ti···O and O/N/C-H···O contacts are in range of 2.11 to 2.43 Å and 1.49 to 2.28 Å, respectively. These values are smaller than the total of van der Waals radii of atoms involved in contacts including Ti and O (3.82 Å), O and H (2.72 Å), C and O (3.22 Å) atoms. Besides, the angles of C-O···Ti and O/N/C-H···O are corresponding to values of 125.9-163.8° and 142.1-164.2°, consistent with binding angles of C=O interactions or hydrogen bonds.[18] Therefore, it can be suggested that the Ti···O and O/N/C-H···O intermolecular interactions are formed following complexation. Also, the geometrical changes for surface and molecules are quite small in comparison with their initial structures. The change of Ti-O, C=O, X-H bonds is *ca.* 0.07 Å, 0.04 Å and 0.07 Å, respectively. The AP, AX molecules tend to arrange in horizontal sequence onto a-TiO_2 to form various intermolecular contacts, similarly to previous studies.[10,11]

3.2. Adsorption energy

In order to evaluate the strength of interactions and the capacity of adsorption of AP, AX on a-TiO_2, the adsorption energies are considered and tabulated in Table 2.

Table 2. Adsorption energy (E_{ads}, kcal.mol^{-1}) of stable configurations

i	1	2	3	4
AP-i	-22.9	-31.1	-18.6	
AX-i	-21.0	-29.3	-17.7	-25.5

The calculated energy values in Table 2 for AP and AX systems range from -17.7 to -31.1 kcal.mol^{-1}. The strength of adhesion of AP, AX on a-TiO$_2$ is comparable to that of enrofloxacin on rutile-TiO$_2$ (110) surface.[10] Accordingly, these processes are evaluated as chemical adsorptions.[19] Moreover, the ability of adsorption of these molecules on TiO$_2$ surfaces is twice weaker than that on vermiculite[11] because of the smaller adsorption energy. This is understood by the significant contributions of electrostatic interactions between Mg^{2+} cation sites on vermiculite and O atoms in >C=O groups of molecules.[11]

In addition, the stability of configurations in this work is contributed by both Ti···O attractive interactions and O/N/C-H···O hydrogen bonds as displayed in Figure 2. The contributions of two Ti···O intermolecular contacts of >C=O groups and one N-H···O hydrogen bond lead to the considerable stability of **AP2**, **AX2** in comparison to the rest of configurations for AP, AX systems. In case of **AX4**, one Ti···O interaction is formed by Ti$_{5f}$ site and O in –OH group. As given in Table 2, **AX4** is slightly less stable than **AX2** of 4 kcal.mol^{-1}. This result is due to the proton affinity (PA) at B3LYP/6-31++G(d,p) level and charge density at O atoms in C=O group (PAs are in range of 200-216 kcal.mol^{-1}) are higher than that in –OH group (PA is *ca.* 185 kcal.mol^{-1})[11], leading to the fact that Ti···O (C=O) becomes more stable than Ti···O (OH). Besides,

the difference of E_{ads} values for **AP2** and **AX2** is small *ca.* 2 kcal.mol^{-1}. The approximation of PA and deprotonation enthalpy (DPE) values at O atoms and N-H bond in AP and AX molecules as analyzed in previous report makes the equally stable interactions in these configurations. In general, the adsorption capacity of AP and AX onto a-TiO$_2$ is slightly different.

On the other hand, one Ti···O and one O-H···O contacts (**AP1**, **AX1**) or two Ti···O and one C-H···O contacts (**AP3**, **AX3**) have an important role to their stability. The adsorption energy increases in the order of **AP2** < **AP1** < **AP3** and **AX2** < **AX4** < **AX1** < **AX3** for AP, AX systems, respectively. This is understood by basing on the ability of forming stable hydrogen bonds in configurations. The strength of hydrogen bonds increases in going from C-H···O to N-H···O and finally to O-H···O as estimated in *ref.11*. Following this report, the significant contribution of O-H···O hydrogen bonds leads to the more stable configurations of **AP1** and **AX1** as compared to **AP3**, **AX3**, respectively. In summary, the stability of obtained configurations depends on the arrangement of molecules on surface to form stable interactions, and resulted by significant contribution of Ti···O electrostatic interactions and addition of O/N-H···O hydrogen bonds.

3.3. AIM and NBO analyses

To gain an insight into existence and role of adhesive interactions, we perform topology analysis for the first layer of **AP2**, **AX2** most stable structures as displayed in Figure 3. The electron density transfers between molecules and surface are investigated further to confirm the existence of intermolecular contacts. Some characteristic parameters are gathered in Table 3.
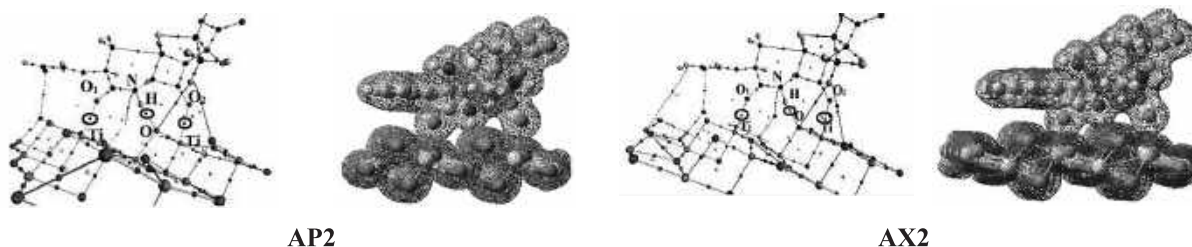

AP2   AX2

Figure 3. The topological geometries of the first-layered structures for AP2 and AX2

Table 3. The topological analysis ($\rho(r)$, electron density; $\nabla^2(\rho(r))$, Laplacian of electron density; $H(r)$, total of electron density energy; all in au) and electron density transfers (EDT, e), hyper-conjugation energy (E^2, kcal.mol^{-1}) for the first-layered structures

	BCPs	$\rho(r)$	$\nabla^2(\rho(r))$	$H(r)$	EDT	Transfers	E^2
	$O_1\cdots Ti$	0.056	0.325	0.004		$\pi(C=O)/LP(O) \rightarrow LP^*(Ti)/\sigma^*(Ti-O)$	23.6
AP-2	$O_2\cdots Ti$	0.039	0.204	0.004	0.116	$\pi(C=O)/LP(O) \rightarrow LP^*(Ti)/\sigma^*(Ti-O)$	16.9
	$N-H\cdots O$	0.052	0.153	-0.008		$LP(O) \rightarrow \sigma^*(N-H)$	22.6
	$O_1\cdots Ti$	0.058	0.340	0.003		$\pi(C=O)/LP(O) \rightarrow LP^*(Ti)/\sigma^*(Ti-O)$	22.6
AX-2	$O_2\cdots Ti$	0.037	0.194	0.004	0.102	$\pi(C=O)/LP(O) \rightarrow LP^*(Ti)/\sigma^*(Ti-O)$	17.0
	$N-H\cdots O$	0.050	0.147	-0.007		$LP(O) \rightarrow \sigma^*(N-H)$	19.3

The calculated results imply that the Ti⋯O and N-H⋯O intermolecular contacts exist. The electron density ($\rho(r)$) and Laplacian of electron density ($\nabla^2(\rho(r))$) values at Ti⋯O BCPs range from 0.037 to 0.058 au and from 0.194 to 0.340 au, respectively. Similarly, the $\rho(r)$ and $\nabla^2(\rho(r))$ at N-H⋯O BCPs correspond to 0.050-0.052 au and 0.147-0.153 au. All these values are within the region of $\rho(r)$ and $\nabla^2(\rho(r))$ for noncovalent interactions.[16] Therefore, the Ti⋯O and N-H⋯O are regarded as noncovalent interactions and hydrogen bonds, respectively. The H(r) values at BCPs of N-H⋯O contacts are slightly negative, indicating that these interactions have a small part of covalency in nature. Besides, the stability of AP2, AX2 is additionally contributed by O⋯C weak interactions with the small $\rho(r)$ of 0.01 au. The $\rho(r)$ values at Ti⋯O and N-H⋯O BCPs in AP2 and AX2 are mostly approximate, therefore, strength of interactions in these configurations is nearly equal. As a result, the stability of AP2 and AX2 is considered to be approximate.

Furthermore, the formation of adsorptive interactions is clarified by electron density transfers from molecules to surface and vice versa. As given in Table 3, the EDT is small positive, of 0.1 e due to the prominent transfers from molecules to surface. The existence of Ti⋯O contacts is confirmed by the transfers of electron density from lone pair of O atoms (LP(O)) and bonding orbitals of C=O (π(C=O)) in molecules to unoccupied lone pair of Ti sites (LP*(Ti)) and anti-bonding orbitals of Ti-O (σ^*(Ti-O)) of

a-TiO$_2$. Similarly, the N-H⋯O hydrogen bonds is formed by the electron density transfers from LP(O) of surface to σ^*(N-H) of molecules. These transfers of electron density are evaluated clearly by hyper-conjunction energies (E^2) (*cf. Table 3*). Accordingly, the E^2 values for formations of Ti⋯O and N-H⋯O interactions are in the range of 17-24 kcal.mol^{-1} and 19-23 kcal.mol^{-1}, respectively. Moreover, the stability of interactions is determined by electrostatic term based on charge densities at sites of molecules and surface. The high charge density at sites leads to forming the strong electrostatic interactions. Following NBO calculations, the charge densities at O, H atoms (in **AP, AX**) and Ti, O sites (in a-TiO$_2$) are in the ranges of -0.65 to -0.74 e; 0.47 to 0.48 e; 1.61 to 1.72 e; -0.97 to -0.98 e, respectively. Therefore, Ti⋯O interactions are stronger than N-H⋯O hydrogen bonds upon complexations. This result also leads to the fact that EDT values are positive values for **AP2** and **AX2**.

4. CONCLUSIONS

In the present work, we investigated the adsorption of ampicillin and amoxicillin molecules on a anatase-TiO$_2$ (101) surface (a-TiO$_2$) using density functional theory calculations. Obtained results show that the adhesion of these antibiotics on the a-TiO$_2$ surface is regarded as chemical adsorption. The stable configurations are contributed significantly by Ti⋯O electrostatic interactions along with O/N/C-H⋯O hydrogen

bonds. The existence and role of adsorptive interactions in the most stable configurations **AP2** and **AX2** are clarified on the basis of the electron density transfers and charge density distribution. Generally, the adsorption of ampicillin onto the a-TiO$_2$ surface is slightly stronger than that of amoxicillin, *ca.* 2 kcal.mol^{-1}. The arrangement of molecules on the a-TiO$_2$ surface tends to be preferable in a horizontal sequence and occurs on a large surface area to form stable interactions.

REFERENCES

1. C. Deiana, E. Fois, G. Martra, S. Narbey, F. Pellegrino and G. Tabacchi. On the Simple Complexity of Carbon Monoxide on Oxide Surfaces: Facet-Specific Donation and Backdonation Effects Revealed on TiO$_2$ Anatase Nanoparticles, *Chem. Phys. Chem.*, **2016**, *17*, 1-6.

2. F. D. Angelis, C. D. Valentin, S. Fantacci, A. Vittadini and A. Selloni. Theoretical Studies on Anatase and Less Common TiO$_2$ Phases: Bulk, Surfaces, and Nanomaterials, *Chem. Rev.*, **2014**, *114*, 9708-9753.

3. A. G. Thomas and K. L. Syres. Adsorption of organic molecules on rutile TiO$_2$ and anatase TiO$_2$ single crystal surfaces, *Chem. Soc. Rev.*, **2012**, *41*, 4207-4217.

4. G. Wu, C. Zhao, X. Zhou, J. Chen, Y. Li, Y. Chen. The interaction between HCHO and TiO$_2$ (101) surface without and with water and oxygen molecules, *Appl. Surf. Sci.*, **2018**, *455*, 410-417.

5. C. L. Pang, R. Lindsay and G. Thornton. Chemical reactions on rutile TiO$_2$ (110), *Chem. Soc. Rev.*, **2008**, *37*, 2328-2353.

6. G. Wu, C. Zhao, C. Guo, J. Chen, Y. Zhang, Y. Li. DFT study on the interaction of TiO$_2$ (001) surface with HCHO molecules, *Appl. Surf. Sci.*, **2018**, *428*, 954-963.

7. M. Sowmiya and K. Senthilkumar. Adsorption of proline, hydroxyproline and glycine on anatase (001) surface: a first-principle study, *Theor. Chem. Acc.*, **2016**, *135*, 12(1-8).

8. P. K. Samanta, N. J. English. Opto-electronic properties of stable blue photosensitisers on a TiO$_2$ anatase (101) surface for efficient dye-sensitised solar cells, *Chem. Phys. Lett.*, **2019**, *731*, 136624 (1-5).

9. X. Liu, D. Yang, Y. Li, Y. Gao and W.-T. Liu. Anisotropic Adsorption of 2-Phenylethyl Alcohol on a Rutile (110) Surface, *J. Phys. Chem. C*, **2019**, *123*, 29759-29764.

10. N. N. Tri, H. Q. Dai, N. T. Trung. Chemisorption of enrofloxacine on rutile TiO$_2$ (110) surface: A theoretical investigation, *Vietnam J. Scie. Tech.*, **2019**, *57*(4), 449-456.

11. N. N. Tri, M. T. Nguyen, N. T. Trung. A molecular level insight into adsorption of β-lactam antibiotics on vermiculite surface, *Surf. Scie.*, **2020**, *695*, 121588 (1-8).

12. E. R. Lewars. *Computational Chemistry*, Springer, Germany, 2016.

13. J. Hafner. Ab-Initio Simulations of Materials Using VASP: Density-Functional Theory and Beyond, *J. Comput. Chem.*, **2008**, *29*, 2044–2078.

14. J. P. Perdew, K. Burke, M. Ernzerhof. Generalized Gradient Approximation Made Simple, *Phys. Rev. Lett.*, **1996**, *77*, 3865-3868.

15. M. J. Frisch, G. W. Trucks and et al. *Gaussian 09 (Revision B.01)*, Wallingford, CT: Gaussian, 2010.

16. R. F. W. Bader. *Atoms in molecules: A quantum theory*, Oxford: Oxford University Press, 1990.

17. F. Weinhold and et al. *GenNBO 5.G*, Theoretical Chemistry Institute, University of Wisconsin: Madison, WI, 2001.

18. H. Q. Dai, N. N. Tri, N. T. T. Trang and N. T. Trung. Remarkable effects of substitution on stability of complexes and origin of the C–H/O(N) hydrogen bonds formed between acetone's derivative and CO$_2$, XCN (X = F, Cl, Br), *RSC Adv.*, **2014**, *4*, 13901-13908.

19. K. Oura, V. G. Lifshits, A. A. Saranin, A. V. Zotov, M. Katayama. *Surface Science: An Introduction*, Springer, Germany, 2003.

Khảo sát ảnh hưởng nhiệt độ nung đến khả năng quang xúc tác phân hủy dung dịch tetracycline của vật liệu TiO$_2$ đồng pha tạp (C, N, S)

Nguyễn Thị Lan[1,*], Võ Hoàng Anh[1], Nguyễn Văn Thắng[1], Lê Thị Cẩm Nhung[1], Lê Thị Thanh Thúy,[1] Nguyễn Phi Hùng[2]

[1]*Khoa Khoa học Tự nhiên, Trường Đại học Quy Nhơn, Việt Nam*
[2] *Viện Khoa học giáo dục, Trường Đại học Quy Nhơn, Việt Nam*

TÓM TẮT

Vật liệu nano TiO$_2$ pha tạp C, N, S (TH-TiO$_{2-a}$) được tổng hợp bằng phương pháp thủy nhiệt sử dụng tiền chất là TiOSO$_4$ và thiourea điều chế từ quặng Ilmenite. Ảnh hưởng của nhiệt độ nung đến khả năng phân hủy quang xúc tác tetracyline của vật liệu TH-TiO$_{2-a}$ được khảo sát một cách hệ thống. Hoạt tính quang xúc tác được đánh giá qua sự phân hủy kháng sinh tetracycline nồng độ 30 mg/L dưới ánh sáng khả kiến. Kết quả thực nghiệm cho thấy nhiệt độ nung đóng một vai trò quan trọng trong quá trình kết tinh của vật liệu. Mẫu TiO$_2$ pha tạp C, N, S nung ở 500°C trong không khí, thời gian 1 giờ cho hiệu quả quang xúc tác tốt nhất. Đáng chú ý là việc nung mẫu ở 500°C đã làm tăng độ tinh thể hóa, loại bỏ hiệu quả dư lượng chất hữu cơ trên bề mặt, điều khiển kích thước hạt thích hợp, và cải thiện đáng kể hoạt tính quang xúc tác.

Từ khóa: *TiO$_2$, Anatase, quang xúc tác, tetracycline.*

Tác giả liên hệ chính.
Email: nguyenthilan@qnu.edu.vn

Influence of the annealing temperature on the tetracycline photocatalytic degradation of (C, N, S) co-doped TiO$_2$ materials

Nguyen Thi Lan[1,*], Vo Hoang Anh[1], Nguyen Van Thang[1], Le Thi Cam Nhung[1],
Le Thi Thanh Thuy[1], Nguyen Phi Hung[2]

[1]*Faculty of Natural Sciences, Quy Nhon University, Vietnam*
[2]*Institute of Education Science, Quy Nhon University, Vietnam*

ABSTRACT

C, N and S co-doped TiO$_2$ samples (TH-TiO$_{2-a}$) were synthesized by the hydrothermal method using thiourea and TiOSO$_4$ prepared from Ilmenite ore. The influence of the annealing temperature on the tetracycline photocatalytic degradation of the TH-TiO$_{2-a}$ samples was systematically investigated. Photocatalytic activities of all samples were evaluated by the photodegradation of tetracycline antibiotics at 30 mg/L under the visible light irradiation. It is found that the annealing temperature plays an important role in the crystallization of the samples. The C, N and S co-doped TiO$_2$ sample annealed at 500°C in air for 1 hour exhibits the best photocatalytic activity. It is worth mentioning that annealing at 500°C improves crystallinity, effectively eliminates organic residues on the surface, controls particle sizes, and significantly improves the photocatalytic activity of the obtained sample.

Keywords: *TiO$_2$, Anatase, Photocatalyst, Tetracycline.*

1. INTRODUCTION

Titanium and titanium dioxide are applied in many fields such as aerospace, space, ship, chemicals (production of paper, paint, plastic, rubber, ink, soap, cosmetic, pharmaceutical,...), food processing, medical instruments, photo-catalytic treatment of organic pollutants,...[1] Currently, there are many methods for fabricating nano-sized TiO$_2$ such as chemical vapor deposition (CVD), oxidation of titanium tetrachloride (known as the chloride process), sol-gel method and thermal treatment or hydrolysis of titanium alkoxides.[2-5] However, these methods require not only expensive titanium sources but also stringent synthesis conditions, and thereby hindering them from practical applications.

Ilmenite ore is known as a cheap, non-toxic and widely distributed material source in the world, in which Vietnam is one of the countries with large reserves of titanium. Therefore, it is highly desirable to fabricate nano-sized TiO$_2$ from Ilmenite ore. Nowadays, titanium dioxide (TiO$_2$) is mainly used for the photocatalytic decomposition of organic pollutants. However, the band gap of TiO$_2$ is wide (3.0 eV for the rutile phase and 3.2 eV for the anatase phase), which is not suitable for the visible light irradiation, so TiO$_2$ only shows photocatalytic activities in the ultraviolet light region.

Besides, the rate of electron-hole recombination in TiO$_2$ is relatively fast, limiting its photocatalytic efficiency. To overcome these

Corresponding author.
Email: nguyenthilan@qnu.edu.vn

disadvantages, many studies have been done such as denaturing TiO_2 by Fe, Mn, Zn...[6, 7] or combining with other materials to prepare composites such as TiO_2/MoO_3,[8] TiO_2/MoS_2,[9] ... In particular, many recent reports have demonstrated that the introduction of non-metals such as N, C, S, P and halogen into the TiO_2 structure leads to an increase in the photocatalytic activities of TiO_2 in the visible light region. In fact, the modification of TiO_2 with carbon increases the photosensitivity of the obtained material[10] while the modification with nitrogen and sulfur reduces the band gap energy of TiO_2, thereby improving the photocatalytic activities under the illumination of the visible light.[11] However, these studies only use commercial TiO_2 sources, very few studies on TiO_2 denaturation from Ilmenite ore.[12] In this present work, C, N and S co-doped TiO_2 samples were synthesized from $TiOSO_4$ (prepared from Ilmenite ore of Binh Dinh) by the hydrothermal method. The hydrothermal method was chosen for this work due to the fact that this method has a number of advantages over other methods, such as low energy consumption, low cost, high speed of reaction, and simple and environmentally friendly process. It should be noted that, this method allows controlling the structure of the obtained materials by tailoring hydrothermal parameters. Therefore, in this work, the influence of the annealing temperature on the photodegradation of tetracycline was systematically investigated.

2. EXPERIMENT

2.1. Materials synthesis

Starting materials: Binh Dinh Ilmenite ore, concentrated sulfuric acid H_2SO_4 (Guangdong, China), distilled water, thiourea CH_4N_2S (Guangdong China), tetracycline $C_{22}H_{24}O_8N_2 \cdot HCl$ (Institute of Drug Quality Control Ho Chi Minh City, Vietnam).

2.1.1. Synthesis of TH-TiO₂

2.27 grams of $TiOSO_4$ prepared from Binh Dinh Ilmenite ore [13] was put into Teflon flask before adding a defined amount of thiourea (with molar ratio of thiourea to TiO_2 = 2:1) and 180 mL of distilled water. The Teflon flask was put in the autoclave and dried at 180°C for 12 hours.[14] After the hydrothermal process, the autoclave was naturally cooled to room temperature. The obtained white precipitate was filtered and washed several times with distilled water until the water has a constant pH before annealing at different temperatures (400, 500, 600, and 700 °C) for 1 hour. The obtained materials were denoted as TH-TiO_{2-a} (a = 400, 500, 600 and 700°C).

2.1.2. Synthesis of TiO₂

The TiO_2 sample was also prepared under the same conditions as the TH-TiO_{2-a} without the addition of thiourea.

2.2. Photocatalytic evaluation

The photocatalytic activities of the obtained samples were evaluated through the photocatalytic decomposition of tetracycline (30 mg/L) with the catalytic concentration of 0.6 g/L. The mixture was magnetically stirred in the dark for 30 minutes to reach an absorption-desorption equilibrium, before illuminating with a 60 W filament lamp (filter cutoff $\lambda > 420$). The remaining tetracycline concentrations as a function of the iradiation time was determined by HPLC-UV method at a wavelength of about 355 nm.

Tetracycline photodegradation efficiency of the obtanined samples is calculated by the following formula:

$$H = \frac{C_o - C}{C_o}.100 \qquad (1)$$

In which, C_o is the initial concentration of tetracycline and C is the remaining concentration of tetracycline after each corresponding time.

2.3. Materials characterization

The phase identification of the obtained materials as done by X-ray diffraction method using a D8 Advance Brucker diffractometer operated at 40 kV and 100 mA with Cu-Kα radiation ($\lambda_{K\alpha}$ = 1,5406Å),

and the crystallite size is calculated from the Debye - Scherrer equation $\bar{r} = \dfrac{0,89 \cdot \lambda}{\beta \cdot \cos\theta}$ (2). The infrared (IR) spectra were recorded on GX-PerkinElmer. Scanning Electron Microscopy (SEM) and Transmission Electron Microscopy (TEM) images were taken on Nova Nano SEM 450 and JEOL JEM 1010, respectively to investigate the surface morphology of the obtained samples. Specific surface area of all samples was measured on Micromeritics ASAP 2000 Equipment at the Institute of Chemistry, Vietnam Academy of Science and Technology. Tetracycline concentration was determined by HPLC-UV method at $\lambda = 355$ on Thermo Scientific series 3300 HPLC (Thermo Scientific Technologies, CA, USA).

3. RESULTS AND DISCUSSION

3.1. Materials characterization

IR spectra in the absorption range from 400 - 4000 cm^{-1} for TH-TiO$_{2-500}$ và TiO$_{2-500}$ samples are shown in Figure 1. IR spectra of both samples has a absorption band around 400 - 700 (860) cm^{-1} due to the valence oscillation of Ti-O-Ti bond.[15, 16] The peaks at 3550 and 1518 cm^{-1} are due to the valence and deformation fluctuations of the hydroxy group of water adsorbed on the surface of the samples.[17, 18] The peak around 1400 cm^{-1} of the doped sample is attributed to the oscillation of NH_4^+ ions.[19] Compared with the TiO$_{2-500}$ sample, the TH-TiO$_{2-500}$ sample has additional absorption peaks at 2982, 2320, 1992, 1712, 1634, and 1050 cm^{-1}, suggesting the formation of new bonds on the TiO$_2$ crystal lattice structure after doping. The peak at 1050 cm^{-1} corresponds to the oscillation of the Ti-O-S bond [20-22] and thereby indicating that S atoms has entered in the TiO$_2$ crystal structure. In addition, the peak at 1992 cm^{-1} of the TH-TiO$_2$ sample may be due to the formation of new compounds during the thiourea decomposition. The characteristic peak observed at 2320 cm^{-1} is related to the oscillation of the C = O group of CO_2 present in the air. The band around 1712 cm^{-1}

corresponds to the carbonyl group, the peak at 2982 cm^{-1} corresponds to the C-H bond [18, 23, 24] and the peak near 1634 cm^{-1} is due to the bending vibration of the N-H bond.[25, 26] These peaks were not observed in the TiO$_{2-500}$ sample, suggesting that C, N, S has been successfully doped into the crystal structure of TiO$_2$.
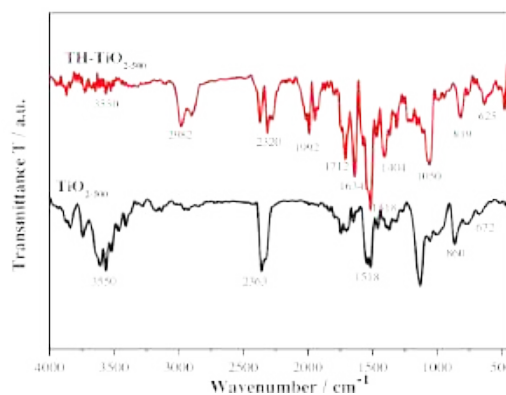
Figure 1. IR spectra for TH-TiO$_{2-500}$ and TiO$_{2-500}$

The influence of the annealing temperature on the crystal structure of TH-TiO$_{2-a}$ samples annealed at between 400-700°C was investigated by XRD and the obtained XRD patterns are shown in Figure 2.
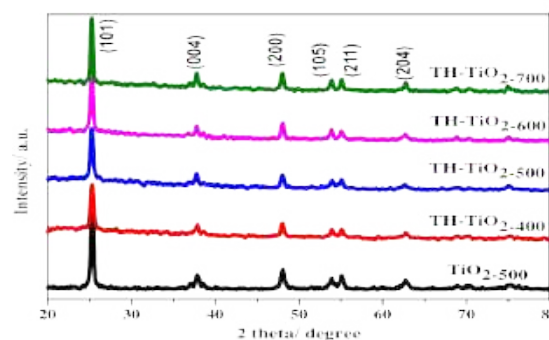

Figure 2. XRD patterns for TH-TiO$_{2-a}$ annealed at different temperatures

It is found that all TH-TiO$_{2-a}$ samples crystallized in the anatase phase, no rutile or brookite phases are observed. As the annealing temperature increases from 400 to 700°C, the (101) peak intensity increases and the spectral line half width at the (101) plane became narrower, resulting in a larger crystallite size. This proves that TiO$_2$ anatase gradually crystallizes as the annealing temperature increases.[27] The average crystallite size of the TH-TiO$_{2-a}$ samples was

calculated by Scherrer's equation and presented in Table 1. It should be noted that the average crystallite size of all doped samples is smaller than of the un − doped sample TiO_{2-500}. This suggests that co-doping C, N and S into TiO_2 prevents the grain growth which is in good agreement with previous reports. [20, 21, 28, 29]

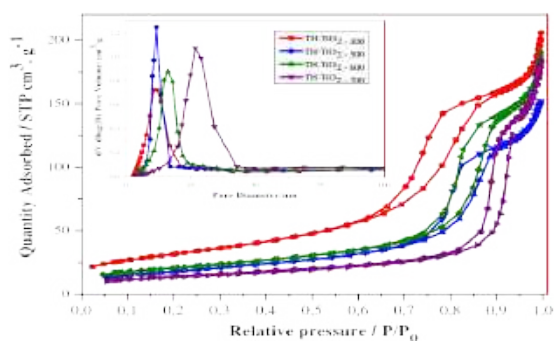The specific surface area and porosity of the obtained samples were determined by the BET method and their results are presented in Figure 3.

Figure 3. N_2 adsorption-desorption isotherms at 77 K and pore diameter distribution curves of the TH-TiO_{2-a} samples according to BJH

The N_2 adsorption-desorption isotherms at 77K of the TH-TiO_{2-a} samples shown in Figure 3 belong to IV type with type H1 loops (according to IUPAC classification). This is characteristics for material systems with the mesoporous structure formed by the particles intersecting with cylindrical-like canal systems or created by agglomeration of spherical particles.

The specific surface area determined according to the BET method for the TH-TiO_{2-a} samples annealed at 400 - 700 °C is 73.47, 92.25, 65.20 and 47.35 m^2/g, respectively.

The N_2 adsorption-desorption isotherms of the TH-TiO_{2-a} samples begin to condense at the relative pressure P/P_0 of about 0.7-0.95, in which TH-TiO_{2-500} sample has capillary condensation phenomenon at a relatively lower pressure than the others, indicating that this sample has narrower capillaries. This is consistent with data on the average pore diameter calculated by the BJH method of the TH-TiO_{2-a} samples (13.81, 10.46, 15.13 and 24.30 nm for

the sample annealed at 400, 500, 600 and 700°C, respectively).

In addition, it is found that except for the TH-TiO_{2-500} sample the specific surface area of the TH-TiO_2 samples decreases as a function of the annealing temperature. This may be because a change in the annealing temperature results in a change in the particle size and consequently a change in the specific surface area. Indeed, Carp et al reported that the activation energy decreases for increasing the annealing temperature, so the growth speed of the grain is large and the particle size increases rapidly at the high temperature. In contrast, at the low annealing temperature, the activation energy is large, so the particle growth rate is slower. As a result, the particle size of the obtained materials is small.[30]

The surface morphology of the TH-TiO_{2-a} samples are characterized by TEM and SEM methods and the results are shown in Figure 4 and Figure 5 (a-e). As can bee seen in Figure 4, the obtained samples have a structured morphology, the particles are spherical, quite uniform. SEM images also indicate that as the annealing temperature increases, the particle size of the TH-TiO_{2-a} samples increases.
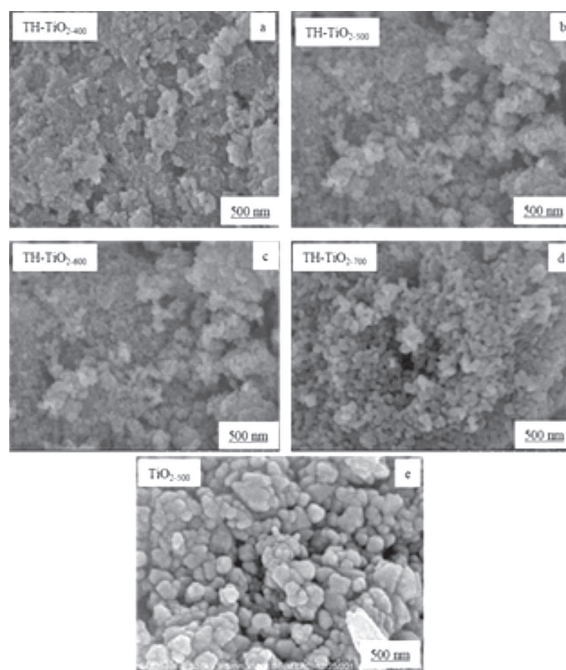

Figure 4. SEM images of TH-TiO_{2-400} (a), TH-TiO_{2-500} (b), TH-TiO_{2-600} (c), TH-TiO_{2-700} (d), and TiO_{2-500} (e)
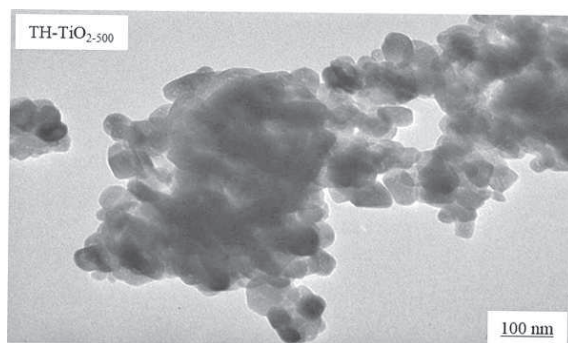
Figure 5. TEM image of TH-TiO$_{2-500}$

UV-Vis DRS was employed to investigate the influence of the annealing temperature on the light adsorption of the TH-TiO$_{2-a}$ samples and their UV-Vis DRS spectra are shown in Figure 6. As can be seen in Figure 6 there is a redshift as a function of the annealing temperature. This may be due to the evaporation of the doped elements as the annealing temperature increases.[31] Particularly, TH-TiO$_{2-400}$ và TH-TiO$_{2-500}$ samples have strong absorption in the visible light region. This is because the three doping of C, N and S has narrowed the band gap energy of TiO$_2$ and created many carriers [21, 32]. The band gap energy of all samples was calculated based on the Kubelka-Munk equation and shown in Table 1 and Figure 7. Accordingly, the band gap of the TH-TiO$_{2-a}$ samples annealed at 400 – 700°C is 2.86, 2.88, 3.02, and 3.05, respectively, which is lower than that of the TiO$_{2-500}$ sample (3.2 eV). The decrease in the band gap energy after doping may be attributed to the occurrence of the hybridized states of doped elements (C, N and S) in the band gap of the doped samples.[32, 33]
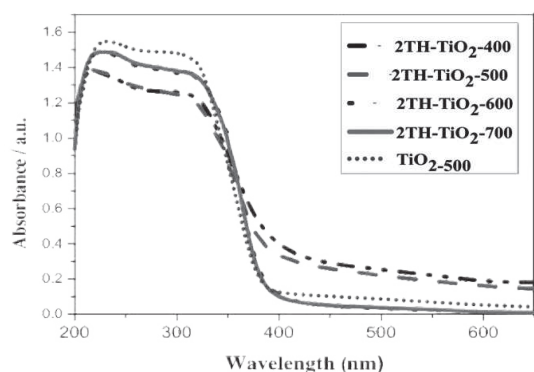
Figure 6. UV-Vis- DRS spectra for TH-TiO$_{2-a}$ (a = 400, 500, 600, 700 °C) and TiO$_{2-500}$.
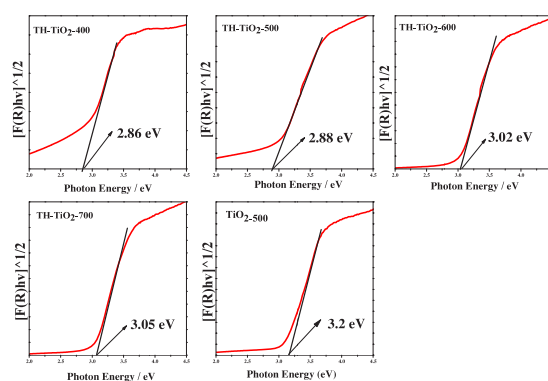
Figure 7. Kubelka-Munk function versus photon energy for band gap estimation

Table 1. Specifications of TH-TiO$_{2-a}$ và TiO$_{2-500}$ materials

Sample	Crystallite size (nm)	Specific surface area (m^2/g)	Band gap energy (eV)
TH-TiO$_{2-400}$	7.47	73.47	2.86
TH-TiO$_{2-500}$	9.52	92.25	2.88
TH-TiO$_{2-600}$	9.79	65.20	3.02
TH-TiO$_{2-700}$	13.40	47.35	3.05
TiO$_{2-500}$	14.39	36.01	3.20

3.2. Photocatalytic activities

The tetracycline photocatalytic degradation of TH-TiO$_{2-a}$ and TiO$_{2-500}$ samples is shown in Figure 8. It is worth mentioning that all C, N and S co-doped TiO$_2$ samples yield a higher photodegradation efficience than the undoped TiO$_2$ sample. Particularly, the TH-TiO$_{2-500}$ show the best photocatalytic activity under the visible light irradiation (96%). This is because the TH-TiO$_{2-500}$ sample has the largest specific surface area and thereby having more active sites for the adsorption and decomposition of the tetracycline.
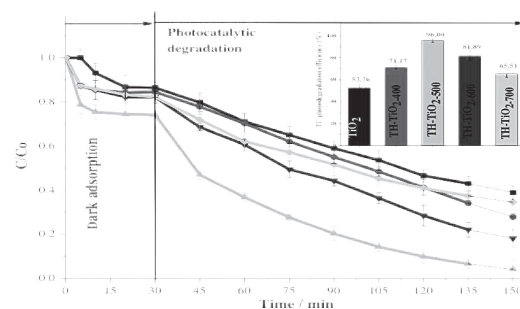

Figure 8. The change in C/Co as a function of time for TiO$_{2-500}$ và TH-TiO$_{2-a}$ (a = 400, 500, 600 and 700°C), tetracycline concentration of 30 mg/L

4. CONCLUSIONS

C, N and S co-doped TiO_2 photocatalysts prepared by the hydrothermal method were characterized by modern physical methods. The effect of the annealing temperature on the tetracycline photodegradation of the TH-TiO_{2-a} samples was systematically investigated. Experimental results indicate that the annealing temperature has a great influence on the particle size and band gap energy and thereby affecting the photocatalytic activities of the TH-TiO_{2-a} samples. It is found that all co-doped samples show the high photocatalytic performance in the visble light region. This suggests that (C, N, S) co-doped TiO_2 photocatalysts are promising candidates for the photocatalytic decomposition of toxic organic substances in the wastewater under the visible light irradiation.

REFERENCES

1. X. Chen and S. S. Mao. Titanium dioxide nanomaterials: synthesis, properties, modifications, and applications, *Chemical reviews,* **2007,** *107* (7), 2891-2959.

2. A. Molea, V. Popescu, N. A. Rowson and A. M. Dinescu. Influence of pH on the formulation of TiO_2 nano-crystalline powders with high photocatalytic activity, *Powder technology,* **2014,** *253,* 22-28.

3. F. Li, H. Li, L.-x. Guan and M.-m. Yao. Nanocrystalline Co^{2+}/F^- codoped TiO_2–SiO_2 composite films for environmental applications, *Chemical Engineering Journal,* **2014,** *252,* 1-10.

4. H. H. Tseng, W. W. Lee, M.-C. Wei, B.-S. Huang, M.-C. Hsieh and P.-Y. Cheng. Synthesis of TiO_2/SBA-15 photocatalyst for the azo dye decolorization through the polyol method, *Chemical Engineering Journal,* **2012,** *210,* 529-538.

5. A. Molea, V. Popescu and N. A. Rowson. Effects of I-doping content on the structural, optical and photocatalytic activity of TiO_2 nanocrystalline powders, *Powder technology,* **2012,** *230,* 203-211.

6. L. Xiao, J. Zhang, Y. Cong, B. Tian, F. Chen and M. Anpo. Synergistic effects of doped Fe^{3+} and deposited Au on improving the photocatalytic activity of TiO_2, *Catalysis letters,* **2006,** *111* (3-4), 207-211.

7. P. N. Kapoor, S. Uma, S. Rodriguez and K. J. Klabunde. Aerogel processing of MTi_2O_5 (M = Mg, Mn, Fe, Co, Zn, Sn) compositions using single source precursors: Synthesis, characterization and photocatalytic behavior, *J. Mol. Catal. A Chem,* **2005,** *229* (1-2), 145–150.

8. T. V. Sviridova, L. Y. Sadovskaya, E. Shchukina, A. Logvinovich, D. Shchukin and D. Sviridov. Nanoengineered thin-film TiO_2/h-MoO_3 photocatalysts capable to accumulate photoinduced charge, *Journal of Photochemistry and Photobiology A: Chemistry,* **2016,** *327,* 44-50.

9. M. Sabarinathan, S. Harish, J. Archana, M. Navaneethan, H. Ikeda and Y. Hayakawa. Highly efficient visible-light photocatalytic activity of MoS_2–TiO_2 mixtures hybrid photocatalyst and functional properties, *RSC Advances,* **2017,** *7* (40), 24754-24763.

10. H. Liu, Y. Wu and J. Zhang. A new approach toward carbon-modified vanadium-doped titanium dioxide photocatalysts, *ACS applied materials & interfaces,* **2011,** *3* (5), 1757-1764.

11. G. Zhang, Y. C. Zhang, M. Nadagouda, C. Han, K. O'Shea, S. M. El-Sheikh, A. A. Ismail and D. D. Dionysiou. Visible light-sensitized S, N and C co-doped polymorphic TiO_2 for photocatalytic destruction of microcystin-LR, *Applied Catalysis B: Environmental,* **2014,** *144,* 614-621.

12. S. Wahyuningsih, A. H. Ramelan, R. M. I. Munifa, L. N. M. Z. Saputri and U. Chasanah. Synthesis of TiO_2 nanorods from titania and titanyl sulfate produced from Ilmenite dissolution by hydrothermal method, *Journal of Physics: Conference Series,* **2016,** *776.*

13. Lan Nguyen Thi, Anh Vo Hoang, An Hoang Duc, Hung Nguyen Phi, Nhiem Dao Ngoc,

Van Thang Bui, Lieu Pham Khac and Khieu Dinh Quang. Synthesis of CNS-Tridoped TiO_2 from Vietnam Ilmenite Ore and Its Visible Light-Driven-Photocatalytic Activity for Tetracyline Degradation, *Journal of Nanomaterials,* **2020,** *2020,*1-14.

14. M. Wu, B. Yang, Y. Lv, Z. Fu, J. Xu, T. Guo and Y. Zhao. Efficient one-pot synthesis of Ag nanoparticles loaded on N-doped multiphase TiO_2 hollow nanorod arrays with enhanced photocatalytic activity, *Applied Surface Science,* **2010,** *256* (23), 7125-7130.

15. H. Zhang, Z. Xing, Y. Zhang, Z. Li, X. Wu, C. Liu, Q. Zhu and W. Zhou. Ni^{2+} and Ti^{3+} co-doped porous black anatase TiO_2 with unprecedented-high visible-light-driven photocatalytic degradation performance, *RSC Advances,* **2015,** *5* (129), 107150-107157.

16. M. A. Mohamed, W. Salleh, J. Jaafar, A. Ismail and N. A. M. Nor. Photodegradation of phenol by N-Doped TiO_2 anatase/rutile nanorods assembled microsphere under UV and visible light irradiation, *Materials Chemistry and Physics,* **2015,** *162,* 113-123.

17. Y. K. Chen, Y. F. Lin, Z. W. Peng and J. L. Lin. Transmission FT-IR study on the adsorption and reactions of lactic acid and poly (lactic acid) on TiO_2, *The Journal of Physical Chemistry C,* **2010,** *114* (41), 17720-17727.

18. Y. Li, C. Xie, S. Peng, G. Lu and S. Li. Eosin Y-sensitized nitrogen-doped TiO_2 for efficient visible light photocatalytic hydrogen evolution, *Journal of Molecular Catalysis A: Chemical,* **2008,** *282* (1-2), 117-123.

19. G. Zhang, Y. C. Zhang, M. Nadagouda, C. Han, K. O'Shea, S. M. El-Sheikh, A. A. Ismail and D. D. Dionysiou. Visible light-sensitized S, N and C co-doped polymorphic TiO_2 for photocatalytic destruction of microcystin-LR, *Applied Catalysis B: Environmental,* **2014,** *144,* 614-621.

20. N. Yao, C. Wu, L. Jia, S. Han, B. Chi, J. Pu and L. Jian. Simple synthesis and characterization of mesoporous (N, S)-codoped TiO_2 with enhanced visible-light photocatalytic activity, *Ceramics International,* **2012,** *38* (2), 1671-1675.

21. C. Han, M. Pelaez, V. Likodimos, A. G. Kontos, P. Falaras, K. O'Shea and D. D. Dionysiou. Innovative visible light-activated sulfur doped TiO_2 films for water treatment, *Applied Catalysis B: Environmental,* **2011,** *107* (1-2), 77-87.

22. R. Venkatkarthick, D. J. Davidson, S. Vasudevan, G. Sozhan and S. Ravichandran. An Investigation of Interfacial and Photoelectrochemical Performance of Thermally Prepared C, N - codoped TiO_2 Photoanodes for Water Splitting, *Chemistry Select,* **2017,** *2* (1), 288-294.

23. X. Y. Zhang and X. L. Cui. Preparation and photocatalytic hydrogen evolution performance of C, N co-doped nano TiO_2 photocatalysts, *Acta Physico-Chimica Sinica,* **2009,** *25* (9), 1829-1834.

24. J. Geng, D. Yang, J. Zhu, D. Chen and Z. Jiang. Nitrogen-doped TiO_2 nanotubes with enhanced photocatalytic activity synthesized by a facile wet chemistry method, *Materials Research Bulletin,* **2009,** *44* (1), 146-150.

25. M. Primet, P. Pichat and M. V. Mathieu. Infrared study of the surface of titanium dioxides. I. Hydroxyl groups, *The Journal of Physical Chemistry,* **1971,** *75* (9), 1216-1220.

26. X. F. Lei, X. X. Xue, H. Yang, C. Chen, X. Li, M. C. Niu, X. Y. Gao and Y. T. Yang. Effect of calcination temperature on the structure and visible-light photocatalytic activities of (N, S and C) co-doped TiO_2 nano-materials, *Applied Surface Science,* **2015,** *332,* 172-180.

27. J. Lv, T. Sheng, L. Su, G. Xu, D. Wang, Z. Zheng and Y. Wu. N, S co-doped-TiO_2/fly ash beads composite material and visible light photocatalytic activity, *Applied Surface Science,* **2013,** *284,* 229-234.

28. M. Khan, J. Xu, N. Chen and W. Cao. First principle calculations of the electronic and optical properties of pure and (Mo, N) co-doped anatase TiO_2, *Journal of Alloys and Compounds,* **2012,** *513,* 539-545.

29. O. Carp, C. L. Huisman and A. Reller. Photoinduced reactivity of titanium dioxide, *Progress in solid state chemistry,* **2004,** *32* (1-2), 33-177.

30. T. Matsumoto, N. Iyi, Y. Kaneko, K. Kitamura, S. Ishihara, Y. Takasu and Y. Murakami. High visible-light photocatalytic activity of nitrogen-doped titania prepared from layered titania/isostearate nanocomposite, *Catalysis today,* **2007,** *120* (2), 226-232.

31. F. Dong, W. Zhao and Z. Wu. Characterization and photocatalytic activities of C, N and S co-doped TiO_2 with 1D nanostructure prepared by the nano-confinement effect, *Nanotechnology,* **2008,** *19*(36), 365-607.

32. Y. L. Kuo, T. L. Su, F. C. Kung and T. J. Wu. A study of parameter setting and characterization of visible-light driven nitrogen-modified commercial TiO_2 photocatalysts, *Journal of hazardous materials,* **2011,** *190* (1-3), 938-944.

Trực quan hóa dữ liệu sức khỏe thời gian thực ứng dụng nền tảng mã nguồn mở Thingsboard

Ngô Văn Tâm*, Nguyễn Đức Thiện

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

TÓM TẮT

Trực quan hóa dữ liệu một chủ đề đầy thách thức và đang thu hút được sự quan tâm của cộng đồng khoa học và các nhà phát triển ứng dụng, đặc biệt là trong lĩnh vực chăm sóc sức khỏe. Với tính chất đa dạng và phức tạp của dữ liệu thu thập từ một lượng lớn các thiết bị cảm biến y sinh, việc biểu diễn dữ liệu một cách trực quan có vai trò hết sức quan trọng. Trên cơ sở ứng dụng nền tảng mã nguồn mở Thingsboard, bài báo này đề xuất giải pháp để trực quan hóa dữ liệu sức khỏe bệnh nhân theo thời gian thực, góp phần hỗ trợ người dùng và đội ngũ y bác sĩ theo dõi tình trạng sức khỏe của bệnh nhân và đưa ra các chẩn đoán phù hợp, kịp thời.

Từ khóa: *Trực quan hóa dữ liệu, chăm sóc sức khỏe từ xa, Thingsboard, IoT.*

*Corresponding author.
Email: ngovantam@qnu.edu.vn

Real-time healthcare data visualisation using open-source platform Thingsboard

Ngo Van Tam[*]**, Nguyen Duc Thien**

Faculty of Engineering and Technology, Quy Nhon University, Vietnam

ABSTRACT

Data visualisation is a challenging topic, attracting much attentions from the scientific community as well as application developers, especially in the healthcare domain. Due to a diverse and complex nature of data collected from a huge number of biomedical sensors, visual representation of data is of utmost importance. Based upon an open source platform namely Thingsboard, this paper proposes a solution to visualise healthcare data in real-time, thus allowing doctors, nurses and experts to monitor patients' health conditions and provide timely accurate diagnoses.

Keywords: *Data visualisation, remote healthcare, Thingsboard, IoT.*

1. INTRODUCTION

The rapid growth in the number of health sensing and monitoring devices, etc., has been generating a huge amount of data. According to IDC's prediction, by 2020, the data generated from healthcare applications can be up to 40 ZB.[1] Due to the diverse and complex nature of the collected data including biometric data, clinical symptoms, electronics health records (EHR) to MRI /CT diagnostic medical imaging data etc., the data visualisation is playing an ulmost important role.

Data visualisation is the process of converting data in numerical form (numerical data) into non-numerical data without reducing the entropy of the original data. In other words, *data visualisation is the main tool for creating images, diagrams, or animations to convey meaningful knowledge and messages to users,*

meeting specific tasks and requirements. With the explosion of data, data visualisation is a challenging topic and attracting much attention from the scientific community and application developers as well.[2,3]

In the current context, the healthcare industry is looking for modern solutions to solve problems related to display, storage, data processing, and analysis. Especially, remote healthcare is actually getting a lot of benefits from technologies and data visualisation solutions. By using existing tools for visualizing health data, doctors and medical staffs can easily and remotely measure and monitor the patients' bio-physical parameters (e.g., heart rate, SpO2 - peripheral blood oxygen saturation, and ECG - Electrocardiogram), especially patients in remote areas, or disabled patients, elderly patients who are unable to move to medicare centers, resulting

---

*Corresponding author.
Email: ngovantam@qnu.edu.vn

an improvement in the reliability of decision support systems.

In recent years, many studies in the area of medical data visualisation have been proposed. Liu et. al, proposed a data visualisation method to describe a relationship between heterogeneous data stored in databases using three types of data structure such as graph, tree, and graph-tree structure.[4] The authors in the study[2] used a graphical tool based on the GeoJSON (Geo Javascript Object Notation) standard to classify the patients' health data. These data were represented by using different colored circles. This method also deployed MongoDB database for storing the large amount of data. Based on a pre-defined and standardlised table of colors, doctors can diagnose the current health status of the patients.

Several research groups have been approaching a combination of technologies of web, mobile, and open-source platforms to develop medical data visualisation methods.[5-7] The authors in the study[5] introduce a multi-dimensional medical data display system from wearable devices on patients by combining smartphones and Web browser-based environments like the Google SDK, Android SDK, Google's IntelliJ. In this system, the collected data will be re-structured to visualise over time (e.g., hours, days, or months), or be interpolated and displayed through different shapes and colors. In the paper,[6] based on the application of virtual reality and the Internet, Xu et. al, propose a lightweight progressive coding architecture to visualize patients' medical data. The original images will be digitised, processed, and visualised to help doctors to provide patients with diagnoses.

To visualise data in IoT applications, selecting an appropriate platform plays a very important role.[2, 6 - 10] In the IoT architecture, IoT platform is responsible for performing main functions such as connecting IoT devices, collecting, monitoring, managing, and analysing data. In recent years, many commercial and open-source platforms have been released to meet the increasing demands of IoT applications. Notably, open-source platforms such as Thingsboard, ThingSpeak, Thingio, Site where, WSo2, Kaa IoT, DeviceHive, Zetta, and Blynk are attracting attention from researchers and application developers in IoT field in general and health sector in particular.[9, 10] Compared with the above open-source platforms, Thingsboard is considered as an effective platform to address the needs of data collection, processing, visualisation, and device management.[10]

The data visualisation methods in [2,6-8] have partially addressed the need to display medical data. However, there are some basic limitations in these methods as follows:

- Firstly, the methods have not met the need of representing data in real time [7] - an extremely important requirement in monitoring patient survival indicators, particular in patients in remote areas, or elderly/disabled patients;

- Secondly, the above solutions only provide measurement features, transfer patient health data through traditional communication methods such as LAN, and display locally on the patients' wearable device.[8] These systems still lack an integration of new technologies, which is a big challenge to develop large-scale data display systems;[6]

- Thirdly, end users such as clinicians, medical staffs, and researchers, etc., are usually not experts in the area of information technology and communications. However, the above solutions are still quite complicated for end-users.[2,6,7] In this context, it is necessary to deploy solutions with an easy-to-use interface in order to help them having quick captures of the patient's health status.

To address the above limitations, in this work, the authors will adopt the open-source Thingsboard platform and develop a system for monitoring and visualising patient health data in real-time. This system helps patients

to understand their health conditions. More important, the system supports doctors and medical staffs to monitor the health status of patients and make appropriate diagnoses in timely fashion. The proposed system addresses the following basic requirements:

- Ability to monitor in real-time;

- Compatibility with a wide range of sensor devices from different manufacturers and data platforms;

- User-friendly interface;

- Being able to adapt to many different types of services, thus effectively deploying available resources.

- Ability to deploy at different scales such as: *(i)* on mobile devices of patients; *(ii)* in patients' homes (including people in remote and isolated areas, places where health care and medical facilities are scarce); *(iii)* at the doctor's office, and *(iv)* at health centers and hospitals.
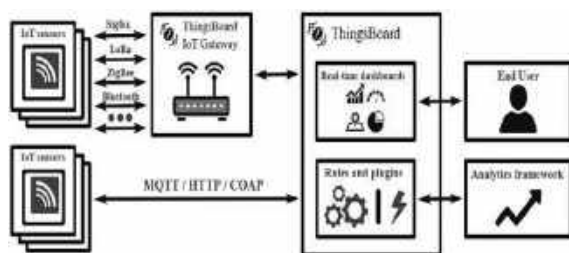
Figure 1. Open source platform hingsboard.[11]

The rest of the paper is organised as follows: Section 2 describes the open-source Thingsboard platform. Section 3 proposes a real-time monitoring and visualisation system for patient health data based on the Thingsboard platform. The results of the proposed system are illustrated in Section 4. The conclusion of the paper will be presented in section 5.

2. THINGSBOARD PLATFORM

2.1. Overview of Thingsboard

In recent years, the open-source Thingsboard platform is being considered as an effective solution for collecting, processing, visualising data, and managing devices for IoT applications.

Thingsboard allows connecting devices through industry-standard IoT protocols such as MQTT (Message Queuing Telemetry Transport), CoAP (Constrained Application Protocol), and HTTP (HyperText Transfer Protocol). In addition, ThingsBoard allows an integration of devices connected to legacy and third-party systems with existing protocols. As depicted in Figure 1. Thingsboard includes some following basic features:

- *Remote data collection:* Thingsboard supports remote data collection and storage in a reliable way. Collected data can be accessed using custom websites or server-side APIs;

- *Data visualisation:* Thingsboard provides many utilities to visually display the collected data. Thingsboard also allows to create own gadgets like Google map widgets, real-time graphs, …;

- *Device management:* Thingsboard enables to register and manage the IoT devices. It allows monitoring of client-side device properties and server-side provisioning. Thingsboard also provides API for server-side applications to send RPC (Remote Procedure Calls) commands to devices and vice versa;

- *Dashboard: is used* to display data and control devices remotely in real- time;

- *Manage alerts:* Thingsboard introduces a tool to create and manage alerts related to entities in the system. In this context, Thingsboard allows to monitor real-time alarms and alarm propagation to the related entities hierarchy.

2.2. The basic communication protocols in Thingsboard

2.2.1. MQTT protocol

MQTT is a Publish-Subscribe messaging protocol. Because this protocol uses a low bandwidth, it is an ideal protocol for IoT applications. In a system using MQTT protocol, many clients connect to a server (i.e., in MQTT, the server is called MQTT Broker). Each client will register to follow the information channel

(topic) or transmit its data to that information channel. This registration process is called "subscribe" and sending data of a client to the information channel is called "publishing". If there are data updates on the given channel (the data can come from different clients), the clients who have subscribed to this channel will receive the updated data.

2.2.2. CoAP (Constrained Application Protocol)

CoAP is a dedicated Internet application protocol for bound devices, as defined in RFC 7252. CoAP is designed to connect devices on the same bound network such as low power networks and lossy networks; or between devices and common nodes on the Internet, or between devices on different bounded networks connecting by the Internet.

CoAP is a service layer protocol designed for use in resource-restricted Internet devices, such as wireless sensor network nodes. CoAP is designed to be easily converted into HTTP in order to integrate with web applications and meet specific requirements such as multicast support, a very low overhead, and simplicity. Therefore, CoAP can be easily deployed in devices with small memory capacity and low power, especially applications in IoT, wireless sensors, and M2M (Machine-to-Machine). CoAP can run on most devices that support UDP (User Datagram Protocol) or similar UDP. To secure the data transmissions, CoAP uses Datagram Transport Layer Security (DTLS).

3. A REAL - TIME HEALTH DATA MONITORING AND VISUALISATION SYSTEM

3.1. System model

In this paper, we build up a real-time monitoring and visualisation system for patient health data combined with the open-source Thingsboard platform. The basic idea of this system is to collect basic bio-physical parameters such as heart rate, SpO2, and ECG. Supported by the Thingsboard platform, the collected data will be managed and visualised in real-time, thereby supporting for inspection and monitoring. The proposed system is illustrated in Figure 2 with some basic subsystems as follows:
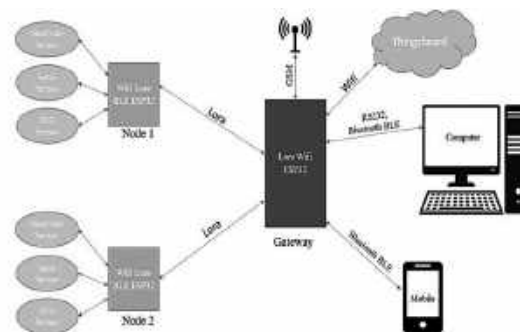
Figure 2. Proposed real-time health data monitoring and visualising system.

- *Sensing subsystem*: This subsystem includes some non-invasive sensors including MAX30100 [12] (i.e., to measure heart rate and SpO2), and AD8232 [13](i.e., to measure ECG). The detailed information of these sensors are as follows:

○ The MAX30100 module is a product of the Maxim cooperation, which uses the common optical measurement method today. This sensor allows an indirect and non-invasive measurement of oxygen saturation in peripheral blood and the patient's heart rate. The patients just wear the measurement device on their body. The device provides a measurement with a high accuracy, durability, and low noise;

○ The AD8232 sensor is used to measure the electrical activity of the heart. This electrical activity can be graphed as an ECG electrocardiogram. This is a method of tracking activity, speed, and beat of the heart. When the heart is active, it will generate a variation in electric current. Here, ECG is a curve that records those changes and used for diagnosing various heart conditions.

- *Wireless and communication processing subsystem:* This subsystem includes microcontrollers and wireless transceivers in order to ensure the communications between the Gateway and Nodes

using LoRa technology; or from the Gateway to the Thingsboard using WiFi:

○ Heltec Wifi Lora 32 Kit is based on a combination of ESP32 chip; Tensilica LX6 processor, speed 240MHz, 520 KB SRAM; WiFi standard 802.11b/g/; two Bluetooth modes (traditional Bluetooth and Bluetooth BLE); and LoRa SX1278 chip. The LoRa SX1278 operates at frequency of 433MHz with the maximum distance of 5 km. The 0.96-inch OLED display is built-in; lithium battery charging circuit and UART CP2102 USB chip support development environment on Arduino;

○ Gateway requires relay communication between Node and server by using two wireless communication technologies of LoRa and WiFi. ESP32 module is chosen because its synchronous design and programming is similar to Arduino. ESP32 chip is used as a powerful microcontroller and integrated WiFi interface, which meets the requirements of the proposed system.

- *Display subsystem* uses a 0.96-inch OLED screen to display measured parameters from the sensing subsystem and the status of peripheral devices. At the same time, the sensed data is also visualised on the Thingsboard platform.

(a). Heart-beat and SpO2 MAX30100 [12]
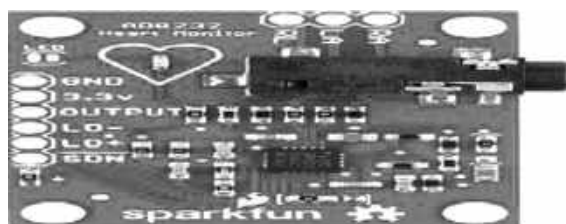
(b). ECG AD8232 [13]

Figure 3. Sensing devices.

Table 1. SpO2 reference table [14]

SpO2 (%)	Meaning	Warning
97 - 99	Oxygen in the blood – good	
94 - 96	Medium blood oxygen	Need more oxygen to breathe
90 - 93	Low blood oxygen	Need to be monitored promptly
< 90	Extremely low blood oxygen	Emergency

Table 2. Heart-beat reference table by age (beats/minute). [15]

Age	Awake Mode	Sleep Mode
< 3 months	85 - 200	80 - 160
3 months - 2	100 - 190	75 - 160
2 - 10	60 - 140	60 - 90
> 10	60 - 100	50 - 90

3.2. Monitoring and alerting the patient's health status
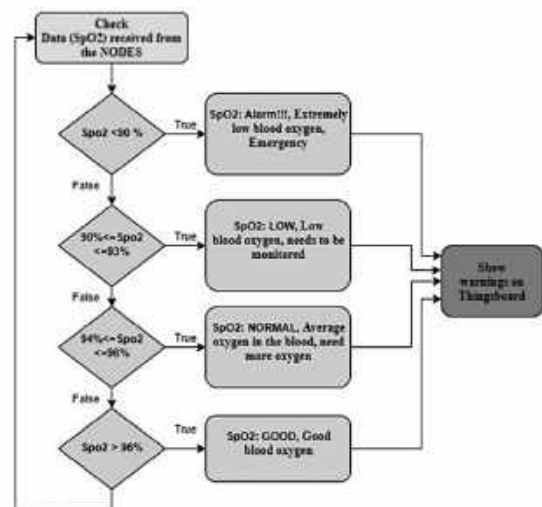
Figure 4. SpO2 level monitoring algorithm

The most important content of the paper is to generate alerts based on the patient health data collected by the sensing subsystem as described in section 3.1. In particular, this paper also develops a SpO2 level alert function as shown in Table 1. The warnings displayed on

the dashboard of Thingsboard platform will help the medical team to capture the patient's status timely. The SpO2 monitoring algorithm illustrated in Figure 4 has four levels: Level 1 - ALARM (Emergency), Level 2 - LOW (Low), Level 3 - NORMAL (Normal), and level 4 - GOOD (Good).

3.3. Connecting with Thingsboard

In this paper, the authors collect three types of data as described in section 3.1 and send them to Thingsboard using MQTT protocol. It is a publish /subscribe (message) messaging protocol, uses low bandwidth, has high reliability, and is capable of operating under unstable transmission conditions. In this case, many station nodes (Client) connect to an MQTT Broker. Each client will subscribe to several channels, such as "/client1/channel1", "/client1/channel2", called "subscribe". Each client will receive data when any other station sends data to the registered channel. When a client sends data to that channel, it's called "publish".
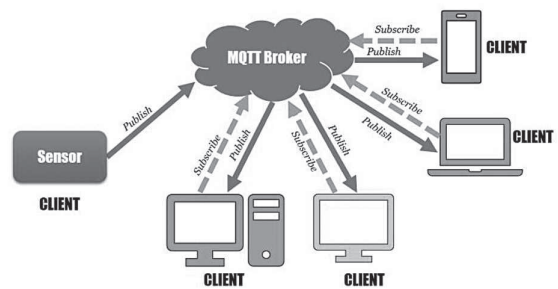
Figure 5. MQTT Broker[16]

4. RESULTS
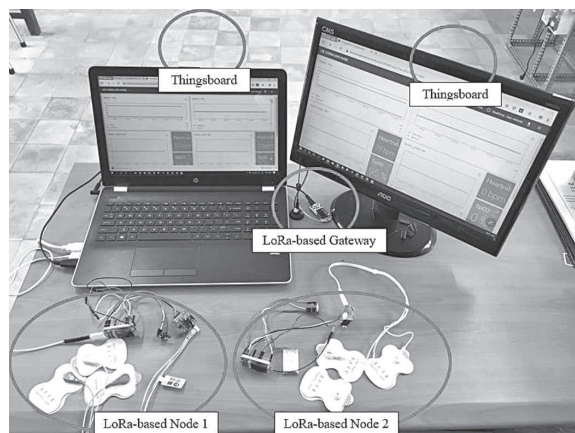
Figure 6. A testbed of the real-time health data monitoring and visualisation system.
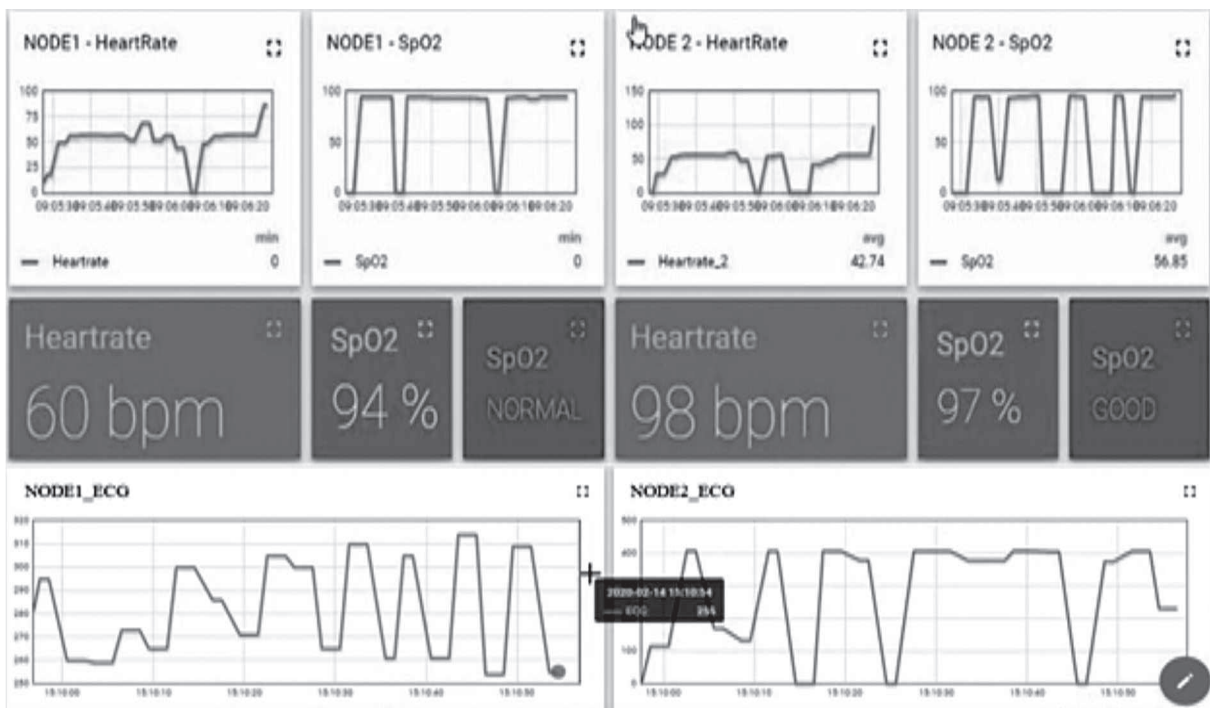
Figure 7. Visualising health data in real-time with Thingsboard platform.

In this work, the authors develop an experimental testbed of the real-time health data monitoring and visualisation system as shown in Figure 6. The testbed consists of two nodes (NODE) and a Gateway using LoRa technology:

- Each NODE will integrate two types of sensors MAX30100 and AD8232. The nodes are mounted on the different patients: a 30-year-old male patient lying on a bed (i.e., NODE 1 is powered by using a grid power) and a 38-year-old male patient wearing a device (i.e., NODE 2 is powered by a battery) moves within a 300m radius of the Gateway.

- Gateway is responsible for collecting data from NODE 1 and 2. At the same time, the Gateway will forward data to Thingsboard for further processing and display.

Patients' health data will be visualised on the Thingsboard's dashboard in real-timeas depicted in Figure 7. Heart rate, SpO2, and ECG values of each patient are displayed in two basic forms: graph and numerical. Figure 8 shows a piece of ECG signal extracted from NODE 1.

As shown Figure 7, the warning function (i.e., based on the SpO2 indicator) has been integrated into the dashboard. If SpO2 values respectively measured from NODEs 1 and 2 are 94% (NORMAL) and 97% (GOOD), doctors can diagnose the patient's condition and suggest a solution for patients to breathe more oxygen. Clearly, the real-time data visualisation plays an important role in supporting clinicians to monitor patients' health status and provide suitable diagnoses a timely.
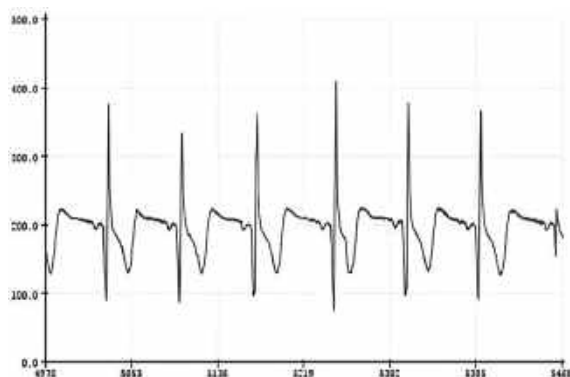
Figure 8. Extracting ECG signal from NODE 1.

5. CONCLUSION

This paper built a model for monitoring and visualizing patient health data in real-time based on the open-source Thingsboard platform. The proposed system is expected to assist doctors, medical staffs to monitor the patients' health status. The results of the paper show that Thingsboard is suitable platform for data visualisation in real-time. More important, the results also confirm the importance of data visualisation in the domain of healthcare.

Based on the preliminary results of the paper, the authors will integrate more sensors and build up new systems with more functionalities, especially data analysis supported byartificial intelligence (AI). New integrations can improve the ability of healthcare for people in the future.

REFERENCES

1. IDC - Analyze the Future, <https://www.kdnuggets.com/2012/12/idc-digital-universe-2020.html>, Retrieved August 25, 2019.

2. Antonino Galletta, Lorenzo Carnevale, Alessia Bramanti, and Maria Fazio. An Innovative Methodology for Big Data Visualization for Telemedicine, *IEEE Trans*, **2019**, *15*(1), 490 – 497.

3. Shah Nazir, Muhammad Nawaz Khan, Sajid Anwar, Awais Adnan, Shahla Asadi, Sara Shahzad, and Shaukat Ali. Big Data Visualization in Cardiology - A Systematic Review and Future Directions, *IEEE Access*, **2019**, *7*, 115945 – 115958.

4. Q. Liu, X. Guo, H. Fan, and H. Zhu. *A novel data visualization approach and scheme for supporting heterogeneous data*, Proc. IEEE 2nd Inf. Technol., Netw., Electron. Autom. Control Conf, China, Dec. 2017.

5. T. M. Frink, J. V. Gyllinsky, and K. Mankodiya. *Visualization of multidimensional clinical data from wearables on the web and on apps*, Proc. IEEE MIT Undergraduate Res. Technol. Conf., Cambridge, Nov. 2017.

6. Xu, Gaowei et al. An IoT-Based Framework of Webvr Visualization for Medical Big Data in Connected Health, *IEEE Access*, **2019**, *7*, 173866 – 173874.

7. J. G. Stadler, K. Donlon, J. D. Siewert, T. Franken and N. Lewis. Improving the Efficiency and Ease of Healthcare Analysis Through Use of Data Visualization Dashboards, *Big Data*, **2016**, *4*(2), 129-135.

8. A. El Attaoui, M. Hazmi, A. Jilbab, and A. Bourouhou. Wearable Wireless Sensors Network for ECG Telemonitoring Using Neural Network for Features Extraction, *Wireless Personal Communications*, **2019**, *111*, 1955-1976.

9. K. Vandikas and V. Tsiatsis. *Performance evaluation of an IoT platform*, in 8th IEEE International Conference on Next Generation Mobile Apps, Services and Technologies (NGMAST), United Kingdom, 2014.

10. A. A. Ismail, H. S. Hamza, and A. M. Kotb. *Performance Evaluation of Open Source IoT Platforms*, IEEE Global Conference on Internet of Things (GCIoT), Egypt, 2018.

11. ThingsBoard Open source IoT Platform, <https://thingsboard.io/> , Retrieved December 02, 2020.

12. Datasheet MAX30100, <https://datasheets.maximintegrated.com/en/ds/MAX30100.pdf>, Retrieved December 02, 2020.

13. SparkFun Single Lead Heart Rate Monitor - AD8232, <https://learn.sparkfun.com/tutorials/ad8232-heart-rate-monitor-hookup-guide/all>, Retrieved December 02, 2020.

14. What is SpO2 in a normal person?, <https://www.vinmec.com/vi/tin-tuc/thong-tin-suc-khoe/chi-so-spo2-o-nguoi-binh-thuong-la-bao-nhieu/?link_type=related_posts>, Retrieved December 02, 2020.

15. Kirpalani H and Luang L H. *Manual of Pediatric Intensive Care*, People's Medical Publishing House, USA, 2009.

16. Comparing MQTT and HTTP: Which protocol is best for IoT, <https: //smartfactoryvn.com/technology/ internet-of-things/so-sanh-mqtt-va-http-giao-thuc-nao -la-tot-nhat-cho-iot/>, Retrieved August 25, 2019.

CONTENTS