



วารสารวิทยาศาสตร์ประยุกต์

THE JOURNAL OF APPLIED SCIENCE

ISSN 1513-7805 [Print] ISSN 2586-9663 [Online]

ปีที่ 19 ฉบับที่ 1 [2563] Vol.19, No.1 [2020]

THE JOURNAL IS COVERED BY:



ASEAN
CITATION
INDEX

Available online at <https://www.tci-thaijo.org/index.php/JASCI>

Consulting Editors

from King Mongkut's University of Technology North Bangkok, Thailand

Assoc.Prof.Dr. Surapun Yimman

Assoc.Prof.Dr. Saowanit Sukparungsee

Assoc.Prof.Dr. Kitti Bodhipadma

Editor

from King Mongkut's University of Technology North Bangkok, Thailand

Assoc.Prof. Narumol Kreua-ongarjnukool

Co-Editor

from King Mongkut's University of Technology North Bangkok, Thailand

Asst.Prof.Dr. Peerapong Pornwongthong

Dr. Nonchanutt Chudpooti

Associate Editor

from King Mongkut's University of Technology North Bangkok, Thailand

Assoc.Prof.Dr. Yupaporn Areepong

Assoc.Prof.Dr. Nutsuda Sumonsiri

Asst.Prof.Dr. Walaiporn Prissanaroon Ouajai

Asst.Prof.Dr. Tanapat Anusas-amornkul

Asst.Prof.Dr. Seksan Sirisubtawee

Asst.Prof.Dr. Suvimol Phanyaem

Asst.Prof.Dr. Theerawut Phusantisampan

Asst.Prof.Dr. Jintawat Tanamatayarat

Dr. Sukanya Thepwatee

Editorial Advisory Board

Prof.Dr. Pranom Chantaranothai
Khon Kaen University, Thailand

Prof.Dr. Yongwimon Lenbury
Mahidol University, Thailand

Prof.Dr. Sangvorn Kitthawee
Mahidol University, Thailand

Prof.Dr. Vanchai Riewruja
King Mongkut's Institute of Technology Ladkrabang, Thailand

Assoc.Prof.Dr. Montri Maleewong
Kasetsart University, Thailand

Assoc.Prof.Dr. Dudsadee Uttapap
King Mongkut's University of Technology Thonburi, Thailand

Assoc.Prof.Dr. Chalermchai WongsAree,
King Mongkut's University of Technology Thonburi, Thailand

Assoc.Prof.Dr. Sa-ad Riyajan
Thammasat University, Thailand

Assoc.Prof.Dr. Sittiwat Lertsiri
Mahidol University, Thailand

Assoc.Prof.Dr. Mariena Ketudat-cairns
Suranaree University of Technology, Thailand

Editorial Advisory Board (cont.)

Assoc.Prof.Dr. Wararit Panichkitkosolkul
Thammasat University, Thailand

Assoc.Prof.Dr. Sanae Rujivan
Walailak University, Thailand

Asst.Prof.Dr. Kuntpong Woraratpanyast,
King Mongkut's Institute of Technology Ladkrabang, Thailand

Asst.Prof.Dr. Suphakant Phimoltare
Chulalongkorn University, Thailand

Col.Asst. Prof.Panadda Hatthachote
Phramongkutklao College of Medicine, Thailand

Dr. Chivarat Muangphat
King Mongkut's University of Technology Thonburi, Thailand

Dr. Elvin James Moor
King Mongkut's University of Technology Thonburi, Thailand

Coordinator and Manager

from King Mongkut's University of Technology North Bangkok, Thailand

Mr. Phirom Pabu

Mr. Piyawat Sutha

Miss.Wilaiporn Khunkaew

Miss.Wilawan Sae-jeng

Miss.Kitsiya Chuchuausuwan

Contact info

Faculty of Applied Science

King Mongkut's University of Technology North Bangkok

1518 Pracharat 1 Road, Wongsawang, Bangsue, Bangkok 10800 : Thailand

Phone: 02-555-2000#4211

Website

www.tci-thaijo.org/index.php/JASCI

JOURNAL POLICY

Focus and Scope

The Journal of Applied Science is an academic journal published biannually by the Faculty of Applied Science, King Mongkut's University of Technology North Bangkok. The Journal publishes original research and review papers either in Thai or English covering all areas of applied science and technology. The journal will not accept articles, which have been published or are being considered for publication by another journal, nor should papers published here be submitted to other journals.

"The Journal of Applied Science does not have policy to collect publication fee"

"Each articles must be evaluated by two peers (double blinded) before accepted for publication"

"Article must be revised and sent back to the journal within 4 weeks after the return for revision unless the article will be rejected"

"The Journal of Applied Science published both as hard -copies [ISSN 1513-7805 (Print)] and electronic journal [ISSN 2586-9663 (Online)] available on ThaiJO system"

Scope of the Journal

To publish academic, research and review articles covering all area of both basic and applied sciences and technology including pure and applied mathematics, statistics, chemistry and applied chemistry, materials science, physics and industrial physics, environmental science and technology, biotechnology, agro-industrial and food technology, medical science and applications, health and beauty technology, computer and informatic sciences and materials science

Peer Review Process

Each article must be double blind peer reviewed by at least 2 reviewers from the related field.

Language

Both Thai and English

Publication Frequency

Twice a year (Biannual)

First Issue: January to June

Second Issue: July to December

Courtesy by the Faculty of Applied Science,
King Mongkut's University of Technology North Bangkok

Remarks

Authors have to be responsible for any legal effects that may occur due to their opinions expressed in the articles.

TABLE OF CONTENT

The community structure of beneficial and harmful arthropod fauna in rice field ecosystems in lower Northern Thailand Grant Berry, Keerati Tanruean, Tiwtawat Napiroon, Weeranut Saetang, Marisa Penpo, Supansa Duangrod and Pisit Poolprasert.....	1
A Study of Information Technology Risk Management of Government and Business Organizations in Thailand using COSO-ERM based on the COBIT 5 Framework Sakchai Tangprasert.....	13
Robustness of CUSUM-Tukey's Control Chart for Detecting of Changes in Process Dispersion Suganya Phantu and Saowanit Sukparungsee.....	25
Asymptotic confidence ellipses for the re-parameterized inverse Gaussian distribution Wikanda Phaphan.....	36
New ratio estimators for population mean in simple random sampling using robust regression Nuanpan Lawson.....	51
The predictions of a daily stock price direction from the Thai news content by using natural language processing Wikanda Phaphan and Aunchana Pimpisal.....	59
A mathematical model of melioidosis transmission in humans and animals with the effect of a rainfall factor Pornwilai Chankitkan and Puntani Pongsumpun.....	80
Daily Wind Speed Forecast Model at an Altitude of 120 Meters, Pak Phanang District, Nakhon Si Thammarat Province Warangkhan Riansut.....	95
The spatial and temporal variations and trends of sea surface temperature in the Gulf of Thailand from satellite imageries Tawatchai Na-u-dom, Anukul Buranapratheprat and Siraporn Tong-u-dom.....	110
The application of wastes from modified tapioca starch plant to produce the renewable energy using one stage anaerobic digestion system Piyavadee Srivachai and Chumaporn Rodsrida.....	124

Research Article

The community structure of beneficial and harmful arthropod fauna in rice field ecosystems in lower Northern Thailand

Grant Berry¹, Keerati Tanruean², Tiwtawat Napiroon², Weeranut Saetang², Marisa Penpo², Supansa Duangrod² and Pisit Poolprasert^{2*}

¹ New Cambridge International School, Phitsanulok, 65000, Thailand

² Biology Program, Faculty of Science and Technology, Pibulsongkram Rajabhat University, Phitsanulok, 65000, Thailand

*E-mail: poolprasert_p@psru.ac.th

Received: 02/12/2019; Revised: 22/03/2020; Accepted: 31/03/2020

Abstract

Studies to understand the diversity of insect and spider species related to rice cultivation and determine their ecological guilds can provide information concerning the composition and structure of such ecosystems, which can be applied to integrated pest management. Herein, the purpose of this current survey was to monitor the community structure of beneficial and harmful arthropod including insects and spiders which inhabit rice paddy fields. Six major locations of rice-growing areas were selected in lower northern Thailand, including Nakhon Sawan, Phetchabun, Phichit, Phitsanulok, Sukhothai and Uttaradit provinces. In each survey site, six sampling plots of paddy field were chosen and insects and spiders were caught using sweeping nets at 20 times/sampling plot during 2017-2019. All samples were identified based on morphology. A total of 6,459 individuals of insects and spiders were collected, representing 102 species in 69 families and 11 orders. The rice paddies of Phitsanulok showed the highest species richness (73 species). In this regard, the species diversity (H') for all insects and spiders of all rice-growing areas was considered relatively high (3.038). Meanwhile, the Evenness (J') and Simpson index (1-D) in all rice paddies from the pooled data set out in this survey were 0.765 and 0.919, respectively. Phichit) $H' = 3.234$ (showed the highest species diversity index, whereas Phetchabun and Sukhothai) $H' = 2.895$ (had the lowest levels of diversity. Significant differences in species diversity indices of six rice-growing sites were detected ($P < 0.05$) in all comparisons. The insect order of Coleoptera had the highest recorded number of species (21 species or 20.59%). It was found that the insect pests namely, *Edwardsiama hermanni* (F. Cicadellidae), *Dyscinetus morator* (F. Scarabaeidae) and *Nilaparvata lugens* (F. Delphacidae) were the dominant species. Meanwhile, *Micraspis discolor* (F. Coccinellidae), *Agriocnemis pygmaea* (F. Coenagrionidae) *Oxyopes javanus* (F. Oxyopidae), *Temelucha philippinensis* (F. Ichneumonidae), *Tropobracon schoenobii* (F. Braconidae) and *Goniozus nr. triangulifer* (F. Bethyilidae) were mainly exhibited as natural enemies. Additionally, the distribution of insects and spiders was aggregated for all rice-planting areas, as indicated by the values of variance to mean ratio ($S^2/\bar{x}$) and the index of aggregation ($I\delta$). This pattern could result in increase the efficiency of natural enemies for controlling those arthropod pests.

Keywords: Species composition, rice pests, natural enemies, rice ecosystem

Introduction

In terms of numbers and the diversity of species, arthropods are the most biologically successful group of animals (Triplehorn and Johnson, 2005). Almost three-quarters of all described animal species are insects and spiders (Borror et al., 1989; Gullan and Cranston, 2004; Triplehorn and Johnson, 2005). Their numbers far exceed all other terrestrial animal species, and are found in almost every terrestrial habitat on the Earth's surface. They have diversified to fill almost every environmental niche imaginable (Putman, 1983; Gullan and Cranston, 1994), making them one of the most crucial components of most terrestrial ecosystems. In agroecosystems, biodiversity performs a variety of ecological services beyond the production of food, including recycling of nutrients, regulation of microclimate and local hydrological processes, suppression of undesirable organisms and detoxification of noxious chemicals (Scherr and McNeely 2008; Fritz et al., 2011; Ane and Hussain, 2015). Likewise, the rice field agroecosystem is able to serve as a repository for the beneficial arthropods and harmful pests in rice paddy during all phases of rice growth (vegetative, reproductive and ripening) (Siregar et al., 2017).

Rice (*Oryza sativa* L.) is a staple cereal that nurtures the lives of the Thai people and of more than half the world's population. In economic terms, rice is a commodity ensuring prosperity and food security. In agriculture, rice is the second most cultivated cereal in the world, after wheat and more than 90% of the world's production is consumed in Asia. Nonetheless, to date, the total planting areas and rice production in Thailand has been declining overtime. The rice fields are considered as temporary humid areas characterized by rapid physical, chemical and biological alterations and they host a rich biodiversity especially of arthropods. In rice field ecosystems arthropods are found in an intermediary position in the food-chain and include herbivores, saprophytes, parasites and the predators of other animals. Subsequently they themselves are consumed as food by other predators, parasitoids and parasites (Siregar et al., 2017). Over 800 species of insects are able to damage the rice plant in several ways, although the majority of them caused minor damage (Pathak, 1968; 1970; Pathak and Khan, 1994; Way and Heong, 1994). Rice cultivation is a rapid changing ecosystem due to the change of activities during plantation. These activities affect the pest population and also the abundance of their natural enemies in the rice field (Tsutsui et al., 2018; Jauharlina et al., 2019). Some of the insect pests that are known to harm the rice plantation are brown plant hopper *Nilaparvata lugens* (Stål), rice shoot fly *Atherigona oryzae* (Malloch), white grub, *Leucopholis rorida* (Fabr.), white stem borer *Scirpophaga innotata* (Walker), and rice bug *Leptocoris oratorius* (Thunberg) (Thorburn, 2015). As a result of heavy insect pressure, the control of rice pests has relied heavily on insecticides (Magallona, 1989). Nonetheless, overuse and misuse of these insecticides have led to increasing problems of insecticide resistance, resurgence of pests and introduction of new pests, as well as beneficial insects' destruction (Pathak and Khan, 1994).

Natural control of insect pests is one of the strategies of the integrated pest management concept (IPM), which has been widely adopted in Thailand since the early 1970s (Wright et al., 2005). It has played a vital role in controlling most potential rice pests (Jervis, and Kidd, 1996). Several species of insect pests and natural enemies (predators & parasites) have been surveyed and reported in Thailand (Jairin et al., 2001; Wongsiri and Kovitvadhi, 1967; Wongsiri et al., 1981). Species diversity concerning abundance and species richness is a quantitative parameter of community structure (Magurran, 1988; May, 1975). Understanding of the species present and their role in the ecosystem could be significant in deciding whether or not to apply pesticides (van Emden and Williams, 1974). In general, species diversity and

complexity of association among species are important to the stability of the community. In view of the changing cultural practices the pest and natural enemy population dynamics needs closely monitoring for suitable amendments in rice IPM strategies. Information on distribution patterns of insect pests is used for data analysis to determine appropriate sampling plans and sample size, and to construct sequential sampling programs. Reliable sampling plans are necessary for periodic screening of pest-population densities where timely pest management decisions are essential (Kuno, 1991). Because the number of samples could vary with the distribution pattern of the pest, it is significant to determine the distribution pattern of rice key pests for specific areas and different seasons (Southwood, 1978). The data acquired are still insufficient to form a foundation and for developing an effective integrated pest management system in Thailand. The purpose of this current study was to determine species composition, together with determining the distribution pattern of arthropods (insects and spiders) that act as pests, predators, parasites or others in an organic rice paddy field.

Materials and methods

The experimental areas are located in the rice paddies of six major locations of lower Northern Thailand including Nakhonsawan, Phetchabun, Phichit, Phitsanulok, Sukhothai and Uttaradit provinces, respectively (Figure 1).

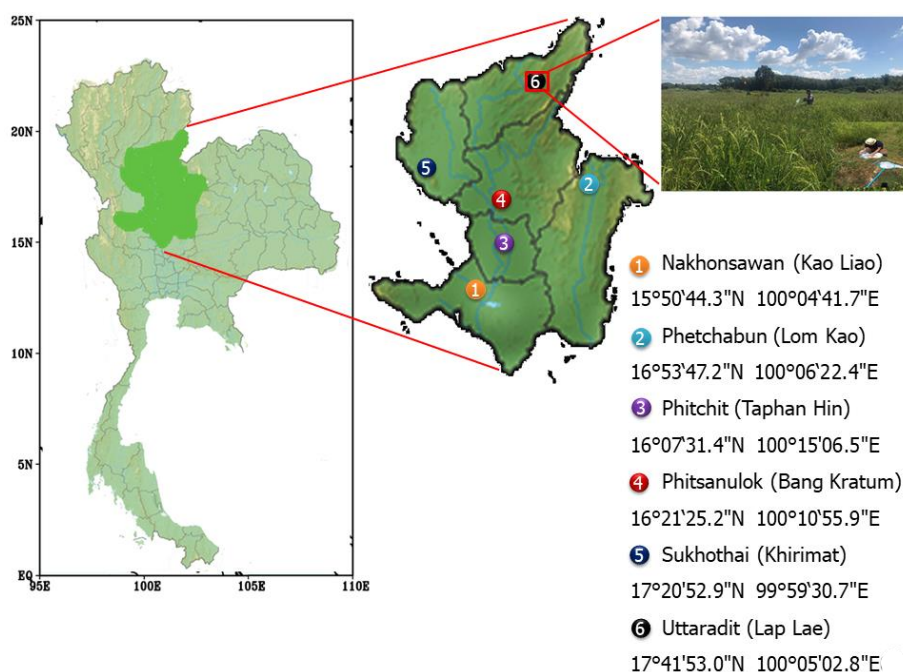


Figure 1 Map showing collection localities of insects and spiders inhabiting rice paddies in lower Northern Thailand, consisting of 1) Nakhonsawan, 2) Phetchabun, 3) Phichit, 4) Phitsanulok, 5) Sukhothai and 6) Uttaradit provinces.

The Khao Dawk Mali 105, a rice variety/line, was cultivated by farmers whose permission was sought to use as sample sites in this study. Samplings were conducted for each site across three years (2017-2019). Six plots were randomly sampled each having an area of

$\pm 1 \text{ hm}^2$ (100 m $\times$ 100 m each). Rice insect pests and natural enemies, predators and parasites, were gathered from each plot in the rice paddy during flowering stage. Airborne samples were collected using sweeping nets at 20 times/sampling surveyed spot. The surveyed line in each rice plot was walked as a diagonal. In addition to airborne samples, aquatic samples were also gathered from the water edge, surface and below the surface up to 30 cm. All insects and spiders collected were preserved in 95% alcohol. Afterwards, all samples were sorted and identified morphologically using the following references: Pathak and Khan (1994), Maloney et al. (2003), Rouyaree (2001) and Shepard et al. (1987).

For analysis of the insects and spiders, the assemblage data from rice paddies were gathered from six localities. All assemblage parameters (i.e. species richness, diversity) in this study were computed based on the average density of each arthropod species in each locality. The Shannon diversity index (H') was computed to measure diversity within a locality

(Shannon, 1948) using the formula $H' = -\sum_{i=1}^s p_i \ln p_i$ where s is number of species and P_i is the

proportion of the total sample belonging to i^{th} species. The diversity index criteria are as follows: H' less than or equal to 1 ($H' \leq 1$) means low diversity; H' greater than 1 and less than or equal to 3 ($1 < H' \leq 3$) means moderate diversity; and H' greater than or equal to 3 ($H' \geq 3$) means high diversity, respectively. Pielou's evenness index (J') was used to quantify differences in relative abundance using the formula $J' = H' / \ln S$, where S is the number of species (Ludwig & Reynolds, 1988). The evenness index (J') is categorized as follows: J' less than or equal to 0.05 ($0.00 < J' \leq 0.50$) indicates a depressed community; J' higher than 0.05 but less than or equal to 0.75 ($0.50 < J' \leq 0.75$) indicates an unstable community; and J' higher than 0.75 but less than or equal to 1.00 ($0.75 < J' \leq 1.00$) indicates a stable community, respectively. The Simpson diversity index (1-D) was also employed to measure the community diversity (Simpson, 1949) using the formula $D = 1 - [\sum n(n-1)] / [N(N-1)]$, where n is the number of individuals displaying one trait and N = the total number of all individuals. Index values (1-D) range from 0 - 1 by the following categories: the value less than 0.05 ($0.00 < 1-D < 0.50$) shows low dominance; the value higher than 0.05 but less than or equal to 0.75 ($0.50 < 1-D \leq 0.75$) shows moderate dominance; and the value higher than 0.75 but less than or equal to 1.00 ($0.75 < 1-D \leq 1.00$) shows high dominance, respectively. For species composition or relative abundance of different arthropods found, the similarity in species composition among the flowering stage of rice was determined by a cluster analysis with Sørensen distance measurement (Bray and Curtis, 1957; Sørensen, 1948). The distribution pattern of arthropods was statistically classified by calculating the Variance-to-mean ratio ($S^2/\bar{x}$) and Morisita's Index (I_d) (Morisita, 1959; 1962). For the distribution pattern, the value higher than 1.0 (> 1.0) shows the aggregated distribution; the value equal to 1.0 ($= 1.0$) indicates the random distribution; and the value lower than 1.0 (< 1.0) means the regular pattern of arthropods.

A pairwise t-test of Shannon diversity index (H') among different study sites of rice-growing areas was compared using the SPSS program (IBM Corp., 2017). The cluster analysis was carried out using PC-ORD 5.10 software (McCune and Mefford, 2006). A rarefaction model was used to compare arthropod diversity among these stages of rice growth. The Rarefaction Value and its 95% confidence interval were computed by EcoSim version 7.72 software (Gotelli and Entsminger, 2004). Afterwards, the values of every survey area were plotted as a function of sampling effort. With this plot, significant difference in species diversity is indicated by an absence of overlap in the confidence interval of rarefaction curves among eight different rice growth stages at maximum sampling effort (Colwell et al., 2004)

Results and discussion

From the rice paddies during the flowering stage in all six rice-planting areas over three years of surveys (2017-2019), a total of 6,459 individual rice pests and their natural enemies from 69 families (11 orders) were sorted from 36 sampling plots. The diversity and abundance of arthropod fauna varied in the paddy ecosystem as well as the growth stages of the paddy cultivation (Siregar et al., 2017). As results, of all the rice paddy-utilizing insects and spiders found, approximately 20.59% (21 out of 102 species) were Coleoptera, followed by Hemiptera (18 species), Diptera (17 species), Hymenoptera (15 species), Lepidoptera (14 species), Araneae (6 species), Orthoptera (5 species), Odonata (3 species) and the insect orders of Blattodea, Ephemeroptera and Thysanoptera held the lowest recorded species (only one species each), respectively (Figure 2).

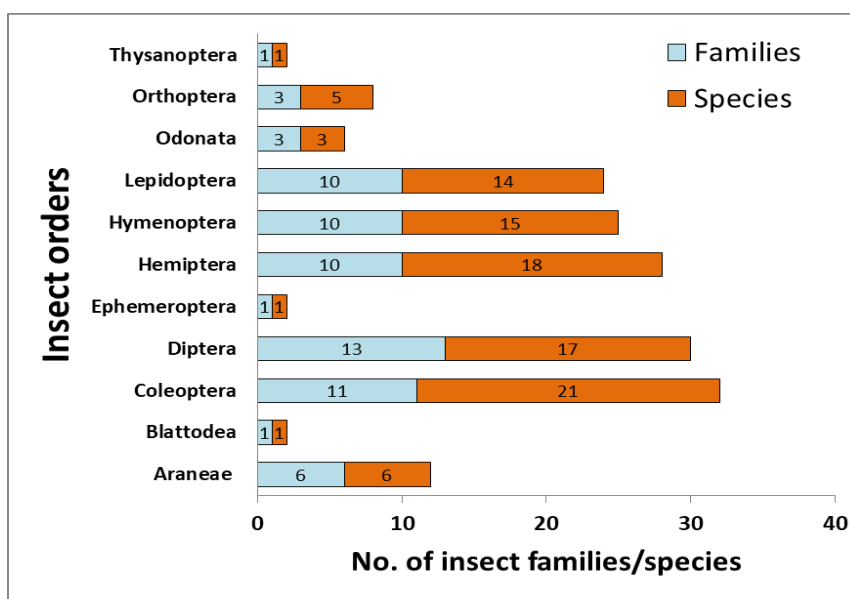


Figure 2 Taxonomic diversity of families and species of insects and spiders in rice field ecosystems in lower northern Thailand using the sweeping technique.

The total number of insect and spider species in each study site, from the highest to the lowest, was as follows: 73 species in Phitsanulok; 67 species in Nakhonsawan; 62 species in Uttaradit; 59 species in Phichit; 56 species in Phetchabun and 52 species in Sukhothai (Table 1). These findings correspond to Sorapongpaisal et al. (2011) who carried out research regarding species diversity of rice insect pests and natural enemies in an organic rice paddy field at Environmental Entomology Research and Development Center, Kasetsart University, Kampaengsaen Campus, Nakhon Pathom province. It was found that 52 species of insects and spiders comprising 20 species of insect pests, 25 species of insect natural enemies and 7 species of spiders were encountered. In this case, *Micraspis discolor* from the family Coccinellidae (O. Coleoptera) showed the highest number of individuals for the entire surveys, probably due to the high adaptive capacity of this lady beetle. It was revealed that the longevity of the *M. discolor* varied from 35 to 53 days with an average of 40.20 ± 1.00 days and 47.50 ± 0.82 days in male and female beetles, respectively (Chowdhury et al., 2008). During its lifespan, *M.*

discolor predated on rice insect pests such the brown planthopper (BPH), *Nilaparvata lugens* at the various rice stage of its development (Rouyaree, 2001; Sorapongpaisal et al., 2011), but it was most predominant at the active tillering or flowering stage of the paddy plants.

Table 1 Insect and spider diversity in different study sites.

Study sites	Ecological indices				
	S	H'	H _{max}	E	1-D
Nakhonsawan	67	3.110	4.205	0.740	0.935
Phetchabun	56	2.895	3.497	0.877	0.897
Phichit	59	3.234	4.078	0.793	0.940
Phitsanulok	73	3.055	4.290	0.712	0.919
Sukhothai	52	2.895	3.951	0.733	0.911
Uttaradit	62	3.040	4.127	0.737	0.911
Overall	102	3.677	4.625	0.795	0.945

Note: S = Number of species; H' = Shannon Wiener Index of diversity; H_{max} = lnS; E = Evenness (H'/H_{max}); 1-D = Simpson's diversity index

Apart from its longevity, *M. discolor* is more adaptive than any other insects because some factors such temperature might be influenced by its age-specific fecundity. Reproduction is a vital factor governing its ability to successfully invade new habitats (Omkar and Pervez, 2000; 2002). The age-specific fecundity is the most overlooked aspect of reproduction in ladybeetles, although female age plays a crucial role in progeny production and influences reproductive vigour (Jalali et al., 1999). Herein, the influence of temperature on age-specific fecundity of *M. discolor* was determined. It was revealed that the oviposition peak tended to shift towards younger females and the oviposition rate rose with increase in temperature from 20 °C to 27 °C (Omkar and Pervez, 2002). In the paddy fields observed in lower northern Thailand, the *M. discolor* collection contained as many as 1,068 individuals, or (16.535%) among insects and spiders collected. From these results, *M. discolor* could be improved in agricultural ecosystems to achieve biological control of pests.

Moreover, the three most common arthropod groups (Relative Abundance; RA) are 1) pests including *Edwardsiama hermanni* (O. Hemiptera: F. Cicadellidae) (4.149%), *Dyscinetus morator* (O. Coleoptera: F. Scarabaeidae) (2.013%) and *Nilaparvata lugens* (O. Hemiptera: F. Delphacidae) (1.858%), 2) predators including *Micraspis discolor* (O. Coleoptera Coccinellidae) (16.535%), *Agriocnemis pygmaea* (O. Odonata: F. Coenagrionidae) (4.443%) and *Oxyopes javanus* (O. Araneae: F. Oxyopidae) (5.032%) and 3) parasites including *Temelucha philippinensis* (O. Hymenoptera: F. Ichneumonidae) (0.897%), *Tropobracon schoenobii* (O. Hymenoptera: F. Braconidae) (0.712%) and *Goniozus nr. triangulifer* (O. Hymenoptera: F. Bethyridae) (0.650%), respectively. On the other hand, it is quite different from the relative abundance of insect and spider species on Boro rice ranked as ladybird beetle (45.85%)>damselfly (12.95%)> ant (9.46%)> wolf spider (8.24%)> jumping spider (4.90%)> ground beetle (4.36%)> wasp (4.22%)> tachinid fly (3.63%)> lynx spider (2.76%)> carabid beetle (2.18%)> mirid bug (1.40%); whereas in Aman rice it was ranked as ladybird beetle (39.30%)> damselfly (16.72%)> longjawed spider (14.85%)> dipteran fly (7.07%)> jumping spider (6.70%)> lynx spider (4.71%)> ant (3.91%)> wasp (3.80%)>carabid beetle (2.89%). Surprisingly, of all 11 beneficial insect and spider species, 8 species were beneficial insects, whereas 3 species were mite species (Rahman et al., 2017). In actual fact, ninety nine species

of parasites and 88 species of predators of rice insect pests have so far been discovered in Bangladesh (Wahiduzzaman, 1993). In this case, because rice fields are strictly monoculture with less diverse ecosystem consequently the diversity of beneficial insect and spider under different orders was relatively low (Rahman et al., 2017).

In general, twenty out of 70 rice insect species are considered as pests in the rice paddy fields of Thailand (Jaroenpol, 2008). Furthermore, over 800 species of insect in rice fields have been reported worldwide (Ane and Hussain, 2015). Of these, about 100 species attack rice and the rest are beneficial insects (Pathak, 1970). However, with more than 100 predatory and parasitic insect species –natural enemies- in the rice field, only 5-6 insect pests were found. This indicated that the ability of natural enemies to suppress insect pests to be under the economic threshold level (ETL) or economic injury level (EIL) in the organic rice paddy field (Jaroenpol, 2008).

Insect and spider (alpha) diversity is also shown in Table 1. The Shannon-Wiener index (H') for all insects and spiders of all sampling rice plots obtained by calculating the pooled data set of this study was considered relatively high (3.677), with high evenness ($E = 0.795$), and dominance by the Simpson index was high for all samplings ($1-D = 0.945$). H' values in each study sites ranged from 2.859 to 3.234. The highest index (H') was in Phitchit ($H' = 3.324$), following by Nakhonsawan ($H' = 3.110$), Phitsanulok ($H' = 3.055$), Uttaradit ($H' = 3.040$); and Phetchabun and Sukhothai ($H' = 2.895$), respectively. A set of individual t-tests of index values (H') among the five different locations of rice-growing areas was compared. Results of the comparison in each set of H' values were significantly different ($P < 0.05$) (Table 2).

Table 2 Grouped comparison of differences of species diversity indices (H') obtained from rice-growing areas, using the pairwise t -test.

Paires	t	df	Sig.
Na-Phet	-3.014	122.994	0.005*
Na-Phic	1.913	118.041	0.004*
Na-Phit	-0.9163	102.882	0.004*
Na-Su	-3.286	115.308	0.004*
Na-Utt	-1.001	127.632	0.005*
Phet-Phic	5.556	107.748	0.004*
Phet-Phit	2.880	92.844	0.003*
Phet-Su	0.001	104.746	0.004*
Phet-Utt	-3.272	117.817	0.004*
Phic-Phit	-3.7991	114.238	0.002*
Phic-Su	-6.278	110.215	0.003*
Phic-Utt	-0.282	105.045	0.003*
Phit-Su	-3.351	101.581	0.002*
Phit-Utt	2.421	113.253	0.004*
Su-Utt	2.421	113.253	0.004*

Note: Na = Nakhonsawan; Phet = Phetchabun; Phic = Phichit; Phit = Phitsanulok; Su = Sukhothai and Utt = Uttaradit

* = Significant difference ($p < 0.05$) and ns = not significantly different.

Differences in the arthropod (insect and spider) communities were evaluated according to the Bray-Curtis similarity (Figure 3) from qualitative-quantitative point of view. Hierarchical classification divided the arthropod communities into two groups. First group represented the

arthropod communities at Nakhonsawan and Phetchabun. The arthropods in Phichit (with a high dominance (%RA) of *Prodiplosis longifila* (Diptera: Cecidomyiidae) (12.619%), *Limnogonus nitidus* (Hemiptera: Gerridae) (12.024%) and *Agriocnemis pygmaea* (Odonata: Agrionidae) (8.929%) joined these two communities. The insect and spider communities of Phitsanulok, Uttaradit and Sukhothai represented the second group. It was found that the two surveyed sites in Phitsanulok and Uttaradit formed a sister group in which the communities of arthropod species were more similar to each other than those found at the various other sites. *Chaoborus astictopus* (Diptera: Culicidae) eudominated in the arthropod communities at Phitsanulok and Uttaradit. Meanwhile, *Chaoborus astictopus* (Diptera: Culicidae) (19.303%), *Agriocnemis pygmaea* (Odonata: Agrionidae) (11.642%) and *Prodiplosis longifila* (Diptera: Cecidomyiidae) (11.045%), demonstrated as a predominance in the arthropod communities at Sukhothai.

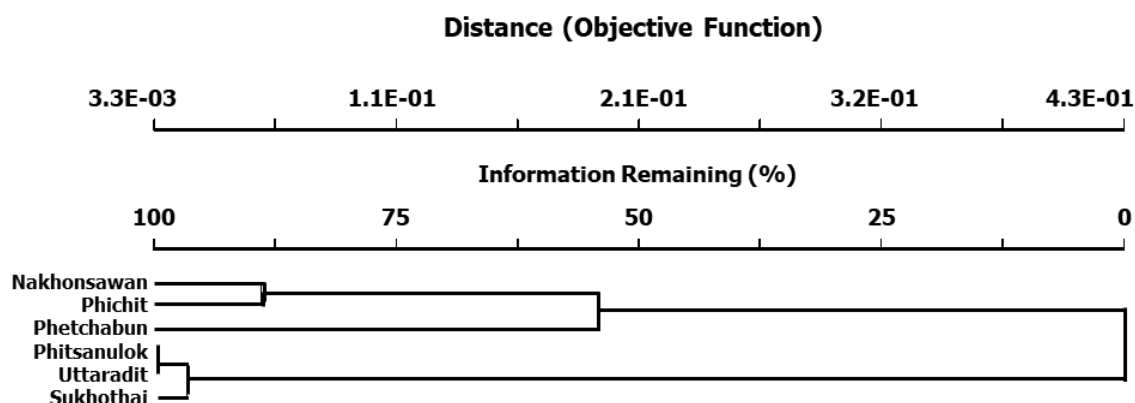


Figure 3 Dendrogram of the communities of insects and spiders in the various sites of rice-growing areas, obtained by cluster analysis.

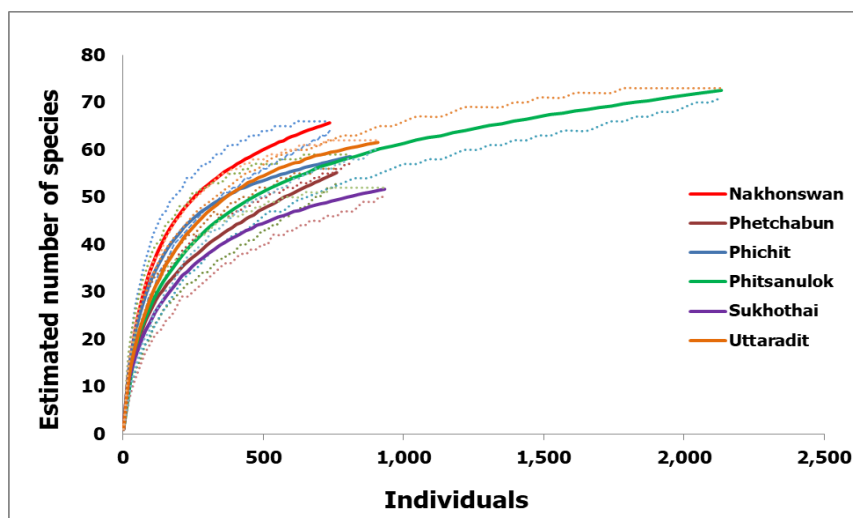


Figure 4 Species richness in the different localities of rice-growing areas by rarefaction curve with $\pm 95\%$ confidence intervals shown as dotted lines.

Additionally, the accumulation curves (observed and expected) in the several localities of rice paddy (Figure 4) exhibited a divergence between 95% confidence interval of these curves, with the curve gradually increasing when the survey ended. However, no such species-accumulation curve illustrated reaching an asymptote when all survey sites were pooled. Some beneficial insect and spider species are awaiting discovery. To obtain more arthropod fauna or the occurrence of them in the rice paddy during all stages of rice growth, i.e. seedling, tillering and heading should be further initiated. Additionally, numerous subplots or different techniques, namely sweep net, malaise trap or light trap should also be considered for rice arthropod sampling.

Table 3 Evaluation of spatial distribution pattern of insects and spiders for rice paddy field ecosystem in different locality.

Localities	Variance to mean ratio ($S^2/\bar{x}$)	Morisita's Index ($I\delta$)	Distribution pattern
Nakhonsawan	17.62	1.61	Clumped
Phetchabun	30.43	1.70	Clumped
Phichit	17.18	1.73	Clumped
Phitsanulok	22.31	4.67	Clumped
Sukhothai	19.67	2.14	Clumped
Uttaradit	31.02	1.93	Clumped
Total	268.02	5.18	Clumped

For the spatial distribution pattern of insects and spiders, the distribution indices -- Variance-to-mean ratio ($S^2/\bar{x}$) and Morisita's Index ($I\delta$) -- held significantly greater than 1.0 for all sites, indicating clumped distribution (Table 3). Most rice arthropods were exhibited an aggregated pattern of dispersal. This result was similar to the report of Poolprasert and Jongitvimol (2014) which recorded that all rice insect populations on the paddy plants at different phases of its development displayed clumped or aggregated distributions. This pattern might be caused by several factors such as intrinsic behavior of the individuals, response to the food and habitat resources distribution, level of water and hill density of the rice paddy field (Gogoi and Bora, 2012). In addition, a resource concentration in some areas has been proposed as the general cause of assemblage of most living things (Ricklefs and Miller, 2000; van Emden and Williams, 1974).

Conclusion

The current results indicated that such biodiversity parameters as richness, abundance, diversity and composition of these beneficial insect and spider and harmful pest communities differed among different localities of rice-growing zones. This might be caused from rice paddy management or other physical factors. However, all indices exhibited relatively high values in the rice field ecosystem in lower northern Thailand. Indicating that the rice ecosystem could be utilized as a great reservoir of beneficial insect and mite species and therefore it would foster a natural biological control process in rice-based agroecosystems throughout Thailand, especially in the lower northern region. The overall spatial distribution form of insects and spiders observed in this study mainly showed an aggregated distribution pattern. Subsequently, this pattern is able to enrich the efficiency of natural enemies for controlling those rice arthropod pests. In addition, the use of a sampling strategy that incorporates spatial distribution would be beneficial to provide the essential information for the development of an appropriate sampling

plan and an optimum sample size for rice integrated pest management (IPM). It would also reduce the cost of pest management strategies, together with decreasing the toxicity risk from synthetic insecticide application.

Acknowledgements

This research was financially supported by the Thailand (NRCT) and the Higher Education Research Promotion (HERP). The authors would like to thank the Faculty of Science and Technology, Pibulsongkram Rajabhat University for providing facilities and laboratory supports.

References

- Ane, N.U. & Hussain, M. (2015). Diversity of insect pests in major rice growing areas of the world. *Journal of Entomology and Zoology Studies*, 4(1), 36–41.
- Bray, J.R. & Curtis, J.T. (1957). An ordination of upland forest communities of southern Wisconsin. *Ecological Monographs*, 27, 325–349. <https://doi.org/10.2307/1942268>
- Borror, D.J., Triplehorn, C.A. & Johnson, N.F. (1989). *An Introduction to the Study of Insects*. Harcourt Brace College Publishing, New York; 873 pp.
- Chowdhury, S.P., Ahad, M.A., Amin, M.R. & Hasan, S.M. (2008). Biology of ladybird beetle *Micraspis discolor* (Fab.) (Coccinellidae: Coleoptera). *International Journal of Sustainable Crop Production*, 3(3), 39–44.
- Colwell, R.K. Mao, C.X. & Chang, J. (2004). Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology*, 85(10), 2717–2727. <https://doi.org/10.1890/03-0557>
- Fritz, L.L., Heinrichs, E.A., Machado, V., Andreis, T.F., Pandolfo, M., de Salles, S.M., de Oliveira, J.V. & Fiuza, L.M. (2011). Diversity and abundance of arthropods in subtropical rice growing areas in the Brazilian south. *Biodiversity and Conservation*, 20, 2211. <https://doi.org/10.1007/s10531-011-0083-3>
- Gogoi, H. & Bora, D. (2012). Spatial distribution of *Nymphula depunctalis* Guenée larvae (Lepidoptera: Pyralidae), an early vegetative pest of *Oryza sativa* L. *Academic Journal of Entomology*, 5(1), 41–46.
- Gotelli, N.J. & Entsminger, G.L. (2004). *EcoSim: Null Models Software for Ecology V. 7*. Acquired Intelligence Inc. & Kesey-Bear. Jericho, VT 05465, 2004. Accessed November 10, 2019. <http://garyentsminger.com/ecosim/index.htm>.
- Gullan, D.J. & Cranston, P.S. (1994). *The Insect: An Outline of Entomology*. Chapman and Hall, London, 491 pp.
- Gullan, D.J. & Cranston, P.S. (2004). *The Insect: An Outline of Entomology*, 3rd edn, Blackwell Science, 505 pp.
- IBM Corp. (2017). IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.
- Jairin, J., Phengrat, K., Khumma, S., Phomraksa, T. & Teangdeerith, S. (2001). The survey of natural enemies on rice insect pests in lower part of northeastern Thailand. *Thai Journal of Agricultural Science*, 19(1), 71–83. (in Thai).
- Jalali, S.K., Singh, S.P. & Biswas, S.R. (1999). Effect of temperature and female age on the development and progeny production of *Cryptolaemus montrouzieri* Mulsant (Coleoptera: Coccinellidae). *Entomon*, 24(3), 293–296.

- Jaroenpol, P. (2008). *Integrated Pest Management*. Nakhon Ratchasima province: Agriculture Department, Huai Tha-land district. (in Thai)
- Jauharlina, J., Hasnah, H. & Taufik, M.I. (2019). Diversity and Community Structure of Arthropods on Rice Ecosystem in Aceh. *AGRIVITA Journal of Agricultural Science*, 41(2), 316–324.
- Jervis, M. & Kidd, N. (1996). *Insect Natural Enemies: Practical Approaches to Their Study and Evaluation*. London: Chapman and Hall, 1996.
- Kuno, E. (1991). Sampling and analysis of insect populations. *Annual Review of Entomology*, 36, 285–304. <https://doi.org/10.1146/annurev.en.36.010191.001441>
- Ludwig, J. A., & Reynolds, J. F. (1988). *Statistical ecology: A primer on methods and computing*. New York: Wiley.
- Maloney, D., Drummond, F. A. & Alford, R. (2003). *Spider Predation in Agroecosystems: Can Spiders Effectively Control Pest Populations*. Orono: Maine Agricultural and Forest Experiment Station.
- Magallona, E.D. (1989). Effects of insecticides in rice ecosystems in Southeast Asia, in Bourdeau, P., Haines, J.A., Klein, W. & Krishna Mmti, C.R. (eds.). *Ecotoxicology and Climate*. New York: John Wiley & Sons, pp. 265–297.
- Magurran, A.E. (1988). *Ecological Diversity and Its Measurement*. New Jersey: Princeton University Press.
- May, R.M. (1975). Patterns of species abundance and diversity, in M. L. Cody, and J. M. Diamond, (eds.). *Ecology and Evolution of Communities*. Cambridge: Harvard University Press, pp. 81–120.
- McCune, B. & Mefford, M.J. (2006). *PC-ORD: Multivariate Analysis of Ecological Data V. 5*. Oregon: MjM Software, Gleneden Beach.
- Morisita, M. (1959). Measuring of the dispersion of individuals and analysis of the distributional patterns. *Memoirs of the Faculty of Science Kyushu University Series E*, 2, 215–235.
- Morisita, M. (1962). I_s-index, a measure of dispersion of individuals. *Researches on Population Ecology* 4(1): 1–7.
- Omkar & Pervez, A. (2000). Sexual dimorphism in *Propylea dissecta* (Mulsant), (Coccinellidae: Coleoptera). *Journal of Aphidology*, 14: 139–140.
- Omkar & Pervez, A. (2002). Influence of temperature on age specific fecundity of a ladybeetle, *Micraspis discolor* (Fabricius). *Insect Science and its Application*, 22, 61–65.
- Pathak, M.D. (1968). Application of Insecticide to paddy water for more effective rice pest control. *International Pest Control*, 10, 12–17.
- Pathak, M. D. (1970). Insect pests of rice and their control, in IRRI. (ed.). *Rice Production Manual*. Los Baños: University of the Philippines, College of Agriculture in cooperation with the International Rice Research Institute, pp. 171–198.
- Pathak, M. D. & Khan, Z. R. (1994). *Insect Pest of Rice*. Manila: International Rice Research Institute.
- Poolprasert, P. & Jongitvimol, T. (2014). Arthropod communities inhabiting organic rice agroecosystem. *International Conference on Agriculture and Medical Science (AEMS-2014)*. London, UK. 1-5. <http://dx.doi.org/10.15242/IICBE.C714014>
- Putman, R.J. (1983). *Carrion and Dung: The Decomposition of Animal Wastes*. Edward Arnold. London, UK. 62 pp.
- Rahman, M., Maleque, M.A., Uddin, M.S. & Ahmed, J. (2017). Abundance and Diversity of Beneficial Insect and Spider Species on Rice Ecosystem in Sylhet Region. *Journal of the Sylhet Agricultural University*, 4 (1), 63-70.
- Rouyaree, S. (2001). *Learning How to Manage the Integrated Rice Insect Pest*. Bangkok: Thai Agriculture Cooperative Press. (in Thai)
- Ricklefs, R.E. & Miller, G. L. (2000). *Ecology*, 4th ed. New York: W.H. Freeman and Company.

- Scherr, S.J. & McNeely, J.A. (2008) Biodiversity conservation and agricultural sustainability: towards a new paradigm of 'ecoagriculture' landscapes. *Philosophical Transactions of the Royal Society B*, 363, 477–494.
- Shepard, B.M., Barrion, A.T. & Lisinger, J. A. (1987). *Helpful Insect, Spiders, and Pathogens: Friends of the Rice Farmer*. Manila: International Rice Research Institution.
- Simpson, E. H. 1949. Measurement of diversity. *Nature*, 163, 688. <http://dx.doi.org/10.1038/163688a0>
- Siregar, A.Z., Tulus & Lubis, K.S. (2017). Diversity of Pest Insects in Paddy Field Cultivation: A Case Study in Lae Parira, Dairi. *International Journal of Trend in Research and Development*, 4(5), 2394-9333.
- Shannon, C.E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 272 379-423. <https://doi.org/10.1002/j.1538-7305.1948.tb00917.x>
- Sorapongpaisal, W., Tanasinchayakul, S., Promwong, W., Wootisarn, C. and Dokchan, P. (2011). Species diversity of rice insect pests and natural enemies in organic rice paddy field. *Journal of Agriculture*, 27(1), 39–48. (in Thai).
- Sørensen, T.A. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content, and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskab, Biologiske Skrifter*, 5, 1–34.
- Southwood, T.R.E. (1978). *Ecological Methods, with Particular reference to the study of insect populations*, 2nd ed. London: Chapman and Hall.
- Thorburn, C. (2015). The rise and demise of integrated pest management in rice in Indonesia. *Insects*, 6(2), 381–408. <https://doi.org/10.3390/insects6020381>
- Triplehorn, C.A., & Johnson, N. F. (2005). *Borror and DeLong's introduction of the study of insect*. 7th ed. Peter Marshall Publishing, California; 864 pp.
- Tsutsui, M. H., Kobayashi, K. & Miyashita, T. (2018). Temporal trends in arthropod abundances after the transition to organic farming in paddy fields. *PLOS ONE*, 13(1), e0190946. <https://doi.org/10.1371/journal.pone.0190946>
- van Emden, H.F. & Williams, G. (1974). Insect stability and diversity in agro-ecosystems. *Annual Review of Entomology*, 19, 455–475. <https://doi.org/10.1146/annurev.en.19.010174.002323>
- Wahiduzzaman M. (1993). *Studies of the insect predatory behavior of the blackdrongon rice ecosystem*. MS Thesis. Institute of Post graduate Studies in Agriculture. Gazipur, 59p.
- Way, M.J. & Heong, K.L. (1994). The role of biodiversity in the dynamics and management of insect pests of tropical irrigated rice: a review. *Bulletin of Entomological Research*, 84, 567–587. <https://doi.org/10.1017/S000748530003282X>
- Wongsiri, T. & Kovitvadhi, K. (1967). Insect pests of rice in Thailand, in IRRI, (ed.). *Major Insect Pests of the Rice Plant: Proceedings of a Symposium at the International Rice Research Institute, September, 1964*. Maryland: Johns Hopkins Press, pp. 571–574.
- Wongsiri, T., Wongsiri, N., Tirawat, C., Navavichit, S., Lewvanich, A. & Yasumatsu, K. (1981). Abundance of natural enemies of rice insect pests in Thailand, in N. Nōgyō and K. Sentā, (ed.). *International Symposium on Problems of Insect Pest Management in Developing Countries: Proceedings of a Symposium on Tropical Agriculture Research, Kyoto, August 6-7, 1980*. Japan: Tropical Agriculture Research Center, Ministry of Agriculture, Forestry and Fisheries, pp. 131–149.
- Wright, M.G., Hoffmann, M.P., Kuhar, T.P., Gardner, J. & Pitcher, S.A. (2005). Evaluating risks of biological control introductions: a probabilistic risk-assessment approach. *Biological Control*, 35, 338–347. <https://doi.org/10.1016/j.biocontrol.2005.02.002>

Research Article

A Study of Information Technology Risk Management of Government and Business Organizations in Thailand using COSO-ERM based on the COBIT 5 Framework

Sakchai Tangprasert

Department of Mathematics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok
1518 Pracharat 1 Road, Wongsawang, Bangsue, Bangkok, 10800, Thailand
E-mail: Sakchai.t@sci.kmutnb.ac.th
Received: 20/01/2020; Revised: 03/06/2020; Accepted: 15/06/2020

Abstract

Information technology (IT) risk management plays an important role in controlling security and building confidence in using IT system services. Many organizations have focused on risk management for IT security. In this study the committee of sponsoring organizations of the treadway commission (COSO) enterprise risk management (ERM) was used in risk management by identifying risk factors in accordance with the requirements of COBIT 5 areas that need to be controlled. In addition, 2 experiments were conducted with 3 government organizations and 3 business organizations in order to evaluate the performance of IT security control. The first experiment identified the risk levels with risk assessment, which provided the risk levels that need risk response in accordance with COBIT 5 framework implementation life cycle. From the second experiment, it was found that the risk management using all 7 phases of the COBIT 5 implementation life cycle decreased the risk levels for both government and business organizations. However, moderate and low risk levels were still observed which need to be managed in order to keep them at a very low level. In addition, it was found that the risks of government and Business organizations were different, which was a result of differences in obstacles and the context of the problems encountered in risk management. The findings of this study provide the guidelines for developing a framework for IT risk management in the future.

Keywords: IT Risk Management, COSO-ERM, COBIT 5 framework

Introduction

In a globalized world, business competition due to both domestic and international factors affects the changes and drives the workflow of both government and business organizations. In order to speed up the operations of the organizations, reduce workloads and increase operational performance (Tangprasert, 2019), IT systems have been used for data storage, communication, operations including providing services to outsiders. In addition to the benefits, on the other hand, there are many risks (Prasetyo & Sucahyo, 2014) affecting the operations, information, and services of the organization. In order to gain confidence and acceptance of IT services, it is necessary to manage the risks (Wijanarka, 2014), such as risks from human error, employee compliance violation, losing important data and IT system

disruption and so on. The risks mentioned above negatively affect the image and also result in loss of budget and

business opportunity. This study therefore adopted the COSO-ERM (COSO, 2004) principle to define the risk criteria for risk assessment and risk response. This research applied COBIT 5 areas for determining guidelines and topics in risk assessment (Su & Zhao, 2011) which is applicable and consistent with information technology risk management of both government and Business organizations (Khrisna & Harlili, 2014). This research has been experimental with 3 government organizations and 3 Business organizations with experimental twice. The 1st experimental for risk assessment of information technology in accordance with the framework of COBIT 5 areas. After the organization was reported the risk of information technology and risk management in accordance with the principles of risk response, then performed the 2nd experimental to assess the result of risk change. The aim of this study to determine the valuation of issues, effects and solutions to be a guideline for indicating the conclusions from this research to continue developing the technology risk prevention framework for government and Business organizations.

IT risk management

IT risk management is a risk management process that involves the use of IT systems to identify threats, find the guidelines for preventing the impact of system errors, data retention, and securing and getting the organization ready to deal with the problems all the time whether it is hardware, software, or data problems (Tangprasert, 2019). These problems are related to personnel at all levels in the organization, from operators, managers to top executives. Everyone must be involved in the organization's operations to ensure standardized processes in order to achieve work performance and the objectives of the organization (Foley, 2009). The important risk management principles of COSO-ERM (COSO, 2004) are as follows (Fang, 2005):

Risk identification

Risk identification or event identification is a process of identifying events in which the organization considered that they may hold potential risks (Islam & Dong, 2008) and may affect the strategic plans and work processes and cause damage to the organization. In risk identification, both external risk factors (Zhu et al., 2012) such as technological changes, competitive conditions, customers, trading partners, economic conditions, inflation, politics and law and internal risk factors such as management policies, employee performance, work processes and IT systems used must be analyzed.

Risk assessment

Risk assessment or risk analysis is an assessment or analysis of the likelihood and impact of the potential events in the organization's operations. Each event affects the organization at a different level and with a different probability of occurrence (Roberts, 2017). Therefore, the risk matrix consists of 2 dimensions (Xiaosong et al., 2009). Dimension 1 is likelihood which considers the probability level of occurrence. Dimension 2 is impact which involves an analysis of the level of impact of an event. The risk matrix in this study is shown in Figure 1.

		Impact				
		1 = Negligible	2 = Minor	3 = Moderate	4 = Significant	5. Severe
Likelihood	5 = Very Likely	Moderate 5 x 1 = 5	Moderate 5 x 2 = 10	High 5 x 3 = 15	Extreme 5 x 4 = 20	Extreme 5 x 5 = 25
	4 = Likely	Low 4 x 1 = 4	Moderate 4 x 2 = 8	High 4 x 3 = 12	High 4 x 4 = 16	Extreme 4 x 5 = 20
	3 = Possible	Low 3 x 1 = 3	Low 3 x 2 = 6	Moderate 3 x 3 = 9	High 3 x 4 = 12	High 3 x 5 = 15
	2 = Unlikely	Very Low 2 x 1 = 2	Low 2 x 2 = 4	Low 2 x 3 = 6	Moderate 2 x 4 = 8	High 2 x 5 = 10
	1 = Rare	Very Low 1 x 1 = 1	Very Low 1 x 2 = 2	Low 1 x 3 = 3	Low 1 x 4 = 4	Moderate 1 x 5 = 5

Figure 1. Risk Matrix

Dimension 1 Likelihood: level 1 = rare, the event has almost no chance of occurring or the event likely to occur once in 1-2 years; level 2 = unlikely, the event can occur at some time or likely to occur once in 6-12 months; level 3 = possible, the event might occur several times or likely to occur once in 3-6 months; level 4 = likely, the event will probably occur on a monthly basis and level 5 = very likely, the event is likely to occur on a daily or weekly basis.

Dimension 2 Impact: level 1 = negligible, very small damage, the impact may not be observed, or affects or only results in unreasonable waste of resources; level 2 = minor, minor damage causing slight delay in minutes in work or little budget damage; level 3 = moderate, moderate damage causing 1-2 hour delay in work and damage to the budget at an extent that need to be acknowledge or approved by the management; level 4 = major, critical damage causing delays in work for more than 1-2 hours or a day or more or causing damage to the budget at an extent that needed to be acknowledged or approved by the management, or the organization loses business opportunities and potential benefits; level 5 = severe or catastrophic, disaster, damage is very high causing system disruption for many days or permanent disruption which is unable to continue or very difficult for recover.

From the risk matrix, the risk can be classified into 5 levels, consisting of 1 - 3 = very low, 4 - 6 = low, 7 - 10 = moderate, 11 - 16 = high and 17 - 25 = extreme to provide a basis for risk identification and further organization's risk response.

Risk response

In risk response and management, the actions that should be taken to manage the risks are chosen according to the risk assessment results (Jing et al., 2014), considering the level of risk occurring and the value of risk management to the risk tolerance level (Jung & Kim, 2015). There are 4 risk response strategies (Gonen, 2011) as follows:

1) Risk avoidance involves managing an extremely high-level risk that is unacceptable and decision must be made in the operations, such as canceling the project or activity, purchasing new materials as replacement and using other alternative processes and so on.

2) Risk sharing involves distributing or transferring the risk to others in order to share responsibility.

3) Risk Reduction involves improving work systems or designing the new working methods in order to reduce the chance of occurrence or the impact to an acceptable level of the organization.

4) Risk acceptance involves accepting the risk because it is not worthwhile to manage, control or prevent the risk, since it is likely to happen less and has little effect (Dhukaram et al., 2011).

COBIT 5 areas

The Cobit 5 Framework is an internationally accepted guideline for control and IT organization management to achieve organization objectives that help enterprises optimum value from IT management by balancing of risk and benefit to make the most of resource. (ISACA, 2012) The benefit of Cobit 5 Framework are applicable to almost every organization including government and business organization that different in process and business scale.

The COBIT 5 framework is the business framework for IT governance that aims to supervise and manage the organization's IT systems and to reduce the potential risks of IT Systems (Susanti & Sucahyo, 2016). It can facilitate the IT security management and increase performance to maximize benefits (Wolden et al., 2015). The COBIT 5 framework defines the scope of control with COBIT 5 areas as follows (ISACA, 2012):

Area 1 align, planning and organize (APO): Corporate planning and management allows corporate executives to align IT and business strategies, encourages workers to gain knowledge and understanding of IT risks and management strategies for potential IT risks in the organization.

Area 2 build, acquire and implement (BAI): The provision and installation of IT projects to be developed or implemented which can be used to solve problems or support the operations of the organization on time according to the specified budget.

Area 3 deliver, service and support (DSS): Delivery and maintenance of the new information system to be used so that they can work efficiently and meet the organization's needs and to effectively control changes that may affect the operations of the organization.

Area 4 monitor, evaluate and assess (MEA): Monitoring of IT services can support operations in accordance with the goals and mission of the organization so that IT systems can be used in the operations of the organization for better performance and safety, including confidentiality, integrity and availability.

Area 5 evaluate, direct and monitor (EDM): The evaluation of IT system performance can detect potential problems before occurrence and can support the organization's mission. Control, compliance, and performance risks are monitored and reported to the top executives.

All COBIT 5 areas allows the organization to assess their operations, supervision and management of IT systems in order to identify the risks that need to be controlled to keep them at the risk tolerance (Jung & Kim, 2015). This results in effective and safe use of IT systems for the benefit of all stakeholders and organization (ISACA, 2012).

Experimental design

This study is an experimental research for assessing the risks of information security management system in order to ensure the reliable and comprehensive survey results. The tools used in this study included interview, observation and record to find facts according to

predetermined objectives (Park et al., 2007). For the diversity of risk assessment and for the purposes of comparison to find the guidelines for risk management, 6 organizations including 3 government organizations and 3 business organizations. To show the various the organization culture and different processes in both of government and business including the number of staff that effect the organization risk and problem solving. In addition, period of time and budget are the important factors on risk management in the different employee scale that selected in range 1) 100 people 2) 100-500 people 3) 500 people respectively from 1) office level, 2) division level and 3) department level of ministry respectively. This study conducted the same method in government and business organization namely 1) limited partnership, 2) company limited (DBD, 2019) and 3) the public limited company (PLC) (SET, 2019) respectively. They were selected in this study using purposive sampling. Interview and data collection were performed with 5 people from each of these 6 organizations, a total of 30 people.

The sample consisted of personnel who were involved in the management and supervision of IT systems at different levels such as chief information officer (CIO), IT consultant, IT manager, system engineer, system administrator and IT support. Two experiments were conducted, the first experiment was conducted in the second quarter of 2018 and the second experiment was conducted in the second quarter of 2019. The experiments were conducted using 3 methods: 1) interview, 2) observation and 3) survey based on empirical evidence as an experimental framework for this study. Risk assessment was conducted by 3 people; researcher, IT external auditor and IT manager of each organization.

The implementation life cycle (Youssfi et al., 2014)

In the first experiment, COBIT 5 areas were used for risk identification in order to cover the scopes of COBIT 5 areas and risk assessment was also performed to determine risk level of the organization and to manage the risk response using appropriate strategies. COBIT 5 implementation life cycle was used in order to apply the risk response strategies to control existing risks. After completion of the first experiment, the second experiment was carried out with the same experimental process as the first experiment. COBIT 5 implementation life cycle consists of 7 phases (ISACA, 2012) as shown in Figure 2.



Figure 2. COBIT 5 framework implementation life cycle (ISACA, 2012)

Phase 1: The starting point for raising awareness and building a consensus in the need to solve problems at both the management and the operator levels.

Phase 2: Defining the scope of operations as a framework for achieving organizational goals and IT goals.

Phase 3: Setting targets for improvement, specify details that can be used as a comprehensive and rapid solution to problems

Phase 4: Planning an integrated problem-solving process to put into action by defining the projects that are in line with the organization's business, as well as having backup plans for dealing with changes.

Phase 5: The comprehensive resolution process after consideration is established as a daily operation.

Phase 6: For sustainable operations, measures, evaluations and monitoring of operations must be defined in accordance with the organization's business.

Phase 7: Review of the overall success of the implementation initiative, specify the supervision and management of IT systems that should be added and promoted for continuous development.

The multiple iterations of the processes in this cycle will lead to effective IT governance and management.

Experimental results

Risk assessment was conducted by 3 risk management experts without the personnel of the organization in order to prevent prejudice and bias which might affect the assessment. The average risk levels of each organization and the overall average risk levels across these 3 organizations were also determined using the results from all 3 experts. Risk assessment in the first experiment provided the risk levels of each organization. The risk levels of these organization were calculated by average risk from researcher, IT external auditor and IT manager as shown in table 1.

Table 1. The risk assessment results in the first experiment

COBIT 5 areas	Government organizations						Business organizations					
	A	B	C	$\bar{X}$	S.D.	Risk	D	E	F	$\bar{X}$	S.D.	Risk
1: APO	9.3	7.7	10.3	9.1	1.7	Moderate	15.7	15.7	20.3	17.2	3.1	Extreme
2. BAI	9.0	8.7	8.3	8.7	0.7	Moderate	20.0	23.3	20.3	21.2	3.8	Extreme
3. DSS	14.3	12.3	10.7	12.4	2.2	High	13.3	15.7	18.7	15.9	2.8	High
4. MEA	20.3	21.7	23.3	21.8	3.1	Extreme	9.3	9.3	9.0	9.2	2.5	Moderate
5. EDM	21.7	23.3	25.0	23.3	2.4	Extreme	14.3	12.3	15.7	14.1	2.4	High

According to the risk assessment in the first experiment, government and business organizations had different risk levels in each area due to many factors and difference in operating processes as follows:

Area 1 align, planning and organize (APO): The government organizations in this study had distinct organization chart and IT master plan defining development plans, problem prevention, or the use of IT systems over the long term of 3-5 years. However, they still lack concrete implementation and there was no an annual IT action plan in accordance with such long-term policy plan. In addition, there was an IT security plan which could identify the risks

affecting IT systems but there was no business continuity management (BCM) operation to prevent risks and to provide a framework for operations in case of system errors or disruptions. In addition, it was found these organizations had no IT policies that specify rules, regulations, including penalties for the use of IT systems, resulting in the moderate level of risk with an average of 9.1. For business organizations, it was found that the risk level was higher than to those of the government organizations because most of them had organization charts, but they had not been updated, making it difficult to determine the chain of command and there was no policy related to complete IT systems. In addition, the policies were not complete, including the IT master plan, IT action plan, IT security plan, BCM and IT Policy, resulting in the extreme level of risk with an average of 17.2.

Area 2 build, acquire and implement (BAI): According to the results, the government organizations had drafted the terms of reference (TOR), procurement of IT projects and specifications and installation procedures were clearly defined in the TOR. However, budget requests and TOR assignments were often delayed and unable to keep up with rapidly changing IT systems, resulting in the moderate level of risk with an average of 8.7. As for the business organizations, the risk observed was a result of procurement due to unclear specifications and some projects did not have an agreement covering the services of vendor. As a result, the installation of IT Systems was not in line with the objectives, additional costs and longer installation time than scheduled. This resulted in the extreme level of risk with an average of 21.2.

Area 3 deliver, service and support (DSS): For the government organizations in this study, the deliveries were inspected by the commission under the terms of the TOR, making them met the specifications. However, personnel and budgets were inadequate to maintain the system and the maintenance contracts for many systems expired and the contract has not been renewed. This resulted in the high level of risk with an average of 12.4. However, the opposite findings were observed for the business organizations. As there was no specific delivery committee appointed, and there was no documentation for inspecting the completeness of the delivery. However, with the use of this system, the person responsible for monitoring, problem solving and maintenance were appointed so that IT Systems could meet the operational needs of the organization. This resulted in the high level of risk with an average of 15.9.

Area 4 monitor, evaluate and assess (MEA): It was found that in the government organizations in this study, IT services were not monitored and evaluated. In addition, these organizations used the file sharing system but there was no data classification policy specifying the type and priority of the data etc. This resulted in an extreme level of risk with an average of 21.8. On the other hand, as for the business organizations, key performance indicators (KPI) or objectives and key results (OKR) were defined for assessing the information services. However, there was still a problem with data classification policy as the government organizations. This resulted in the moderate level of risk with an average of 9.2.

Area 5 evaluate, direct and monitor (EDM): The government organizations in this study did not periodically monitor the IT systems to ensure their availability. In addition, the established IT security plan had not been used to control the risks in concrete, and there was no BCM plan including business continuity plan (BCP) testing which led to lack of confidence that the system will work when problems or errors actually occur. This resulted in the extreme level of risk with an average of 23.3. For private sector organizations, IT systems were monitored regularly and there was also the preventive maintenance (PM). Risk management and BCM were also the problems as government organizations. This resulted in the high level of risk with an average of 14.1.

The first risk assessment was performed with the real situation to show the problem issue and risk in these organizations. Before assessment, baseline study was recored and applied COBIT 5 areas as a guideline of IT software management. After the risk assessment in the first experiment, the organization managed their risk response by selecting the appropriate strategies in each area using the COBIT 5 implementation life cycle in order to apply such risk response strategies in controlling the existing risks. After the completion of all operations, the second experiment was conducted and the results by the risk levels of these organization were calculated by average risk from researcher, IT external auditor and IT manager as shown in table 2.

Table 2. The risk assessment results in the second experiment

COBIT 5 areas	Government organization						Business organization					
	A	B	C	$\bar{X}$	S.D.	Risk	D	E	F	$\bar{X}$	S.D.	Risk
1: APO	4.3	2.3	2.3	3.0	1.2	Very Low	5.3	4.7	5.7	5.2	0.9	Low
2. BAI	2.3	2.3	2.7	2.4	1.2	Very Low	10.7	8.7	9.7	9.7	1.5	Moderate
3. DSS	5.7	7.7	4.3	5.9	1.7	Low	5.7	4.7	5.3	5.2	0.6	Low
4. MEA	7.7	9.3	11.3	9.4	1.8	Moderate	2.7	3.3	2.3	2.8	0.9	Very Low
5. EDM	8.3	6.7	7.7	7.6	1.7	Moderate	4.3	2.7	4.7	3.9	1.2	Low

Area 1 align, planning and organize (APO): For the government organizations, the annual IT action plan was prepared and implemented in accordance with the IT master plan. BCM was also implemented and IT policy was established. An average risk level of the government organizations decreased by 3.0 and was at very low level. For the business organizations, it was found that the organization chart was reviewed and updated and with various IT policies. However, some policies were incomplete and did not comply with actual operations. An average risk level of the business organizations decreased by 5.2 and was at a low level.

Area 2 build, acquire and implement (BAI): For the government organizations, after an annual IT action plan was established and in line with the IT master plan, each organization could plan its budget requests and establish TOR in time. An average risk level of the government organizations decreased by 2.4 and was at a very low level. For the business organizations, it was found that the contract with the vendors was entered into but there was still a clear specification as well as the output from the operation in accordance with the organization's objectives. An average risk level of the business organizations decreased by 9.7 and was at a moderate level.

Area 3 deliver, service and support (DSS): For the government organizations, it was found that the workload was assigned to the personnel for the maintenance of information systems. However, it did not cover all existing systems. An average risk level decreased of the government organizations by 5.9 and was at a low level. The business organizations defined the clear delivery criteria and the inspection committees were assigned for the high-value projects. An average risk level of the business organizations decreased by 5.2 and was at a low level.

Area 4 monitor, evaluate and assess (MEA): The government organizations defined the IT service assessment criteria and data classification policies. However, in the implementation, the lack of cooperation from management and the lack of enforcement with

the operators were observed. An average risk level of the government organizations decreased by 9.4 and was at a moderate level. The similar results were observed for the business organizations in this study. It was found that the business organizations defined the data classification policy, but the management gave priority to this policy and instructed the operators to strictly comply with the new policy. An average risk level of the business organizations decreased by 2.8 and was at a very low level.

Area 5 Evaluate, Direct and Monitor (EDM): It was found that the government organizations in this study assigned the person responsible for periodic system monitoring. However, the lack of cooperation and implementation were observed in risk control according to IT security plan and business continuity plan. An average risk level of the government organizations decreased by 7.6 and was at a moderate level. The business organizations monitored their IT system and risk prevention and BCM has been implemented but not yet complete. An average risk level the business organizations decreased by 3.9 and was at a low level.

Discussions and conclusions

From both experiments, before and after using COBIT 5 framework implementation life cycle according to the requirements of the COBIT 5 framework in all 5 areas, the comparison of risk assessment results between government and business organizations is shown in figure 3.

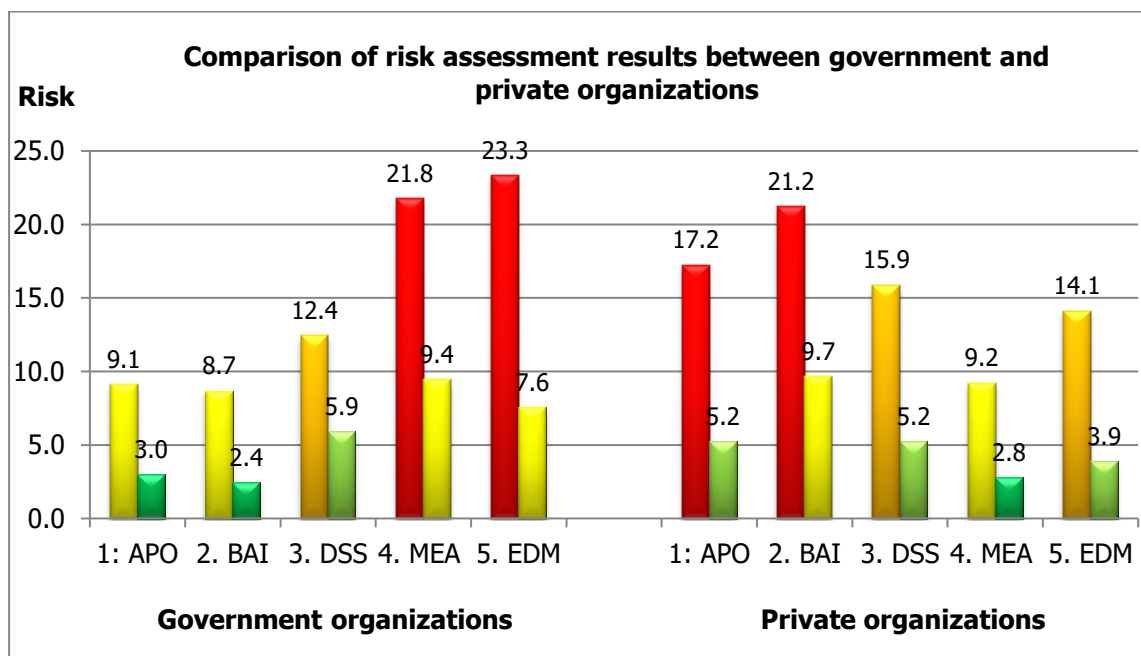


Figure 3 the comparison of risk assessment results between government and business organizations

The study of IT risk management in government and business organizations in Thailand using COSO-ERM based on COBIT 5 framework showed that the use of COBIT 5 implementation life cycle in risk control reduced organizational risks. In the first experiment, the

author observed that management and personnel in all 6 organizations recognized the importance of IT risk management. However, they lacked knowledge and understanding of IT system risk management in order to prevent the potential risks caused by errors or omissions of IT governance and to enhance the organization's business performance. After COBIT 5 areas were used to determine the extent of control along with the risk assessment in the first experiment, the government organizations showed a lower risk level in area 1 APO than that of the business organization. This is because government policies require government organizations to establish and prepare regulations, policies and plans, unlike business organizations that do not focus on policies, resulting in an extreme level of risk. For area 2 BAI, it was found that the government organization defined TOR with the qualifications and delivery, while the business organization did not define a clear TOR and scope of IT system development, resulting in an extreme level of risk and higher than that of the government organization. For area 3 DSS, both government and business organizations showed a high level of risk. The government organizations had IT delivery inspection, but the system administrator was not assigned. While the business organization did not have IT delivery inspection, but the system administrator was assigned. For area 4 MEA, it was found that IT service in the government organizations was monitored and evaluated, however there was no data classification police, resulting in an extreme level of risk. On the other hand, the business organizations focused on evaluation of performance of personnel at all levels and IT service with clear measures using KPI or OKR. For area 5 EDM, it was found that the government and business organizations had different levels of risk. The government organizations showed extreme level of risk because IT systems in these organization were not evaluated and there was no periodic system monitoring. In addition, the BCM was not given priority and awareness. While the business organizations gave priority to IT system monitoring and preventive maintenance (PM) and took regular actions.

From the second experiment, it was found that, the risk management using all 7 phases of the COBIT 5 implementation life cycle decreased the risk levels of both government and business organizations. However, it can be seen that there are still moderate and low levels of risk that need to be managed and controlled. The risk tolerance of all these 6 organizations should be at very low level for all areas or at low level for some areas. In this study, the differences in problems encountered by the government organizations were observed. In the initial stage of IT system development, the moderate level of risk was observed in area 1 APO and area 2 BAI. After the development of IT systems, there was no maintenance management, raising awareness of cooperation in the inspection and audit and surveillance for potential problems, resulting in an extreme level of risk in area 4 MEA and area 5 EDM. This is different from the business organizations which encountered problems in the initial stage of IT system procurement and development project. However, person in charge was assigned and the IT system maintenance was performed periodically, as can be seen from the extreme risk levels in area 1 APO and area 2 BAI. At the completion of IT system development, the risk levels in area 3 DSS, area 4 MEA and area 5 EDM decreased. The related factors consist of personnel who are ready to solve problems and sufficient time and budget for risk management. These 3 factors are very important factors allowing the organization to decide in risk response to keep the level of risk at its risk tolerance.

From risk identification in risk management indicated the different of organization scale has effected on the different risk management results which appeared on the cooperation and risk management acceptance especially small organization could manage and solve easier than large organization because of the good employee communication in a smaller group. The benefit of organization after applied risk management brought IT security system in that

organization. IT General Control was controlled by IT Governance who implement IT policy especially the confidence of director staff and other institutions related IT problem and business continuity management.

Future work

Risk assessment of government and business organizations reveals obstacles and context of the problems encountered in risk management and also provides the guidelines for developing the framework to be used in risk management in the future.

References

- COSO. (2004). *Enterprise risk management – integrated framework (Executive summary and framework) The committee of sponsoring organizations of the treadway commission (COSO)*
- DBD. (2019). DBD data warehouse+ The source of thailand's business information: department of business development, ministry of commerce. *Department of business development, ministry of commerce*. Department of business development, ministry of commerce. Retrieved June 26, 2019, 2019, from <https://www.dbd.go.th>
- Dhukaram, A. V., Baber, C., Elloumi, L., Beijnum, B. van & Stefanis, P. De. (2011). *End-user perception towards pervasive cardiac healthcare services: Benefits, acceptance, adoption, risks, security, privacy and trust*. Proceedings of the 2011 5th International conference on pervasive computing technologies for healthcare (Pervasive health) and workshops.
- Fang H. (2005). *IC research about the AIS under the environment of the networks*. Proceedings of the 7th international conference on Electronic commerce, Xi'an, China.
- Foley, S.N. (2009). *Security risk management using internal controls*. Proceedings of the first ACM workshop on Information security governance, Chicago, Illinois, USA.
- Gonen, A. (2011). *Optimal risk response plan of project risk management*. Proceedings of the 2011 IEEE International conference on industrial engineering and engineering management.
- ISACA. (2012). *Cobit 5 a business framework for the governance and management of enterprise IT*. United states of america: ISACA knowledge center.
- Islam, Shareeful & Dong, Wei. (2008). *Human factors in software security risk management*. Proceedings of the first international workshop on Leadership and management in software architecture, Leipzig, Germany.
- Jing, Z., Zhao, Y., Qiang, H, & Tian, W. (2014). *Risk response strategy research of water and sediment regulation system running schemes in the yellow river*. Proceedings of the 26th Chinese control and decision conference (2014 CCDC).
- Jung, Yoonhyuk & Kim, Seongcheol. (2015). Response to potential information technology risk: Users' valuation of electromagnetic field from mobile phones. *Telematics and informatics*, 32(1), 57-66. doi: <https://doi.org/10.1016/j.tele.2014.03.002>
- Khrisna, A, & Harlili. (2014). *Risk management framework with COBIT 5 and risk management framework for cloud computing integration*. Proceedings of the 2014 International conference of advanced informatics: Concept, theory and application (ICAICTA).
- Park, J., MacRae, H., Musselman, L.J., Rossos, P., Hamstra, S.J., Wolman, S., & R.K., Richard K. (2007). Randomized controlled trial of virtual reality simulator training: transfer to live patients. *The American Journal of Surgery*, 194(2), 205-211. doi: <http://dx.doi.org/10.1016/j.amjsurg.2006.11.032>

- Prasetyo, S., & Sucahyo, Y. G. (2014). *Information security risk management planning: A case study at application module of state asset directorate general of state asset ministry of finance*. Proceedings of the 2014 International conference on advanced computer science and information system.
- Roberts, D. T. (2017). *Applying risk assessment at the worker level*. Proceedings of the 2017 Petroleum and chemical industry technical conference (PCIC).
- Tangprasert, S. (2019, 13 - 15 December 2019). *Internal control for information technology security controls with COBIT 5 framework in enterprises*. Proceedings of the International conference robotics, informatics, and intelligence control technology (RIIT) 2019, Bangkok, Thailand.
- SET. (2019). Companies or securities in focus. Retrieved on June 6, 2019, from <https://www.set.or.th>
- Su, X., & Zhao, X. (2011). *Analysis on effects of risk management level on internal control*. Proceedings of the 2011 IEEE 18th International conference on industrial engineering and engineering management.
- Susanti, R. Y. & Sucahyo, Y. G. (2016). *Information technology governance evaluation and processes improvement prioritization based on COBIT 5 framework at secretariat general of the Indonesian house of representatives*. Proceedings of the 2016 4th International conference on information and communication technology (ICoICT).
- Wijanarka, H. (2014). *IT risk management to support the realization of IT value in public organizations*. Proceedings of the 2014 International conference on ICT for smart society (ICISS).
- Wolden, Mark, Valverde, Raul & Talla, Malleswara. (2015). The effectiveness of COBIT 5 information security framework for reducing cyber attacks on supply chain management system. *IFAC-PapersOnLine*, 48(3), 1846-1852. doi: <https://doi.org/10.1016/j.ifacol.2015.06.355>
- Xiaosong, L., Shushi, L., Wenjun, C. & Songjiang, F. (2009). *The application of risk matrix to software project risk management*. Proceedings of the 2009 International forum on information technology and applications.
- Yousfi, K., Boutahar, J. & Elghazi, S. (2014). *IT Governance implementation: A tool design of COBIT 5 roadmap*. Proceedings of the 2014 Second world conference on complex systems (WCCS).
- Zhu, Y., Shi, L. & Hipel, K. W. (2012). *The identification of risk factors in brownfield redevelopment: An empirical study*. Proceedings of the 2012 IEEE International conference on systems, man, and cybernetics (SMC).

Research Article

Robustness of CUSUM-Tukey's Control Chart for Detecting of Changes in Process Dispersion

Suganya Phantu¹ and Saowanit Sukparungsee^{2*}

¹Faulty of Science, Energy, and Environment, King's Mongkut University of Technology North Bangkok, Rayong 21120, Thailand.

²Department of Applied Statistics, Faulty of Applied Science, King's Mongkut University of Technology North Bangkok, Bangkok 10800, Thailand.

*E-mail: saowanit.s@sci.kmutnb.ac.th

Received: 02/04/2020; Revised: 25/04/2020; Accepted: 10/05/2020

Abstract

The aims of this paper have investigated the robustness of Cumulative Sum – Tukey's chart (CUSUM-Tukey) to detecting process dispersion when the assumption of process distribution has deviated from the normal distribution. The performance of control charts is commonly measured by an out of control Average Run Length (ARL_1) by given in control Average Run Length (ARL_0) = 370 and 500. The performance of CUSUM-Tukey is compared with Tukey's chart (Tukey) and Cumulative Sum control chart (CUSUM) where the best performance of the control chart will give the minimum value of ARL_1 . The numerical results are figured out by the Monte Carlo simulation method to approximate ARL which found that the CUSUM-Tukey's control chart is superior to Tukey and CUSUM control charts for all case studies of distributions and all spans for calculating moving range.

Keywords: variation, nonparametric control chart, monitoring, Monte Carlo simulation.

Introduction

Statistical Process Control (SPC) is a collection of different technologies that help differentiate between the common cause and the special cause of variation in the response of a quality characteristic of interest in a process. One of the most effective techniques of SPC is the control chart which can be separate the common causes of the special causes of variation. It is used to monitor a changed parameter in process, such as scale, location and dispersion parameter. Shewhart (1931) first introduced the concept to use a control chart to bring the process back in a normal situation when it is out of control due to special cause variation. The Shewhart control chart normally accommodates just the current sample information, therefore they are sometimes called memoryless control charts. Similarly, Cumulative Sum (CUSUM) control chart (Page, 1954) and Exponentially Weighted Moving Average (EWMA) control chart (Roberts, 1959) are used to detect small to moderate shifts and also relax to normality assumption of the population. The structure of these charts allows them to use the past information along with current information which makes them namely memory or time-weighted control charts. In 2004, Alemi applied the standard form of the diagram constructing a control chart for individual observations with independent and identically distributed (i.i.d.) called Tukey's chart (Tukey). It is an effective alternative to the individual and moving range

control chart (X-MR) for individual process monitoring when the data follow the skew distributions and/or have some outliers.

The monitoring of the process output requires early detection of the shifts of scale/location and dispersion parameters. To monitor those parameters, the stabilized dispersion parameter would be preferred rather than the scale/location parameter, the reason bring dependency on location structure on the dispersion parameter. Therefore, the dispersion parameter is of more immediate concern in process monitoring (Montgomery, 2009).

For efficient monitoring of process dispersion, Torng and Lee (2008) calculated the ARL values of the Tukey's chart under several distributions. For a small sample, the Tukey's chart performed very well when compared with other control charts. Torng et al. (2009) presented the economic design of the Tukey's chart. Lee (2011) applied the asymmetrical control limit (ACL) to the Tukey's chart. The ACL-Tukey's chart offered lower ARL values than the symmetrical control chart (SCL-Tukey) for skewed distributions. Sukparungsee (2012) evaluated the performance of Tukey's chart which its performance is superior to classical EWMA and Shewhart control charts. Lee et al. (2013) used the ACL to economic design Tukey to get the optimum performance. Sukparungsee (2013) also applied ACL to Tukey's chart to evaluate the performance under non-normal distributed data and Tukey's chart has better AARL performance for both cases of ACL-Tukey and SCL-Tukey. Khaliq et al. (2014) investigated the comparative analysis to judge the performance of Tukey's chart versus X/MR chart under the several probability models and Tukey's chart was the best choice in many cases. There are several studies presented that this chart is a good alternative to the traditional Shewhart and X/MR charts for monitoring when the data are skewed distribution. Khaliq et al. (2016) introduced the EWMA design for the Tukey's chart so-called mixed Tukey-EWMA chart. The concept of this chart was designed to have more sensitive for a small sustained shift in process location than the Tukey's chart. This design also performs better than the classical EWMA when the data follow the skewed distribution. Khaliq and Riaz (2017) designed the CUSUM structure of the Tukey's chart for small and sustained shifts. The performance of that chart is superior to the classical CUSUM chart when data follow the skewed distribution. For asymmetric distribution, this design has similar run length performance when compared with the classical CUSUM. Raiz et al. (2017) introduced mixed Tukey EWMA-CUSUM design which was more sensitive to small and moderate amounts of shift. Later, Mongkoltawat et al. (2017) presented Exponentially Weighted Moving Average-Tukey's chart (EWMA-Tukey) for moving range and range which was superior to EWMA and Tukey control charts for all magnitudes of changes. Thitisowaranon et al. (2019) study the robustness of mixed CUSUM-Tukey's chart for detecting process dispersion which was superior to Tukey, CUSUM, and EWMA-Tukey for all cases of asymmetric distribution.

The control charting system is normally practiced in two distinct stages: Phase I (ARL_0) and Phase II (ARL_1). In Phase I, the key concern is to understand the process and to access process stability, making sure that the process is operating at the intended target under some natural causes of variation. Phase I also involves the estimation of the parameters as well as setting up or estimating the control limits. In Phase II, the control chart is used to monitor the process online to detect shifts happening in the process so that any corrective actions can be taken quickly. The performance of the charts based on moving range is evaluated by assessing the average run length (ARL) under normality and in the presence of various types of contaminations employing simulation.

Thus, this paper focuses on the robust CUSUM-Tukey control charts based on the moving range for monitoring the process dispersion parameter. Particularly, their design structures and performances are studied under different environments and in the presence of

special causes in the dispersion parameter of the process. The remaining of the paper is organized as follows. Section 2 gives a review of the control chart and property. In Section 3, the performance evaluation and analysis are addressed. In Section 4, the detection properties of the CUSUM-Tukey's chart are verified by evaluating the average run length (ARL). Finally, Section 5 conclusion and summary of results are discussed.

Control chart and property

1. Tukey's control chart

In 2004, Alemi first purposed Tukey's control chart which is one of the nonparametric control charts by applying the principle of Box plot to obtain the control limits. The control limits of Tukey's chart are proposed by Torng and Lee (2008) as following

$$\begin{aligned} UCL &= F^{-1}(0.75) + K_1 \times IQR \\ LCL &= F^{-1}(0.25) - K_1 \times IQR \end{aligned} \quad (1)$$

where UCL and LCL are upper and lower control limits, respectively. The $F^{-1}(0.75)$ and $F^{-1}(0.25)$ are the third quartile (Q_3) and the first quartile (Q_1) and IQR is interquartile range ($IQR = Q_3 - Q_1$). Usually, the value K_1 is a constant as a coefficient of control limits which given to 1.5 when the assumption of the process is the normal distribution (Ryan, 2000).

2. Cumulative Sum Control Chart

The CUSUM chart was introduced by Page (1954), and it is used cumulative deviation from the target value. It is a favorable tool to detect the small to moderate shifts. The CUSUM chart is based on the following two CUSUM statistics:

$$\begin{aligned} C_t^+ &= \max[0, (X_t - \mu_0) - k_1 + C_{t-1}^+] \\ C_t^- &= \max[0, -(X_t - \mu_0) - k_1 + C_{t-1}^-] \end{aligned} \quad (2)$$

where k_1 is the reference value, t is time or sample number. The C_t^+ and C_t^- are the upper and lower CUSUM statistics, respectively, and the initial value of C_0^+ and C_0^- are zero. The CUSUM statistics C_t^+ and C_t^- plotted against to upper control limit (K_2) and lower control limit is given to be zero, respectively.

3. Cumulative Sum – Tukey's Control Chart

The mixed control charts are combined CUSUM and Tukey's control chart by using the concept that the CUSUM control chart is represented as statistic while the parameter of mean and variance are substituted with nonparametric as quartile and interquartile, respectively. Then, the mixed CUSUM-Tukey statistics are denoted in the form of two statistics C_t^+ and C_t^- as:

$$\begin{aligned} C_t^+ &= \max[0, (x_t - Q_3^s) - k_2 + C_{t-1}^+] \\ C_t^- &= \max[0, -(Q_1^s - x_t) - k_2 + C_{t-1}^-] \end{aligned} \quad (3)$$

where k_2 is the reference value of CUSUM-Tukey's chart Q_1^s and Q_3^s are adjusted from Q_1 and Q_3 by Khaliq et al. (2017) as follows:

$$\begin{aligned} Q_3^s &= (Q_3 - 0.75Q_3) + 0.5Q_2 + 0.25Q_1 \\ Q_1^s &= (Q_1 - 0.75Q_1) + 0.5Q_2 + 0.25Q_3. \end{aligned} \quad (4)$$

The upper control limit of CUSUM-Tukey's chart is K_3 and lower control limit is given to be zero. The value of Q_1, Q_2 and Q_3 are defined as the first, second and third quartile, respectively.

Performance evaluation and analysis

A sequence of points plotted on a control chart until an out of control signal is identified is known as a run and a series of points in a run is named as Run Length (RL). Typically, RL is expected to be higher while out of control RL anticipated being as small as possible for in control process. Several measures based on RL are presented in the literatures to evaluate the performance of the control chart.

Average Run Length (ARL)

The ARL is generally used to assess the performance of control chart for specific shifts. This value refers to the average number of points plotted on a chart until an out of control signal is identified. There are two technical terms used in the control chart that are in control and out of control ARL. The in and out of control ARL are usually denoted by ARL_0 and ARL_1 , respectively. The best performance in order to detect of a change will give the minimal value of ARL_1 . The ARL can be calculated as:

$$ARL = \frac{\sum_{j=1}^m RL_j}{m}. \quad (5)$$

Using Monte Carlo simulation with m replications, we computed the aforementioned RL properties of Tukey, CUSUM and CUSUM-Tukey control charts. The numerical result is reported on Tables 1 - 8 in the form of ARL_s when given $ARL_0=370$ and 500 by varying the magnitudes of shifts in dispersion parameter.

Performance measure evaluation and comparison

The performance of control charts is approximated by the Monte Carlo simulation technique with $m=10^5$ replications to order to get the RL properties for detection of a change in process dispersion based on the moving range with R programming. The comparative analysis of the CUSUM-Tukey's chart, Tukey and CUSUM charts are studied and addressed which the best performance of control charts will give the minimal value of ARL_1 . The numerical results of ARL_s when observations are distributed as Lognormal(0,1) with the span value to calculate a moving range (MR) is given to 2, 3, 5 and 10 is presented on Table 1, Table 2, Table 3 and Table 4, respectively for the case of Lognormal distribution. Tables 5 - 8 shows the numerical results of the comparative performance of control charts when observations are Logistic(6,2) distributed with MR=2, 3, 5 and 10, respectively. These results are arranged in tabular using $ARL_0 = 370$ and 500, magnitudes of shift size (δ) are 1, 1.2, 1.4, 1.6, 1.8, 2, 2.5 and 3, and the number of observations are 2,000.

We have compared the performance of CUSUM-Tukey control chart of detection of change in process dispersion with Tukey and CUSUM control charts when the observations are distributed Lognormal and Logistic. The ARL_s of the control chart are approximated by the Monte Carlo simulation technique. The numerical results show that CUSUM-Tukey's chart is superior to the Tukey and CUSUM control charts for all cases. The comparative performance of Lognormal distribution with the span of MR = 2, 3, 5 and 10 also shown on Figures 1 – 4, respectively. On Figures 5 – 8, the performance of CUSUM-Tukey, Tukey and CUSUM control charts are compared underlying Logistic distribution with the span of MR = 2, 3, 5 and 10,

respectively. Obviously, the performance of CUSUM-Tukey control chart is robust to Lognormal and Logistic distributions because the performance of CUSUM-Tukey can detect of change for process dispersion not only symmetric distributions but also asymmetric distributions.

Table 1. Comparison of ARL for Lognormal (0,1) and MR =2.

δ	ARL ₀ = 370			ARL ₀ = 500		
	Tukey $K_1=9.77$	CUSUM $K_2=16.175$	CUSUM-Tukey $K_3=3.81$	Tukey $K_1=8.27$	CUSUM $K_2=14.08$	CUSUM-Tukey $K_3=1.655$
1	370.66 ± 0.52*	370.14 ± 0.52	370.103 ± 0.52	500.31 ± 0.70	500.22 ± 0.59	500.04 ± 0.43
1.2	155.39 ± 0.22	125.21 ± 0.52	57.673 ± 0.28	200.91 ± 0.68	393.78 ± 0.55	83.47 ± 0.28
1.4	84.36 ± 0.12	78.36 ± 0.52	19.174 ± 0.11	157.67 ± 0.67	312.57 ± 0.54	64.40 ± 0.20
1.6	53.87 ± 0.08	45.69 ± 0.53	9.752 ± 0.05	102.61 ± 0.70	248.24 ± 0.52	48.40 ± 0.13
1.8	37.91 ± 0.05	28.86 ± 0.53	4.485 ± 0.04	86.58 ± 0.71	199.97 ± 0.48	25.42 ± 0.10
2	28.81 ± 0.04	23.71 ± 0.52	1.475 ± 0.02	74.57 ± 0.47	160.66 ± 0.43	18.51 ± 0.83
2.5	17.72 ± 0.03	13.72 ± 0.52	0.675 ± 0.01	65.18 ± 0.33	94.79 ± 0.33	3.35 ± 0.02
3	12.86 ± 0.02	8.26 ± 0.52	0.232 ± 0.01	49.33 ± 0.07	55.42 ± 0.26	1.48 ± 0.01

* standard deviation of run length.

Table 2. Comparison of ARL for Lognormal(0,1) and MR =3.

δ	ARL ₀ = 370			ARL ₀ = 500		
	Tukey $K_1=9.87$	CUSUM $K_2=13.49$	CUSUM-Tukey $K_3=9.09$	Tukey $K_1=10.89$	CUSUM $K_2=15.10$	CUSUM-Tukey $K_3=10.321$
1	370.75 ± 0.97	370.35 ± 0.53	370.49 ± 0.50	500.63 ± 0.61	500.54 ± 0.70	500.28 ± 0.71
1.2	330.49 ± 0.17	255.83 ± 0.22	212.98 ± 0.19	406.30 ± 0.51	378.63 ± 0.38	293.50 ± 0.40
1.4	270.92 ± 0.09	184.42 ± 0.12	144.80 ± 0.06	381.51 ± 0.35	246.32 ± 0.33	150.17 ± 0.25
1.6	247.33 ± 0.06	153.69 ± 0.08	106.18 ± 0.03	328.80 ± 0.32	192.37 ± 0.31	97.51 ± 0.20
1.8	195.27 ± 0.04	137.89 ± 0.05	93.18 ± 0.02	275.63 ± 0.29	111.15 ± 0.30	57.33 ± 0.09
2	128.11 ± 0.03	98.77 ± 0.04	51.86 ± 0.01	213.14 ± 0.24	91.26 ± 0.25	32.65 ± 0.02
2.5	118.75 ± 0.02	77.70 ± 0.03	30.69 ± 0.01	183.22 ± 0.19	81.24 ± 0.15	11.15 ± 0.01
3	94.26 ± 0.02	32.89 ± 0.02	10.34 ± 0.00	102.15 ± 0.15	42.31 ± 0.09	3.24 ± 0.01

Table 3. Comparison of ARL for Lognormal(0,1) and MR =5.

δ	ARL ₀ = 370			ARL ₀ = 500		
	Tukey $K_1=8.54$	CUSUM $K_2=10.23$	CUSUM-Tukey $K_3=9.13$	Tukey $K_1=7.483$	CUSUM $K_2=13.21$	CUSUM-Tukey $K_3=8.23$
1	370.82 ± 0.38	370.92 ± 0.31	370.55 ± 0.28	500.88 ± 0.64	500.52 ± 0.71	500.83 ± 0.68
1.2	268.54 ± 0.38	248.55 ± 0.32	194.41 ± 0.20	476.47 ± 0.62	340.23 ± 0.01	202.21 ± 0.23
1.4	215.03 ± 0.37	155.55 ± 0.33	117.51 ± 0.12	353.77 ± 0.61	241.41 ± 0.23	119.69 ± 0.12

1.6	191.21 ± 0.34	123.67 ± 0.35	85.11 ± 0.08	338.04 ± 0.60	173.85 ± 0.31	86.95 ± 0.08
1.8	143.34 ± 0.31	94.453 ± 0.376	68.60 ± 0.06	226.84 ± 0.60	120.34 ± 0.48	40.29 ± 0.06
2	101.00 ± 0.30	48.815 ± 0.38	20.56 ± 0.05	217.64 ± 0.57	90.09 ± 0.65	28.89 ± 0.04
2.5	98.15 ± 0.29	37.815 ± 0.38	6.33 ± 0.03	193.27 ± 0.55	55.88 ± 0.64	6.85 ± 0.03
3	77.38 ± 0.27	17.54 ± 0.38	1.84 ± 0.03	125.45 ± 0.48	16.47 ± 0.61	1.41 ± 0.03

Table 4. Comparison of ARL for Lognormal(0,1) and MR =10.

δ	ARL ₀ = 370			ARL ₀ = 500		
	Tukey $K_1=9.6$	CUSUM $K_2=11.78$	CUSUM-Tukey $K_3=10.02$	Tukey $K_1=10.73$	CUSUM $K_2=12.92$	CUSUM-Tukey $K_3=8.785$
1	370.81 ± 0.33	370.57 ± 0.53	370.21 ± 0.52	500.32 ± 0.38	500.43 ± 0.70	500.17 ± 0.61
1.2	342.01 ± 0.33	215.36 ± 0.29	187.25 ± 0.32	421.87 ± 0.35	377.29 ± 0.38	241.75 ± 0.33
1.4	301.02 ± 0.32	124.46 ± 0.15	107.21 ± 0.22	401.55 ± 0.31	246.24 ± 0.35	139.54 ± 0.30
1.6	294.75 ± 0.32	77.22 ± 0.12	61.02 ± 0.19	342.76 ± 0.31	194.88 ± 0.33	97.28 ± 0.27
1.8	251.34 ± 0.31	49.10 ± 0.11	32.74 ± 0.15	254.64 ± 0.29	101.58 ± 0.31	58.39 ± 0.24
2	195.36 ± 0.30	27.85 ± 0.09	19.21 ± 0.12	175.36 ± 0.25	82.69 ± 0.24	21.01 ± 0.22
2.5	157.22 ± 0.29	13.53 ± 0.06	6.23 ± 0.08	113.33 ± 0.22	57.26 ± 0.20	13.18 ± 0.14
3	112.36 ± 0.29	5.20 ± 0.02	1.72 ± 0.03	100.18 ± 0.20	21.04 ± 0.15	4.22 ± 0.10

Table 5. Comparison of ARL for Logistic(6,2) and MR =2.

δ	ARL ₀ = 370			ARL ₀ = 500		
	Tukey $K_1=2.503$	CUSUM $K_2=17.83$	CUSUM-Tukey $K_3=2.395$	Tukey $K_1=2.64$	CUSUM $K_2=10.5$	CUSUM-Tukey $K_3=1.549$
1	370.95 ± 0.53	370.60 ± 0.52	370.183 ± 0.03	500.89 ± 0.71	500.49 ± 0.22	500.00 ± 0.05
1.2	322.06 ± 0.51	203.14 ± 0.29	136.771 ± 0.01	465.77 ± 0.71	399.25 ± 0.16	375.77 ± 0.04
1.4	285.23 ± 0.53	111.56 ± 0.16	82.334 ± 0.01	403.58 ± 0.73	331.77 ± 0.12	263.90 ± 0.03
1.6	244.11 ± 0.53	61.13 ± 0.09	38.826 ± 0.01	359.40 ± 0.71	283.79 ± 0.09	201.99 ± 0.03
1.8	190.94 ± 0.53	33.55 ± 0.05	21.551 ± 0.01	296.38 ± 0.71	247.80 ± 0.08	154.45 ± 0.02
2	135.34 ± 0.53	18.43 ± 0.03	13.954 ± 0.01	231.46 ± 0.67	219.78 ± 0.06	67.50 ± 0.02
2.5	107.24 ± 0.53	4.11 ± 0.01	4.004 ± 0.01	177.25 ± 0.63	171.32 ± 0.04	27.89 ± 0.02
3	73.14 ± 0.53	0.92 ± 0.00	0.248 ± 0.01	146.53 ± 0.72	140.25 ± 0.03	7.37 ± 0.01

Table 6. Comparison of ARL for Logistic(6,2) and MR =3.

δ	ARL ₀ = 370			ARL ₀ = 500		
	Tukey $K_1=2.41$	CUSUM $K_2=12.32$	CUSUM-Tukey $K_3=5.104$	Tukey $K_1=2.567$	CUSUM $K_2=17.43$	CUSUM-Tukey $K_3=5.021$
1	370.66 ± 0.38	370.21 ± 0.53	370.34 ± 0.56	500.75 ± 0.76	500.16 ± 0.80	500.06 ± 0.75
1.2	344.69 ± 0.36	270.75 ± 0.33	176.43 ± 0.19	456.67 ± 0.03	460.19 ± 0.76	387.82 ± 0.27

1.4	335.53 ± 0.36	211.24 ± 0.31	108.16 ± 0.10	330.71 ± 0.01	328.68 ± 0.52	240.55 ± 0.16
1.6	306.68 ± 0.34	193.95 ± 0.30	78.84 ± 0.07	311.79 ± 0.01	305.47 ± 0.35	204.75 ± 0.09
1.8	255.41 ± 0.35	127.31 ± 0.28	64.79 ± 0.05	271.08 ± 0.00	296.60 ± 0.29	171.88 ± 0.05
2	201.16 ± 0.32	90.23 ± 0.21	36.10 ± 0.04	141.13 ± 0.00	157.71 ± 0.10	122.12 ± 0.03
2.5	197.43 ± 0.30	73.26 ± 0.20	23.91 ± 0.03	119.98 ± 0.00	132.56 ± 0.05	82.97 ± 0.02
3	135.00 ± 0.22	20.21 ± 0.12	12.96 ± 0.02	107.79 ± 0.00	92.57 ± 0.02	36.89 ± 0.01

Table 7. Comparison of ARL for Logistic(6,2) and MR =5.

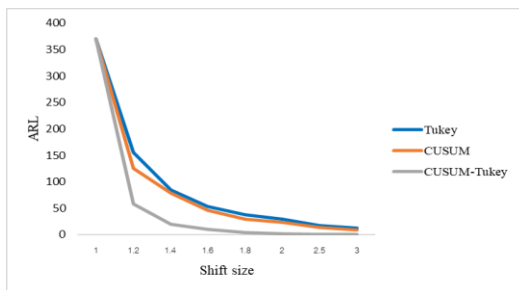
δ	ARL ₀ = 370			ARL ₀ = 500		
	Tukey $K_1=1.54$	CUSUM $K_2=10.43$	CUSUM-Tukey $K_3=4.701$	Tukey $K_1=2.09$	CUSUM $K_2=12.83$	CUSUM-Tukey $K_3=4.925$
1	370.45 ± 0.34	370.24 ± 0.35	370.12 ± 0.31	500.84 ± 0.80	500.41 ± 0.47	500.12 ± 0.35
1.2	330.22 ± 0.33	305.81 ± 0.31	245.56 ± 0.31	427.52 ± 0.72	382.78 ± 0.41	318.02 ± 0.31
1.4	284.38 ± 0.27	275.63 ± 0.30	184.64 ± 0.28	316.74 ± 0.51	315.43 ± 0.38	234.84 ± 0.27
1.6	217.41 ± 0.21	205.22 ± 0.24	111.11 ± 0.20	308.10 ± 0.49	293.48 ± 0.36	144.73 ± 0.16
1.8	199.81 ± 0.15	154.66 ± 0.16	92.63 ± 0.15	291.48 ± 0.43	216.33 ± 0.28	101.25 ± 0.14
2	136.91 ± 0.10	119.25 ± 0.10	51.35 ± 0.12	171.25 ± 0.38	187.68 ± 0.21	71.26 ± 0.11
2.5	99.10 ± 0.07	74.51 ± 0.09	17.26 ± 0.08	145.32 ± 0.32	102.75 ± 0.20	14.25 ± 0.08
3	65.24 ± 0.03	31.25 ± 0.06	8.00 ± 0.02	102.05 ± 0.30	89.04 ± 0.16	2.58 ± 0.02

Table 8. Comparison of ARL for Logistic(6,2) and MR =10.

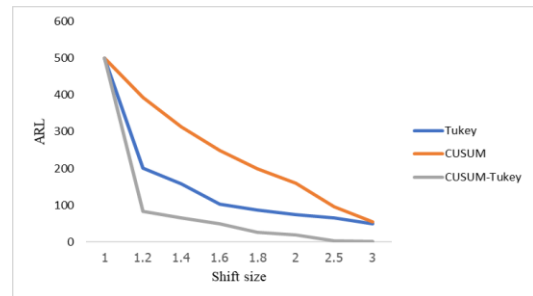
δ	ARL ₀ = 370			ARL ₀ = 500		
	Tukey $K_1=2.53$	CUSUM $K_2=15.45$	CUSUM-Tukey $K_3=6.432$	Tukey $K_1=2.7$	CUSUM $K_2=19.60$	CUSUM-Tukey $K_3=5.392$
1	370.72 ± 0.39	370.22 ± 0.34	370.16 ± 0.37	500.49 ± 0.51	500.16 ± 0.70	500.12 ± 0.52
1.2	335.21 ± 0.34	245.09 ± 0.32	210.41 ± 0.23	434.04 ± 0.50	465.22 ± 0.62	397.15 ± 0.51
1.4	320.21 ± 0.32	228.03 ± 0.29	131.40 ± 0.16	401.26 ± 0.45	414.21 ± 0.61	377.84 ± 0.45
1.6	284.37 ± 0.31	209.63 ± 0.26	114.48 ± 0.12	365.39 ± 0.41	349.18 ± 0.52	269.40 ± 0.38
1.8	251.04 ± 0.29	179.86 ± 0.21	78.85 ± 0.11	326.95 ± 0.25	302.76 ± 0.41	139.21 ± 0.32
2	223.37 ± 0.21	145.56 ± 0.19	34.98 ± 0.10	295.01 ± 0.19	274.35 ± 0.32	93.25 ± 0.27
2.5	192.76 ± 0.15	99.92 ± 0.15	13.57 ± 0.04	186.48 ± 0.15	124.53 ± 0.29	30.57 ± 0.14
3	114.75 ± 0.09	83.20 ± 0.10	4.03 ± 0.02	94.17 ± 0.10	75.45 ± 0.24	15.62 ± 0.09

Conclusion

The comparative studies of the performance of Tukey, CUSUM and CUSUM-Tukey control charts for detecting a change in variation by using the moving range. The mixed CUSUM-Tukey control chart is superior to Tukey and CUSUM control charts for all magnitudes of shift in dispersion parameter on both cases of Lognormal(0,1) and Logistic(6,2) distributions. Therefore, the mixed CUSUM-Tukey control chart is robust to detect of dispersion parameter not only symmetry but also asymmetry distributions. In addition, the CUSUM-Tukey control chart robust to monitor the process dispersion when unknown population distributions and ARL_0 is changed to be 500. Because of the mixed CUSUM-Tukey control chart is a nonparametric control chart which parameter do not need to be known.

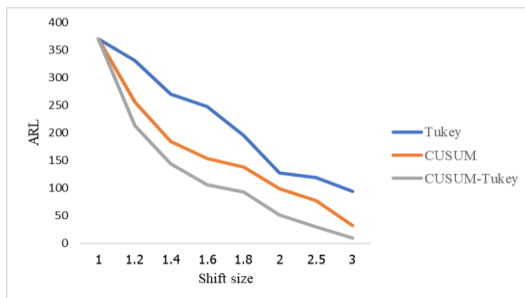


(a) $ARL_0 = 370$

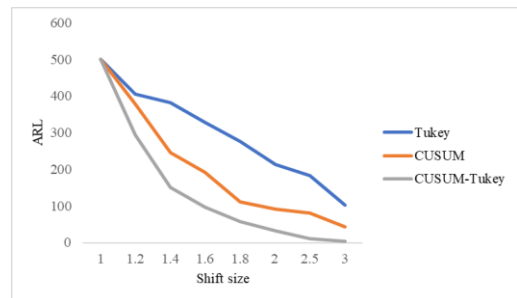


(b) $ARL_0 = 500$

Figure 1. Comparison of ARL for Lognormal(0,1) when MR = 2.

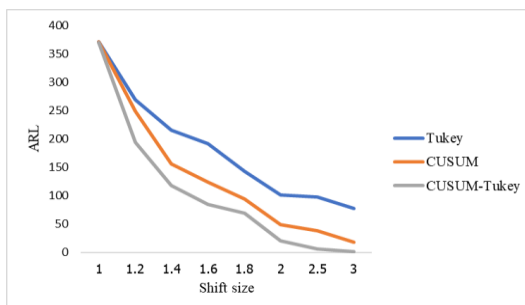


(a) $ARL_0 = 370$

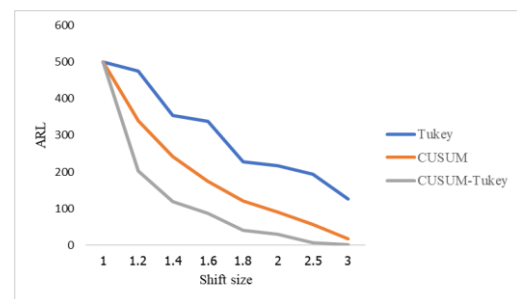


(b) $ARL_0 = 500$

Figure 2. Comparison of ARL for Lognormal(0,1) when MR = 3.

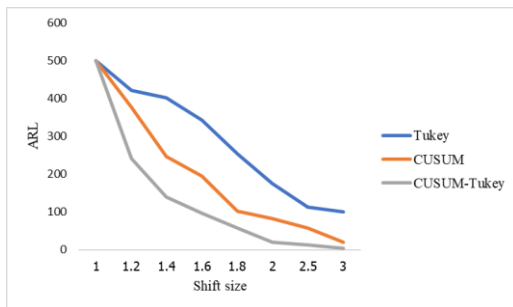


(a) $ARL_0 = 370$

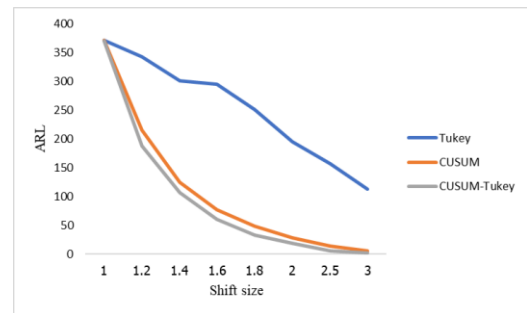


(b) $ARL_0 = 500$

Figure 3. Comparison of ARL for Lognormal(0,1) when MR = 5.

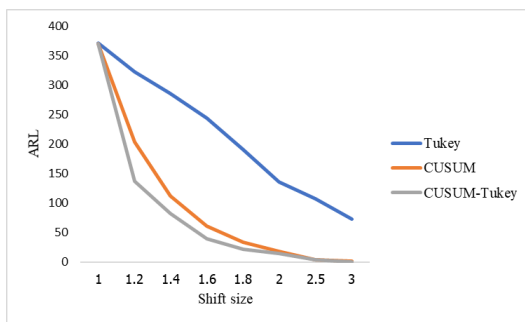


(a) $ARL_0 = 370$

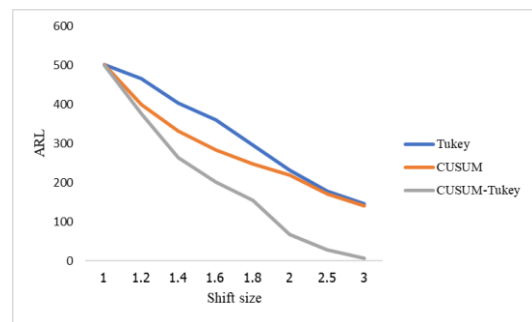


(b) $ARL_0 = 500$

Figure 4. Comparison of ARL for Lognormal(0,1) when MR = 10.

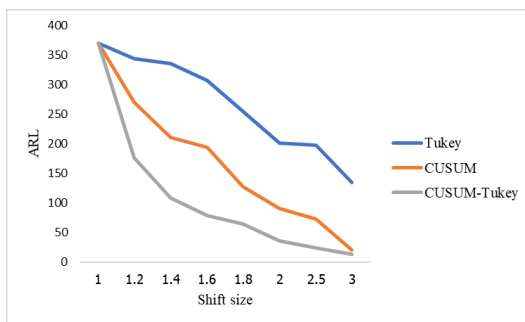


(a) $ARL_0 = 370$

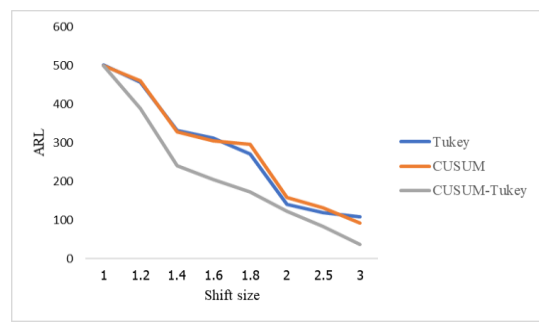


(b) $ARL_0 = 500$

Figure 5. Comparison of ARL for Logistic(6,2) when MR = 2.

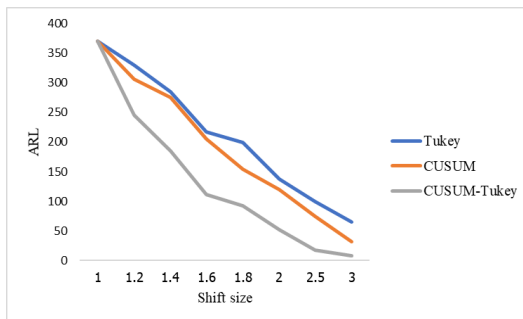


(a) $ARL_0 = 370$

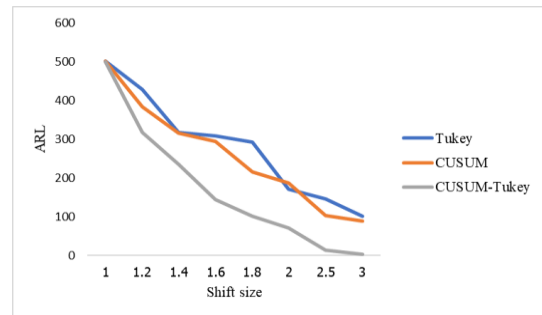


(b) $ARL_0 = 500$

Figure 6. Comparison of ARL for Logistic(6,2) when MR = 3.

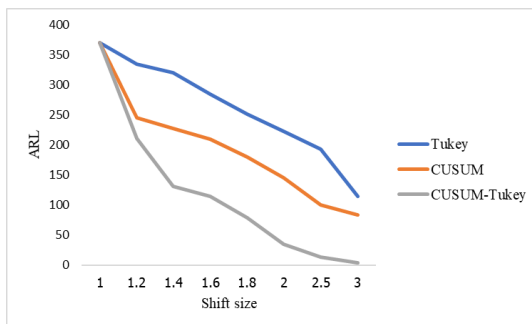


(a) $ARL_0 = 370$

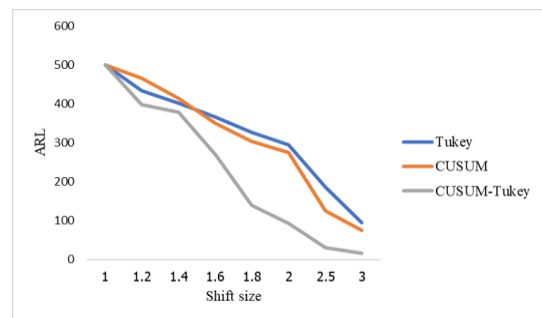


(b) $ARL_0 = 500$

Figure 7. Comparison of ARL for Logistic(6,2) when MR = 5.



(a) $ARL_0 = 370$



(b) $ARL_0 = 500$

Figure 8. Comparison for ARL for Logistic(6,2) when MR = 10.

Acknowledgments

The authors would like to express their sincere gratitude to the Department of Applied Statistics for allowance and financial support the Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Thailand for supporting research grants with contract no. 6345107.

References

- Alemi, F. (2004). Tukey's control chart. *Quality Management in Health Care*, 13(4): 216-221. DOI = <http://doi.10.1097/00019514-200410000-00004>.
- Khaliq, Q. U. A., Riaz, M. & Alemi, F. (2014). Performance of Tukey's and individual/moving range control charts. *Quality and Reliability Engineering*, 31(6): 1063-1077. DOI = <http://doi.10.1002/qre.1664>.
- Khaliq, Q. U. A., Riaz, M. & Ahmad, S. (2016). On designing a new Tukey-EWMA control chart for process monitoring. *The International Journal of Advanced Manufacturing Technology*, 82: 1-23. DOI= <https://doi.org/10.1007/s00170-015-7289-6>.
- Khaliq, Q. U. A. & Riaz, M. (2017). Robust Tukey-CUSUM control chart for process monitoring. *Quality and Reliability Engineering*, 32(3): 933-948. DOI = <http://doi.org/10.1002./qre.1804>.

- Khaliq, Q. U. A., Riaz, M. & Gui, S. (2017). Mixed Tukey EWMA-CUSUM control chart and its applications. *Quality Technology and Quantitative Management*, 14(4): 378-411. DOI = <http://doi.org/10.1080/16843703.2017.1304034>.
- Lee, P. H. (2011). The effects of Tukey's control chart with asymmetrical control limits on monitoring of production processes. *African Journal of Business Management*, 5(11): 4044-4050. DOI= <http://doi.10.5897/AJBM10.1389>.
- Lee, P. H., Kuo, T. I. & Lin, C. S. (2013), Economically optimum design of Tukey's control chart with asymmetrical control limits for controlling process mean of skew population distribution. *European Journal of Industrial Engineering*, 7(6): 687-703. DOI = <http://doi.org/10.1504/EJIE.2013.058391>.
- Montgomery, D. C. (2009). *Introduction to Statistical Quality Control (6th ed.)*. USA: John Wiley & Sone Inc.
- Mongkoltawat, P., Sukparungsee, S. & Areepong, Y. (2017). Exponentially weighted moving average-Tukey's control charts for moving range and range. *The Journal of KMUTNB*, 27(4): 843-854. DOI = <http://doi.10.14416/j.kmutnb .2017.11.002>.
- Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1): 100-114. DOI = <http://doi:10.1093/biomet/41.1-2.100>.
- Riaz, M., Mahmood, T., Abbasi, S. A., Abbas N. & Ahmad, S. (2017). Linear profile monitoring using EWMA structure under ranked set schemes. *The International Journal of Advanced Manufacturing Technology*, 91: 2751-2775. DOI = 10.1007/s00170-016-9608-y.
- Roberts, S. W. (1959). Control chart tests based on geometric moving average. *Technometrics*, 42(1): 239-250. DOI= <http://dx.doi.org/10.1080/00401706.1959.10489860>.
- Ryan, T.P. (2000). Statistical methods for quality improvement. *Wiley series in probability and Mathematics*, Wiley. 1-14, 71-211.
- Shewhart, W. A. (1931). *Economic Control of Quality of manufactured Product*. New York: D. Van Nostrand Company Inc.
- Sukparungsee, S. (2012). Robustness of Tukey's Control Chart in Detecting a Changes of Parameter of Skew Distributions. *International Journal of Applied Physics and Mathematics*, 2(5): 379-382. DOI = <http://doi.10.7763/IJAPM.2012.V2.140>.
- Sukparungsee, S. (2013). Asymmetric Tukey's control chart robust to skew and non-skew process observation. *International Journal of Mathematical Computation, Physical, Electrical and Computer Engineering*, 7(8): 51-54. DOI = <http://doi.10.5281/zenodo .1086603>.
- Thitisoowaranon. R., Sukparungsee, S. & Areepong, Y. (2019). A mix Cumulative sum-Tukey's control chart for detecting process dispersion. *The Journal of KMUTNB*, 29(3): 507-517. DOI = <http://doi.10.14416/l.kmutnb.2019.04.004>.
- Torng, C. C. & Lee, P. H. (2008). ARL performance of the Tukey's control chart. *Communication in Statistics-Simulation and Computation*, 33(9): 1904-1913. DOI = <http://doi.org/10 .1080/03610910802263141>.
- Torng, C. C., Lee, H. N. & Tseng, C. C. (2009). An Economic Design of Tukey's control chart. *Journal of Chinese Institute of Industrial Engineers*, 26(1): 53-59, DOI = <http://dx.doi.org/10.1080/10170660909509121>.

Research Article

Asymptotic confidence ellipses for the re-parameterized inverse Gaussian distribution

Wikanda Phaphan^{1*}

¹Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand

*E-mail: wikanda.p@sci.kmutnb.ac.th

Received: 27/04/2020; Revised: 11/06/2020; Accepted: 18/06/2020

Abstract

The re-parameterized inverse Gaussian distribution is a very useful distribution for statistics and it is applied to various fields, such as physics, engineering, biology, etc. It is also appropriate to analyze the right-skewed data. In this research, we were interested in studying the Fisher-information matrix and we wanted to find the covariance matrix to form asymptotic confidence ellipses of parameters for the re-parameterized inverse Gaussian distribution. We compared the coverage probability with the confidence coefficient of 0.98 of confidence ellipses with sample sizes of $n = 10, 30, 50, 100, 1,000$, and $10,000$. The parameters of $\lambda = 0.5, 1, 5, 10, 50$ and $\theta = 0.5, 1, 5, 10, 50$. The data were simulated by the composition method which they were repeated 10,000 rounds in each case with RStudio programming. The simulation results indicate that the value of coverage probability had been close to the confidence coefficient of 0.98 at the sample size of 10,000.

Keywords: inverse Gaussian distribution, confidence ellipses, Fisher-information matrix, covariance matrix

Introduction

The inverse Gaussian distribution became known in statistics in the twentieth century from the research of Schrödinger (1915) and Tweedie (1957). This distribution is an alternative distribution suitable for an analysis of the right-skewed data. Therefore, it is popular and is applied to various fields such as physics, engineering, biology, finance, etc. As Folks & Chhikara (1978) explained in their research, the inverse Gaussian distribution is the continuous probability distribution which the probability density functions of inverse Gaussian distribution contain two parameters are μ and λ .

$$f_{IG}(x, \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}; x > 0, \quad (1)$$

where $\mu > 0$ and $\lambda > 0$, they represent a symbol of $X \sim IG(\mu, \lambda)$. And μ is a location parameter or an average parameter, λ is a shape parameter. Thus, the average of distribution is $E(X) = \mu$ and the variance of distribution is $Var(X) = \mu^3 / \lambda$. Furthermore, the skewness is

$3\sqrt{\mu/\lambda}$. The maximum likelihood estimator of μ is $\tilde{\mu} = \bar{x}$ and the maximum likelihood estimator of λ is $\tilde{\lambda} = \left(\sum_{i=1}^n x_i^{-1} / n - 1/\bar{x} \right)^{-1}$, where $\bar{x} = \sum_{i=1}^n x_i / n$.

Besides, the inverse Gaussian distribution still has a similar property with and bears a relation to the Birnbaum-Saunders distribution which was first introduced by Birnbaum & Saunders (1969). This Birnbaum-Saunders distribution was the failure distribution and was often used for reliability analysis. Afterward, the parameters of the Birnbaum-Saunders distribution were again adapted by Ahmed et al. (2008) in order to be suitable for use. Furthermore, they solved the relationship equation between the latest parameters and former parameters, including various properties of this distribution. Jangphanish & Budsaba (2013) considered the parameter estimation of the inverse Gaussian distribution based upon new parameters as Ahmed et al. (2008).

Therefore, the probability density function of the inverse Gaussian distribution with the new parameters can be written in this formula:

$$f_{IG}(x; \lambda, \theta) = \frac{\lambda}{\theta \sqrt{2\pi}} \left(\frac{\theta}{x} \right)^{\frac{3}{2}} \exp \left[-\frac{1}{2} \left(\sqrt{\frac{x}{\theta}} - \lambda \sqrt{\frac{\theta}{x}} \right)^2 \right] ; x > 0, \quad (2)$$

where $\lambda > 0$ and $\theta > 0$, they represent a symbol of $X \sim IG(\lambda, \theta)$, λ is the thickness of the sample, and θ is the nominal treatment loading on the sample. The maximum likelihood

estimators of λ is $\hat{\lambda} = \frac{n}{T\bar{x} - n}$ and the maximum likelihood estimators of θ is $\hat{\theta} = \frac{T\bar{x}^2 - \sum_{i=1}^n x_i}{n}$

where $T = \sum_{i=1}^n \frac{1}{x_i}$.

The two types of estimation of population parameters are point estimation and interval estimation. The estimation of the population parameters, appealing to a study due to containing a single statistical value, is the point estimation. The interval estimation or the confidence interval estimation is the estimation of interesting population parameters because it contains an interval value which is between two values by containing the point estimation at the center between the two values.

In the past decade, numerous studies have given special attention to the confidence interval for one-dimensional space; few researchers are interested in studying the confidence interval for two-dimensional space (which is called the elliptical confidence interval).

From the previous research into the inverse Gaussian distribution, we found that there was no research work on the confidence interval of asymptotic ellipses for the re-parameterized inverse Gaussian distribution. Consequently, we are interested in doing a study into the confidence interval of asymptotic ellipses for the re-parameterized inverse Gaussian distribution based on distribution as Jangphanish & Budsaba (2013).

A brief review of the re-parameterized inverse Gaussian distribution

The basic parameters of the re-parameterized inverse Gaussian distribution are λ and θ . This distribution is a group of continuous probability distributions with values of random variables from 0 to ∞ . The probability density functions of the random variable with the re-parameterized inverse Gaussian distribution that proposed by Jangphanish & Budsaba (2013) is

$$f(x; \lambda, \theta) = \frac{\lambda}{\theta \sqrt{2\pi}} \left(\frac{\theta}{x} \right)^{\frac{3}{2}} \exp \left\{ -\frac{1}{2} \left(\lambda \sqrt{\frac{\theta}{x}} - \sqrt{\frac{x}{\theta}} \right)^2 \right\}; \quad x > 0, \quad (3)$$

where $\lambda > 0$ is the thickness of the machine element and $\theta > 0$ is the nominal treatment pressure on the machine.

The re-parameterized inverse Gaussian distributions, which consists of two parameters, are λ and θ . The likelihood functions of the re-parameterized inverse Gaussian distributions depend on the two parameters. They can be written in this formula:

$$L(x; \lambda, \theta) = \frac{\lambda^n \theta^{\frac{n}{2}}}{(\sqrt{2\pi})^n \prod_{i=1}^n x_i^{\frac{3}{2}}} \exp \left\{ -\sum_{i=1}^n \frac{\lambda^2 \theta}{2x_i} + n\lambda - \frac{\sum_{i=1}^n x}{2\theta} \right\}. \quad (4)$$

Therefore, the logarithm of the probability function can be written in this formula:

$$l(x; \lambda, \theta) = \ln L(x; \lambda, \theta) = n \ln \lambda + \frac{n}{2} \ln \theta - n \ln \sqrt{2\pi} - \sum_{i=1}^n \frac{3}{2x_i} - \frac{\lambda^2 \theta}{2} \sum_{i=1}^n \frac{1}{x_i} + n\lambda - \frac{\sum_{i=1}^n x}{2\theta},$$

and the maximum likelihood estimators of parameters λ and θ can be written in this form

$$\hat{\theta} = \frac{T\bar{X}^2 - \sum_{i=1}^n x_i}{n} \quad \text{and} \quad \hat{\lambda} = \frac{n}{T\bar{X} - n}, \quad (5)$$

$$\text{where} \quad T = \sum_{i=1}^n \frac{1}{x_i}, \quad \bar{X} = \frac{\sum_{i=1}^n x_i}{n},$$

and $E(X)$ is the first moment of the re-parameterized inverse Gaussian distribution, therefore it is in this form: $E(X) = \lambda\theta$,

$$\begin{aligned} \text{and} \quad \text{Var}(X) &= E(X^2) + (E(X))^2 \\ &= \lambda\theta^2(\lambda+1) - \lambda^2\theta^2 \\ &= \lambda^2\theta^2 + \lambda\theta^2 - \lambda^2\theta^2 = \lambda\theta^2. \end{aligned} \quad (6)$$

Theoretical results

1. The Fisher information matrix for the re-parameterized inverse Gaussian distribution

Give a random variable X to be the re-parameterized inverse Gaussian distribution with parameters (λ, θ) . The Fisher Information Matrix is Θ when $\Theta = (\mu, \lambda)$ is a two-dimensional vector of the parameter standing for $I(\Theta)$ as shown in below

$$I(\Theta) = I(\lambda, \theta) = -E \begin{bmatrix} \frac{\partial^2}{\partial \lambda^2} \ln f(x; \lambda, \theta) & \frac{\partial^2}{\partial \lambda \partial \theta} \ln f(x; \lambda, \theta) \\ \frac{\partial^2}{\partial \theta \partial \lambda} \ln f(x; \lambda, \theta) & \frac{\partial^2}{\partial \theta^2} \ln f(x; \lambda, \theta) \end{bmatrix}. \quad (7)$$

The logarithm of the probability function of the re-parameterized inverse Gaussian distribution is

$$\begin{aligned} \ln f(x; \lambda, \theta) &= \ln \left[\frac{\lambda \theta^{\frac{1}{2}}}{\sqrt{2\pi} x^{\frac{3}{2}}} \exp \left\{ -\frac{\lambda^2 \theta}{2x} + \lambda - \frac{x}{2\theta} \right\} \right] \\ &= \ln(\lambda) + \frac{1}{2} \ln(\theta) - \ln \sqrt{2\pi} - \ln \left(x^{\frac{3}{2}} \right) - \frac{\lambda^2 \theta}{2x} + \lambda - \frac{x}{2\theta}. \end{aligned} \quad (8)$$

From $E(X) = \lambda\theta$, thus we can estimate the expectation of reciprocal $E(1/X)$ by Taylor expansion around $E(X)$ of a variable, so we obtain

$$\begin{aligned} E\left(\frac{1}{X}\right) &\approx E\left(\frac{1}{E(X)} - \frac{1}{E(X)^2}(X - E(X)) + \frac{1}{E(X)^3}(X - E(X))^2\right) \\ &= \frac{1}{E(X)} + \frac{1}{E(X)^3} \text{Var}(X) \\ &= \frac{1}{\lambda\theta} + \frac{1}{\lambda^3\theta^3} \lambda\theta^2 = \frac{1}{\lambda\theta} + \frac{1}{\lambda^2\theta} = \frac{1}{\lambda^2\theta}(\lambda+1). \end{aligned} \quad (9)$$

Hence, $E \begin{bmatrix} \frac{\partial^2}{\partial \lambda^2} \ln f(x; \lambda, \theta) & \frac{\partial^2}{\partial \lambda \partial \theta} \ln f(x; \lambda, \theta) \\ \frac{\partial^2}{\partial \theta \partial \lambda} \ln f(x; \lambda, \theta) & \frac{\partial^2}{\partial \theta^2} \ln f(x; \lambda, \theta) \end{bmatrix}$ will be in these formulas:

$$\begin{aligned} E\left(\frac{\partial^2}{\partial \theta^2} \ln f(x; \lambda, \theta)\right) &= -\frac{1}{2\theta^2} - \frac{E(x)}{\theta^3} = -\frac{1}{2\theta^2} - \frac{\lambda\theta}{\theta^3} \\ &= -\frac{1}{2\theta^2} - \frac{\lambda}{\theta^2} = -\frac{(2\lambda+1)}{2\theta^2} \end{aligned} \quad (10)$$

$$\begin{aligned} E\left(\frac{\partial^2}{\partial \lambda^2} \ln f(x; \lambda, \theta)\right) &= -\frac{1}{\lambda^2} - \theta E\left(\frac{1}{x}\right) \\ &= -\frac{1}{\lambda^2} - \theta \left(\frac{1}{\lambda^2\theta}\right)(\lambda+1) \\ &= -\frac{1}{\lambda^2} - \frac{\theta}{\lambda\theta} - \frac{\theta}{\lambda^2\theta} = -\frac{1}{\lambda^2} - \frac{1}{\lambda} - \frac{1}{\lambda^2} \\ &= -\frac{1}{\lambda} - \frac{2}{\lambda^2} = -\frac{1}{\lambda^2}(\lambda+2) \end{aligned} \quad (11)$$

$$\begin{aligned}
 E\left(\frac{\partial^2}{\partial \lambda \partial \theta} \ln f(x; \lambda, \theta)\right) &= -\lambda E\left(\frac{1}{x}\right) \\
 &= -\frac{1}{\lambda \theta}(\lambda + 1)
 \end{aligned} \tag{12}$$

Therefore, the Fisher Information Matrix for the re-parameterized inverse Gaussian distribution is in this form as shown.

$$\begin{aligned}
 I(\Theta) &= -\begin{bmatrix} -\frac{1}{\lambda^2}(\lambda + 2) & -\frac{(\lambda + 1)}{\lambda \theta} \\ -\frac{(\lambda + 1)}{\lambda \theta} & -\frac{(2\lambda + 1)}{2\theta^2} \end{bmatrix} \\
 &= \begin{bmatrix} \frac{(\lambda + 2)}{\lambda^2} & \frac{(\lambda + 1)}{\lambda \theta} \\ \frac{(\lambda + 1)}{\lambda \theta} & \frac{(2\lambda + 1)}{2\theta^2} \end{bmatrix}.
 \end{aligned} \tag{13}$$

When Θ is the two-dimensional vector of parameters and $\Theta = (\lambda, \theta)$, so we will obtain the Fisher information matrix for sample sizes n are

$$I_n(\Theta) = \begin{bmatrix} \frac{n(\lambda + 2)}{\lambda^2} & \frac{n(\lambda + 1)}{\lambda \theta} \\ \frac{n(\lambda + 1)}{\lambda \theta} & \frac{n(2\lambda + 1)}{2\theta^2} \end{bmatrix}. \tag{14}$$

2. The covariance matrix

The value of the covariance matrix and the value of the inverse of the Fisher information matrix are equal. They stand for $\Lambda = I_n^{-1}(\Theta) = I_n^{-1}(\lambda, \theta)$. Consequently, when n comes to ∞ , it will be the following equations

$$\begin{aligned}
 I_n^{-1}(\Theta) &= \frac{1}{\left(\frac{n(\lambda + 2)}{\lambda^2} \times \frac{n(2\lambda + 1)}{2\theta^2}\right) - \left(\frac{n(\lambda + 1)}{\lambda \theta} \times \frac{n(\lambda + 1)}{\lambda \theta}\right)} \begin{bmatrix} \frac{n(2\lambda + 1)}{2\theta^2} & -\frac{n(\lambda + 1)}{\lambda \theta} \\ -\frac{n(\lambda + 1)}{\lambda \theta} & \frac{n(\lambda + 2)}{\lambda^2} \end{bmatrix} \\
 &= \frac{1}{\frac{n^2(2\lambda + 1)(\lambda + 2)}{2\lambda^2\theta^2} - \frac{n^2(\lambda + 1)^2}{\lambda^2\theta^2}} \begin{bmatrix} \frac{n(2\lambda + 1)}{2\theta^2} & -\frac{n(\lambda + 1)}{\lambda \theta} \\ -\frac{n(\lambda + 1)}{\lambda \theta} & \frac{n(\lambda + 2)}{\lambda^2} \end{bmatrix}.
 \end{aligned} \tag{15}$$

Consider this formula

$$\frac{n^2(2\lambda + 1)(\lambda + 2)}{2\lambda^2\theta^2} - \frac{2}{2} \times \frac{n^2(\lambda + 1)^2}{\lambda^2\theta^2} = \frac{n^2(2\lambda + 1)(\lambda + 2)}{2\lambda^2\theta^2} - \frac{2n^2(\lambda + 1)^2}{2\lambda^2\theta^2}$$

$$\begin{aligned}
 &= \frac{n^2(2\lambda+1)(\lambda+2) - 2n^2(\lambda+1)^2}{2\lambda^2\theta^2} \\
 &= \frac{n^2[2\lambda^2+4\lambda+\lambda+2]}{2\lambda^2\theta^2} - \frac{2n^2[\lambda^2+2\lambda+1]}{2\lambda^2\theta^2} \\
 &= \frac{n^2[2\lambda^2+5\lambda+2] - 2n^2[\lambda^2+2\lambda+1]}{2\lambda^2\theta^2} \\
 &= \frac{2n^2\lambda^2+5n^2\lambda+2n^2-2n^2\lambda^2-4n^2\lambda-2n^2}{2\lambda^2\theta^2} \\
 &= \frac{2n^2\lambda^2-2n^2\lambda^2+5n^2\lambda-4n^2\lambda+2n^2-2n^2}{2\lambda^2\theta^2} \\
 &= \frac{n^2\lambda}{2\lambda^2\theta^2} = \frac{n^2}{2\lambda\theta^2}.
 \end{aligned} \tag{16}$$

That is

$$\begin{aligned}
 \Lambda = I_n^{-1}(\Theta) &= \frac{2\lambda\theta^2}{n^2} \begin{bmatrix} \frac{n(2\lambda+1)}{2\theta^2} & -\frac{n(\lambda+1)}{\lambda\theta} \\ -\frac{n(\lambda+1)}{\lambda\theta} & \frac{n(\lambda+2)}{\lambda^2} \end{bmatrix} \\
 &= \begin{bmatrix} \frac{\lambda(2\lambda+1)}{n} & -\frac{2\theta(\lambda+1)}{n} \\ -\frac{2\theta(\lambda+1)}{n} & \frac{2\theta^2(\lambda+2)}{n\lambda} \end{bmatrix}.
 \end{aligned} \tag{17}$$

3. An asymptotic normal distribution

Consider the order of random variable from maximum likelihood $\hat{\Theta}_n^{(MLE)}$ using the Delta method. It will be this form

$$\begin{aligned}
 \sqrt{n}(\hat{\Theta}_n^{(MLE)} - \Theta) &\xrightarrow{d} X \sim N_2(0, I^{-1}(\Theta)) \\
 \frac{(\hat{\Theta}_n^{(MLE)} - \Theta)}{\sqrt{I_n^{-1}(\Theta)}} &\xrightarrow{d} Z \sim N_2(0, I_2).
 \end{aligned}$$

From the value of the inverse of the Fisher information matrix is equal to the value of the covariance matrix, it will be

$$\frac{(\hat{\Theta}_n^{(MLE)} - \Theta)}{\sqrt{\Lambda}} \xrightarrow{d} N_2(0, I_2). \tag{18}$$

Therefore,

$$\left(\Lambda^{\frac{1}{2}} \right)^{-1} \left(\hat{\Theta}_n^{(MLE)} - \Theta \right) \xrightarrow{d} Z \sim N_2(0, I_2), \quad (19)$$

where $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is an identity matrix and Z_1, Z_2 are random variables with an average normal distribution of 0 and the variance of 1, and they are independent. Thus, $Z' = (Z_1, Z_2)$ is the bivariate normal distribution $N_2(0, I_2)$ as it shows

$$\begin{aligned} \left(\hat{\Theta}_n^{(MLE)} - \Theta \right)' (\Lambda)^{-\frac{1}{2}} (\Lambda)^{-\frac{1}{2}} \left(\hat{\Theta}_n^{(MLE)} - \Theta \right) &= \left(\hat{\Theta}_n^{(MLE)} - \Theta \right)' \Lambda^{-1} \left(\hat{\Theta}_n^{(MLE)} - \Theta \right) \\ &= Z'Z = Z_1^2 + Z_2^2 = \sum_{i=1}^2 Z_i^2. \end{aligned}$$

This distribution is Chi-Square distribution with two degrees of freedom (χ_2^2).

4. The region of the confidence for the re-parameterized inverse Gaussian distribution

From Duangchana & Budsaba (2014), Hansoongnern et al. (2018) and Phaphan & Pongsart (2019) we will obtain the confidence region of $100(1-\alpha)\%$ for the parameter Θ of Two-Dimensional Normal Distribution. It is an ellipse as shown in this equation.

$$\left(\hat{\Theta}_n^{(MLE)} - \Theta \right)' \Lambda^{-1} \left(\hat{\Theta}_n^{(MLE)} - \Theta \right) \leq \chi_2^2(\alpha). \quad (20)$$

Since covariance matrix of the re-parameterized inverse Gaussian distribution is

$$\Lambda^{-1} = I_n(\Theta). \quad (21)$$

We will obtain the confidence region of $100(1-\alpha)\%$ for the parameter $\Theta' = (\lambda, \theta)$ of the re-parameterized inverse Gaussian distribution as it indicates

$$\begin{bmatrix} \hat{\lambda}_n^{(MLE)} - \lambda & \hat{\theta}_n^{(MLE)} - \theta \end{bmatrix} \begin{bmatrix} \frac{n(\lambda+2)}{\lambda^2} & \frac{n(\lambda+1)}{\lambda\theta} \\ \frac{n(\lambda+1)}{\lambda\theta} & \frac{n(2\lambda+1)}{2\theta^2} \end{bmatrix} \begin{bmatrix} \hat{\lambda}_n^{(MLE)} - \lambda \\ \hat{\theta}_n^{(MLE)} - \theta \end{bmatrix} \leq \chi_{(2)}^2(\alpha) \quad (22)$$

$$\frac{n(\lambda+2)}{\lambda^2} \left(\hat{\lambda}_n^{(MLE)} - \lambda \right)^2 + \frac{2n(\lambda+1)}{\lambda\theta} \left(\hat{\lambda}_n^{(MLE)} - \lambda \right) \left(\hat{\theta}_n^{(MLE)} - \theta \right) + \frac{n(2\lambda+1)}{2\theta^2} \left(\hat{\theta}_n^{(MLE)} - \theta \right)^2 \leq \chi_{(2)}^2(\alpha). \quad (23)$$

Simulation results

In this topic, we studied the coverage probability of asymptotic confidence ellipses for the re-parameterized inverse Gaussian distribution. We defined parameter of $\lambda = 0.5, 1, 5, 10, 50$ and parameters of $\theta = 0.5, 1, 5, 10, 50$ by simulating random numbers of the re-parameterized inverse Gaussian distribution with the composition method (Ngamkham, 2019) of the sample sizes of $n = 10, 30, 50, 100, 1,000, 10,000$. We took the random numbers which we obtained to

estimate the parameters of λ and θ with the maximum likelihood method (Jangphanish & Budsaba, 2013), and afterward they were redone 10,000 rounds in each case. Lastly, asymptotic confidence ellipses were constructed and calculated the value of coverage probability of asymptotic confidence ellipses with the confidence coefficient of 0.98 by using RStudio programming as indicated in Tables 1-6 and Figures 1-6)Here it shows some cases(.

Table 1. A sample size of $n = 10$, an average of parameter estimates of λ , an average of parameter estimates of θ and the value of Coverage Probability (CP)

λ	θ	The average estimates		CP
		$\hat{\lambda}$	$\hat{\theta}$	
0.5	0.5	0.8675	0.5452	0.5823
	1	0.8575	1.0897	0.5937
	5	0.8545	5.4169	0.5906
	10	0.8719	10.3855	0.5974
	50	0.8508	54.4864	0.5907
1	0.5	1.5615	0.5036	0.6373
	1	1.5807	0.9974	0.6337
	5	1.5689	4.938	0.6478
	10	1.5628	9.9506	0.6407
	50	1.5689	49.7005	0.6502
5	0.5	7.2783	0.4597	0.6436
	1	7.3671	0.9119	0.6457
	5	7.2711	4.5582	0.6517
	10	7.2696	9.1681	0.6509
	50	7.2986	45.7389	0.6506
10	0.5	14.4873	0.4558	0.6105
	1	14.5794	0.9016	0.6151
	5	14.4262	4.5457	0.6084
	10	14.378	9.0936	0.618
	50	14.423	45.6088	0.6182
50	0.5	71.0387	0.4527	0.4894
	1	72.0167	0.9048	0.4791
	5	70.3464	4.5332	0.4969
	10	71.1927	9.0702	0.4932
	50	71.7472	44.921	0.4858

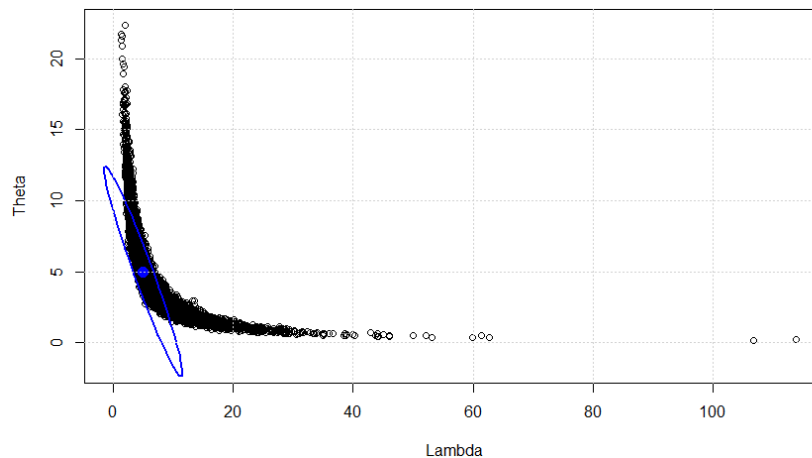


Figure 1. Asymptotic confidence ellipses in the case of $n=10$, $\lambda=5$, $\theta=5$

Table 2 . A sample size of $n=30$, an average of parameter estimates of λ , an average of parameter estimates of θ and the value of Coverage Probability (CP)

λ	θ	The average estimates		CP
		$\hat{\lambda}$	$\hat{\theta}$	
0.5	0.5	0.5905	0.5186	0.7504
	1	0.5937	1.0262	0.7474
	5	0.5914	5.1701	0.7484
	10	0.5956	10.2686	0.743
	50	0.5909	52.1257	0.7439
1	0.5	1.1473	0.4996	0.7897
	1	1.1474	0.9978	0.7949
	5	1.1508	4.9867	0.7892
	10	1.1518	9.9833	0.7917
	50	1.1459	50.0676	0.7916
5	0.5	0.4874	5.5757	0.8024
	1	5.5935	0.9745	0.7957
	5	5.568	4.8754	0.8055
	10	5.611	9.7014	0.7985
	50	5.5712	48.8262	0.8056
10	0.5	11.2136	0.4822	0.7612
	1	11.1245	0.9716	0.7676
	5	11.1737	4.8381	0.7652
	10	11.105	9.7371	0.7672
	50	11.1026	48.6128	0.7687
50	0.5	55.5276	0.4852	0.6323
	1	55.5863	0.9677	0.6355
	5	55.6313	4.8275	0.635
	10	55.4025	9.7275	0.6293
	50	56.0297	47.9986	0.6304

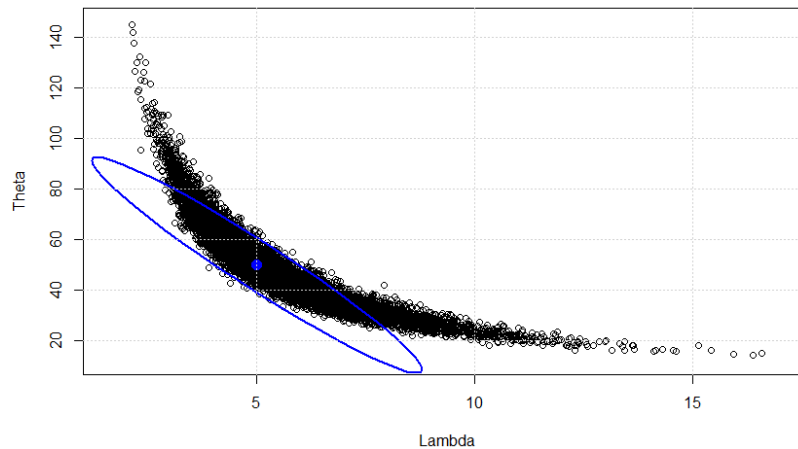


Figure 2. Asymptotic confidence ellipses in the case of $n = 30$, $\lambda = 5$, $\theta = 50$

Table 3 . A sample size of $n = 50$, an average of parameter estimates of λ , an average of parameter estimates of θ and the value of Coverage Probability (CP)

λ	θ	The average estimates		CP
		$\hat{\lambda}$	$\hat{\theta}$	
0.5	0.5	0.5529	0.513	0.7964
	1	0.5544	1.0175	0.8063
	5	0.5544	1.0175	0.8063
	10	0.5539	10.2085	0.7962
	50	0.5514	51.2268	0.8001
1	0.5	1.09	0.4969	0.84
	1	1.0868	1.001	0.8395
	5	1.0899	4.9634	0.8409
	10	1.0901	9.9562	0.8381
	50	1.0918	49.7151	0.8394
5	0.5	5.3281	0.4928	0.8463
	1	5.3495	0.9805	0.8478
	5	5.3338	4.9295	0.8468
	10	5.3388	9.8379	0.8522
	50	5.3318	49.2313	0.8523
10	0.5	10.6684	0.4902	0.8183
	1	10.6326	0.9857	0.8159
	5	10.6656	4.91	0.823
	10	10.6833	9.8181	0.8114
	50	10.6307	49.2032	0.8219
50	0.5	53.2004	0.4907	0.691
	1	53.2717	0.9796	0.6989
	5	53.1505	4.9103	0.6975
	10	53.3072	9.7789	0.7016
	50	53.3408	48.9673	0.6922

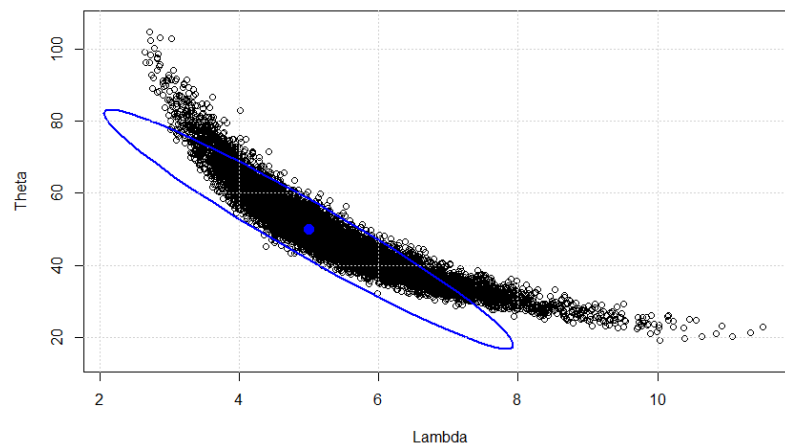


Figure 3. Asymptotic confidence ellipses in the case of $n = 50$, $\lambda = 5$, $\theta = 50$

Table 4 . A sample size of $n = 100$, an average of parameter estimates of λ , an average of parameter estimates of θ and the value of Coverage Probability (CP)

λ	θ	The average estimates		CP
		$\hat{\lambda}$	$\hat{\theta}$	
0.5	0.5	0.5254	0.5052	0.8599
	1	0.528	1.0048	0.8579
	5	0.5278	5.0348	0.8547
	10	0.5249	10.1248	0.8659
	50	0.5248	50.6336	0.8694
1	0.5	1.0388	0.5013	0.9001
	1	1.0394	1.001	0.8976
	5	1.0419	4.9992	0.8939
	10	1.0423	10.0013	0.8894
	50	1.0415	49.9095	0.8994
5	0.5	5.1529	0.4974	0.8987
	1	5.1646	0.9922	0.8991
	5	5.1495	4.97	0.9015
	10	5.1658	9.9094	0.8984
	50	5.1621	49.6401	0.8967
10	0.5	10.2967	0.4969	0.8713
	1	10.3127	0.992	0.8733
	5	10.3233	4.9551	0.874
	10	10.3107	9.9127	0.8848
	50	10.3062	49.5991	0.8824
50	0.5	51.5636	0.4952	0.7704
	1	51.4611	0.9921	0.7747
	5	51.5803	4.9467	0.7748
	10	51.4586	9.9216	0.7716
	50	51.6396	49.429	0.7787

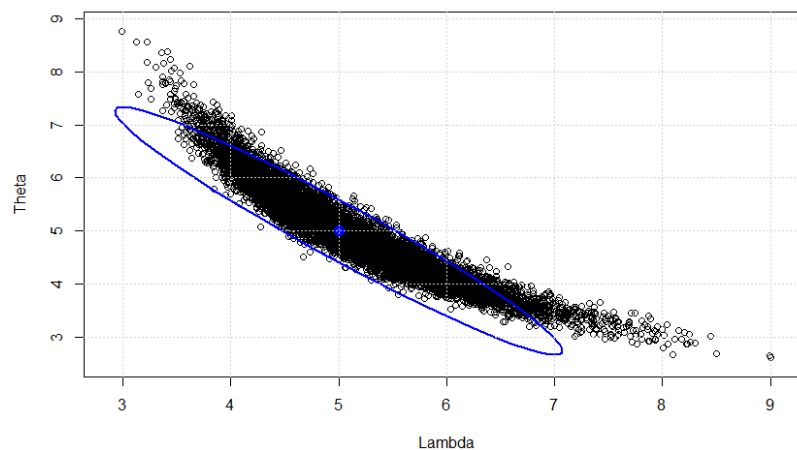


Figure 4. Asymptotic confidence ellipses in the case of $n=100$, $\lambda=5$, $\theta=5$

Table 5 . A sample size of $n=1,000$, an average of parameter estimates of λ , an average of parameter estimates of θ and the value of Coverage Probability (CP)

λ	θ	The average estimates		CP
		$\hat{\lambda}$	$\hat{\theta}$	
0.5	0.5	0.5023	0.5008	0.9625
	1	0.5026	1.0007	0.961
	5	0.5023	5.0072	0.9602
	10	0.5026	10.0084	0.9594
	50	0.5028	50.021	0.9596
1	0.5	1.0033	0.5003	0.9685
	1	1.0036	1.0008	0.9684
	5	1.0041	5.0027	0.9654
	10	1.0047	9.9927	0.9676
	50	1.0035	50.0385	0.9668
5	0.5	5.0171	0.4994	0.9668
	1	5.0161	0.9991	0.966
	5	5.0173	4.9944	0.9685
	10	5.0204	9.9834	0.9655
	50	5.0188	49.9296	0.9687
10	0.5	10.0401	0.4992	0.9616
	1	10.0263	0.9996	0.9654
	5	10.0251	4.9985	0.9634
	10	10.0278	9.9965	0.9627
	50	10.0287	49.9667	0.9651
50	0.5	50.1816	0.4992	0.9322
	1	50.1523	0.999	0.9301
	5	50.1736	4.9928	0.9376
	10	50.1553	9.9891	0.9328
	50	50.1492	49.9536	0.9316

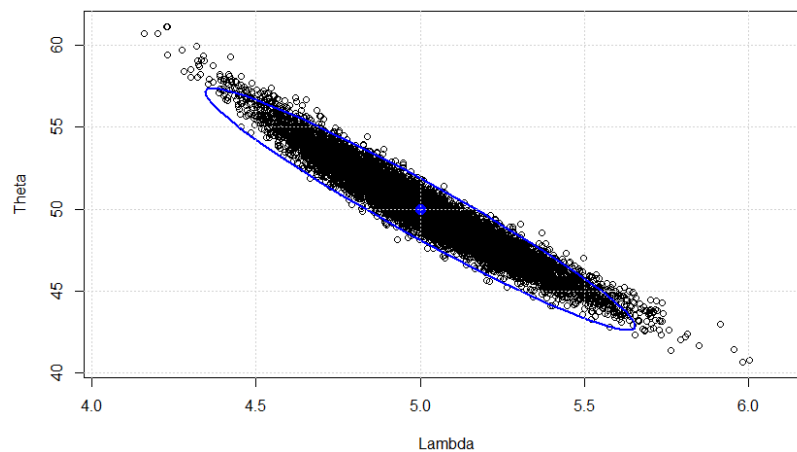


Figure 5. Asymptotic confidence ellipses in the case of $n=1,000$, $\lambda=5$, $\theta=50$

Table 6 . A sample size of $n=10,000$, an average of parameter estimates of λ , an average of parameter estimates of θ and the value of Coverage Probability (CP)

λ	θ	The average estimates		CP
		$\hat{\lambda}$	$\hat{\theta}$	
0.5	0.5	0.5003	0.4999	0.9776
	1	0.5001	1.0005	0.9759
	5	0.5002	5.0002	0.9776
	10	0.5003	9.9975	0.9768
	50	0.5004	49.985	0.9807
1	0.5	1.0006	0.4999	0.9772
	1	1.0005	0.9997	0.9783
	5	1.0002	5.0013	0.9787
	10	1.0005	9.9972	0.9787
	50	1.0005	49.9921	0.9791
5	0.5	5.0003	0.5001	0.9788
	1	5.0024	0.9998	0.9799
	5	5.0033	4.9981	0.9782
	10	5.0021	9.9982	0.977
	50	5.002	49.9908	0.9761
10	0.5	10.0004	0.5001	0.9762
	1	10.0025	1.0001	0.977
	5	10.002	5.0002	0.9787
	10	10.003	9.9994	0.9765
	50	10.0031	49.9937	0.9788
50	0.5	50.0165	0.4999	0.9758
	1	50.0279	0.9996	0.9734
	5	50.0095	5.0002	0.972
	10	50.0122	9.9997	0.9712
	50	50.0095	50.0144	0.9773

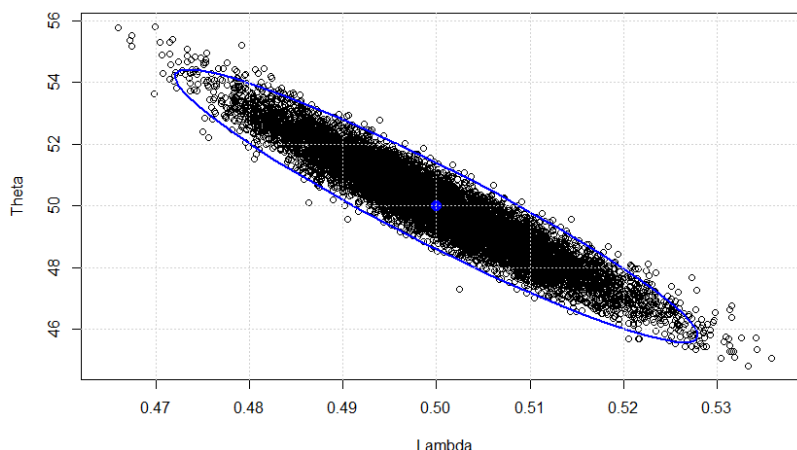


Figure 6. Asymptotic confidence ellipses in the case of $n=10,000$, $\lambda=0.5$, $\theta=50$

Discussion

The research results of simulation as shows in Table 1-6 revealed that the parameter estimation of λ that had been introduced had more values than the actual values)overestimate(. Also, from Figure 2: $n=30$, $\lambda=5$, $\theta=50$ showed that several parameter estimations came out of the ellipse, but when the sample size had increased to 1,000 as shown in Figure 5: $n=1,000$, $\lambda=5$, $\theta=50$, the number of parameter estimations coming out of the ellipse decreased. Similarly, from Figure 1: $n=10$, $\lambda=5$, $\theta=5$ were found that several parameter estimations came out of the ellipse, but when the sample size had increased to 10,000 as it indicates in Figure 6: $n=10,000$, $\lambda=5$, $\theta=5$, the number of parameter estimations coming out of the ellipse reduced significantly.

The research indicated that the coverage probability of asymptotic confidence ellipses would have been increasing value when the sample size of n rose up. As we calculated, the value of coverage probability had been close to the confidence coefficient of 0.98 at the sample size of $n=10,000$ which gave us the maximum coverage probability. Consequently, the value of coverage probability of asymptotic confidence ellipses for the re-parameterized inverse Gaussian distribution would have been increasing when the sample size was large.

In this research, we use the maximum likelihood estimates that have more errors when sample sizes are small. Therefore, the value of coverage probabilities are very low in the cases of small sample sizes.

Besides, for n less than or equal to 100, the parameter of $\lambda=5$ and the various parameter of θ gave us the maximum coverage probability (as shown in Table 1-4). Therefore, the various parameter of λ were able to cause the value of coverage probability of asymptotic confidence ellipses for the re-parameterized inverse Gaussian distribution.

Conclusion

In the theoretical part, we calculate the Fisher information matrix for sample sizes n in closed form for a probability density function under the re-parameterized inverse Gaussian distribution and find the covariance matrix which equals an inverse of the Fisher information

matrix for the sample size n to construct a normal approximation that gives elliptical confidence regions with center $\Theta = (\lambda, \theta)$ for all 150 cases.

For the simulation part, the objective of this part is conducted for an asymptotic confidence ellipses construction for parameters of the re-parameterized inverse Gaussian distribution and we compare the coverage probabilities of confidence ellipses of parameters for a re-parameterized inverse Gaussian distribution with the confidence coefficient 0.98. We consider the parameters of $\lambda = 0.5, 1, 5, 10, 50$, $\theta = 0.5, 1, 5, 10, 50$ and the sample sizes n as 10, 30, 50, 100, 1,000, 10,000. The data were generated by a simulation technique using RStudio programming and the experimental is repeated 10,000 times to obtain the coverage probability. The simulation results indicated that the coverage probability of asymptotic confidence ellipses would have been increasing value when the sample size of n increase and the various values of λ can cause the coverage probability values of asymptotic confidence ellipses to increase or decrease. In connection with this research, we could continue applying to contemporary problems for real life data, mostly in Engineering. Also, we would like to suggest other distribution such as Weibull distribution, log-normal distribution, extreme value distribution, Gompertz distribution and log-logistic distribution for constructing the confidence ellipses.

Acknowledgments

This research was funded by the Faculty of Applied Science, King Mongkut's University of Technology North Bangkok in 2019. Contract no. 6245101

References

- Ahmed, S.E., Budsaba, K., Lisawadi, S., & Volodin, A. (2008). Parametric estimation for the Birnbaum-Saunders lifetime distribution based on new parametrization. *Thailand Statistician*, 6(2), 213-240.
- Birnbaum, Z. W., & Saunders, S. C. (1969). A new family of lifetime distribution. *Applied Probability*, 6, 319-327.
- Duangchana, N. & Budsaba, K. (2014). Asymptotic confidence ellipses of parameters for the inverse Gaussian distribution. *Thammasat International Journal of Science and Technology*, 19(2), 22-29.
- Folks, J. L. & Chhikara, R. S. (1978). The inverse Gaussian distribution and its statistical application - a review. *Journal of the Royal Statistical Society. Serie B*, 40, 263-289.
- Hansongnorn, K., Sakonthawichot, B., Sittisareesmer, A. & Phaphan, W. (2018). Estimation of asymptotic confidence ellipses for Birnbaum-Saunders distribution. *Burapha Science Journal*, 23(2), 1029-1043. (in Thai)
- Jangphanish, K. & Budsaba, K. (2013). Parameter estimation for re-parameterized inverse Gaussian distribution. *Thammasat International Journal of Science and Technology*, 18(1), 43-53.
- Ngamkham, T. (2019). On the crack random numbers generation procedure. *Lobachevskii Journal of Mathematics*, 40(8), 1204-1217.
- Phaphan, W. & Pongsart, T. (2019). Asymptotic confidence ellipses for length-biased inverse Gaussian distribution. *The Journal of KMUTNB*, 29(2), 332-341. (in Thai)
- Schrödinger, E. (1915). Zur theorie der fall-und steigversuche an teilchen mit brownischer bewegung. *Physikalische Zeitschrift*, 16, 289-295.
- Tweedie, M.C.K. (1957). Statistical properties of inverse Gaussian distributions I. *The Annals of Mathematical Statistics*, 28, 362-377.

Research Article

New ratio estimators for population mean in simple random sampling using robust regression

Nuanpan Lawson^{1*}

¹ Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand.

*E-mail: nuanpan.n@sci.kmutnb.ac.th

Received: 29/03/2020; Revised: 27/04/2020; Accepted: 11/05/2020

Abstract

Using the traditional least square regression estimator we can validate the assumption of normality of the residual error when outliers occur in the data. In this paper, alternative ratio estimators for estimating population mean using robust regression are proposed for use in the case where data is contaminated with outliers, under simple random sampling. The bias and mean square error of the proposed estimators have been investigated. A simulation study has been conducted to compare the efficiency of the proposed estimators with traditional estimators. The results show that the proposed estimators perform better than the existing estimators.

Keywords: ratio estimator, simple random sampling, robust regression, Taylor series approximation, Huber M-estimate.

Introduction

Auxiliary information could be utilised for estimating population mean of the variable of interest if the auxiliary variable is related to the variable of interest. Ratio and regression estimators use the benefit of related auxiliary variables and the variable of interest in order to increase the efficiency of the population mean estimators. Several researchers developed ratio and regression estimators in their study using known parameters of auxiliary variables. For example, Sisodia & Dwivedi (1981) proposed to use the benefit of a correlation coefficient to develop a new ratio estimator for estimating population mean. Later Singh & Tailor (2003) also improve the ratio estimator by adding the correlation coefficient into the traditional one. (see also Subramani & Kumarapandiyam, 2012; Jaroengratikun & Lawson, 2019; Lawson, 2019). Some researchers proposed to use the regression coefficient to develop new ratio estimators. For instance, Nangsue (2009) proposed new ratio and regression estimators in the case of missing data. One of Nangsue's estimators utilizing the regression coefficient is as follows.

$$\hat{Y}_1 = \left[\bar{y} + b(\bar{X} - \bar{x}) \right] \left(\frac{\bar{X}}{\bar{x}} \right)^b, \quad (1)$$

where $\bar{y}$ and $\bar{x}$ are the sample means of the study variable X and auxiliary variable Y respectively. $\bar{X}$ is the population mean of auxiliary variable is the sample b and X regression coefficient.

Later, Soponviwatkul & Lawson (2017) proposed new ratio estimators following Sisodia & Dwivedi (1981), Singh & Tailor (2003) and Nangsue (2009). Soponviwatkul & Lawson (2017) estimators performs well when conditions are satisfied.

However, estimation of regression coefficients using an ordinary least square method is not suitable for the data that is contaminated with outliers. The outliers will affect the normal distribution assumption of the traditional least square regression method. Robust regression is an alternative method of estimation dealing with data with outliers in order to increase the efficiency of the estimators using an idea of weighting method by assigning different weights to observations. Some researches proposed to use robust regression in the estimation of the regression coefficients in order to increase the efficiency when data is contaminated by outliers. For example, Kadilar et al. (2007) proposed using robust regression in the estimation of population mean using ratio estimators in the case of where data is contaminated under simple random sampling. They considered the ratio estimators proposed by Sisodia & Dwivedi (1981), Singh & Kakran (1993), Upadhyaya & Singh (1999) and Kadilar & Cingi (2004) in their study. Kadilar et al. (2007) applied robust regression called Huber M- estimator to estimate the coefficient in ratio estimator. The Huber M- estimator assigns different weights to the observations dependent on the residuals, observations with larger residuals get smaller weights. Kadilar et al. (2007) estimators are as follows.

$$\hat{Y}_{pr1} = [\bar{y} + b_{rob}(\bar{X} - \bar{x})] \left(\frac{\bar{X}}{\bar{x}} \right), \quad (2)$$

$$\hat{Y}_{pr2} = [\bar{y} + b_{rob}(\bar{X} - \bar{x})] \left(\frac{\bar{X} + C_x}{\bar{x} + C_x} \right), \quad (3)$$

$$\hat{Y}_{pr3} = [\bar{y} + b_{rob}(\bar{X} - \bar{x})] \left(\frac{\bar{X} + \beta_{2(x)}}{\bar{x} + \beta_{2(x)}} \right), \quad (4)$$

$$\hat{Y}_{pr4} = [\bar{y} + b_{rob}(\bar{X} - \bar{x})] \left(\frac{\beta_{2(x)}\bar{X} + C_x}{\beta_{2(x)}\bar{x} + C_x} \right), \quad (5)$$

$$\hat{Y}_{pr5} = [\bar{y} + b_{rob}(\bar{X} - \bar{x})] \left(\frac{C_x\bar{X} + \beta_{2(x)}}{C_x\bar{x} + \beta_{2(x)}} \right), \quad (6)$$

where C_x is the coefficient of variation of X , $\beta_{2(x)}$ is the coefficient of kurtosis of X and b_{rob} is the sample regression coefficient estimated by the Huber M-estimator. Huber (1981)

proposed to minimize the summation of outlier function. The objective function is $\min \sum_{i=1}^n \rho(e_i)$

and Huber's function (ρ) is given by $\rho(e) = \begin{cases} \frac{e^2}{2} & , |e| \leq k \\ k|e| - \frac{k^2}{2} & , |e| > k \end{cases}$

The mean square error of the Kadilar et al. (2007) estimators are given by

$$MSE(\hat{Y}_{pri}) = \frac{(1-f)}{n} \bar{Y}^2 \left[W_i^2 C_X^2 + C_Y^2 + b_{rob}^2 K^2 C_X^2 - 2W_i \rho C_X C_Y + 2W_i b_{rob} K C_X^2 - 2b_{rob} K \rho C_X C_Y \right], \quad (7)$$

$$\text{where } W_1 = \frac{\bar{X}}{\bar{X}}, \quad W_1 = \frac{\bar{X}}{\bar{X} + C_x}, \quad W_3 = \frac{\bar{X}}{\bar{X} + \beta_{2(x)}}, \quad W_4 = \frac{\beta_{2(x)} \bar{X}}{\beta_{2(x)} \bar{X} + C_x},$$

$$W_5 = \frac{C_x \bar{X}}{C_x \bar{X} + \beta_{2(x)}} \text{ and } K = \frac{\bar{X}}{\bar{Y}}, \quad i=1,2,\dots,5$$

Later Zaman & Bulut (2018) compared the efficiency of the estimators proposed by Kadilar et al. (2007) using the Huber M- estimator with other estimators using different methods of robust regression with each. Recently, Zaman (2019) proposed a combined estimator which combined the estimators proposed by Zaman & Bulut (2018) when data is contaminated with an outlier and it is found that the new estimators are more efficient than the old ones under certain conditions.

In this paper, the alternative ratio estimators for estimating population mean using robust regression in the case that data is contaminated with outliers under simple random sampling, following the work of Nangsue (2009) and Kadilar et al. (2007), have been proposed. The proposed population mean estimators are discussed in Section 2. A simulation study is shown in Section 3 to compare the efficiency of the proposed estimators with the traditional estimators. Finally conclusion and discussion are given in section 4.

Methods

Motivated by Nangsue (2009) and Kadilar et. al (2007), we proposed new ratio estimators for use in the case of the existence of contaminated data by using a robust regression method to estimate the sample regression coefficient under simple random sampling. The proposed estimators are as follows.

$$\hat{Y}_{N1} = \left[\bar{y} + b_{rob} (\bar{X} - \bar{x}) \right] \left(\frac{\bar{X}}{\bar{x}} \right)^{b_{rob}}, \quad (8)$$

$$\hat{Y}_{N2} = \left[\bar{y} + b_{rob} (\bar{X} - \bar{x}) \right] \left(\frac{\bar{X} + C_x}{\bar{x} + C_x} \right)^{b_{rob}}, \quad (9)$$

$$\hat{Y}_{N3} = [\bar{y} + b_{rob}(\bar{X} - \bar{x})] \left(\frac{\bar{X} + \beta_{2(x)}}{\bar{x} + \beta_{2(x)}} \right)^{b_{rob}}, \quad (10)$$

$$\hat{Y}_{N4} = [\bar{y} + b_{rob}(\bar{X} - \bar{x})] \left(\frac{\beta_{2(x)}\bar{X} + C_x}{\beta_{2(x)}\bar{x} + C_x} \right)^{b_{rob}}, \quad (11)$$

$$\hat{Y}_{N5} = [\bar{y} + b_{rob}(\bar{X} - \bar{x})] \left(\frac{C_x\bar{X} + \beta_{2(x)}}{C_x\bar{x} + \beta_{2(x)}} \right)^{b_{rob}}, \quad (12)$$

From equations (8) to (12), we can write a general form for the proposed estimators as follows.

$$\hat{Y}_N = [\bar{y} + b_{rob}(\bar{X} - \bar{x})] \left(\frac{A\bar{X} + D}{A\bar{x} + D} \right)^{b_{rob}}, \quad (13)$$

where C_x is the coefficient of variation of X , $\beta_{2(x)}$ is the coefficient of the kurtosis of X , A and D are parameters of an auxiliary variable X and b_{rob} is the sample regression coefficient estimated by the Huber M-estimator.

Note that: where there are no outliers (an outlier is usually a data point that is larger than the third quartile or smaller than the first quartile), the proposed family of estimators in equation (13) can also be used by replacing b_{rob} with b , a sample regression coefficient estimated by the ordinary least square method.

The bias and mean square error of the proposed estimators are investigated under simple random sampling without replacement using a Taylor series approximation.

$$\text{Let } e_0 = \frac{\bar{y} - \bar{Y}}{\bar{Y}} \text{ then } \bar{y} = \bar{Y}(1 + e_0) \text{ and let } e_1 = \frac{\bar{x} - \bar{X}}{\bar{X}} \text{ then } \bar{x} = \bar{X}(1 + e_1), \text{ such that}$$

$$E(e_0) = E(e_1) = 0, E(e_0^2) = \frac{1-f}{n} C_y^2, E(e_1^2) = \frac{1-f}{n} C_x^2 \text{ and } E(e_0 e_1) = \frac{1-f}{n} C_{xy} = \frac{1-f}{n} \rho C_y C_x.$$

Expressing equation (13) in terms of e 's, we have

$$\begin{aligned} \hat{Y}_N &= [\bar{Y}(1 + e_0) + b_{rob}(\bar{X} - \bar{X}(1 + e_1))] \left(\frac{A\bar{X} + D}{A\bar{X}(1 + e_1) + D} \right)^{b_{rob}}, \\ &= [\bar{Y}(1 + e_0) + b_{rob}(\bar{X} - \bar{X}(1 + e_1))] (1 + W e_1)^{-b_{rob}}, \\ &= \bar{Y} - W b_{rob} e_1 \bar{Y} + \frac{b_{rob}(b_{rob} + 1)}{2} W^2 e_1^2 \bar{Y} + \bar{Y} e_0 - W b_{rob} \bar{Y} e_0 e_1 + \frac{b_{rob}(b_{rob} + 1)}{2} W^2 e_1^2 \bar{Y} e_0 \\ &\quad - b_{rob} \bar{X} e_1 + b_{rob}^2 \bar{X} W e_1^2. \end{aligned}$$

Up to the first degree of approximation using a Taylor series, the approximate bias of the proposed estimator is given by

$$\text{Bias}(\hat{Y}_N) \approx (\hat{Y}_N - \bar{Y})$$

$$\begin{aligned} &= E(-Wb_{rob}e_1\bar{Y} + \frac{b_{rob}(b_{rob}+1)}{2}W^2e_1^2\bar{Y} + e_0\bar{Y} - Wb_{rob}e_0e_1\bar{Y} + \frac{b_{rob}(b_{rob}+1)}{2}W^2e_0e_1^2\bar{Y} \\ &\quad - b_{rob}e_1\bar{X} + b_{rob}^2We_1^2\bar{X}) \\ &= \frac{(1-f)}{n}\bar{Y}\left[\left(\frac{b_{rob}(b_{rob}+1)}{2}W^2 + b_{rob}^2KW\right)C_X^2 - Wb_{rob}\rho C_X C_Y\right], \end{aligned} \quad (14)$$

The approximate MSE of the proposed estimator $\hat{Y}_N$ to the first order of approximation is given by

$$\begin{aligned} \text{MSE}(\hat{Y}_N) &\approx E(\hat{Y}_N - \bar{Y})^2 \\ &= E(W^2b_{rob}^2e_1^2\bar{Y}^2 + e_0^2\bar{Y}^2 + b_{rob}^2e_1^2\bar{X}^2 - 2Wb_{rob}e_0e_1\bar{Y}^2 + 2Wb_{rob}^2e_1^2\bar{X}\bar{Y} - 2b_{rob}e_0e_1\bar{X}\bar{Y}) \\ &= \frac{(1-f)}{n}\bar{Y}^2\left[W^2b_{rob}^2C_X^2 + C_Y^2 + b_{rob}^2K^2C_X^2 - 2Wb_{rob}\rho C_X C_Y + 2Wb_{rob}^2KC_X^2 - 2b_{rob}K\rho C_X C_Y\right] \\ &= \frac{(1-f)}{n}\bar{Y}^2\left[C_Y^2 + (W^2b_{rob}^2 + b_{rob}^2K^2 + 2Wb_{rob}^2K)C_X^2 - 2(Wb_{rob} + b_{rob}K)\rho C_X C_Y\right], \end{aligned} \quad (15)$$

where $W = \frac{A\bar{X}}{A\bar{X} + D}$ and $K = \frac{\bar{X}}{\bar{Y}}$.

To see the performance of the proposed estimators, the mean square error of the proposed estimators in equation (15) will be compared with the mean square error of the Kadilar et al. (2007) estimators in equation (7). The proposed estimators are more efficient than the estimators of Kadilar et al. (2007) if the conditions below are satisfied. The details are as follows:

The proposed general estimator $\hat{Y}_N$ is more efficient than the estimator $\hat{Y}_{pri}$, if

$$\text{MSE}(\hat{Y}_N) < \text{MSE}(\hat{Y}_{pri})$$

$$W^2b_{rob}^2C_X^2 + C_Y^2 + b_{rob}^2K^2C_X^2 - 2Wb_{rob}\rho C_X C_Y + 2Wb_{rob}^2KC_X^2 - 2b_{rob}K\rho C_X C_Y < W^2C_X^2 + C_Y^2 + b_{rob}^2K^2C_X^2 - 2W\rho C_X C_Y + 2Wb_{rob}KC_X^2 - 2b_{rob}K\rho C_X C_Y$$

$$W^2C_X^2(b_{rob}^2 - 1) - 2W\rho C_X C_Y(b_{rob} - 1) + 2WK C_X^2(b_{rob}^2 - b_{rob}) < 0$$

$$(b_{rob} - 1)\left[W^2C_X^2(b_{rob} + 1) - 2W\rho C_X C_Y + 2WK C_X^2b_{rob}\right] < 0$$

For $b_{rob} - 1 > 0$ that is $b_{rob} > 1$ then

$$W^2C_X^2(b_{rob} + 1) - 2W\rho C_X C_Y + 2WK C_X^2b_{rob} < 0$$

$$b_{rob} < \frac{2W\rho C_X C_Y - W^2 C_X^2}{2WKC_X^2 + W^2 C_X^2}$$

Similarly, for $b_{rob} - 1 < 0$ that is $b_{rob} < 1$ then

$$b_{rob} > \frac{2W\rho C_X C_Y - W^2 C_X^2}{2WKC_X^2 + W^2 C_X^2},$$

$$\text{where } W = \frac{A\bar{X}}{A\bar{X} + D} \text{ and } K = \frac{\bar{X}}{\bar{Y}}.$$

Results and discussion

The proposed estimators are used in simulation data in order to see the performance of the proposed estimators against the estimators of Kadilar et al. (2007). A population of (X, Y) is generated by the bivariate normal distribution with mean equal to 500 and 40 respectively. The coefficient of variation of X and Y are equal to 0.13 and 0.25 respectively. The correlation coefficient between X and Y is equal to 0.9. A sample size $n = 5, 20, 30, 50$ is selected using simple random sampling without replacement from the population size $N = 100$. There is an outlier in the sample set of data. The simulation is repeated 10,000 times using the R program. The bias and mean square error for all estimators is calculated. The results are presented in Table 1 below.

Table 1 The bias and mean square error of the proposed estimators and existing estimators.

Estimators	$n = 5$		$n = 20$		$n = 30$		$n = 50$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{Y}_{pr1}$	6.767	46.366	3.237	11.581	1.540	2.892	0.628	0.593
$\hat{Y}_{pr2}$	7.053	50.117	2.980	9.991	1.398	2.494	0.568	0.530
$\hat{Y}_{pr3}$	6.822	47.063	3.186	11.254	1.511	2.809	0.616	0.580
$\hat{Y}_{pr4}$	6.777	46.488	3.228	11.523	1.535	2.877	0.626	0.590
$\hat{Y}_{pr5}$	6.767	46.367	3.237	11.580	1.539	2.892	0.629	0.593
$\hat{Y}_{N1}$	0.970	8.492	0.307	1.413	0.226	0.783	0.137	0.324
$\hat{Y}_{N2}$	0.405	7.952	0.167	1.393	0.141	0.779	0.099	0.326
$\hat{Y}_{N3}$	0.862	8.338	0.279	1.405	0.209	0.781	0.129	0.324
$\hat{Y}_{N4}$	0.951	8.463	0.302	1.411	0.223	0.783	0.136	0.324
$\hat{Y}_{N5}$	0.969	8.492	0.307	1.413	0.226	0.783	0.137	0.324

From Table 1 we can see that all proposed estimators perform a lot better than the estimators of Kadilar et al. (2007) in both bias and mean square error. The mean square errors of all proposed estimators are similar. The bias and mean square error are smaller when

sample size increases. The $\hat{Y}_{N2}$ seems to perform the best in term of bias and mean square error in this scenario. The bias and mean square errors for the proposed estimators are similar when sample size increased.

Conclusion

Outliers can affect the interpretation of results if they are not deal with before analysis. In this paper new ratio estimators have been proposed using the benefit of robust regression to solve a problem of contaminated data. We also presented the proposed estimators in a general form that can be useful with some known auxiliary variables. We have shown in theory that the mean square error of the proposed estimators are smaller than the existing estimators in certain conditions. A simulation study was also used to support the theory that the alternative ratio estimators for estimating population mean and the proposed estimators performed well when compare to the existing estimators. The Huber M-estimator has been used to estimate the regression coefficient for the robust regression method. The other robust regression methods, e.g., the Hampel (1971), Turkey (1977) methods can also be applied in the proposed estimators when data is contaminated with outliers. The alternative estimators may be beneficial in estimating population mean data with outliers.

Acknowledgement

This work was funded by The Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Thailand. I would like to thank the unknown referees for offering useful suggestions for improving this paper.

References

- Hampel, F. R. (1971). A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, 42 (6):1887–96. doi:10.1214/aoms/1177693054
- Huber, P. (1981). *Robust Statistics*, Wiley, New York.
- Jaroengratikun, U. & Lawson, N. (2019). A combined family of ratio estimators for population mean using an auxiliary variable in simple random sampling, *Journal of Mathematical and Fundamental Sciences*, 51(1), 1-12. DOI: 10.5614/j.math.fund.sci.2019.51.1.1
- Kadilar, C., & H. Çingi. (2004). Ratio estimators in simple random sampling. *Applied Mathematics and Computation*, 151 (3), 893–902. doi:10.1016/S0096-3003(03)00803-8
- Kadilar, C., M. Candan & H. Çingi. (2007). Ratio estimators using robust regression. *Hacettepe Journal of Mathematics and Statistics* 36 (2):181–88.
- Lawson, N. (2019). Ratio estimators of population means using quartile function of auxiliary variable using double sampling, *Songklanakarin Journal of Science and Technology*, 41(1), 117-122.
- Nangsue, N. (2009). Adjusted ratio and regression type estimators for estimation of population mean when some observations are missing. *International Scholarly and Scientific Research & Innovation*, 3(5), 334-337.
- Sisodia, B. V. & Dwivedi, V. K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable. *Journal Indian Society of Agricultural Statistics*, 33(2), 13-18.
- Singh, H. P., & Kakran, M. S. (1993). A modified ratio estimator using known coefficient of kurtosis of an auxiliary character. (Unpublished manuscript).

- Singh, H.P. & Tailor, R. (2003). Use of known correlation coefficient in estimating the finite population mean, *Statistics in Transition*, 6, pp. 555-560.
- Soponviwatkul, K. & Lawson, N. (2017). New ratio estimators for estimating population mean in simple random sampling using a coefficient of variation, correlation coefficient and a regression coefficient. *Gazi University Journal of Science*, 30(4), 610-621.
- Subramani, J. & Kumarapandiyan, G. (2012). Modified ratio estimators for population mean using function of quartiles of auxiliary variable. *Bonfring International Journal of Industrial Engineering and Management Science*, 2 (2): 19-23.
- Tukey, J. W. (1977). *Exploratory data analysis*. MA: Addison-Wesley.
- Upadhyaya, L. N. & Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal* 41 (5):627-36. doi:10.1002/(SICI)1521-4036(199909)41:5%3c627::AID-BIMJ627%3e3.0.CO;2-W
- Zaman, T. (2019). Improvement of modified ratio estimators using robust regression methods. *Applied Mathematics and Computation*, Elsevier, 348(C), 627-631. <https://doi.org/10.1016/j.amc.2018.12.037>
- Zaman, T. & Bulut, H. (2018). Modified ratio estimators using robust regression methods, *Communications in Statistics -Theory and Methods*, 0(0), 1-10. <https://doi.org/10.1080/03610926.2018.1441419>

Research Article

การพยากรณ์ทิศทางของราคาหุ้นรายวันจากข้อความข่าวภาษาไทย โดยใช้วิธีการประมวลผลภาษาธรรมชาติ

The predictions of a daily stock price direction from the Thai news content by using natural language processing

วิกานดา ผาพันธุ์^{1*} และ อัญชนา พิมพ์สาล¹

Wikanda Phaphan^{1*}, and Aunchana Pimpisal¹

¹ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

¹Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok

*E-mail: wikanda.p@sci.kmutnb.ac.th

Received: 16/04/2020; Revised: 14/05/2020; Accepted: 15/06/2020

บทคัดย่อ

ปัจจัยที่กระทบต่อราคาของหุ้นในตลาดหลักทรัพย์แห่งประเทศไทยนั้นมีอยู่หลายปัจจัย ข่าวสารต่าง ๆ ก็เป็นปัจจัยหนึ่งที่มีผลกระทบต่อราคาของหุ้น ผู้วิจัยจึงเกิดแนวคิดที่จะพยากรณ์ทิศทางของราคาหุ้นรายวันจากข้อความข่าวโดยใช้วิธีการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) เพื่อให้นักลงทุนสามารถคาดคะเนทิศทางของราคาหุ้นก่อนที่จะตลาดหลักทรัพย์แห่งประเทศไทยเปิดโดยศึกษาข้อความข่าวจากแหล่งข่าวต่างๆ และใช้การตัดคำ (Tokenizer) จาก Library pythainlp ในโปรแกรมภาษาไพธอน ver.3.7.1 จากนั้นสร้างแบบจำลองโดยใช้ตัวแบบการจำแนก (classification model) เพื่อหาแบบจำลอง (model) และวิธีการตัดคำ (Tokenizer) ที่มีค่าความถูกต้องแม่นยำ (accuracy) สูงสุดเพื่อใช้พยากรณ์ทิศทางของราคาหุ้นรายวัน ซึ่งในงานวิจัยนี้ได้พยากรณ์ทิศทางของราคาหุ้นทั้งหมด 3 วัน คือวันที่ 5, 6 และ 7 กุมภาพันธ์ 2563 โดยสุ่มหุ้นอย่างละ 1 ตัว ด้วยการสุ่มตัวอย่างอย่างง่าย (SRS) จากหุ้น 5 กลุ่ม คือ กลุ่ม ICT กลุ่ม ENER กลุ่ม HEALTH กลุ่ม COMM และ กลุ่ม BANK ผลวิจัยพบว่า กลุ่ม ICT สุ่มได้หุ้นของบริษัท อินทัช โฮลดิ้งส์ จำกัด (INTUCH) ตัวแบบ Gradient Boosting Classifier เป็นตัวแบบที่มีความเหมาะสมมากที่สุดและเปรียบเทียบสถานะค่าพยากรณ์กับค่าจริงได้ความถูกต้องร้อยละ 100 ในส่วนของกลุ่ม ENER สุ่มได้หุ้นของบริษัท ไทยออยล์ จำกัด (TOP) และกลุ่ม HEALTH สุ่มได้หุ้นของบริษัท โรงพยาบาลบำรุงราษฎร์ จำกัด (BH) นั้น ไม่สามารถสรุปตัวแบบที่มีความเหมาะสม

มากที่สุดได้และเปรียบเทียบสถานะค่าพยากรณ์กับค่าจริงได้ความถูกต้องร้อยละ 66.67 ในกลุ่ม COMM สุ่มได้หุ้นของบริษัท ซีพี ออลล์ จำกัด (CPALL) และกลุ่ม BANK คือธนาคารกสิกรไทย จำกัด (มหาชน) (KBANK) ตัวแบบ K-Neighbors Classifier เป็นตัวแบบที่ความเหมาะสมมากที่สุดและเปรียบเทียบสถานะค่าพยากรณ์กับค่าจริงได้ความถูกต้องร้อยละ 66.67

คำสำคัญ: วิธีการประมวลผลภาษาธรรมชาติ, ตัวแบบการจำแนก, หุ่น, Pythainlp, การวิเคราะห์ข้อความข่าว

Abstract

Factors affecting a stock price in the Stock Exchange of Thailand have several factors, including the various news. Hence, the concept of the daily stock price direction prediction from the Thai news content using the natural language processing is studied so investors are able to forecast the stock price direction before the Stock Exchange of Thailand operates. We made a study of the Thai news content with the tokenizer in Python version 3.7.1 from library pythainlp and then classification model was used for finding the most accurate values of the model and the tokenizer. This study was carried out the forecast of stock price direction in three days: 5th, 6th, and 7th February, 2020. One stock randomly chosen used simple random sampling from the following five stock groups: the ICT group, the ENER group, the HELTH group, the COMM group, and the BANK group. The results revealed that the stock of Intouch Holdings Company (INTUCH) randomly chosen by the ICT group is an efficient Gradient Boosting Classifier model when it is compared with forecasting and actual values of 100 %, the stocks of Thai Oil Public Company Limited (TOP) and Bumrungrad Hospital (BH), randomly chosen by the ENER group and the HELTH group respectively, are not able to give us efficient models when they are compared with forecasting and actual values of 66.67 %. In addition, the stocks of CP ALL public company limited of COMM group and the stock of Kasikornbank Public Company Limited of the BANK group are efficient KNeighbors Classifier models when they are compared with forecasting and actual values of 66.67 %.

Keywords: natural language processing, classification model, stock, Pythainlp, news analysis

บทนำ

ตลาดหลักทรัพย์เป็นแหล่งซื้อขายแลกเปลี่ยนหลักทรัพย์ระยะยาว ทำหน้าที่เป็นตลาดทุน เพื่อให้บริษัทมหาชน จำกัด ซึ่งถือว่าเป็นตลาดรอง (Secondary Market) สามารถระดมเงินทุนเพิ่มเติมจากสาธารณะได้ โดยหลักทรัพย์ระยะยาวจะประกอบไปด้วยตราสารหนี้ และตราสารทุนซึ่งประกอบไปด้วย หุ้นสามัญ หุ้นบุริมสิทธิ ใบสำคัญแสดงสิทธิแบบต่าง ๆ ใบสำคัญแสดงสิทธิอนุพันธ์ หุ้นกู้ และ หน่วยลงทุน เป็นต้น

ในปัจจุบันการซื้อขายหลักทรัพย์ในตลาดหลักทรัพย์แห่งประเทศไทยเป็นที่นิยมและรู้จักกันอย่างกว้างขวาง เนื่องจากเป็นแหล่งระดมเงินทุนที่มีผลตอบแทนสูงกว่าการลงทุนประเภทอื่น เช่น การซื้อที่ดิน การซื้อทองคำแท่งหรือ การฝากเงินกับธนาคารพาณิชย์ เป็นต้น จึงทำให้มีผู้สนใจเข้ามาลงทุนเพื่อสร้างความมั่งคั่งให้กับตนเองเป็นจำนวนมากแต่การลงทุนในตลาดหลักทรัพย์แห่งประเทศไทยมีอ่อนไหวสูง ไม่ว่าจะเป็นเหตุการณ์ใดๆ เกิดขึ้นก็จะส่งผลกระทบต่อสถานะตลาดได้อย่างรวดเร็ว เช่น ปัญหาเศรษฐกิจ ปัญหาการแพร่ระบาดของโรค ปัญหาการเมือง ความมั่นคง และการทหาร เป็นต้น ซึ่งปัญหาต่างๆ เหล่านี้จะส่งผลให้ราคาหลักทรัพย์มีความผันผวนตลอดเวลา ทำให้ผลตอบแทนไม่เป็นไปตามที่นักลงทุนคาดหวัง ดังนั้นการลงทุนในหลักทรัพย์ควรมีหลักเกณฑ์ในการพิจารณาเพื่อเป็นแนวทางในการตัดสินใจของนักลงทุน

ปัจจัยที่กระทบต่อราคาของหุ้นในตลาดหลักทรัพย์แห่งประเทศไทยที่นักลงทุนนิยมพิจารณานั้นมีอยู่หลายปัจจัย เช่น ผลการดำเนินงานของบริษัท จำนวนการซื้อขายหุ้นในแต่ละวัน รวมทั้งข่าวสารต่างๆ ที่มีผลกระทบต่อจิตวิทยาของตลาดหุ้นและราคาของหุ้น อาทิ ข่าวการลงทุนของบริษัท ข่าวการระบาดของโรค ข่าวการชุมนุมทางการเมือง เป็นต้น ด้วยความสำคัญของข่าวสารต่างๆ นี้ ผู้วิจัยจึงเกิดแนวคิดที่จะพยากรณ์ทิศทางของราคาหุ้นรายวันจากข้อความข่าวภาษาไทย โดยใช้วิธีการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) เพื่อให้ นักลงทุนสามารถคาดคะเนทิศทางของราคาหุ้นก่อนที่ตลาดหลักทรัพย์แห่งประเทศไทยจะเปิดได้ โดยจะทำการศึกษาข้อความข่าวของหุ้นจากตลาดหลักทรัพย์แห่งประเทศไทย (SET) จำนวน 5 ตัว จากหุ้น 5 กลุ่ม ด้วยโปรแกรมภาษาไพธอน ver.3.7.1 บนโปรแกรม Visual Studio Code และใช้โปรแกรม Power BI ช่วยในการแสดงผล

งานวิจัยที่เกี่ยวข้อง

Li et al. (2014) ศึกษาการเพิ่มประสิทธิภาพการพยากรณ์ราคาหุ้นระหว่างวันของ Hong Kong Stock Exchange (HKEx) โดยรวมข้อมูลจาก 2 แหล่งคือ 1. ข้อความข่าว และ 2. ราคาหุ้นในอดีต ซึ่งข้อมูลทั้งสองแหล่งนี้ มีความอิสระต่อกัน มีผลกระทบต่อราคาหุ้นระหว่างวันมาก และมีคุณสมบัติดังนี้ 1. เวลา กล่าวคือข้อมูลข้อความข่าวและราคาหุ้นในอดีตส่วนมีเรื่องของเวลาเข้ามาเกี่ยวข้อง 2. ความถี่ ในการพยากรณ์ราคาหุ้นระหว่างวันเป็นการพยากรณ์ระยะสั้น หากต้องการความถี่ในการพยากรณ์สูง จะต้องควบคุมข้อมูลให้มีลักษณะดังนี้ 1. มีปริมาณซื้อขายมาก และ 2. มีความแตกต่างระหว่างช่วงเวลา โดยการเตรียมข้อความข่าวที่เป็นภาษาจีนนั้น จะแบ่งกลุ่มโดยใช้ซอฟต์แวร์การแบ่งข้อมูลภาษาจีนซึ่งมีคลังคำศัพท์ทางการเงิน จากนั้นทำการลบคำที่ไม่สำคัญออกไป และให้ค่าน้ำหนักด้วย TFIDF (term frequency - inverse document frequency) แล้วเรียงลำดับเวลาของราคาหุ้น รวมทั้งจัดการข้อความความข่าวและราคาหุ้นตอนตลาด HKEx เปิด คือเวลา 10.00-12.30 น. และ 14.30-16.00น. นอกจากนี้คณะผู้วิจัยได้ทำการระบุหัวข้อข่าวทุกๆ 5, 10, 15, 20, 25 และ 30 นาที รวมทั้งได้เพิ่มตัวชี้วัดการวิเคราะห์หุ้นทางเทคนิคคือ Relative Strength Index (RSI), Raw Stochastic Value (RSV), Williams Index, Bias และ Psychological Line (PYS) จากนั้น ทำการลดความซ้ำซ้อนของข้อมูล (normalization) ก่อนสร้างตัวแบบ ส่วนตัวแบบ (model) ที่ใช้ แบ่งเป็น 3 กลุ่มคือ 1. ตัวแบบที่ฝึกสอนด้วยข้อมูลเพียง

แหล่งเดียว (ข้อความข่าวหรือราคาหุ้นในอดีต) 2. ตัวแบบที่ฝึกสอนด้วยข้อมูลจาก 2 แหล่ง โดยใช้การรวมตัวแปร (feature) อย่างง่าย 3. ตัวแบบที่ฝึกสอนด้วยข้อมูลจาก 2 แหล่ง โดยใช้วิธีการรวมกันแบบ Multi-kernel จากการวิจัยพบว่า Multi-Kernel Support Vector Regression (MKSVR) ซึ่งเป็นตัวแบบในกลุ่มที่ 3 เป็นตัวแบบที่มีค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (root mean square error) และค่าเฉลี่ยส่วนเบี่ยงเบนสัมบูรณ์ (mean absolute error) ต่ำสุด

Sadia et al. (2019) ศึกษาการพยากรณ์ราคาหุ้นโดยใช้ข้อมูลราคาหุ้นจากเว็บไซต์ Kaggle และทำการเตรียมข้อมูลโดยการตรวจสอบค่าสูญหาย แปลงข้อมูลเชิงคุณภาพ กำหนดราคาปิดของหุ้นโดยกำหนดประเภทให้ มีค่าเป็น -1 เมื่อราคาปิดของวันนั้นน้อยกว่าราคาปิดของเมื่อวาน และมีค่าเป็น 1 เมื่อราคาปิดของวันนั้นมีค่ามากกว่าราคาปิดของเมื่อวาน รวมถึงการแบ่งข้อมูลฝึกสอน (train) และข้อมูลทดสอบ (test) โดยพิจารณา 2 ตัวแบบคือ 1. Random Forest และ 2. Support Vector Machine และใช้ตาราง Confusion Matrix ในการแสดงค่าความแม่นยำ (accuracy) ผลการวิจัยพบว่าตัวแบบ Random Forest มีค่าความแม่นยำเป็นร้อยละ 80.8 และตัวแบบ Support Vector Machine มีค่าความแม่นยำร้อยละ 78.7

นอกจากนี้ยังมีงานวิจัยอื่นๆ ที่เกี่ยวข้องกับการพยากรณ์ทิศทางของราคาหุ้นโดยใช้ข้อมูลราคาหุ้นในอดีต เช่น งานวิจัยของ Huang et al. (2005) , Kim (2003) และ Tay & Cao (2001) เป็นต้น

วิธีดำเนินการวิจัย

การเก็บรวบรวมข้อมูล

งานวิจัยชิ้นนี้ผู้วิจัยใช้โปรแกรมภาษาไพธอนเป็นเครื่องมือช่วยในการวิเคราะห์ข้อมูลข่าว และทำการสุ่มเลือกหุ้นโดยใช้การสุ่มตัวอย่างอย่างง่าย (simple random sampling : SRS) อย่างละ 1 ตัว จากหุ้น 5 กลุ่ม คือ 1. เทคโนโลยีสารสนเทศและการสื่อสาร (ICT) สุ่มได้หุ้นของบริษัท อินทัช โฮลดิ้งส์ จำกัด (INTUCH) 2. พลังงานและสาธารณูปโภค (ENERG) สุ่มได้หุ้นของบริษัท ไทยออยล์ จำกัด (มหาชน) (TOP) 3. การแพทย์ (HEALTH) สุ่มได้หุ้นของบริษัท โรงพยาบาลบำรุงราษฎร์ จำกัด (มหาชน) (BH) 4. พาณิชยกรรม (COMM) สุ่มได้หุ้นของบริษัท ซีพี ออลล์ จำกัด (มหาชน) (CPALL) 5. ธนาคาร (BANK) สุ่มได้หุ้นของธนาคารกสิกรไทย จำกัด (มหาชน) (KBANK) รวมจำนวน 5 ตัว จาก SET100 ซึ่งเป็นดัชนีราคาหุ้นของตลาดหลักทรัพย์แห่งประเทศไทย (the stock exchange of Thailand) ที่ได้รับการคัดเลือก 100 อันดับแรก เก็บรวบรวมข้อมูลตั้งแต่วันที่ 18 พฤษภาคม 2562 ถึง วันที่ 7 กุมภาพันธ์ 2563 โดยรวบรวมข้อความข่าวจากแหล่งข่าวต่างๆ จำนวน 6 แหล่งข่าวคือ 1. สำนักข่าวกรุงเทพธุรกิจ 2. สำนักข่าวหุ้นอินไซด์ 3. สำนักข่าวข่าวหุ้น 4. สำนักข่าวโพสต์ทูเดย์ 5. สำนักข่าว Innnews และ 6. สำนักข่าวหุ้น สมาร์ท จากลิงค์ข้อมูลข่าวของเว็บไซต์ <https://stock.gapfocus.com/> ซึ่งสามารถค้นหาข่าวของหุ้นแต่ละตัวที่สุ่มเลือกได้ เนื่องจากเก็บรวบรวมข้อมูลเป็นการรวบรวมลิงค์ข้อความของข่าวที่มีจำนวนมากจึงใช้ Developer Tools ของ Google Chrome มาช่วยคัดแยกลิงค์ข้อความของข่าวโดยใช้ Library BeautifulSoup ในโปรแกรมภาษาไพธอนช่วยในการเก็บรวบรวม ดังรูปที่ 1 และ 2 จากนั้นทำความสะอาดข้อมูลโดยทำการลบสัญลักษณ์ต่างๆ เช่น \[.*?]\(), ลบตัวเลข และตัวอักษรภาษาอังกฤษออก รวมถึงการลบเว้นวรรคออกทั้งหมด ดังรูปที่ 3

และผู้วิจัยได้เก็บรวบรวมข้อมูลราคาหุ้นรายวันจากเว็บไซต์ <https://finance.yahoo.com/> โดยใช้ Module HistoricalPrices ใน Library yahoofinance ของโปรแกรมภาษาไพธอนช่วยในการเก็บรวบรวม เพื่อกำหนดสถานะของราคาปิดตลาดในแต่ละวันจากสมการ

$$Diff = Closeprice_i - Closeprice_{i-1} \quad (1)$$

กำหนดให้ $Diff$ คือ ค่าความแตกต่างของราคาหุ้น

$Closeprice_i$ คือ ราคาปิดของหุ้นวันที่ i

$Closeprice_{i-1}$ คือ ราคาปิดของหุ้นวันที่ $i-1$

จากนั้นนำค่า $Diff$ มาหาสถานะ โดยกำหนดให้ $Diff > 0$ หมายถึง ราคาหุ้นมีแนวโน้มเป็นบวก (positive)

$Diff < 0$ หมายถึง ราคาหุ้นมีแนวโน้มเป็นลบ (negative) และ $Diff = 0$ หมายถึง ราคาหุ้นมีแนวโน้มเป็นกลาง (neutral)

```

<a href="https://www.hooninside.com/news-feed/154836/view/" rel="nofollow" target="_blank">...
</a> == $0
<a href="https://www.kaohoon.com/content/348219" rel="nofollow" target="_blank">...</a>
<a href="https://www.kaohoon.com/content/348171" rel="nofollow" target="_blank">...</a>
<a href="https://www.ryt9.com/s/ig05/3105921" rel="nofollow" target="_blank">...</a>
<a href="https://www.ryt9.com/s/ig05/3105922" rel="nofollow" target="_blank">...</a>
<a href="https://www.kaohoon.com/content/348126" rel="nofollow" target="_blank">...</a>
<a href="https://www.bangkokbiznews.com/news/detail/870955" rel="nofollow" target="_blank">...
</a>
<a href="https://www.bangkokbiznews.com/news/detail/870980" rel="nofollow" target="_blank">...
</a>
<a href="https://twitter.com/K_Securities/status/1239553732196491267/" rel="nofollow"
target="_blank">...</a>

```

รูปที่ 1 ลิขัข้อมูลข่าวของ <https://stock.gapfocus.com> จาก Developer Tools ของ Google Chrome

stock_name	date	Title	link	source
TOP	03 Dec 07:01	หุ้นบีทีเอส-อีลิกซ์กลดคอพีน ส่นจระจากราคาโลกขยว	https://www.bangkokbiznews.com/news/detail/856814	www.bangkokbiznews.com
TOP	03 Dec 07:01	รัฐเร่งเลขลงทุน ต่อเติม 'ความเข้มแข็ง'	https://www.bangkokbiznews.com/news/detail/856783	www.bangkokbiznews.com
TOP	03 Dec 06:31	เปิดบันทึกสื่อ 4 ชั่วโมง เจาะเดือนเมษา 3 สัร	https://www.bangkokbiznews.com/news/detail/856823	www.bangkokbiznews.com
TOP	03 Dec 06:01	ไทยเอสดีเคการไฟฟ้านานาชาติ "ราคาหุ้น" 2 - 6 ส.ค. 62 และสรุปสถานการณ์ 25 - 29 พ.ย. 62	https://www.bangkokbiznews.com/news/detail/856752	www.bangkokbiznews.com
TOP	02 Dec 17:32	ราคาน้ำมันดิบเบงโกะไม่ไหว จากความไม่แน่นอนของการเจรจาการค้าสหรัฐฯ - จีน	https://www.thunhoon.com/215955/39/39/	www.thunhoon.com
TOP	02 Dec 16:02	ธนาคารพาณิชย์ "หวังสื่อคำประกันแบบบล็อกเชน" ในเครือกลุ่ม SCG	https://mgronline.com/stockmarket/detail/9620000115371	mgronline.com
TOP	02 Dec 14:32	ไทยเอสดี ราคาหุ้นดิบเบงโกะไม่ไหว จากความไม่แน่นอนของการเจรจาการค้าสหรัฐฯ - จีน วัน/เวลา: 2 ส.ค.	https://www.hooninside.com/news-feed/137527/view/	www.hooninside.com
TOP	02 Dec 14:31	Daily View - P.M. - บล.กสิกรไทย	https://www.kaohoon.com/content/329763	www.kaohoon.com
TOP	02 Dec 12:02	SET ดูปฏิทิน 20 จุด! โบทและฉายจิ้งหรีด 6 สัปดาห์เป็นความเลื่องลือ	https://www.kaohoon.com/content/329699	www.kaohoon.com
TOP	02 Dec 12:01	ความคาดหวังโลกคลั่งกำลังการผลิต + PMI จีนเป็นบวก หนุนบรรยากาศการค้าในช่วงสั้น	https://www.bangkokbiznews.com/news/detail/856709	www.bangkokbiznews.com
TOP	02 Dec 12:01	ทรูเอ็นบีเอ็น "หวังสื่อคำประกันแบบบล็อกเชน" ในเครือกลุ่ม... 2 ส.ค. 62 11:50 น. -ThaiPR.net ธนาคาร	https://www.ryt9.com/s/prg/3073321	www.ryt9.com
TOP	02 Dec 11:32	สหรัฐลดน้ำมันดิบลงเป็นประวัติการณ์ บีซีที ไทยเอสดี จำกัด (มหาชน) รายงานสถานการณ์น้ำมันดิบตลาดปร	https://www.innews.co.th/economy/news_545376/	www.innews.co.th
TOP	02 Dec 11:02	TOP เนย ราคาหุ้นดิบเบงโกะไม่ไหว จากความกังวลสงครามการค้าและการผลิตน้ำมันดิบสหรัฐฯ พุ่งสูงเป็นประวัติการณ์ วัน/	https://www.hooninside.com/news-feed/137527/view/	www.hooninside.com
TOP	02 Dec 11:01	Daily Strategy for Investors on December 2, 2019	https://www.kaohoon.com/content/329685	www.kaohoon.com
TOP	02 Dec 09:32	SET ปิดดัชนีลงทวน ขวัญใจดัชนีหุ้นน้ำมัน	https://www.thunhoon.com/215873/00/49/	www.thunhoon.com
TOP	02 Dec 08:46	รายงานเทคนอลจิสติกส์ ประจำวันที่ 29 พฤศจิกายน 2562 2 ส.ค. 62 07:49 น. -สำนักข่าวอินโฟเควสท์ (IQ) หลั	https://www.ryt9.com/s/ig05/3073154	www.ryt9.com
TOP	02 Dec 08:46	รายงานเทคนอลจิสติกส์ ประจำวันที่ 29 พ.ย. 2562 2 ส.ค. 62 07:53 น. -สำนักข่าวอินโฟเควสท์ (IQ) หลั	https://www.ryt9.com/s/ig05/3073156	www.ryt9.com
TOP	02 Dec 08:01	SET ปิดชีพไฟแดงในวัน 1580-1620 จุด เป็นจุดซื้อเชิงพื้นฐาน ชู CPF, BCH, STPI หนุน	https://www.kaohoon.com/content/329624	www.kaohoon.com
TOP	29 Nov 15:31	"ไทยเอสดี" คว้ารางวัลดีเด่นสาขาความเป็นเลิศด้านการจัดการทรัพยากรบุคคล จาก THAILAND CORPORATE EX	https://mgronline.com/greeninnovation/detail/9620000114617mgronline.com	mgronline.com
TOP	29 Nov 13:32	ไทยเอสดีคว้ารางวัลดีเด่นสาขาความเป็นเลิศด้านการจัดการทรัพยากรบุคคล จาก THAILAND CORPORATE EX	https://www.hooninside.com/news-feed/137354/view/	www.hooninside.com
TOP	29 Nov 13:03	ไทยเอสดี คว้ารางวัลดีเด่นสาขาความเป็นเลิศด้านการจัดการทรัพยากรบุคคล	http://www.efnancethai.com/LastestNews/LatestNewsMain.a	www.efnancethai.com
TOP	29 Nov 12:32	TOP เนย ราคาหุ้นดิบเบงโกะไม่ไหว จากความเครียดระหว่างสหรัฐฯ และจีน วัน/เวลา: 29 พ.ย. 2562 / 10:46:01	https://www.hooninside.com/news-feed/137336/view/	www.hooninside.com

รูปที่ 2 การเก็บรวบรวมลิงค์ของแหล่งข่าว

Date	Text
5/2/2020	'ภาวะตลาดหุ้นไทยเดือนมกราคมที่ผ่านมาไม่ค่อยสดใสเนื่องจากได้รับปัจจัยกดดันทั้งภาวะภัยแล้งและประเด็นการเมืองใน
4/2/2020	'วิเคราะห์สถานการณ์ราคาน้ำมันกพ', 'ราคาน้ำมันดิบปรับตัวลดลงต่อหลังความต้องการใช้น้ำมันในจีนซบเซาจากการระบาดข
3/2/2020	'หุ้นไทยภาคเข้าปีที่จุดลบจุดหรือ', 'กรุงเทพธุรกิจออนไลน์การซื้อขายหลักทรัพย์ภาคเข้าวันจันทร์ทเวลาณดัชนีปิดอยู่ที่
2/2/2020	'ปัดทบทวนแผนลงทุนปียัดเทรด'พลังงานสะอาด', 'ทิศทางธุรกิจพลังงานทั่วโลกที่มุ่งสู่การผลิตและการใช้เชื้อเพลิงที่สะ
1/2/2020	'ไออาร์เร่งขยายปั๊มขายดีเซลบี', 'ไออาร์'ค้ายอดขยายน้ำมันผ่านปั๊มบีโดต่อเนื่องเร่งออกแคมเปญมิดีกระด้นยอขยาย
31/1/2020	'วิเคราะห์สถานการณ์ราคาน้ำมันกพ', 'ราคาน้ำมันดิบปรับตัวลดลงต่อหลังความต้องการใช้น้ำมันในจีนซบเซาจากการระบาดข
30/1/2020	'ปัดทบทวนคดีกำไรหมื่นล้าน', 'ปัดทบทวนคดีกำไรหมื่นล้านบาทเพิ่มขึ้นรับรายได้เพิ่มขึ้นกว่า', 'เมื่อวันทีมคมบริษัทป
29/1/2020	'หุ้นไทยภาคเข้าปีที่จุดลบจุดหรือ', 'กรุงเทพธุรกิจออนไลน์การซื้อขายหลักทรัพย์ภาคเข้าวันพุธมเวลาณดัชนีปิดข
28/1/2020	'พลังงานน้ำมันดีเซลลดสต็อกขึ้นราคาคงเดิม', 'ราคาน้ำมันพุ่งขึ้นดีเซลปรับลดสต็อกขึ้นราคาคงเดิม', 'บมจปตทน้ำมันแ
27/1/2020	'หุ้นไทยภาคเข้าปีที่จุดลบจุดหรือ', 'กรุงเทพธุรกิจออนไลน์การซื้อขายหลักทรัพย์ภาคเข้าวันจันทร์ทเวลาณดัชนีปิด
26/1/2020	'พจนททำเรือศรีสุวรรณ'ถกแหลมม้งเรื่องไกรยีนปชสอมนายกเอื้อกลุ่มทุน', 'พนักงานทททก'ศรีสุวรรณ'หนุนสอมนประมุ
25/1/2020	'สนธิรัตน์เร่งหนุนใช้หวังดันราคาสินค้าเกษตร', 'สนธิรัตน์'ประกาศนโยบายพลังงานดันหน้าผกด้านยกเลิกกสยอสมล์มี
24/1/2020	'หุ้นไทยภาคเข้าปีที่จุดลบจุดหรือ', 'กรุงเทพธุรกิจออนไลน์การซื้อขายหลักทรัพย์ภาคเข้าวันศุกร์มเวลาณดัชนีปิดข
21/1/2020	'ปัดทบทวนจากปรับขึ้นราคาน้ำมันทุกชนิดลดเว้นขยับสตร', 'ปัดทบทวนจากปรับขึ้นราคาน้ำมันทุกชนิดลดเว้นขยับสตรมีผลพวงนี้'
20/1/2020	'ปัดท-มิตชยลงนามสัญญาร่วมทุนเดินหน้าธุรกิจหุ่นยนต์และปัญญาประดิษฐ์', 'ปัดท-มิตชย'ลงนามสัญญาร่วมทุนเดินหน้า
18/1/2020	'หนี้ครัวเรือนระเบิดเวลาเศรษฐกิจปี', 'เทคโนโลยี'กำลังกลายเป็นจุดเริ่มต้นแห่งปัญหาทางการเงินส่วนบุคคลหากไม่มีวินย
17/1/2020	'ท่าเลทองหรือโอกาสทองของใคร', 'แม่โครงการอีอีซีเกิดขึ้นเพียงไม่กี่ปีแต่กลายเป็นท่าเลทองที่เกิดการลงทุนอุตสาหกรรม
16/1/2020	'บอทดตทไฟเขียวลงทุนบีบีเอสลานาพา', 'บอทดตทไฟเขียวลงทุนบีบีเอสลานาพาทยขยายธุรกิจค้าขายและทอ
15/1/2020	'หุ้นไทยภาคเข้าปีที่จุดลบจุดหรือ', 'กรุงเทพธุรกิจออนไลน์การซื้อขายหลักทรัพย์ภาคเข้าวันพุธมเวลาณดัชนีปิดข
14/1/2020	'ปัจจัยภายนอกหนุน', 'คาดปรับตัวขึ้นปีจุดก่อนจะสั่นแอ่นตัวต่อรับเชิงบวกสหรัฐขึ้นเตรียมลงนามข้อตกลงการค้าเฟสแ

รูปที่ 3 ข้อความข่าวในแต่ละวัน

การเตรียมข้อมูล

ขั้นตอนนี้เป็นขั้นตอนการแปลงข้อมูลที่เก็บรวบรวม (raw data) ใต้ให้กลายเป็นข้อมูลที่สามารถนำมาวิเคราะห์ โดยการแปลงข้อมูลนี้จะเป็นกระบวนการที่ใช้เวลามากที่สุด ซึ่งในงานวิจัยนี้มีขั้นตอนดังนี้

1. รวมข้อมูลข้อความข่าวและข้อมูลหุ้นรายวันที่ได้จากขั้นตอนแรก โดยกำหนดข้อมูลสถานะของราคาปิดเป็นตัวแปร y ดังรูปที่ 4

Date	Close	Diff	Close_status	Text
27/11/2019	69.5	-1	negative	'วิเคราะห์สถานการณ์ราคาน้ำมันกพ', 'ราคาน้ำมันดิบปรับตัวลดลงต่อหลังความต้องการใช้น้ำมันในจีนซบเซาจากการระบาดข
28/11/2019	69	-0.5	negative	'อาจหันเหจากข่าวกฎหมายใหม่ของสหรัฐแต่ยังมองโลกโง่', 'สงครามการค้าไม่แยกไปกว่าที่เป็นอยู่', 'เข้าปีน
29/11/2019	69	0	neutral	'เก็งกำไรในหุ้นที่มีปัจจัยบวกในช่วงเดือนธ', 'บรรยากาศลงทุนช่วงสิ้นปีชะลอตัวตามการรื้อการค้า', 'หลังประธานา
2/12/2019	68.25	-0.75	negative	'ความคาดหวังโอบล้อมกำลังการผลิตจีนเป็นบวกหนุนบรรยากาศเก็งกำไรช่วงสิ้น', 'เงินคดต้นสหรัฐปรับลดภาษีลง
3/12/2019	68.5	0.25	positive	'หุ้นบีโอดีเล็ก'กอดคอปันทุนเจรจาการค้าโลกขสว', 'หุ้นกลุ่มบีโอดี'และ'อีเล็กทรอนิกส์'พุ่งกว่าช่วงหนึ่งเด
4/12/2019	68	-0.5	negative	'วิเคราะห์สถานการณ์ราคาน้ำมันกพ', 'ราคาน้ำมันดิบปรับตัวลดลงต่อหลังความต้องการใช้น้ำมันในจีนซบเซาจากการระบาดข
6/12/2019	66.75	-1.25	negative	'สำนักข่าวหุ้นอินโอดี'ความเคลื่อนไหวคดีทุจริตในโอดี'ข่าวรองกรรมการผู้อำนวยการด้านการตลาดบริษัทหลักทรัพย์
9/12/2019	66.75	0	neutral	'กองทุนหุ้นไทย-ลงทุนที่ทวนปีสุดท้าย', 'ปีสุดท้ายของกองทุนรวมระยะยาวยังเป็นตัวชี้วัดสำคัญในการลดห
11/12/2019	64.5	-2.25	negative	'ปัดทบทวนจากปรับขึ้นราคาน้ำมันทุกชนิดลดเว้นขยับสตร', 'ปัดทบทวนจากปรับขึ้นราคาน้ำมันทุกชนิดลดเว้นขยับสตรมีผลพวงนี้'
12/12/2019	65	0.5	positive	'ลิแกเดไม่ลับปฏิรูปปลาน้ำมันโครได้ผลประโยชน์', 'ห้องข่าวเศรษฐกิจธุรกิจหุ้นแล้วสักทุกมิติเรื่องเศรษฐกิจการลง
13/12/2019	67.5	2.5	positive	'วิเคราะห์สถานการณ์ราคาน้ำมันกพ', 'ราคาน้ำมันดิบปรับตัวเพิ่มขึ้นเล็กน้อยจากความหวังปริมาณน้ำมันดิบลดจากข้อ
16/12/2019	68	0.5	positive	'รีอีไอเกียรตินาดีตัวเต็ง', 'เวลาช่วงผ่านไปเร็วเหลือเกินเหลือแค่ปีเดียวอีกไม่กี่วันก็จะสิ้นปีแล้วไม่ทราบว่พร
17/12/2019	67.5	-0.5	negative	'กับขงมีโอเวิ้ลด์แบงก์กระด้นลงทุน', 'กับขสสความร่วมมือกับธนาคารโลกยกระดับการลงทุนเพื่อความยั่งยืน
18/12/2019	69	1.5	positive	'"อรรถพลวิรัตน์"ชิงชัยโอดี', 'คณะกรรมการสรรหาของบอทดตทเลือก'ชัย'อีอีโอดี'ปัดททใหม่หลังชวโรยตัว
19/12/2019	68.25	-0.75	negative	'โดยมีจุดขายคดขาดทุน', 'คาดการณ์ตลาดหุ้นไทยวันนี', 'เรามีมุมมองและคาดดัชนีแนวหน้าจุดแนวรับจุดเป็นหลา
20/12/2019	68	-0.25	negative	'กลุ่มปัดทงตั้งงบแสนล้านลุยธุรกิจไทย', 'ห้องข่าวเศรษฐกิจธุรกิจหุ้นแล้วสักทุกมิติเรื่องเศรษฐกิจการลงท
23/12/2019	70	2	positive	'หุ้นไทยแกว่งตัวกรอบแคบจุด', 'บรคาคาหุ้นไทยแกว่งตัวจุดแคบได้แรงหนุนสหรัฐขึ้นบรรลุลดลงลดการการค้า
24/12/2019	69.5	-0.5	negative	'คาดแกว่งตัวจุดเนื่องจากภาวะตลาดรอปัจจัยใหม่กระตุ้นการลงทุน', 'ตลาดหุ้นวานนี้', 'ปรับขึ้นแรงจุดปิดที่ระดับจุด
25/12/2019	69.5	0	neutral	'หุ้นไทยแกว่งตัวแคบกรอบแคบจุด', 'บลกรูมมองหุ้นไทยวันนี'เคลื่อนไหวตามบททดสอบจุดก่อนจะสั่นแอ่น

รูปที่ 4 ข้อมูลสถานะของราคาปิดรวมกับข้อมูลข้อความข่าว

2. ลบประโยคที่ไม่มีผลต่อการพยากรณ์ทิศทางของหุ้นออก เช่น 'มูลค่าการซื้อขายล้านบาทที่บาทเพิ่มขึ้นบาท', 'มูลค่าการซื้อขายล้านบาทที่บาทลดลงบาท', 'ที่มา' เป็นต้น
3. Word Tokenize เป็นการตัดข้อความข่าวแยกออกมาเป็นคำโดยใช้ Library pythainlp ซึ่งในงานวิจัยนี้ใช้ 9 วิธีคือ

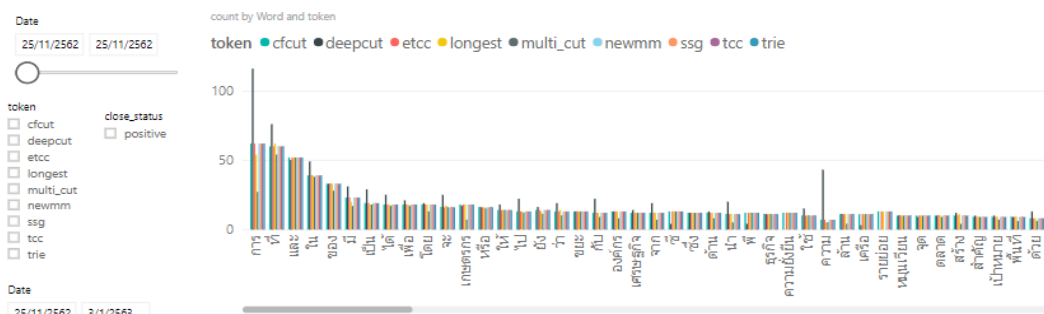
- 1) Cfcut เป็นการแบ่งประโยคภาษาไทยโดยใช้เทคนิคคอนดิชันนอลแรนดอมฟิลด์ (conditional random field) ซึ่งเรียนรู้จากชุดข้อมูลของ TED
- 2) Deepcut เป็นการตัดคำโดยใช้เทคนิค 1D Convolution Neural Network หรือ CNN ซึ่งเป็นโครงข่ายประสาทเทียมชนิดหนึ่งในกลุ่ม Bio-inspired โดยที่ CNN จะจำลองการมองเห็นของมนุษย์ที่มองพื้นที่เป็นที่ย่อยๆ และนำกลุ่มของพื้นที่ย่อยๆ มาผสานกัน
- 3) Longest เป็นการตัดคำโดยวิธีการเทียบคำที่ยาวที่สุด (longest matching) เสนอโดย Sonlertlamvanit (1993) ซึ่งเป็นวิธีที่ใช้พจนานุกรมช่วยในการตัดคำภาษาไทย โดยวิธีนี้จะทำการตรวจสอบหรือแสกนข้อความจากซ้ายไปขวาแล้วนำไปเปรียบเทียบกับพจนานุกรม ถ้าอักษรที่ประกอบกันนั้นไม่สามารถเทียบเป็นคำในพจนานุกรมได้ก็จะทำการลดความยาวลง จนกว่าจะเทียบคำในพจนานุกรมได้แล้วกลับไปยังจุดย้อนกลับ จากนั้นก็เริ่มทำงานที่จุดย้อนกลับอีกครั้งเพื่อทำการตรวจสอบอักษรที่ประกอบกัน (Chaicharoen, 2001) เช่น “ห้วงโยสถานการณ์การระบาดของไวรัสโคโรนา” ข้อความนี้ไม่สามารถเทียบคำพจนานุกรมได้จึงตัดเหลือ “ห้วงโยสถานการณ์การระบาดของไวรัสโคโรนา” ตัดไปเรื่อยๆ จนได้สายอักขระ “ห้วงโย” ซึ่งสามารถเทียบคำในพจนานุกรมได้
- 4) Etc เป็นวิธีการตัดคำโดยใช้เทคนิคการรวมกันของ Forward และ Backward Longest Matching Techniques
- 5) multi_cut เป็นการตัดคำด้วยวิธีการตัดคำเพื่อให้ได้จำนวนคำและคำที่ไม่มีในพจนานุกรมน้อยที่สุด (maximal matching) เป็นระเบียบวิธีฮิวริสติกส์ (heuristic) วิธีหนึ่งที่ใช้พจนานุกรมเทียบคำภาษาไทย พัฒนาโดย Sonlertlamvanit (1993) ซึ่งได้ดัดแปลงมาจากวิธีการเทียบคำที่ยาวที่สุด เริ่มจากการหาทางเลือกของรูปแบบการตัดคำทั้งหมดที่เป็นไปได้เสียก่อน โดยทำย้อนกลับทีละคำ หลังจากได้คำตอบจากวิธีการเทียบคำที่ยาวที่สุดแล้ว จึงเลือกทางเลือกที่มีจำนวนคำน้อยที่สุด (Chaicharoen, 2001)
- 6) tcc เป็นการตัดคำโดยใช้เทคนิคกลุ่มอักขระภาษาไทย หรือ Thai Character Clusters (TCC) ในภาษาไทยมีลักษณะขององค์ประกอบที่หลากหลายมากกว่าเมื่อเปรียบเทียบกับภาษาอังกฤษ เช่น สระ วรรณยุกต์ เสียงวรรณยุกต์ และอักขระพิเศษ เทคนิค TCC นั้นองค์ประกอบต่างๆ ในวลีจะไม่คลุมเครือและสามารถตั้งกฎขึ้นมาได้ เช่น สระที่อยู่ข้างหน้าและอักขระถัดไปถูกจัดกลุ่มให้อยู่ในกลุ่มเดียวกัน เครื่องหมายวรรณยุกต์จะอยู่บนพยัญชนะเสมอ และไม่สามารถแบ่งออกจากพยัญชนะได้ ส่วนสระที่อยู่ตำแหน่งและตำแหน่งก่อนหน้าจะถูกแบ่งให้อยู่ในกลุ่มเดียวกัน เป็นต้น
- 7) newmm เป็นการตัดคำโดยอ้างอิงจากพจนานุกรมโดยใช้เทคนิค Maximal Matching Algorithm และ Thai Character Cluster (TCC)
- 8) ssg เป็นการตัดคำโดยใช้พยางค์ด้วยเทคนิค Conditional Random Field
- 9) trie เป็นการตัดคำโดยใช้พยางค์ด้วยเทคนิคทรี (trie) ในวิทยาศาสตร์คอมพิวเตอร์ Trie เรียกอีกชื่อว่า Digital Tree หรือ Prefix Tree คือโครงสร้างของข้อมูลต้นไม้ที่เป็นลำดับเพื่อเก็บข้อมูลเป็นแบบไดนามิกหรืออาร์เรย์ที่เชื่อมโยงกัน ใช้กับข้อมูลที่เป็นตัวอักษร การทำงานเป็นแผนผังการค้นหาแบบไบนารี โดยข้อมูลอยู่บนเส้นทางการเดินระหว่าง

โหนดราก (root node) ไปยังโหนดใบ (leaf node) ซึ่งใช้ในการเก็บค่าในพจนานุกรม โดยแต่ละกิ่ง (branch) ระหว่างโหนดจะมีตัวอักษรกำกับอยู่

[illegible]

รูปที่ 5 การตัดคำและสถานะของราคาปิด

4. Word frequency และ Word Index คือ การหาความถี่ของของคำแต่ละคำในแต่ละวัน รวมทั้งการสร้าง Index ให้แต่ละคำด้วย เช่น กำหนดให้คำว่า “การ” มี Index เป็น 1 ,คำว่า “ที่” มี Index เป็น 2 ,และคำว่า “มี” มี Index เป็น 3 เป็นต้น ดังรูปที่ 6 และ 7



รูปที่ 6 ความถี่ของคำในแต่ละวันที่ตัดคำด้วยวิธีการตัดคำต่างๆ ทั้ง 9 วิธี

close_status	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
positive	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	1	1	0
positive	0	0	0	0	0	0	0	0	0	0	0	4	8	0	0	0	0	0	0	0	0	0	0
positive	0	0	0	0	0	0	0	0	0	0	2	0	1	0	5	0	0	0	0	0	0	0	0
negative	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	1	0	0	0
negative	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0
neutral	0	0	0	0	0	0	0	0	0	0	0	2	0	3	0	2	0	0	0	0	0	0	0
negative	0	0	0	0	0	0	0	0	0	0	0	0	0	1	8	0	0	0	0	0	0	0	0
positive	0	0	2	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2	0	0	0	0	0
negative	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
negative	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0
neutral	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
negative	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	1	0	0	0
positive	2	0	0	0	0	0	1	0	2	2	0	0	0	0	13	0	0	0	0	0	0	0	0
positive	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0
positive	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
negative	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	1	0
positive	0	0	0	0	0	0	0	0	0	0	0	3	0	5	0	0	0	0	0	0	0	0	0
negative	0	0	0	1	0	0	0	0	0	0	1	0	0	0	6	0	0	0	0	0	0	0	0
negative	0	0	0	0	0	0	0	2	0	0	3	0	1	0	11	0	0	0	0	0	0	0	0

รูปที่ 7 สถานะราคาปิดของหุ้นและความถี่ของแต่ละ Word Index

5. TFIDF (term frequency-inverse document frequency) เป็นกระบวนการที่แปลงความถี่ในเอกสาร โดยให้น้ำหนักคำในเอกสารใดๆ เทียบกับคำในเอกสารทั้งหมด ซึ่งสามารถหาได้จากสมการ

$$TFIDF = \text{Term Frequency} \times \text{Inverse Document Frequency} \quad (2)$$

โดยค่า Term Frequency หรือ tf นั้นมีสมการดังนี้

$$tf(w, D) = f_{wD} \quad (3)$$

กำหนดให้ w คือคำที่ต้องการคำนวณ และ D คือ เอกสารที่สนใจ

ส่วนค่า Inverse Document Frequency หรือ idf มีสมการดังนี้

$$idf(t) = 1 + \log \frac{C}{1 + df(t)} \quad (4)$$

กำหนดให้ C คือ จำนวนเอกสารทั้งหมดที่มีอยู่ และ $df(t)$ คือ จำนวนของเอกสารที่มีค่า t ปรากฏอยู่

close_status	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
positive	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0	0	0	0	0	0.02	0.02	0
positive	0	0	0	0	0	0	0	0	0	0	0	0.06	0.06	0	0	0	0	0	0	0	0	0	0
positive	0	0	0	0	0	0	0	0	0	0	0.03	0.01	0	0.03	0	0	0	0	0	0	0	0	0
negative	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0	0	0	0	0.02	0	0	0	0
negative	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0	0	0	0	0	0	0	0	0
neutral	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0	0.01	0	0.03	0	0	0	0	0	0
negative	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0.05	0	0	0	0	0	0	0	0
positive	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0.03	0	0	0.1	0	0	0	0	0	0
negative	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0	0	0	0	0	0	0	0	0
negative	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0
neutral	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0
negative	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0	0	0	0.01	0	0	0	0
positive	0.01	0	0	0	0	0	0.01	0	0.01	0.01	0	0	0	0	0.02	0	0	0	0	0	0	0	0
positive	0	0	0	0	0	0	0	0	0	0	0	0	0	0.06	0	0	0	0	0	0	0	0	0
positive	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0	0	0	0	0	0	0	0	0
negative	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0	0	0	0	0	0	0	0.02	0
positive	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0.03	0	0	0	0	0	0	0	0	0
negative	0	0	0	0.02	0	0	0	0	0	0	0.01	0	0	0.02	0	0	0	0	0	0	0	0	0
negative	0	0	0	0	0	0	0.01	0	0	0.01	0	0	0	0.02	0	0	0	0	0	0	0	0	0.01

รูปที่ 8 สถานะราคาปิดของหุ้นและค่า TFIDF เพื่อให้น้ำหนักคำในแต่ละ Word Index

6. Training และ Test Dataset ทำการแบ่งข้อมูลเป็น 2 ส่วนคือข้อมูลฝึกสอน (training dataset) จำนวนร้อยละ 80 ของข้อมูลทั้งหมด และข้อมูลทดสอบ (test dataset) จำนวนร้อยละ 20 ของข้อมูลทั้งหมด

การสร้างแบบจำลอง

เป็นขั้นตอนการวิเคราะห์ข้อมูล ในงานวิจัยนี้ได้ใช้เทคนิคการเรียนรู้ของเครื่อง (machine learning) ประเภทการเรียนรู้แบบมีผู้สอน (supervised learning) ซึ่งใช้แบบจำลองการจำแนก (classification model) ทั้งหมด 7 ตัวแบบคือ 1. K-neighbors Classifier, 2. Logistic Regression, 3. Gradient Boosting Classifier, 4. Random Forest Classifier, 5. AdaBoost Classifier, 6. XGB Classifier และ 7. Support Vector Machine (SVM) โดยตัวแบบ K-neighbors Classifier, Logistic Regression, Gradient Boosting Classifier, Random Forest Classifier, AdaBoost Classifier และ Support Vector Machine

(SVM) ใช้ Library sklearn ช่วยในการประมวลผลและตัวแบบ XGB Classifier ใช้ Library xgboost ช่วยในการประมวลผล

การประเมินผล

ในงานวิจัยนี้ใช้ค่าความแม่นยำ (accuracy) เป็นตัววัดประสิทธิภาพของตัวแบบการจำแนก โดยคำนวณจาก Confusion Matrix ซึ่งเป็นตารางค่าพยากรณ์เปรียบเทียบกับข้อมูลจริง และใช้ Library sklearn import cross_val_score ช่วยในการคำนวณ

การใช้งาน

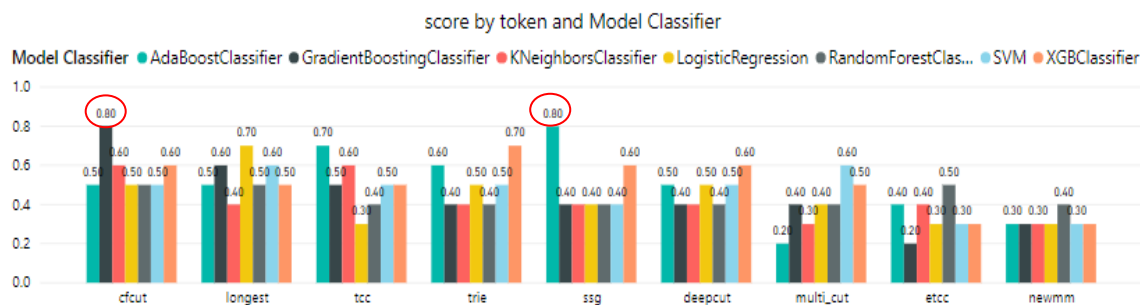
หลังจากเลือกตัวแบบที่มีค่าความแม่นยำสูงที่สุดแล้วจึงนำตัวแบบไปใช้งาน โดยสรุปเป็นขั้นตอนการนำไปใช้งานดังนี้

1. เก็บรวบรวมและเตรียมข้อมูลล่วงหน้าก่อนถึงวันที่ต้องการพยากรณ์ 1 วัน และทำการแบ่งข้อมูลเป็น Training Dataset จำนวน 80% ของข้อมูลทั้งหมดและ Test Dataset จำนวน 20% ของข้อมูลทั้งหมด
2. สร้างแบบจำลองด้วยข้อมูล Training Dataset และใช้ Test Dataset ประเมินประสิทธิภาพของแบบจำลอง (model) เพื่อหาแบบจำลอง และวิธีการตัดคำ (Tokenizer) ที่มีค่าความถูกต้องแม่นยำของการพยากรณ์สูงสุดจากการเรียนรู้ทั้งหมด 5 ครั้ง ซึ่งจะใช้คำสั่ง predict() ใน Library sklearn ทำให้ที่ผลออกมาจะอยู่ในรูปของ ทิศทางบวก (positive), ทิศทางลบ (negative) และเป็นกลาง (neutral)
3. นำตัวแบบที่ได้ในข้อ 2. มาพยากรณ์ทิศทางของราคาหุ้นจากข้อความข่าววันถัดมา

ผลการวิจัย

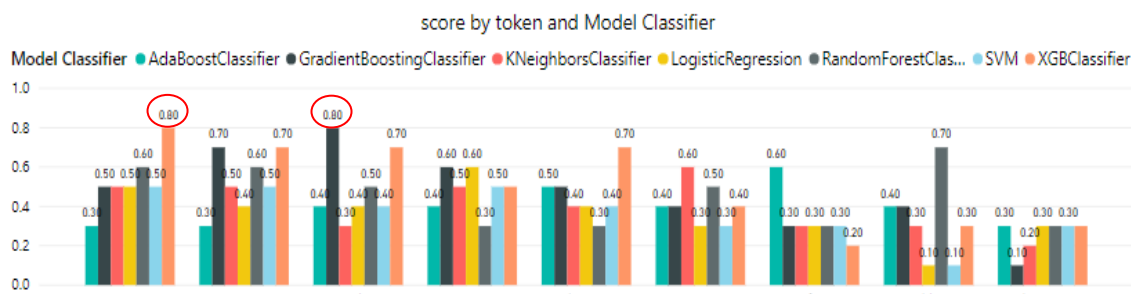
ห้่นกลุ่มเทคโนโลยีสารสนเทศและการสื่อสาร (ICT)

ผู้วิจัยได้สุ่มเลือกห้่นกลุ่มเทคโนโลยีสารสนเทศและการสื่อสาร (ICT) 1 ตัว คือ ห้่นของบริษัท อินทัช โฮลดิ้งส์ จำกัด (INTUCH) ซึ่งมีธุรกิจด้านการลงทุนในธุรกิจโทรคมนาคม สื่อ และเทคโนโลยี ในการวิจัยครั้งนี้ได้ทำการพยากรณ์ทิศทางของราคาหุ้นจากข้อความข่าวจำนวน 3 วัน ซึ่งมีวันที่ 5, 6 และ 7 กุมภาพันธ์ 2563 โดยแบ่งข้อความสำหรับฝึกสอน (train) และทดสอบตัวแบบ (test) เพื่อหาตัวแบบและการตัดคำที่มีค่าความแม่นยำ (accuracy) สูงที่สุดแบบสุ่ม (หมายความว่าถ้าค่าความแม่นยำสูงที่สุดมีหลายตัวแบบ โปรแกรมจะเลือกตัวแบบที่เหมาะสมโดยวิธีการสุ่ม) เพื่อนำมาพยากรณ์ทิศทางของราคาหุ้นบริษัท อินทัช โฮลดิ้งส์ จำกัด (INTUCH) โดยพยากรณ์ซ้ำทั้งหมด 5 ครั้งในแต่ละวัน และเลือกตัวแบบที่ดีที่สุดจาก 5 ครั้งนี้ ได้ผลดังรูปที่ 9, 10 และ 11



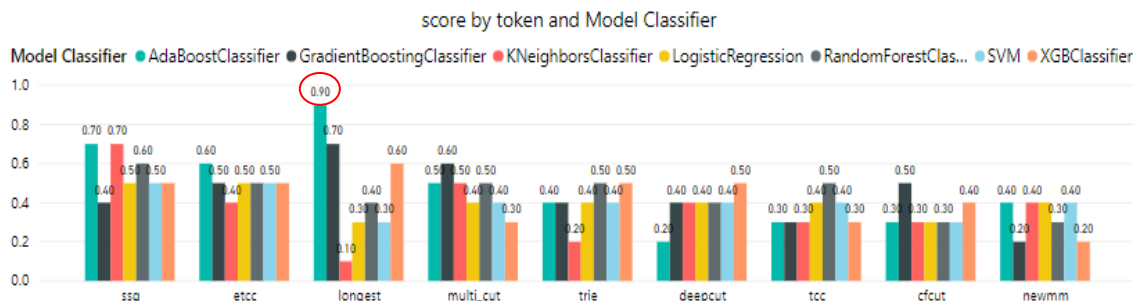
รูปที่ 9 ค่าความแม่นยำ (accuracy) ของหุ่นบริษัท อินทัช โฮลดิ้งส์ จำกัด (INTUCH) เพื่อพยากรณ์ทิศทางของราคาหุ้น วันที่ 5 กุมภาพันธ์ 2563 โดยใช้ข้อมูลข่าวตั้งแต่วันที่ 25 พฤศจิกายน 2562 จนถึง วันที่ 4 กุมภาพันธ์ 2563 และเลือกตัวแบบในการทำซ้ำครั้งที่ 1

จากรูปที่ 9 จะเห็นได้ว่าการทำซ้ำครั้งที่ 1 (จากทั้งหมด 5 ครั้ง) ตัวแบบที่มีค่าความแม่นยำสูงที่สุด (0.8) มี 2 ตัวแบบคือตัวแบบ Gradient Boosting Classifier (วิธีการตัดคำ cfcut) และตัวแบบ AdaBoost Classifier (วิธีการตัดคำ ssg) โปรแกรมจะทำการเลือกตัวแบบที่ดีที่สุดโดยสุ่ม ซึ่งในการประมวลผลครั้งนี้เลือกตัวแบบ Gradient Boosting Classifier เป็นตัวแบบที่ดีที่สุด โดยวิธีการตัดคำคือ cfcut



รูปที่ 10 ค่าความแม่นยำ (accuracy) ของหุ่นบริษัท อินทัช โฮลดิ้งส์ จำกัด (INTUCH) เพื่อพยากรณ์ทิศทางของราคาหุ้น วันที่ 6 กุมภาพันธ์ 2563 โดยใช้ข้อมูลข่าวตั้งแต่วันที่ 25 พฤศจิกายน 2562 จนถึง วันที่ 5 กุมภาพันธ์ 2563 และเลือกตัวแบบในการทำซ้ำครั้งที่ 4

จากรูปที่ 10 จะเห็นได้ว่าการทำซ้ำครั้งที่ 4 ตัวแบบที่มีค่าความแม่นยำสูงที่สุด (0.8) มี 2 ตัวแบบคือตัวแบบ XGB Classifier (วิธีการตัดคำ etcc) และตัวแบบ Gradient Boosting Classifier (วิธีการตัดคำ trie) โปรแกรมจะทำการเลือกตัวแบบที่ดีที่สุดโดยสุ่ม ซึ่งในการประมวลผลครั้งนี้เลือกตัวแบบ Gradient Boosting Classifier และวิธีการตัดคำคือ trie



รูปที่ 11 ค่าความแม่นยำ (accuracy) ของหุ่นบริษัท อินทัช โฮลดิ้งส์ จำกัด (INTUCH) เพื่อพยากรณ์ทิศทางของราคาหุ้น วันที่ 7 กุมภาพันธ์ 2563 โดยใช้ข้อมูลข่าวตั้งแต่วันที่ 25 พฤศจิกายน 2562 จนถึง วันที่ 6 กุมภาพันธ์ 2563 และเลือกตัวแบบในการทำซ้ำครั้งที่ 4

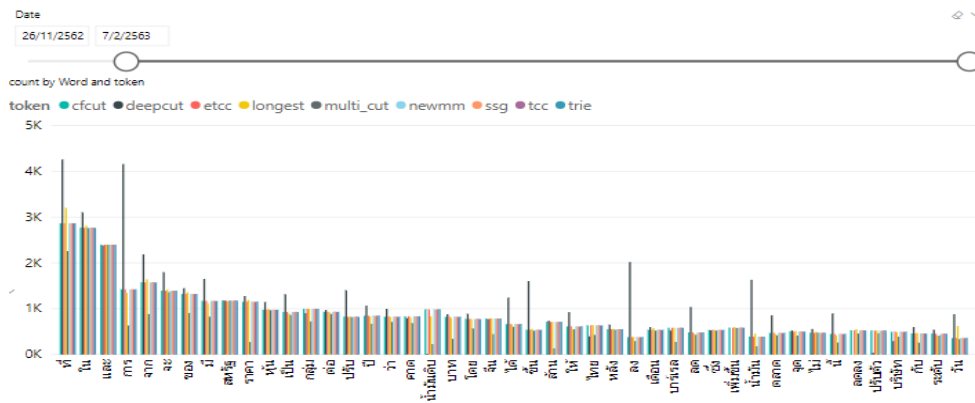
ตารางที่ 1 การพยากรณ์ทิศทางของราคาหุ้นของบริษัท อินทัช โฮลดิ้งส์ จำกัด (INTUCH)

ค่าความแม่นยำ	ตัวแบบการจำแนก	การตัดคำ	หุ้น	วันที่พยากรณ์	ทำซ้ำครั้งที่	ค่าพยากรณ์	ค่าจริง
0.8	GradientBoostingClassifier	cfcut	INTUCH	5/2/2020	1	positive	positive
0.8	GradientBoostingClassifier	trie	INTUCH	6/2/2020	4	negative	negative
0.9	AdaBoostClassifier	longest	INTUCH	7/2/2020	4	negative	negative

ผลการพยากรณ์ทิศทางของราคาหุ้นของบริษัท อินทัช โฮลดิ้งส์ จำกัด (INTUCH) นั้น เมื่อเปรียบเทียบกับข้อมูลจริงของวันที่ 5, 6 และ 7 กุมภาพันธ์ 2563 พบว่ามีความถูกต้องร้อยละ 100 ตามที่ได้แสดงไว้ในตารางที่ 1

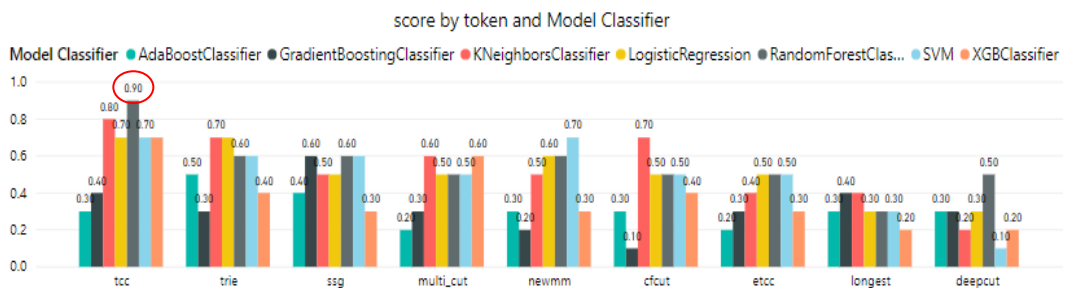
หุ้นกลุ่มพลังงานและสาธารณูปโภค (ENERG)

ผู้วิจัยได้ทำการสุ่มเลือกหุ้นกลุ่มพลังงานและสาธารณูปโภค (ENERG) 1 ตัว คือ หุ้นบริษัท ไทยออยล์ จำกัด (มหาชน) (TOP) ซึ่งเป็นผู้ประกอบการกลั่นและจำหน่ายน้ำมันปิโตรเลียมที่ใหญ่ที่สุดในประเทศไทย และมีโรงงานที่มีประสิทธิภาพสูงในภูมิภาคเอเชียแปซิฟิก มีธุรกิจที่หลากหลาย เช่น ธุรกิจการกลั่นน้ำมัน ธุรกิจปิโตรเคมีและธุรกิจน้ำมันหล่อลื่นพื้นฐาน รวมทั้งธุรกิจธุรกิจไฟฟ้า ธุรกิจสารทำละลาย เป็นต้น จากการตัดคำด้วย Library pythainlp ซึ่งมี 9 วิธี คือ 'cfcut', 'deepcut', 'etcc', 'longest', 'multi_cut', 'newmm', 'ssg', 'tcc' และ 'trie' ความถี่ของคำตั้งแต่วันที่ 26 พฤศจิกายน 2562 จนถึง วันที่ 7 กุมภาพันธ์ 2563 แสดงดังรูปที่ 12

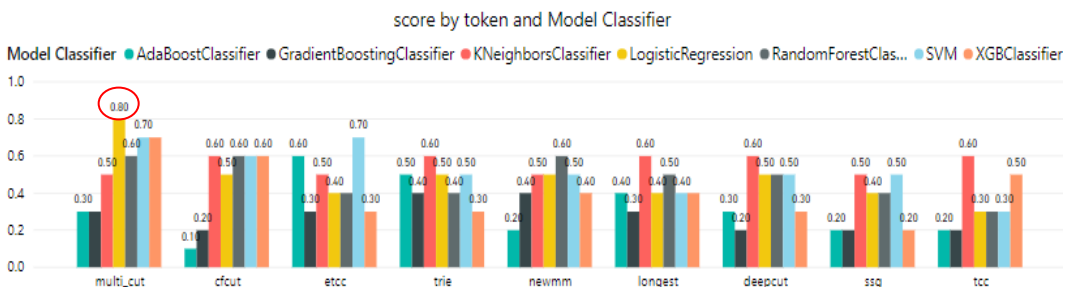


รูปที่ 12 ความถี่ของคำของหุ้นบริษัท ไทยออยล์ จำกัด (มหาชน) (TOP)

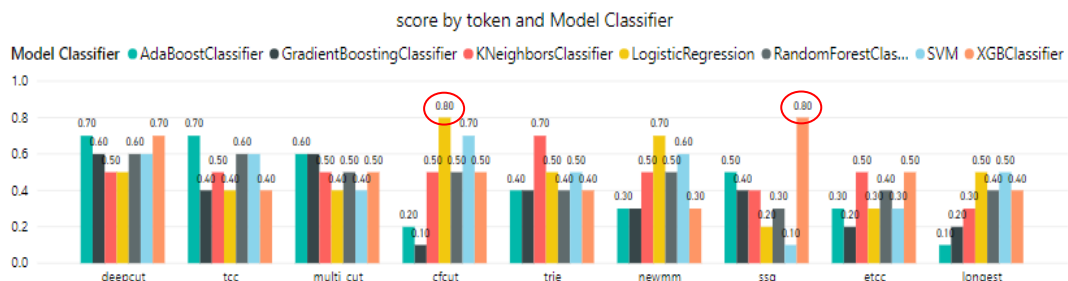
นำข้อมูลมาใช้ในการสร้างตัวแบบการเรียนรู้เพื่อพยากรณ์ทิศทางของราคาหุ้นวันที่ 5, 6 และ 7 กุมภาพันธ์ 2563 โดยพิจารณาจากค่าความแม่นยำของตัวแบบการจำแนกและวิธีการตัดคำโดย Library pythainlp แสดงผลดังรูปที่ 13, 14 และ 15



รูปที่ 13 ค่าความแม่นยำของหุ้นบริษัท ไทยออยล์ จำกัด (มหาชน) (TOP) เพื่อพยากรณ์ทิศทางของราคาหุ้นวันที่ 5 กุมภาพันธ์ 2563 โดยใช้ข้อมูลข่าวตั้งแต่วันที่ 26 พฤศจิกายน 2562 จนถึง วันที่ 4 กุมภาพันธ์ 2563 และเลือกตัวแบบในการทำซ้ำครั้งที่ 2



รูปที่ 14 ค่าความแม่นยำของหุ้นบริษัท ไทยออยล์ จำกัด (มหาชน) (TOP) เพื่อพยากรณ์ทิศทางของราคาหุ้นวันที่ 6 กุมภาพันธ์ 2563 โดยใช้ข้อมูลข่าวตั้งแต่วันที่ 26 พฤศจิกายน 2562 จนถึง วันที่ 5 กุมภาพันธ์ 2563 และเลือกตัวแบบในการทำซ้ำครั้งที่ 1



รูปที่ 15 ค่าความแม่นยำของหุ่นบริษัท ไทยออยล์ จำกัด (มหาชน) (TOP) เพื่อพยากรณ์ทิศทางของราคาหุ้น วันที่ 7 กุมภาพันธ์ 2563 โดยใช้ข้อมูลข่าวตั้งแต่วันที่ 26 พฤศจิกายน 2562 จนถึง วันที่ 6 กุมภาพันธ์ 2563 และเลือกตัวแบบในการทำซ้ำครั้งที่ 3

จากรูปที่ 15 จะเห็นได้ว่าการทำซ้ำครั้งที่ 3 ตัวแบบที่มีค่าความแม่นยำสูงสุด (0.8) มี 2 ตัวแบบคือตัวแบบ Logistic Regression (วิธีการตัดคำ cfcut) และตัวแบบ XGB Classifier (วิธีการตัดคำ ssg) โปรแกรมจะทำการเลือกตัวแบบที่ดีที่สุดโดยสุ่ม ซึ่งในการประมวลผลครั้งนี้เลือกตัวแบบ XGB Classifier และวิธีการตัดคำคือ ssg

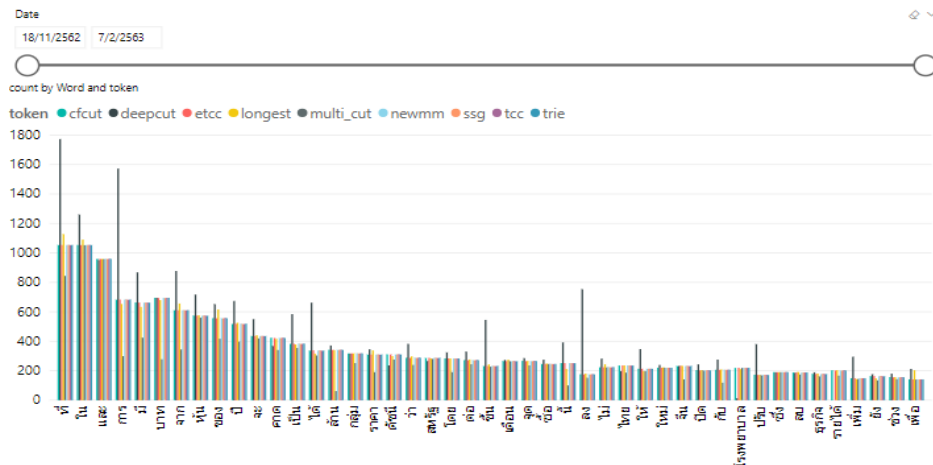
ตารางที่ 2 การพยากรณ์ทิศทางของราคาหุ้นบริษัท ไทยออยล์ จำกัด (มหาชน) (TOP)

ค่าความแม่นยำ	ตัวแบบการจำแนก	การตัดคำ	หุ้น	วันที่พยากรณ์	ทำซ้ำครั้งที่	ค่าพยากรณ์	ค่าจริง
0.9	RandomForestClassifier	tcc	TOP	5/2/2020	2	negative	positive
0.8	LogisticRegression	multi_cut	TOP	6/2/2020	1	positive	positive
0.8	XGBClassifier	ssg	TOP	7/2/2020	3	negative	negative

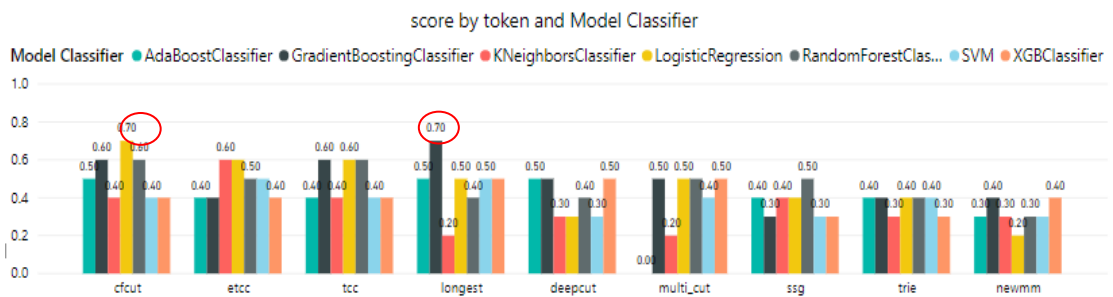
ค่าพยากรณ์ทิศทางของราคาหุ้นบริษัท ไทยออยล์ จำกัด (มหาชน) (TOP) นั้น เมื่อเปรียบเทียบกับข้อมูลจริงของวันที่ 5, 6, 7 กุมภาพันธ์ 2563 มีความถูกต้องร้อยละ 66.67 ตามที่ได้แสดงไว้ในตารางที่ 2

หุ่นกลุ่มการแพทย์ (HEALTH)

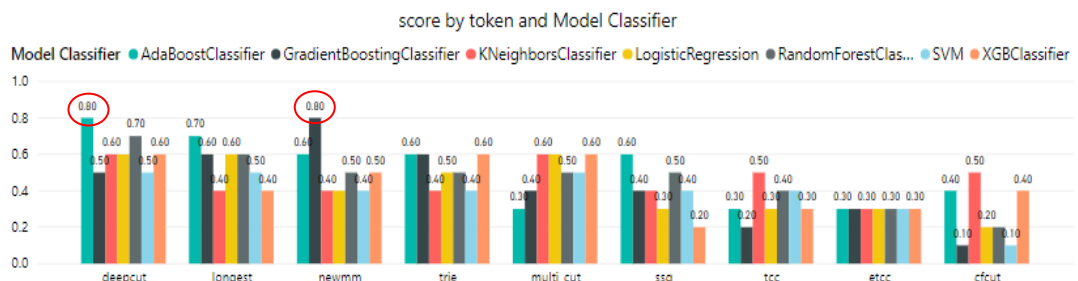
ทางผู้วิจัยได้ทำการสุ่มเลือกหุ่นกลุ่มการแพทย์ (HEALTH) 1 ตัว คือบริษัท โรงพยาบาลบำรุงราษฎร์ จำกัด (มหาชน) (BH) ซึ่งเป็นธุรกิจโรงพยาบาลเอกชนที่ให้บริการทางการแพทย์ทั้งผู้ป่วยในและผู้ป่วยนอก รวมทั้งผู้ป่วยต่างชาติ อีกทั้งยังมีการลงทุนในธุรกิจทางการแพทย์ทั้งในและต่างประเทศ จากการตัดคำด้วย Library pythainlp แสดงความถี่ของคำตั้งแต่วันที่ 18 พฤศจิกายน 2562 จนถึง วันที่ 7 กุมภาพันธ์ 2563 ดังรูปที่ 16 และผลการพยากรณ์ทิศทางของราคาหุ้นบริษัท โรงพยาบาลบำรุงราษฎร์ จำกัด (มหาชน) (BH) แสดงดังรูปที่ 17, 18 และ 19



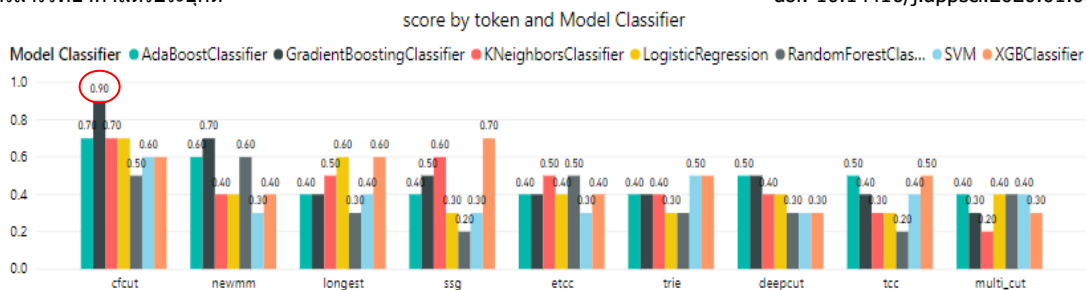
รูปที่ 16 ความถี่ของคำของหุ่นบริษัท โรงพยาบาลบำรุงราษฎร์ จำกัด (มหาชน) (BH)



รูปที่ 17 ค่าความแม่นยำของหุ่นบริษัท โรงพยาบาลบำรุงราษฎร์ จำกัด (มหาชน) (BH) เพื่อพยากรณ์ทิศทางของราคาหุ้น วันที่ 5 กุมภาพันธ์ 2563 โดยใช้ข้อมูลข่าวตั้งแต่วันที่ 18 พฤศจิกายน 2562 จนถึง วันที่ 4 กุมภาพันธ์ 2563 และเลือกตัวแบบในการทำซ้ำครั้งที่ 1



รูปที่ 18 ค่าความแม่นยำของหุ่นบริษัท โรงพยาบาลบำรุงราษฎร์ จำกัด (มหาชน) (BH) เพื่อพยากรณ์ทิศทางของราคาหุ้น วันที่ 6 กุมภาพันธ์ 2563 โดยใช้ข้อมูลข่าวตั้งแต่วันที่ 18 พฤศจิกายน 2562 จนถึง วันที่ 5 กุมภาพันธ์ 2563 และเลือกตัวแบบในการทำซ้ำครั้งที่ 4



รูปที่ 19 ค่าความแม่นยำของหุ่นบริษัท โรงพยาบาลบำรุงราษฎร์ จำกัด (มหาชน) (BH) เพื่อพยากรณ์ทิศทางของราคาหุ้น วันที่ 7 กุมภาพันธ์ 2563 โดยใช้ข้อมูลข่าวตั้งแต่วันที่ 18 พฤศจิกายน 2562 จนถึง วันที่ 6 กุมภาพันธ์ 2563 และเลือกตัวแบบในการทำซ้ำครั้งที่ 1

จากรูปที่ 17 จะเห็นได้ว่าการทำซ้ำครั้งที่ 1 ตัวแบบที่มีค่าความแม่นยำสูงที่สุด (0.7) มี 2 ตัวแบบคือตัวแบบ Logistic Regression (วิธีการตัดคำ cfcut) และตัวแบบ Gradient Boosting Classifier (วิธีการตัดคำ longest) โปรแกรมจะทำการเลือกตัวแบบที่ดีที่สุดโดยสุ่ม ซึ่งในการประมวลผลครั้งนี้เลือกตัวแบบ Logistic Regression และวิธีการตัดคำคือ cfcut

จากรูปที่ 18 จะเห็นได้ว่าการทำซ้ำครั้งที่ 4 ตัวแบบที่มีค่าความแม่นยำสูงที่สุด (0.8) มี 2 ตัวแบบคือตัวแบบ AdaBoost Classifier (วิธีการตัดคำ deepcut) และตัวแบบ Gradient Boosting Classifier (วิธีการตัดคำ newmm) โปรแกรมจะทำการเลือกตัวแบบที่ดีที่สุดโดยสุ่ม ซึ่งในการประมวลผลครั้งนี้เลือกตัวแบบ AdaBoost Classifier และวิธีการตัดคำคือ deepcut

ตารางที่ 3 การพยากรณ์ทิศทางของราคาหุ้นบริษัท โรงพยาบาลบำรุงราษฎร์ จำกัด (มหาชน) (BH)

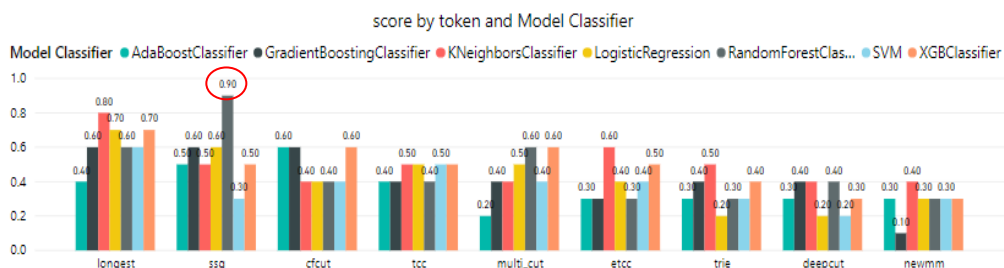
ค่าความแม่นยำ	ตัวแบบการจำแนก	การตัดคำ	หุ้น	วันที่พยากรณ์	ทำซ้ำครั้งที่	ค่าพยากรณ์	ค่าจริง
0.7	LogisticRegression	cfcut	BH	5/2/2020	1	negative	negative
0.8	AdaBoostClassifier	deepcut	BH	6/2/2020	4	negative	negative
0.9	GradientBoostingClassifier	cfcut	BH	7/2/2020	1	negative	positive

ผลการพยากรณ์ทิศทางของราคาหุ้นบริษัท โรงพยาบาลบำรุงราษฎร์ จำกัด (มหาชน) (BH) นั้น เมื่อเปรียบเทียบกับข้อมูลจริงในวันที่ 5, 6 และ 7 กุมภาพันธ์ 2563 พบว่ามีความถูกต้องร้อยละ 66.67 ตามที่ได้แสดงไว้ในตารางที่ 3

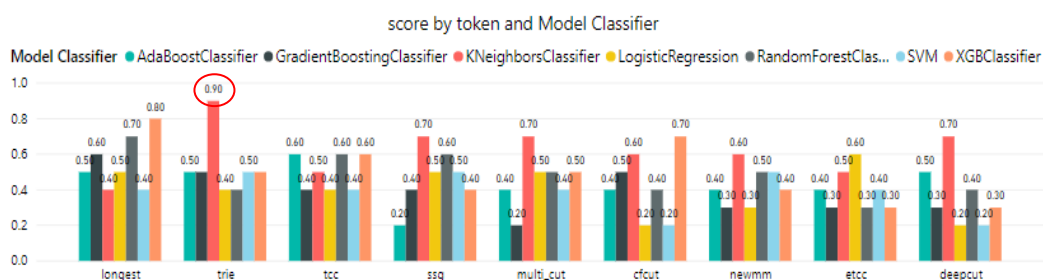
หุ่นกลุ่มพาณิชย์ (COMM)

ทางผู้วิจัยได้ทำการสุ่มเลือกหุ่นกลุ่มพาณิชย์ (COMM) 1 ตัว คือ บริษัท ซีพี ออลล์ จำกัด (มหาชน) (CPALL) ซึ่งเป็นธุรกิจร้านค้าปลีกภายใต้เครื่องหมายการค้า 7-Eleven ในประเทศไทย และลงทุนในธุรกิจผลิตและจำหน่ายอาหาร

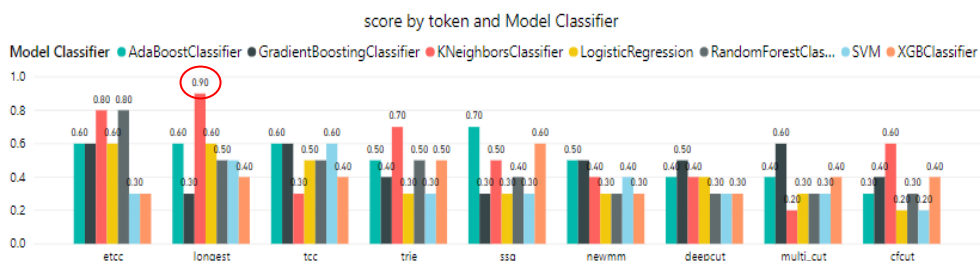
สำเร็จรูปและเบเกอร์รี่ ตัวแทนรับชำระเงินค่าสินค้าและบริการ ผลการเรียนรู้ข้อมูลเพื่อพยากรณ์ทิศทางของราคาหุ้นบริษัท ซีพี ออลล์ จำกัด (มหาชน) (CPALL) แสดงดังรูปที่ 20, 21 และ 22



รูปที่ 20 ค่าความแม่นยำของหุ้นบริษัท ซีพี ออลล์ จำกัด (มหาชน) (CPALL) เพื่อพยากรณ์ทิศทางของราคาหุ้นวันที่ 5 กุมภาพันธ์ 2563 โดยใช้ข้อมูลข่าวตั้งแต่วันที่ 25 พฤศจิกายน 2562 จนถึง วันที่ 4 กุมภาพันธ์ 2563 และเลือกตัวแบบในการทำซ้ำครั้งที่ 1



รูปที่ 21 ค่าความแม่นยำของหุ้นบริษัท ซีพี ออลล์ จำกัด (มหาชน) (CPALL) เพื่อพยากรณ์ทิศทางของราคาหุ้นวันที่ 6 กุมภาพันธ์ 2563 โดยใช้ข้อมูลข่าวตั้งแต่วันที่ 25 พฤศจิกายน 2562 จนถึง วันที่ 5 กุมภาพันธ์ 2563 และเลือกตัวแบบในการทำซ้ำครั้งที่ 2



รูปที่ 22 ค่าความแม่นยำของหุ้นบริษัท ซีพี ออลล์ จำกัด (มหาชน) (CPALL) เพื่อพยากรณ์ทิศทางของราคาหุ้นวันที่ 7 กุมภาพันธ์ 2563 โดยใช้ข้อมูลข่าวตั้งแต่วันที่ 25 พฤศจิกายน 2562 จนถึง วันที่ 6 กุมภาพันธ์ 2563 และเลือกตัวแบบในการทำซ้ำครั้งที่ 3

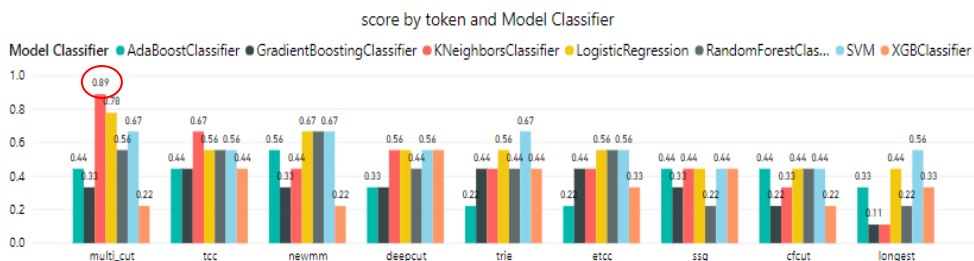
ตารางที่ 4 การพยากรณ์ทิศทางของราคาหุ้นของบริษัท ซีพี ออลล์ จำกัด (มหาชน) (CPALL)

ค่าความแม่นยำ	ตัวแบบการจำแนก	การตัดคำ	หุ้น	วันที่พยากรณ์	ทำซ้ำครั้งที่	ค่าพยากรณ์	ค่าจริง
0.9	RandomForestClassifier	ssg	CPALL	5/2/2020	1	positive	positive
0.9	KNeighborsClassifier	trie	CPALL	6/2/2020	2	negative	positive
0.9	KNeighborsClassifier	longest	CPALL	7/2/2020	3	positive	positive

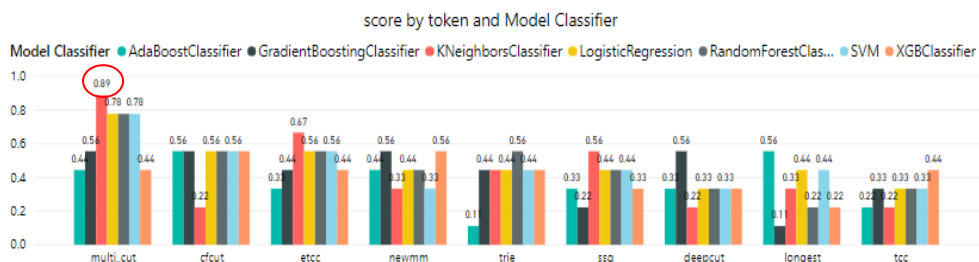
ผลการพยากรณ์ทิศทางของราคาหุ้นบริษัท ซีพี ออลล์ จำกัด (มหาชน) (CPALL) นั้น เมื่อเปรียบเทียบกับข้อมูลจริงของวันที่ 5, 6, 7 กุมภาพันธ์ 2563 มีความถูกต้องร้อยละ 66.67 ตามที่ได้แสดงไว้ในตารางที่ 4

หุ้นกลุ่มธนาคาร (BANK)

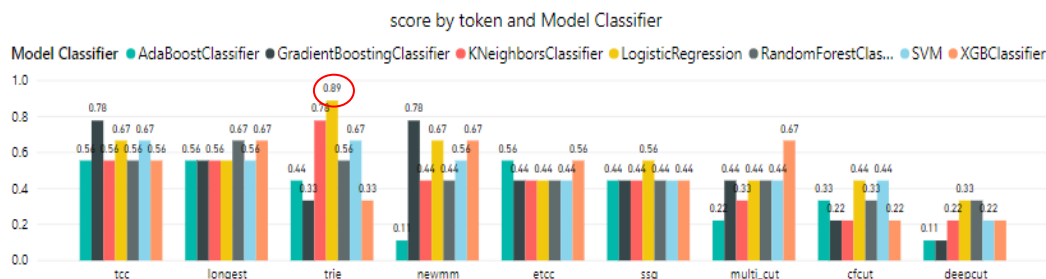
ผู้วิจัยได้ทำการสุ่มเลือกหุ้นกลุ่มธนาคาร (BANK) 1 ตัว คือ ธนาคารกสิกรไทย จำกัด (มหาชน) (KBANK) ซึ่งประกอบธุรกิจธนาคารพาณิชย์ ธุรกิจหลักทรัพย์และธุรกิจที่ได้รับอนุญาตไว้ในพระราชบัญญัติธุรกิจสถาบันการเงินฯ และพระราชบัญญัติหลักทรัพย์และตลาดหลักทรัพย์ฯ และประกาศที่เกี่ยวข้อง ณ วันที่ 31 ธันวาคม 2561 ผลการเรียนรู้ข้อมูลเพื่อพยากรณ์ทิศทางของราคาหุ้นธนาคารกสิกรไทย จำกัด (มหาชน) (KBANK) แสดงดังรูปที่ 23, 24 และ 25



รูปที่ 23 ค่าความแม่นยำของหุ้นธนาคารกสิกรไทย จำกัด (มหาชน) (KBANK) เพื่อพยากรณ์ทิศทางของราคาหุ้นวันที่ 5 กุมภาพันธ์ 2563 โดยใช้ข้อมูลข่าวตั้งแต่วันที่ 28 พฤศจิกายน 2562 จนถึง วันที่ 4 กุมภาพันธ์ 2563 และเลือก ตัวแบบในการทำซ้ำครั้งที่ 1



รูปที่ 24 ค่าความแม่นยำของหุ้นธนาคารกสิกรไทย จำกัด (มหาชน) (KBANK) เพื่อพยากรณ์ทิศทางของราคาหุ้นวันที่ 6 กุมภาพันธ์ 2563 โดยใช้ข้อมูลข่าวตั้งแต่วันที่ 28 พฤศจิกายน 2562 จนถึง วันที่ 5 กุมภาพันธ์ 2563 และเลือก ตัวแบบในการทำซ้ำครั้งที่ 1



รูปที่ 25 ค่าความแม่นยำของหุ่นธนาคารกสิกรไทย จำกัด (มหาชน) (KBANK) เพื่อพยากรณ์ทิศทางของหุ้นราคา วันที่ 7 กุมภาพันธ์ 2563 โดยใช้ข้อมูลข่าวตั้งแต่วันที่ 28 พฤศจิกายน 2562 จนถึง วันที่ 6 กุมภาพันธ์ 2563 และเลือก ตัวแบบในการทำซ้ำครั้งที่ 2

ตารางที่ 5 การพยากรณ์ทิศทางของราคาหุ้นธนาคารกสิกรไทย จำกัด (มหาชน) (KBANK)

ค่าความแม่นยำ	ตัวแบบการจำแนก	การตัดคำ	หุ้น	วันที่พยากรณ์	ทำซ้ำครั้งที่	ค่าพยากรณ์	ค่าจริง
0.888889	KNeighborsClassifier	multi_cut	KBANK	5/2/2020	1	positive	positive
0.888889	KNeighborsClassifier	multi_cut	KBANK	6/2/2020	1	positive	positive
0.888889	LogisticRegression	trie	KBANK	7/2/2020	2	negative	positive

การพยากรณ์ทิศทางของราคาหุ้นธนาคารกสิกรไทย จำกัด (มหาชน) (KBANK) นั้น เมื่อเปรียบเทียบกับข้อมูลจริงของวันที่ 5, 6, 7 กุมภาพันธ์ 2563 พบว่ามีความถูกต้องร้อยละ 66.67 ตามที่ได้แสดงไว้ในตารางที่ 5

วิจารณ์ผลการวิจัย

ในงานวิจัยชิ้นนี้ ผู้วิจัยใช้ Library pythainlp ช่วยในการประมวลผล และใช้การตัดคำ 9 วิธี คือ 'cfcut', 'deepcut', 'etcc', 'longest', 'multi_cut', 'newmm', 'ssg', 'tcc' และ 'trie' จากนั้นแบ่งข้อมูลออกเป็น 2 ส่วน คือ ข้อมูลฝึกสอน (training dataset) 80% ของข้อมูลทั้งหมด และข้อมูลทดสอบ (test dataset) 20% ของข้อมูลทั้งหมด ใช้ข้อมูลฝึกสอนในการสร้างตัวแบบ โดยแบ่งข้อมูลและสร้างตัวแบบซ้ำทั้งหมด 5 ครั้ง ในแต่ละวัน จากตารางที่ 1-5 จะเห็นว่าตัวแบบการจำแนกและวิธีการตัดคำที่เหมาะสมในแต่ละประเภทหุ้นไม่ได้มีความสัมพันธ์กัน เนื่องจากวิธีการตัดคำเป็นอัลกอริทึม (algorithm) ที่ใช้ในการตัดคำในภาษาไทยเพื่อให้ได้ความหมายที่ถูกต้อง ซึ่งอยู่ในขั้นตอนของการเตรียมข้อมูลก่อนนำไปสร้างตัวแบบ โดยคำในข้อความข่าวจะแตกต่างกันไปในแต่ละกลุ่มของหุ้น อีกทั้งการแบ่งข้อมูลออกเป็นข้อมูลฝึกสอนและข้อมูลทดสอบเป็นการแบ่งโดยใช้การสุ่มตัวอย่างอย่างง่ายแบบไม่ใส่คืน ดังนั้นการตัดคำที่เหมาะสมก็จะแตกต่างกันไปในแต่ละครั้งที่ทำซ้ำ นอกจากนี้หากเกิดกรณีที่ตัวแบบมีค่าความแม่นยำเท่ากันผู้วิจัยจะเลือกตัวแบบที่ดีที่สุดโดยสุ่ม ด้วยเหตุผลทั้งหมดที่กล่าวมาข้างต้นนี้ส่งผลให้แบบจำลองการจำแนกที่เหมาะสมกับการพยากรณ์ทิศทางของราคาหุ้นบางครั้งจะมีความแตกต่างกันในแต่ละครั้งที่ประมวลผล

สรุปผลการวิจัย

ในงานวิจัยนี้ได้ทำการสุ่มเลือกหุ้นอย่างละ 1 ตัว จากหุ้น 5 ประเภท ทำการรวบรวมข้อมูลข้อความข่าวของหุ้นแต่ละตัวในแต่ละวันและสถานะของราคาหุ้น ณ ปัจจุบันเทียบกับราคาหุ้นของวันก่อนหน้า โดยใช้วิธีการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) ซึ่งทำการตัดคำโดยใช้วิธีการตัดคำ 9 วิธี ใน Library pythainlp คือ 'cfcut', 'deepcut', 'etcc', 'longest', 'multi_cut', 'newmm', 'ssg', 'tcc' และ 'trie' แล้วให้น้ำหนักคำโดยใช้เทคนิค TFIDF (term frequency-inverse document frequency) ทำการเรียนรู้ข้อมูลฝึกสอน (training dataset) 80% ของข้อมูลทั้งหมด และข้อมูลทดสอบ (test dataset) 20% ของข้อมูลทั้งหมด ใช้แบบจำลองการจำแนก (classification model) ทั้งหมด 7 แบบจำลอง คือ K-neighbors Classifier, Logistic Regression, Gradient Boosting Classifier, Random Forest Classifier, AdaBoost Classifier , XGB Classifier และ Support Vector Machine (SVM) เพื่อหาตัวแบบที่มีค่าความแม่นยำ (accuracy) สูงสุด เพื่อทำการพยากรณ์ทิศทางของราคาหุ้นจากข้อความข่าว ในวันที่ 5, 6 และ 7 กุมภาพันธ์ 2563 ของหุ้นทั้งหมด 5 กลุ่ม โดยแสดงในส่วนของร้อยละความถูกต้องเพื่อเป็นการเปรียบเทียบทิศทางของราคาหุ้นของค่าพยากรณ์และค่าจริง ดังตารางที่ 6

จากตารางที่ 6 จะเห็นว่าตัวแบบการจำแนกที่เหมาะสมนั้นแตกต่างกันในแต่ละกลุ่มของหุ้นและบางครั้งไม่สามารถสรุปได้ว่าตัวแบบใดเป็นตัวแบบที่ดีที่สุด ดังนั้นในการนำไปใช้งานจริงผู้วิจัยจึงเขียนโปรแกรมภาษาไพธอนให้สามารถประมวลผลโดยเลือกตัวแบบที่ดีที่สุดจากทั้งหมด 7 ตัวแบบและเลือกการตัดคำที่ดีที่สุดจากทั้งหมด 9 วิธี เพื่อใช้ในการพยากรณ์ทิศทางของราคาหุ้นจากข้อความข่าวในแต่ละวัน

ตารางที่ 6 สรุปผลการวิจัยของหุ้นแต่ละกลุ่ม

ประเภทของหุ้น	ชื่อหุ้น	ชื่อบริษัท	แบบจำลองที่มีประสิทธิภาพสูงสุด		ร้อยละความถูกต้อง
			ตัวแบบการจำแนก	การตัดคำ	
เทคโนโลยีสารสนเทศและการสื่อสาร (ICT)	INTUCH	บริษัท อินทัช โฮลดิ้งส์ จำกัด	Gradient Boosting Classifier	ไม่สามารถสรุปได้	100
พลังงานและสาธารณูปโภค(ENERG)	TOP	บริษัท ไทยออยล์ จำกัด	ไม่สามารถสรุปได้	ไม่สามารถสรุปได้	66.67
การแพทย์ (HEALTH)	BH	บริษัท โรงพยาบาลบำรุงราษฎร์ จำกัด	ไม่สามารถสรุปได้	cfcut	66.67
พาณิชย์ (COMM)	CPALL	บริษัท ซีพี ออลล์ จำกัด	K-Neighbors Classifier	ไม่สามารถสรุปได้	66.67
ธนาคาร (BANK)	KBANK	ธนาคารกสิกรไทย จำกัด	K-Neighbors Classifier	multi_cut	66.67

ข้อเสนอแนะ

1. เนื่องจากภาษาไทยเป็นภาษาที่มีความซับซ้อนดังนั้นจึงควรมีนักภาษาศาสตร์เข้ามาช่วยพิจารณาในการแยกคำ การจัดกลุ่มคำและการให้ความหมาย
2. ควรมีการปรับปรุงและพัฒนาแบบจำลองเพื่อให้แบบจำลองสามารถทำงานแบบ real-time monitoring เพื่อเป็นเครื่องมือช่วยประกอบการตัดสินใจลงทุนที่มีประสิทธิภาพ
3. ควรมีการหาค่าพารามิเตอร์ที่เหมาะสมในแต่ละตัวแบบ ซึ่งในงานวิจัยชิ้นนี้ใช้ค่า Default ของโปรแกรม ภาษาไพธอนทุกตัวแบบ
4. เนื่องจากเวลาในการวิจัยมีจำกัดจึงไม่ได้มีการคัดคำที่ไม่มีผลเกี่ยวข้องกับการพยากรณ์ออก เช่น คำว่า “การ” “บาท” “ที่” และ “ความ” เป็นต้น ดังนั้นในอนาคตควรศึกษาเพิ่มเติมในเรื่องนี้

เอกสารอ้างอิง

- Chaicharoen, N. (2001). *Computerized integrated word segmentation and part-of-speech tagging of Thai* (M.Sc. thesis). Department of Linguistics, Faculty of Arts, Chulalongkorn University, Bangkok. (in Thai)
- Huang, W., Nakamori, Y. & Wang, S-Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, 32(10), 2513–2522.
- Kim, K-J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1), 307–319.
- Li, X., Huang, X., Deng, X., & Zhu, S. (2014). Enhancing Quantitative Intra-day Stock Return Prediction by Integrating both Market News and Stock Prices Information. *ScienceDirect*, 142, 228-238.
<https://doi.org/10.1016/j.neucom.2014.04.043>
- Sadia, K., Sharma, A., Paul, A., Padhi, S., & Sanyal, S. (2019). Stock Market Prediction Using Machine Learning Algorithms. *International Journal of Engineering and Advanced Technology (IJEAT)*, 8(4), 25-31.
- Sonlertlamvanit, W. (1993). *Thai word wrapping in the translation system*. Bangkok, Thailand: National Electronics and Computer Technology Center. (in Thai)
- Tay, F.E.H. & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, 29(4), 309–317.

Research Article

แบบจำลองทางคณิตศาสตร์การระบาดของโรคmelioidosisจาก สัตว์สู่คน โดยพิจารณาจากผลกระทบปัจจัยปริมาณน้ำฝน

A mathematical model of melioidosis transmission in humans and animals with the effect of a rainfall factor

พรวิไล ชาญกิจกรณ^{1*} และพันธุ์ณี พงศ์สัมพันธ์²

Pornwilai Chankitkan^{1*} and Puntani Pongsumpun²

¹สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครสวรรค์ นครสวรรค์

¹Department of Mathematic and Statistics, Faculty of Science and Technology, Nakhon Sawan Rajabhat University, Nakhon Sawan

²ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพมหานคร

²Department of Mathematic, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok

*E-mail: pornwilai.c@nsru.ac.th

Received: 18/01/2020; Revised: 07/06/2020; Accepted: 15/06/2020

บทคัดย่อ

โรคmelioidosisเป็นโรคที่เกิดจากเชื้อ *Burkholderia pseudomallei* ในบทความวิจัยฉบับนี้เกี่ยวข้องกับ
การศึกษาการแพร่ระบาดของโรคmelioidosisและสร้างแบบจำลองทางคณิตศาสตร์การระบาดของโรคmelioidosis
จากสัตว์สู่คนโดยเพิ่มปัจจัยปริมาณน้ำฝนที่ส่งผลต่อการแพร่ระบาดของโรคmelioidosisกับสัดส่วนประชากร
คนที่ได้รับเชื้อและแสดงอาการ ในบทความวิจัยได้หาจุดสมดุลภายใต้สภาวะไร้โรค สภาวะระบาดเรื้อรังและค่า
สืบพันธุ์พื้นฐาน(R_0) วิเคราะห์ความเสถียรภาพเฉพาะที่ของจุดสมดุลภายใต้สภาวะไร้โรคด้วยเงื่อนไข $R_0 < 1$
และจุดสมดุลภายใต้สภาวะระบาดเรื้อรังด้วยเงื่อนไข $R_0 > 1$ โดยตรวจสอบจากการใช้เงื่อนไขเร้าท์เซอร์วิท
จากนั้นวิเคราะห์ความเสถียรภาพวงกว้างของจุดสมดุลโดยการใช้ฟังก์ชันเลียปูนอฟ และผลลัพธ์เชิงตัวเลขพบว่าเมื่อ
ปัจจัยปริมาณน้ำฝนมีค่ามากจะส่งผลกระทบต่อสัดส่วนประชากรคนที่ได้รับเชื้อและแสดงอาการติดเชื้อเพิ่มมากขึ้น
ดังนั้นปัจจัยปริมาณน้ำฝนจึงมีผลกระทบต่อการระบาดของโรคmelioidosisในประชากรคน
คำสำคัญ: melioidosis, แบบจำลองทางคณิตศาสตร์, จุดสมดุล, ค่าสืบพันธุ์พื้นฐาน, ปริมาณน้ำฝน

Abstract

Melioidosis is caused by *Burkholderia pseudomallei*. This research paper involves the study of the transmission of the melioidosis by formulating the mathematical model of melioidosis transmission in humans and animals. Adding rainfall factor has an effect on proportion of infectious and symptomatic population. This research was conducted to find the disease free, endemic equilibrium points and basic reproductive number (R_0). Local stability analysis for the disease free equilibrium point with the condition $R_0 < 1$ and the endemic equilibrium point with the condition $R_0 > 1$ was investigated using the Routh – Hurwitz condition. Moreover, global stability of such equilibrium points was analyzed using the Lyapunov function. The numerical results show that when the rainfall factor is high, it will increase the proportion of the infectious and symptomatic population. Therefore, a rainfall factor has an impact on the Melioidosis outbreak in the population.

Keywords: melioidosis, mathematical model, equilibrium points, basic reproductive number, rainfall

บทนำ

โรคmelioidosis เป็นโรคที่เกิดจากเชื้อแบคทีเรียชนิดแกรมลบ *Burkholderia pseudomallei* ซึ่งเป็นโรคที่สามารถติดเชื้อได้ทั้งคนและสัตว์โดยพบเชื้อในสิ่งแวดล้อมทั่วไป เช่น ดิน โคลน และแหล่งน้ำ โรคmelioidosis มักจะเกิดในช่วงฤดูฝนหลังจากมีฝนตกกระาะเชื้อที่อยู่ในดินจะถูกชะขึ้นมาที่ผิวดิน เมื่อคนและสัตว์ที่สัมผัสกับบริเวณดังกล่าวจะมีโอกาสติดเชื้อได้ (Mangkala, 2009) การติดเชื้อที่สำคัญคือ ทางผิวหนังที่มีบาดแผล การหายใจ หรือแม้กระทั่งผู้ที่ไม่มีโรคเรื้อรังอันเป็นสาเหตุให้มีภูมิคุ้มกันต้านทานโรคต่ำจะมีความเสี่ยงสูงต่อการติดเชื้อ ส่วนการแพร่ระบาดของเชื้อจากสัตว์สู่คนสามารถติดเชื้อได้จากคนสัมผัสสารคัดหลั่งหลังจากสัตว์ที่ติดเชื้อ ลักษณะอาการของโรคเมื่อได้รับเชื้อมักจะไม่มีอาการจำเพาะแต่อาจมาด้วยอาการไข้สูง ซึ่งระยะฟักตัวของโรคหลังจากได้รับเชื้อโดยเฉลี่ยประมาณ 9 วัน โดยผู้ป่วยในรายที่ติดเชื้อในกระแสเลือดเฉียบพลันมักเสียชีวิตภายใน 1-3 วันหลังเข้ารับการรักษา (Suwanchairrob, 2015) สำหรับสถานการณ์โรคmelioidosis ในประเทศไทย จากข้อมูลระบบเฝ้าระวังรายงาน 506 พบว่าตั้งแต่ พ.ศ. 2546 – 2552 อัตราป่วยโรคmelioidosis อยู่ที่ 1.18 ต่อประชากรแสนคน หลังจากนั้นตั้งแต่ปี พ.ศ. 2553 – 2560 อัตราป่วยเพิ่มขึ้นอยู่ที่ 4.9 ต่อประชากรแสนคน อัตราตายอยู่ที่ร้อยละ 1.01 และตั้งแต่วันที่ 1 มกราคม – 2 พฤศจิกายน 2561 พบรายงานผู้ป่วยโรคmelioidosis จำนวน 2,266 ราย อัตราป่วยเพิ่มขึ้นอยู่ที่ 3.43 ต่อประชากรแสนคน ซึ่งจะพบได้ว่าผู้ที่ป่วยด้วยโรคmelioidosis มักจะป่วยในช่วงเดือน กรกฎาคม - พฤศจิกายน ของทุกปี (Division of Innovation and Research (DIR.), 2019)

จากข้อมูลการแพร่ระบาดของโรคmelioidosis ข้างต้นทำให้เห็นได้ว่ามีแนวโน้มที่จะเกิดโรคดังกล่าวอยู่ จึงทำให้มีบทความวิจัยที่เกี่ยวข้องกับการระบาดของโรคมามากขึ้น โดย Tanwisaid & Surin (2008) ได้ศึกษาปริมาณน้ำฝนรายเดือนและความรุนแรงของโรคmelioidosis ของจังหวัดนครพนมในภาคตะวันออกเฉียงเหนือของ

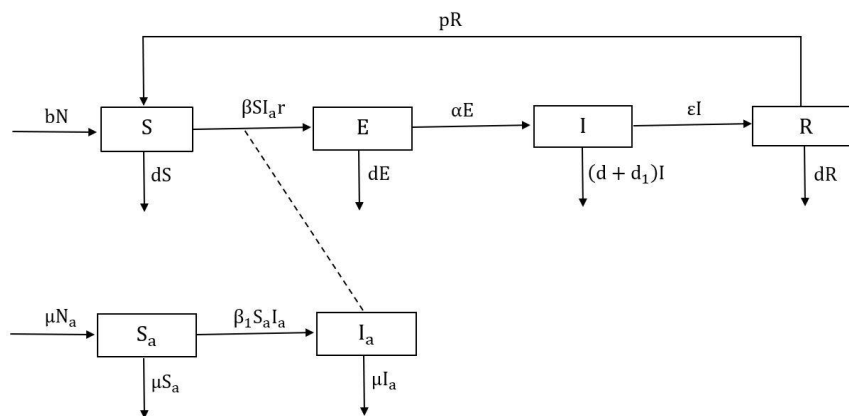
ประเทศไทยของข้อมูลย้อนหลัง 6 ปี จากการศึกษาพบว่า มีความชุกของผู้ป่วยโดยเฉลี่ยอยู่ที่ 9.97 ต่อประชากรแสนคน และจากข้อมูลปริมาณน้ำฝนรายเดือนทำให้ทราบว่าที่ร้อยละ 73.27 ผู้ป่วยได้รับผลกระทบในช่วงฤดูฝนในช่วงเดือน มิถุนายน – พฤศจิกายน ต่อมา Jeenpun (2013) ได้ศึกษาเกี่ยวกับระบาดวิทยาการติดเชื้อ *Burkholderia pseudomallei* ในสวนสัตว์สงขลาเพื่อประเมินความชุกของสัตว์ทั้งในสัตว์และผู้ปฏิบัติงานที่สัมผัสกับสัตว์ซึ่งจากการเก็บตัวอย่างสัตว์จากสัตว์และผู้ปฏิบัติงาน พบว่าความชุกของโรคเมลิออยโดสิสในสัตว์คิดเป็นร้อยละ 1.53 แสดงว่าสัตว์ในสวนสัตว์มีโอกาสสัมผัสกับเชื้อ *B. pseudomallei* ได้จากสิ่งแวดล้อมเช่นเดียวกับการตรวจเจอความชุกในผู้ปฏิบัติงานสัมผัสสัตว์คิดเป็นร้อยละ 1.07 ต่อมา Sumpuntharat et al. (2014) ได้ศึกษาปัจจัยที่มีความสัมพันธ์กับพฤติกรรมป้องกันการติดเชื้อซ้ำของผู้ติดเชื้อเมลิออยด์ โดยในบทความวิจัยพบว่าในการป้องกันการติดเชื้อซ้ำของผู้ติดเชื้อเมลิออยด์ควรส่งเสริมให้ผู้ป่วยรับรู้เกี่ยวกับโอกาสเสี่ยงต่อการเกิดโรคและรับรู้ประโยชน์ในการปฏิบัติพฤติกรรมในการป้องกันโรค ต่อมา Egilmez et al. (2018) ได้ศึกษาเชื้อไวรัสที่ขึ้นอยู่กับอุณหภูมิที่ส่งผลต่อการเกิดเชื้อแบคทีเรียโดยโรคเมลิออยโดสิส ซึ่งบทความวิจัยได้พัฒนาแบบจำลองทางคณิตศาสตร์เพื่อวิเคราะห์ความสัมพันธ์ของเชื้อแบคทีเรียที่พบในน้ำและดินโดยใช้พารามิเตอร์อุณหภูมิของอัตราการเจริญเติบโตของแบคทีเรียที่ทำให้เกิดการติดเชื้อต่อประชากรของ *B. pseudomallei* โดยผลจากการวิเคราะห์แบบจำลองพบว่า การเปลี่ยนแปลงฤดูกาลมีผลต่อแบคทีเรียซึ่งมีโอกาสเสี่ยงสูงที่จะทำให้เกิดโรคเมลิออยโดสิส และ Mahikul et al. (2019) ได้สร้างแบบจำลองการเปลี่ยนแปลงของประชากรตามฤดูกาลเพื่อพยากรณ์การเกิดโรคเมลิออยโดสิส โดยในบทความวิจัยได้สร้างแบบจำลองทางคณิตศาสตร์ SEIRS และใช้ข้อมูลอายุ เพศ พื้นที่ทางภูมิศาสตร์ในช่วงปี พ.ศ. 2551 – 2558 จากนั้นวิเคราะห์แบบจำลองโดยวิธีมาร์คอฟ เซน มอนติคาร์โล โดยผลของการพยากรณ์ของแบบจำลองพบว่าอัตราการเกิดของโรคเมลิออยโดสิสสูงขึ้นถึงสองเท่าเมื่อเทียบกับข้อมูลในปี พ.ศ. 2558

จากปัญหาการแพร่ระบาดของโรคเมลิออยโดสิสและบทความวิจัยที่เกี่ยวข้องผู้วิจัยจึงได้แนวคิดในการสร้างแบบจำลองทางคณิตศาสตร์จากบทความวิจัยของ Jeenpun (2013) และ Mahikul et al. (2019) ผู้วิจัยจึงสร้างแบบจำลองทางคณิตศาสตร์การระบาดของโรคเมลิออยโดสิสจากสัตว์สู่คน โดยแบ่งกลุ่มประชากรออกเป็น 2 กลุ่มคือ กลุ่มประชากรคนแสดงในรูปแบบจำลอง SEIRS ประชากรสัตว์แสดงในรูปแบบจำลอง SI และเนื่องจากโรคเมลิออยด์โดสิสเป็นโรคที่มักเกิดในช่วงฤดูฝนและจากการศึกษางานวิจัยของ Tanwisaid & Surin (2008) ผู้วิจัยจึงสนใจที่จะเพิ่มปัจจัยปริมาณน้ำฝนในกลุ่มประชากรคนเพื่อต้องการวิเคราะห์ปัจจัยปริมาณน้ำฝนที่ส่งผลกระทบต่อการแพร่ระบาดของโรคเมลิออยโดสิสกับสัดส่วนประชากรคนที่ได้รับเชื้อและแสดงอาการ ซึ่งในบทความวิจัยได้หาจุดสมดุลภายใต้สภาวะไร้โรคและสภาวะระบาดเรื้อรัง ค่าสืบพันธุ์พื้นฐาน(R_0) โดย $R_0 < 1$ จะแสดงถึงแนวโน้มเกิดการระบาดในกลุ่มประชากรลดลง และ $R_0 > 1$ จะแสดงถึงแนวโน้มเกิดการระบาดในกลุ่มประชากรเพิ่มขึ้น วิเคราะห์ความเสถียรภาพเฉพาะที่เชิงเส้นกำกับ(Boonpikum & Chinviratit, 2015; Yosyingyong & Viriyapong, 2017) และความเสถียรภาพวงกว้างเชิงเส้นกำกับ(Yongzhen, et al., 2018) ที่สภาวะไร้โรคและสภาวะระบาดเรื้อรัง

วิธีการทดลอง

1. การสร้างแบบจำลองทางคณิตศาสตร์

ในบทความวิจัยได้สร้างแบบจำลองทางคณิตศาสตร์การระบาดของโรคเมลิออยโดสิสจากสัตว์สู่คน โดยแบ่งกลุ่มประชากรออกเป็น 2 กลุ่ม คือ กลุ่มประชากรคน และประชากรสัตว์ ซึ่งมีขนาดคงที่โดยกลุ่มประชากรคนจะแบ่งออกเป็น 4 กลุ่มย่อยคือ กลุ่มประชากรคนที่เสี่ยงต่อการติดเชื้อ $S(t)$ กลุ่มประชากรคนที่ได้รับเชื้อแต่ยังไม่แสดงอาการ $E(t)$ กลุ่มประชากรคนที่ได้รับเชื้อและแสดงอาการ $I(t)$ และกลุ่มประชากรคนที่ฟื้นไข้ $R(t)$ กลุ่มประชากรสัตว์จะแบ่งออกเป็น 2 กลุ่มย่อย คือ กลุ่มประชากรสัตว์ที่เสี่ยงต่อการติดเชื้อ $S_a(t)$ และกลุ่มประชากรสัตว์ที่ติดเชื้อและสามารถแพร่เชื้อได้ $I_a(t)$ โดยในแบบจำลองสามารถอธิบายได้ดังนี้



รูปที่ 1 แนวคิดการสร้างแบบจำลองทางคณิตศาสตร์การระบาดของโรคเมลิออยโดสิสจากสัตว์สู่คน

โดยที่ β เป็นอัตราการแพร่เชื้อในประชากรคน r เป็นปริมาณน้ำฝนโดยที่ $r \in [0.000001, 50]$ α เป็นอัตราการฟักตัวในประชากรคน ϵ เป็นอัตราการฟื้นไข้ของประชากรคน p เป็นความน่าจะเป็นที่จะกลับมาเป็นประชากรคนกลุ่มเสี่ยงต่อการติดเชื้อ b เป็นอัตราการเกิดของประชากรคน d เป็นอัตราการตายของประชากรคน N เป็นจำนวนประชากรคนทั้งหมด d_1 เป็นอัตราการตายจากโรคเมลิออยโดสิสในประชากรคน β_1 เป็นอัตราการแพร่เชื้อในประชากรสัตว์ μ เป็นอัตราการเกิดและการตายของประชากรสัตว์ และ N_a เป็นจำนวนประชากรสัตว์ทั้งหมดซึ่งจากรูปที่ 1 สามารถสร้างแบบจำลองทางคณิตศาสตร์ในรูประบบสมการเชิงอนุพันธ์ดังนี้

กลุ่มประชากรคน

$$\frac{dS}{dt} = bN - \beta S I_a r + pR - dS \quad (1)$$

$$\frac{dE}{dt} = \beta S I_a r - dE - \alpha E \quad (2)$$

$$\frac{dI}{dt} = \alpha E - (d + d_1)I - \varepsilon I \quad (3)$$

$$\frac{dR}{dt} = \varepsilon I - dR - pR \quad (4)$$

$$\text{โดยที่ } N = S + E + I + R$$

กลุ่มประชากรสัตว์

$$\frac{dS_a}{dt} = \mu N_a - \mu S_a - \beta_1 S_a I_a \quad (5)$$

$$\frac{dI_a}{dt} = \beta_1 S_a I_a - \mu I_a \quad (6)$$

$$\text{โดยที่ } N_a = S_a + I_a$$

จากระบบสมการ (1) - (6) กำหนดให้อยู่รูปสัดส่วนต่อไปนี้

$$\bar{S} = \frac{S}{N}, \quad \bar{E} = \frac{E}{N}, \quad \bar{I} = \frac{I}{N}, \quad \bar{R} = \frac{R}{N} \text{ และ } \bar{S}_a = \frac{S_a}{N_a}, \quad \bar{I}_a = \frac{I_a}{N_a} \text{ ซึ่งลดรูปสมการ (1) - (6) ได้ดังนี้}$$

$$\frac{d\bar{S}}{dt} = b - \beta \bar{S} \bar{I}_a N_a r + p(1 - \bar{S} - \bar{E} - \bar{I}) - d\bar{S} \quad (7)$$

$$\frac{d\bar{E}}{dt} = \beta \bar{S} \bar{I}_a N_a r - (d + \alpha) \bar{E} \quad (8)$$

$$\frac{d\bar{I}}{dt} = \alpha \bar{E} - ((d + d_1) + \varepsilon) \bar{I} \quad (9)$$

$$\frac{d\bar{I}_a}{dt} = \beta_1 (1 - \bar{I}_a) \bar{I}_a N_a - \mu \bar{I}_a \quad (10)$$

โดยที่ $\bar{S} + \bar{E} + \bar{I} + \bar{R} = 1$ และ $\bar{S}_a + \bar{I}_a = 1$ จากระบบสมการ (7) - (10) จะมีเซตย่อยไม่แปรผันที่ไม่เป็นลบ โดยกำหนดให้ $\Omega = \{(\bar{S}, \bar{E}, \bar{I}, \bar{I}_a) \in \mathbb{R}_+^4 | 0 \leq \bar{S} \leq 1, 0 \leq \bar{E} \leq 1, 0 \leq \bar{I} \leq 1, 0 \leq \bar{I}_a \leq 1\}$

ผลการทดลอง

1. การวิเคราะห์จุดสมดุล

ในการหาจุดสมดุลจะกำหนดให้จุดสมดุลอยู่ในรูป $(\bar{S}^*, \bar{E}^*, \bar{I}^*, \bar{I}_a^*)$ โดยการนำสมการ (7) - (10) จับเท่ากับ ศูนย์ซึ่งจะทำให้ได้จุดสมดุล 2 จุด คือ จุดสมดุลภายในได้สภาวะไร้โรค $E_0 = \left(\frac{b+p}{p+d}, 0, 0, 0\right)$ และจุดสมดุลภายในได้สภาวะระบาดเรื้อรัง $E_1 = (\bar{S}^*, \bar{E}^*, \bar{I}^*, \bar{I}_a^*)$ โดยที่

$$\bar{S}^* = \frac{b+p-\bar{E}^*p-\bar{I}^*p}{d+\beta\bar{I}_a^*N_a+p}, \quad \bar{E}^* = \frac{\beta\bar{I}_a^*N_a\bar{S}^*}{d+\alpha}, \quad \bar{I}^* = \frac{\alpha\bar{E}^*}{d+d_1+\varepsilon} \text{ และ } \bar{I}_a^* = \frac{\beta_1N_a-\mu}{\beta_1N_a}$$

พิจารณาเมื่อ $\bar{I}_a^* > 0$ จะได้ว่า $R_0 > 1$ และ $\bar{S}^* > 0$ โดยที่ $R_0 = \frac{\beta_1 N_a}{\mu}$

2. การวิเคราะห์ความเสถียรภาพของจุดสมดุล

การวิเคราะห์หาความเสถียรภาพของจุดสมดุลจะพิจารณาจากค่าลักษณะเฉพาะของเมทริกซ์จาโคเบียนของระบบสมการ (7) – (10) ซึ่งสามารถเขียนให้อยู่ในรูปเมทริกซ์จาโคเบียนได้ดังนี้

$$J = \begin{bmatrix} -\beta_1 \bar{I}_a^* N_a r - p - d & -p & -p & -\beta_1 \bar{S} N_a r \\ \beta_1 \bar{I}_a^* N_a r & -(d + \alpha) & 0 & \beta_1 \bar{S} N_a r \\ 0 & \alpha & -((d + d_1) + \varepsilon) & 0 \\ 0 & 0 & 0 & \beta_1 N_a - 2\beta_1 \bar{I}_a^* N_a - \mu \end{bmatrix} \begin{pmatrix} \bar{S}^* \\ \bar{E}^* \\ \bar{I}^* \\ \bar{I}_a^* \end{pmatrix}$$

โดยสามารถหาค่าลักษณะเฉพาะได้จากสมการลักษณะเฉพาะ $\det(J - \lambda I) = 0$ เมื่อ λ เป็นค่าลักษณะเฉพาะ และ I เป็นเมทริกซ์เอกลักษณ์ขนาด $n \times n$ และได้ศึกษาเงื่อนไขที่ทำให้จุดสมดุลภายใต้สภาวะไร้โรคมีความเสถียรภาพเฉพาะที่เชิงเส้นกำกับดังนี้

ทฤษฎีบท 1 ถ้า $R_0 < 1$ แล้วจุดสมดุลภายใต้สภาวะไร้โรค E_0 จะเป็นจุดสมดุลที่มีเสถียรภาพเฉพาะที่เชิงเส้นกำกับ (locally asymptotically stable) ถ้า $R_0 > 1$ จุดสมดุลภายใต้สภาวะไร้โรคจะไม่มีเสถียรภาพ(unstable)

วิเคราะห์ความเสถียรภาพจุดสมดุลภายใต้สภาวะไร้โรค $E_0 = \left(\frac{b+p}{p+d}, 0, 0, 0\right)$ จะได้ค่าลักษณะเฉพาะคือ

$\lambda_1 = -d - p$ และสมการลักษณะเฉพาะ $\lambda^3 + a_1 \lambda^2 + a_2 \lambda + a_3 = 0$ ซึ่งจะได้ว่า

$$\begin{aligned} a_1 &= 2d + d_1 + \varepsilon + \alpha - \beta_1 N_a + \mu \\ &= 2d + d_1 + \varepsilon + \alpha + \mu(1 - R_0) \end{aligned}$$

$$\begin{aligned} a_2 &= d^2 + \varepsilon\alpha - \beta_1 \varepsilon N_a - \beta_1 \alpha N_a + \varepsilon\mu + \alpha\mu + d_1(\alpha - \beta_1 N_a + \mu) \\ &\quad + d(d_1 + \varepsilon + \alpha - 2\beta_1 N_a + 2\mu) \\ &= d^2 + \varepsilon\alpha + d_1\alpha + d(d_1 + \varepsilon + \alpha) + (d_1 + \alpha + \varepsilon + 2d)\mu(1 - R_0) \end{aligned}$$

$$\begin{aligned} a_3 &= (\mu - \beta_1 N_a)(d + d_1 + \varepsilon)(d + \alpha) \\ &= \mu(1 - R_0)(d + d_1 + \varepsilon)(d + \alpha) \end{aligned}$$

$$\begin{aligned} a_1 a_2 &= ((d + d_1 + \varepsilon)(d + \alpha)(2d + d_1 + \varepsilon + \alpha + (1 - R_0)\mu) + (2d + d_1 + \varepsilon + \alpha)(1 - R_0) \\ &\quad (2d + d_1 + \varepsilon + \alpha + (1 - R_0)\mu)) \end{aligned}$$

เพราะว่า $(d + d_1 + \varepsilon)(d + \alpha)(2d + d_1 + \varepsilon + \alpha + (1 - R_0)\mu) > \mu(1 - R_0)(d + d_1 + \varepsilon)(d + \alpha)$

จะได้ว่ารากสมการลักษณะเฉพาะ $\lambda_1 < 0$ และ $\lambda_2, \lambda_3, \lambda_4$ ซึ่งเป็นรากอันดับสามของสมการลักษณะเฉพาะซึ่งสอดคล้องกับเงื่อนไขไฮเร้าเซอร์วิท(Routh - Hurwitz condition) จากการพิจารณา $a_1 > 0, a_2 > 0$ และ $a_1 a_2 > a_3$ (Edelstein-Keshet, 1988; Banerjee, 2014) เมื่อ $R_0 < 1$ ดังนั้นทำให้ได้ว่าจุดสมดุลภายใต้สภาวะไร้โรค E_0 มีความเสถียรภาพเฉพาะที่เชิงเส้นกำกับ

และได้ศึกษาเงื่อนไขของจุดสมดุลภายใต้สภาวะไร้โรคและสภาวะระบาดเรื้อรังที่มีความเสถียรภาพกว้างเชิงเส้นกำกับดังบทนิยาม 1 และทฤษฎีบท 2 ต่อไปนี้

บทนิยาม 1 ให้ X^* เป็นจุดสมดุลของระบบ $X' = F(X)$ เมื่อ F เป็นฟังก์ชันต่อเนื่องและสามารถหาอนุพันธ์ได้ และฟังก์ชันค่าจริงต่อเนื่อง V บนย่านใกล้เคียงที่มีจุดสมดุล $X^* \in \Omega$ จะเป็นฟังก์ชันเลียปูนอฟ (Lyapunov function) ก็ต่อเมื่อ

1. $V > 0$ และ $V(X^*) = 0$
2. $\nabla_X V(X) \cdot F(X) \leq 0$ บน Ω

ทฤษฎีบท 2 ถ้า $R_0 < 1$ แล้วจุดสมดุลภายใต้สภาวะไร้โรค E_0 จะเป็นจุดสมดุลที่มีความเสถียรภาพกว้างเชิงเส้นกำกับ (globally asymptotically stable)

พิสูจน์ พิจารณาฟังก์ชัน $V_1 : \Omega \rightarrow \mathbb{R}_+$ โดยกำหนดให้ $V_1(t) = V_1(\bar{S}, \bar{E}, \bar{I}, \bar{I}_a) = \frac{\bar{I}_a + \bar{I}_a^2}{\mu}$ ซึ่งจะเห็นได้ชัดว่า $V_1(t) \geq 0$ ทุก ๆ $(\bar{S}, \bar{E}, \bar{I}, \bar{I}_a) \in \Omega$ พิจารณา $R_0 < 1$ และ $0 \leq \bar{I}_a \leq 1$ จะได้ว่า

$$\begin{aligned} V_1'(t) &= \left(\frac{1+2\bar{I}_a}{\mu} \right) (\beta_1(1-\bar{I}_a)\bar{I}_a N_a - \mu \bar{I}_a) \\ &= -\bar{I}_a(1+2\bar{I}_a)((1-R_0) + R_0 \bar{I}_a) \leq 0 \end{aligned}$$

ดังนั้นทำให้ได้ว่าฟังก์ชัน V_1 เป็นฟังก์ชันเลียปูนอฟ ก็ต่อเมื่อ $R_0 < 1$ และ $V_1'(t) \leq 0$ จะได้ว่าที่จุดสมดุลภายใต้สภาวะไร้โรค E_0 มีความเสถียรภาพกว้างเชิงเส้นกำกับ

ทฤษฎีบท 3 ถ้า $R_0 > 1$ แล้วจุดสมดุลภายใต้สภาวะระบาดเรื้อรัง E_1 จะเป็นจุดสมดุลที่มีความเสถียรภาพเฉพาะที่เชิงเส้นกำกับ (locally asymptotically stable)

วิเคราะห์ความเสถียรภาพจุดสมดุลภายใต้สภาวะระบาดเรื้อรัง $E_1 = (\bar{S}^*, \bar{E}^*, \bar{I}^*, \bar{I}_a^*)$ จะมีค่าลักษณะเฉพาะคือ

$$\begin{aligned} \lambda_1 &= -\beta_1 N_a + \mu \text{ เมื่อ } \mu < \beta_1 N_a \text{ และสมการลักษณะเฉพาะ } \lambda^3 + c_1 \lambda^2 + c_2 \lambda + c_3 = 0 \text{ โดยที่} \\ c_1 &= 3d + d_1 + \varepsilon + \alpha + \beta_1 N_a - \frac{\beta_1 \mu}{\beta_1} + p \end{aligned}$$

$$\begin{aligned}
 &= 3d + d_1 + \varepsilon + \alpha + p + \frac{\beta r \mu}{\beta_1} (R_0 - 1) \\
 c_2 &= 3d^2 + \varepsilon \alpha + \beta \varepsilon r N_a + \beta r \alpha N_a + \varepsilon p + \alpha p + \beta r N_a p + d_1 (\alpha + \beta r N_a + p) \\
 &\quad + 2d(d_1 + \varepsilon + \alpha + \beta r N_a + p) - \frac{\beta r \mu (2d + d_1 + \varepsilon + \alpha + p)}{\beta_1} \\
 &= 3d^2 + \varepsilon \alpha + \beta \varepsilon r N_a + \beta r \alpha N_a + \varepsilon p + \alpha p + \beta r N_a p + d_1 (\alpha + \beta r N_a + p) \\
 &\quad + 2d(d_1 + \varepsilon + \alpha + p) + \left(\frac{2d}{(2d + d_1 + \varepsilon + \alpha + p)} (R_0 - 1) \right) \left(\frac{\beta r \mu (2d + d_1 + \varepsilon + \alpha + p)}{\beta_1} \right) \\
 c_3 &= d^3 + (d_1 + \varepsilon) \alpha p + d^2 (d_1 + \varepsilon + \alpha + \beta r N_a + p) + \beta r N_a \alpha p \\
 &\quad + \beta r N_a (d_1 + \varepsilon) (\alpha + p) + d (\beta r \alpha N_a + \alpha p + \beta r N_a p + d_1 (\alpha + \beta r N_a + p)) \\
 &\quad + d \varepsilon (\alpha + \beta r N_a + p) - \frac{\beta r \mu (d^2 + \varepsilon \alpha + \varepsilon p + \alpha p + d_1 (\alpha + p) + d (d_1 + \varepsilon + \alpha + p))}{\beta_1} \\
 &= d^3 + \alpha p (d_1 + \varepsilon) + d^2 (d_1 + \varepsilon + \alpha + p) + d (\alpha p + (d_1 + \varepsilon) (\alpha + p)) \\
 &\quad + \frac{\beta r \mu (d^2 + \varepsilon \alpha + \varepsilon p + \alpha p + d_1 (\alpha + p) + d (d_1 + \varepsilon + \alpha + p))}{\beta_1} (R_0 - 1) \\
 c_1 c_2 &= 9d^3 + 3\alpha p (d_1 + \varepsilon) + 9d^2 (d_1 + \varepsilon + \alpha + p) + 7d (\alpha p + (d_1 + \varepsilon) (\alpha + p)) \\
 &\quad + d_1 ((2d + p)(d_1 + 2\varepsilon) + \alpha(d_1 + 2\varepsilon + \alpha) + p^2) + \alpha(\varepsilon^2 + (d_1 + \varepsilon)\alpha \\
 &\quad + (\varepsilon^2 + \alpha^2)p + (\varepsilon + \alpha)p^2 + 2d(\varepsilon^2 + \alpha^2 + p^2) + \beta r N_a (9d^2 + d_1^2 + 2d_1\varepsilon + \varepsilon^2 \\
 &\quad + 3d_1\alpha + 3\varepsilon\alpha + \alpha^2 + 3(d_1 + \varepsilon + \alpha)p + p^2 + 7d(d_1 + \varepsilon + \alpha + p) + \beta r N_a (2d \\
 &\quad + d_1 + \varepsilon + \alpha + p)) - \frac{\beta r \mu (9d^2 + 3\varepsilon\alpha + 3\varepsilon p + 3\alpha p + 3d_1(\alpha + p) + 7d(d_1 + \varepsilon + \alpha + p))}{\beta_1} \\
 &\quad - \frac{\beta r \mu (d_1^2 + 2d_1\varepsilon + \varepsilon^2 + \alpha^2 + p^2)}{\beta_1} - \frac{2\beta^2 r^2 \mu N_a (2d + d_1 + \varepsilon + \alpha + p)}{\beta_1} + \frac{\beta^2 r^2 \mu^2 (2d + d_1 + \varepsilon + \alpha + p)}{\beta_1} \\
 &= 9d^3 + 3\alpha p (d_1 + \varepsilon) + 9d^2 (d_1 + \varepsilon + \alpha + p) + 7d (\alpha p + (d_1 + \varepsilon) (\alpha + p)) \\
 &\quad + d_1 ((2d + p)(d_1 + 2\varepsilon) + \alpha(d_1 + 2\varepsilon + \alpha) + p^2) + \alpha(\varepsilon^2 + (d_1 + \varepsilon)\alpha \\
 &\quad + (\varepsilon^2 + \alpha^2)p + (\varepsilon + \alpha)p^2 + 2d(\varepsilon^2 + \alpha^2 + p^2) + \beta^2 r^2 N_a^2 (2d + d_1 + \varepsilon + \alpha + p) \\
 &\quad - \frac{2\beta^2 r^2 \mu N_a (2d + d_1 + \varepsilon + \alpha + p)}{\beta_1} + \beta r N_a (d_1^2 + 2d_1\varepsilon + \varepsilon^2 + \alpha^2 + p^2) \\
 &\quad - \frac{\beta r \mu (d_1^2 + 2d_1\varepsilon + \varepsilon^2 + \alpha^2 + p^2)}{\beta_1} + \beta r N_a (9d^2 + 3\varepsilon\alpha + 3\varepsilon p + 3\alpha p + 3d_1) \\
 &\quad + 7d(d_1 + \varepsilon + \alpha + p) - \frac{\beta r \mu (9d^2 + 3\varepsilon\alpha + 3\varepsilon p + 3\alpha p + 3d_1(\alpha + p) + 7d(d_1 + \varepsilon + \alpha + p))}{\beta_1}
 \end{aligned}$$

$$\begin{aligned}
 & + \frac{\beta^2 r^2 \mu^2 (2d + d_1 + \varepsilon + \alpha + p)}{\beta_1} \\
 & = 9d^3 + 3\alpha p(d_1 + \varepsilon) + 9d^2(d_1 + \varepsilon + \alpha + p) + 7d(\alpha p + (d_1 + \varepsilon)(\alpha + p)) \\
 & \quad + d_1((2d + p)(d_1 + 2\varepsilon) + \alpha(d_1 + 2\varepsilon + \alpha) + p^2) + \alpha(\varepsilon^2 + (d_1 + \varepsilon)\alpha + (\varepsilon^2 + \alpha^2)p + \\
 & \quad (\varepsilon + \alpha)p^2) \\
 & \quad + 2d(\varepsilon^2 + \alpha^2 + p^2) + (R_0 - 1) \left(\frac{2\beta^2 r^2 \mu N_a (2d + d_1 + \varepsilon + \alpha + p)}{\beta_1} \right) \\
 & \quad + \frac{\beta r \mu (d_1^2 + 2d_1 \varepsilon + \varepsilon^2 + \alpha^2 + p^2)}{\beta_1} (R_0 - 1) + \frac{\beta^2 r^2 \mu^2 (2d + d_1 + \varepsilon + \alpha + p)}{\beta_1} \\
 & \quad + \frac{\beta r \mu (9d^2 + 3\varepsilon \alpha + 3\varepsilon p + 3\alpha p + 3d_1(\alpha + p) + 7d(d_1 + \varepsilon + \alpha + p))}{\beta_1} (R_0 - 1)
 \end{aligned}$$

เพราะว่า $9d^3 + 3\alpha p(d_1 + \varepsilon) + 9d^2(d_1 + \varepsilon + \alpha + p) + 7d(\alpha p + (d_1 + \varepsilon)(\alpha + p))$

$$\begin{aligned}
 & + \frac{\beta r \mu (9d^2 + 3\varepsilon \alpha + 3\varepsilon p + 3\alpha p + 3d_1(\alpha + p) + 7d(d_1 + \varepsilon + \alpha + p))}{\beta_1} (R_0 - 1) > d^3 + \alpha p(d_1 + \varepsilon) \\
 & + d^2(d_1 + \varepsilon + \alpha + p) + d(\alpha p + (d_1 + \varepsilon)(\alpha + p)) \\
 & + \frac{\beta r \mu (d^2 + \varepsilon \alpha + \varepsilon p + \alpha p + d_1(\alpha + p) + d(d_1 + \varepsilon + \alpha + p))}{\beta_1} (R_0 - 1)
 \end{aligned}$$

จะได้ว่ารากสมการลักษณะเฉพาะ $\lambda_1 < 0$ และ $\lambda_2, \lambda_3, \lambda_4$ ซึ่งเป็นรากอันดับสามของสมการลักษณะเฉพาะซึ่งสอดคล้องกับเงื่อนไขเร้าฮอร์วิท จากการพิจารณา $c_1 > 0, c_2 > 0$ และ $c_1 c_2 > c_3$ เมื่อ $R_0 > 1$ ดังนั้นทำให้ได้ว่าจุดสมดุลภายใต้สภาวะระบาดเรื้อรัง E_1 มีความเสถียรภาพเฉพาะที่เชิงเส้นกำกับ

และได้ศึกษาเงื่อนไขของจุดสมดุลภายใต้สภาวะระบาดเรื้อรังที่มีความเสถียรภาพวงกว้างเชิงเส้นกำกับดังบทนิยาม 1 และทฤษฎีบท 4 ดังนี้

ทฤษฎีบท 4 ถ้า $R_0 > 1$ แล้วจุดสมดุลภายใต้สภาวะระบาดเรื้อรัง E_1 จะเป็นจุดสมดุลที่มีเสถียรภาพวงกว้างเชิงเส้นกำกับ (globally asymptotically stable)

พิสูจน์ พิจารณาฟังก์ชัน $V_2 : \Omega \rightarrow \mathbb{R}_+$ โดยกำหนดให้ $V_2(t) = V_2(\bar{S}, \bar{E}, \bar{I}, \bar{I}_a) = \bar{I}_a - \bar{I}_a^* - \ln \frac{\bar{I}_a}{\bar{I}_a^*}$ ซึ่งจะเห็นได้ชัดว่า $V_2(t) \geq 0$ ทุก ๆ $(\bar{S}, \bar{E}, \bar{I}, \bar{I}_a) \in \Omega$ เมื่อพิจารณา $R_0 > 1$ และ $0 \leq \bar{I}_a \leq 1$ จะได้ว่า

$$\begin{aligned}
 V_2'(t) & = \left(\frac{\bar{I}_a - 1}{\bar{I}_a} \right) (\beta_1 (1 - \bar{I}_a) \bar{I}_a N_a - \mu \bar{I}_a) \\
 & = \mu (1 - \bar{I}_a) (1 - R_0 (1 - \bar{I}_a)) \leq 0
 \end{aligned}$$

ดังนั้นทำให้ได้ว่าฟังก์ชัน V_2 เป็นฟังก์ชันเลียปูนอฟ ก็ต่อเมื่อ $0 \leq \bar{I}_a \leq 1$ และ $R_0 > 1$ ซึ่งทำให้ $V_2'(t) \leq 0$ จะได้ว่าที่จุดสมดุลภายใต้สภาวะระบาดเรื้อรัง E_1 มีความเสถียรภาพวงกว้างเชิงเส้นกำกับ

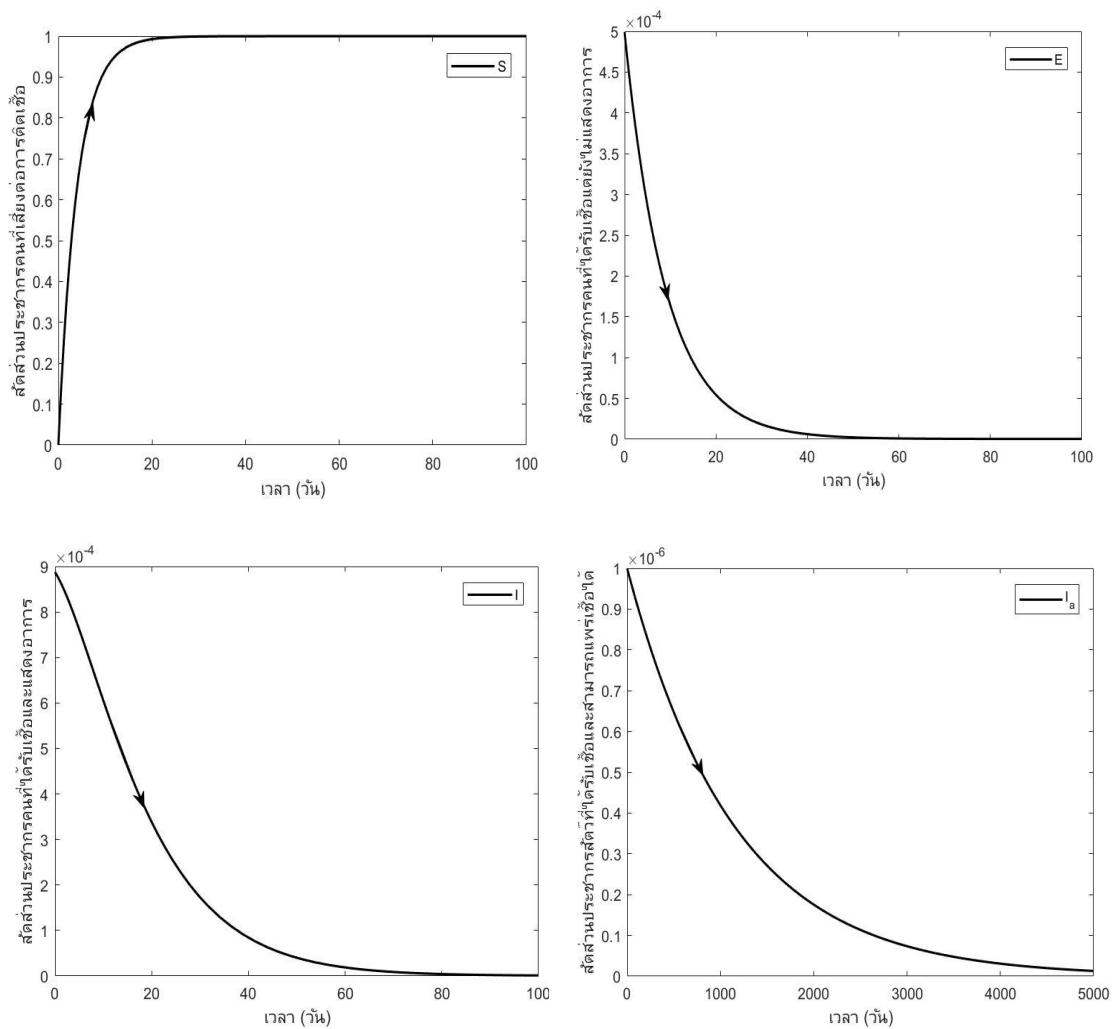
3. การวิเคราะห์เชิงตัวเลข

ในบทความวิจัยผู้วิจัยได้ศึกษาข้อมูลการระบาดของโรคเมลิออยโดสิสในสัตว์และคน (Jeenpun, 2013; Suwanchairob, 2015; Sumpuntharat et al., 2014; Suchatsoonthorn & Sealim, 2019; Muangjaiphet & Suggaravetsiri, 2019) และนำค่าพารามิเตอร์ต่าง ๆ มาใช้ในการวิเคราะห์เชิงตัวเลขซึ่งแสดงดังตารางต่อไปนี้

ตารางที่ 1. ค่าพารามิเตอร์ต่าง ๆ ของโรคเมลิออยโดสิส

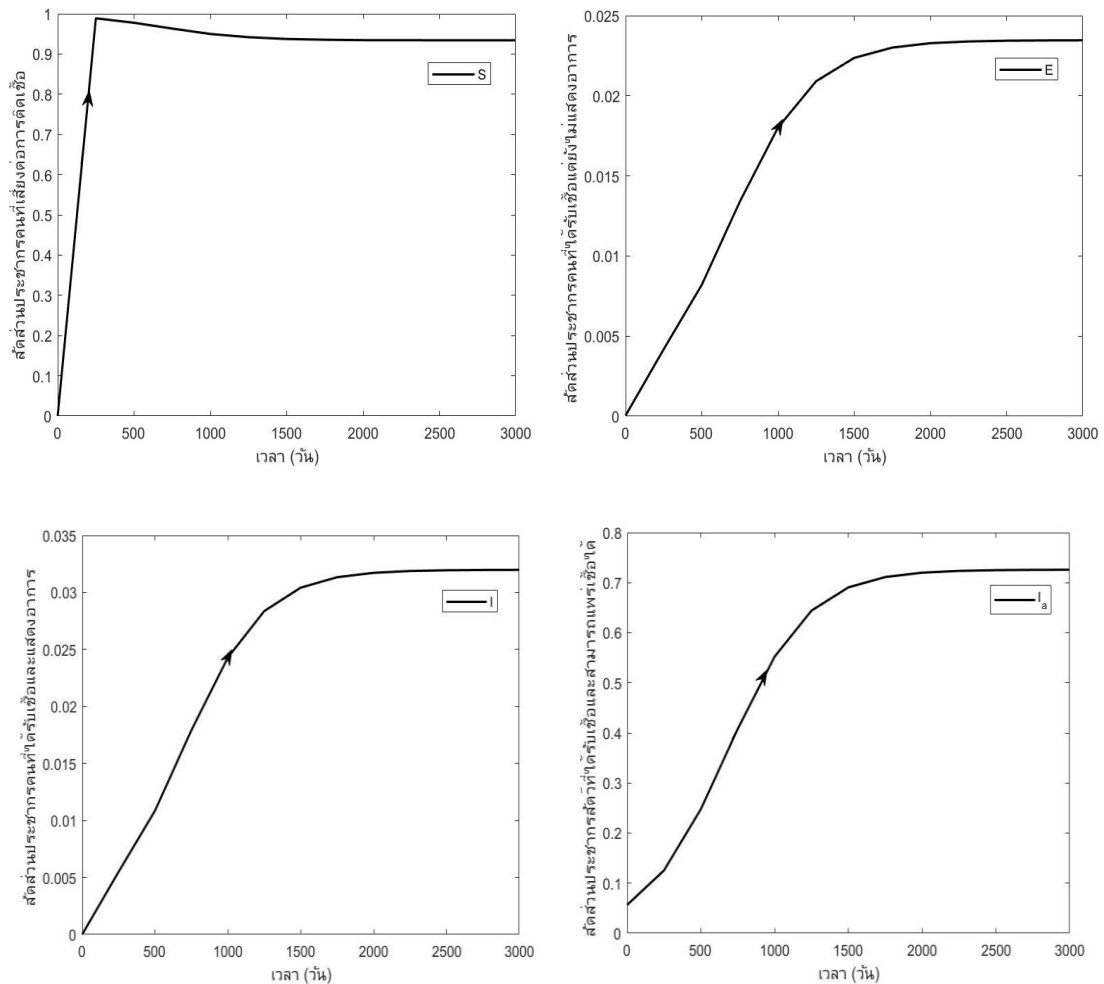
คำอธิบาย	สัญลักษณ์	ค่าพารามิเตอร์	หน่วย
อัตราการเกิดของประชากรคน	b	$1/(365 \times 75.5)$	ต่อคน ต่อวัน
อัตราการตายของประชากรคน	d	$1/(365 \times 75.5)$	ต่อคน ต่อวัน
อัตราการตายจากโรคเมลิออยโดสิสในประชากรคน	d_1	0.01	ต่อคน
ปริมาณน้ำฝน	r	0.001	มิลลิเมตร
อัตราการแพร่เชื้อในประชากรคน	β	$1/13$	ต่อวัน
อัตราการฟักตัวในประชากรคน	α	$1/9$	ต่อวัน
อัตราการฟื้นไข้ของประชากรคน	ε	$1/14$	ต่อวัน
ความน่าจะเป็นที่จะกลับมาเป็นประชากรคนกลุ่มเสี่ยงต่อการติดเชื้อ	p	0.25	
อัตราการแพร่เชื้อในประชากรสัตว์	β_1	0.0001	ต่อวัน
อัตราการเกิดและการตายของประชากรสัตว์	μ	$1/(365 \times 2)$	ต่อตัว ต่อวัน
สถานะไร้โรค			
จำนวนประชากรคนทั้งหมด	N	10	คน
จำนวนประชากรสัตว์ทั้งหมด	N_a	5	ตัว
สภาวะระบาดเรื้อรัง			
จำนวนประชากรคนทั้งหมด	N	100	คน
จำนวนประชากรสัตว์ทั้งหมด	N_a	50	ตัว

พิจารณาที่สถานะไร้โรคสัดส่วนประชากรที่จุดสมดุลภายใต้สถานะไร้โรค E_0 จะแสดงดังนี้



รูปที่ 2 กราฟแสดงสัดส่วนประชากรที่เสี่ยงต่อการติดเชื้อ ประชากรที่ได้รับเชื้อแต่ยังไม่แสดงอาการ ประชากรที่ได้รับเชื้อและแสดงอาการ และประชากรที่หายดีและสามารถแพร่เชื้อได้ในสภาวะไร้โรค จากรูปที่ 2 แสดงให้เห็นได้ว่าผลเฉลยเชิงตัวเลขจะเข้าสู่จุดสมดุลที่ $E_0 = (1, 0, 0, 0)$ โดยที่ $R_0 = 0.365$

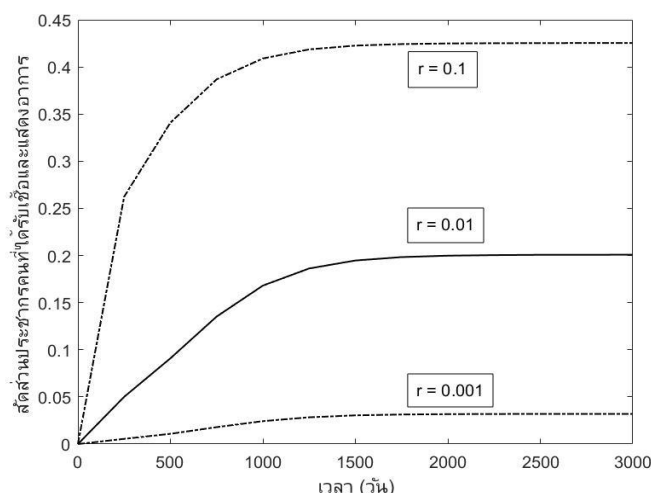
พิจารณาที่สภาวะระบาดเรื้อรังสัดส่วนประชากรที่จุดสมดุลภายใต้สภาวะระบาดเรื้อรัง E_1 จะแสดงดังนี้



รูปที่ 3 กราฟแสดงสัดส่วนประชากรคนที่เสี่ยงต่อการติดเชื้อ ประชากรคนที่ได้รับเชื้อแต่ยังไม่แสดงอาการ ประชากรคนที่ได้รับเชื้อและแสดงอาการ และประชากรสัตว์ที่ได้รับเชื้อและสามารถแพร่เชื้อได้
ในสภาวะระบาดเรื้อรัง

จากรูปที่ 3 แสดงให้เห็นได้ว่าผลเฉลยเชิงตัวเลขจะเข้าสู่จุดสมดุลที่ $E_1 = (0.9341, 0.0235, 0.0320, 0.7260)$ โดยที่ $R_0 = 3.65$

พิจารณาระดับปริมาณน้ำฝนที่ส่งผลกระทบต่อการแพร่ระบาดของโรคเมลิออยโดสิสกับสัดส่วนประชากรคนที่ได้รับเชื้อและแสดงอาการซึ่งจะแสดงดังนี้



รูปที่ 4 กราฟแสดงการเปรียบเทียบปัจจัยระดับปริมาณน้ำฝนที่ส่งผลกระทบต่อการแพร่ระบาดของโรคเมลิออยโดสิสกับสัดส่วนประชากรคนที่ได้รับเชื้อและแสดงอาการ

จากรูปที่ 4 ได้จำลองค่าปริมาณน้ำฝนโดยเพิ่มระดับปริมาณน้ำฝนเป็น 0.001, 0.01 และ 0.1 มิลลิเมตรตามลำดับ และคำนวณผลเฉลยเชิงตัวเลขทำให้ได้ว่าที่ระดับปริมาณน้ำฝนที่มากขึ้นจะส่งผลให้เกิดการระบาดของโรคเพิ่มขึ้น โดยจากการวิเคราะห์สัดส่วนประชากรคนที่ได้รับเชื้อและแสดงอาการที่มีค่าเพิ่มขึ้นเป็น 0.0320 , 0.2008 และ 0.4253 ตามลำดับ ซึ่งจากผลลัพธ์ดังกล่าวจะสอดคล้องกับบทความวิจัยของ Tanwisaid & Surin (2008) ที่ศึกษาเกี่ยวกับปริมาณน้ำฝนรายเดือนและความรุนแรงของโรคเมลิออยโดสิสของจังหวัดนครพนม และได้ข้อสรุปว่าปริมาณน้ำฝนในช่วงฤดูฝนส่งผลให้เกิดความชุกของการเกิดโรคเมลิออยโดสิส ดังนั้นจึงเป็นเหตุผลหนึ่งในการสนับสนุนว่าปัจจัยปริมาณน้ำฝนส่งผลกระทบต่อการระบาดของโรค

วิจารณ์และสรุปผลการทดลอง

ในบทความวิจัยสร้างแบบจำลองทางคณิตศาสตร์การระบาดของโรคเมลิออยโดสิสจากสัตว์สู่คนโดยพิจารณาจากผลกระทบปัจจัยปริมาณน้ำฝนซึ่งกำหนดให้จำนวนประชากรคนและสัตว์มีขนาดคงที่ ในบทความวิจัยได้แบ่งกลุ่มประชากรออกเป็น 2 กลุ่ม คือ กลุ่มประชากรคนแสดงในรูปแบบจำลอง SEIRS ประชากรสัตว์แสดงในรูปแบบจำลอง SI โดยจากแบบจำลองได้พิจารณาจุดสมดุลภายใต้สภาวะไร้โรค E_0 และสภาวะระบาดเรื้อรัง E_1 และค่าสืบพันธุ์พื้นฐาน $R_0 = \frac{\beta_1 N_a}{\mu}$

ผลการวิจัยพบว่าผลลัพธ์เชิงทฤษฎีในการวิเคราะห์ความเสถียรภาพที่จุดสมดุลภายใต้สภาวะไร้โรค E_0 สอดคล้องกับเงื่อนไขเร้าเซอร์วิท เมื่อ $R_0 < 1$ ทำให้ได้ว่าจุดสมดุลภายใต้สภาวะไร้โรคมีความเสถียรภาพเฉพาะที่เชิงเส้นกำกับ และผลลัพธ์การวิเคราะห์เชิงตัวเลขเข้าสู่จุดสมดุล $E_0 = (1,0,0,0)$ โดยที่ $R_0 = 0.365$ จุดสมดุลภายใต้สภาวะระบาดเรื้อรัง E_1 สอดคล้องกับเงื่อนไขเร้าเซอร์วิท เมื่อ $R_0 > 1$ ทำให้ได้ว่าจุดสมดุลภายใต้สภาวะระบาดเรื้อรังมีความเสถียรภาพเฉพาะที่เชิงเส้นกำกับ และผลลัพธ์การวิเคราะห์เชิงตัวเลขเข้าสู่จุดสมดุล $E_1 = (0.9341, 0.0235, 0.0320, 0.7260)$ โดยที่ $R_0 = 3.65$ และได้วิเคราะห์ความเสถียรภาพวงกว้างเชิงเส้นกำกับโดยผลลัพธ์ที่ได้สอดคล้องกับทฤษฎีของ Lyapunov – Lasalle

จากผลการวิเคราะห์การจำลองปริมาณน้ำฝนโดยเพิ่มระดับปริมาณน้ำฝนเป็น 0.001, 0.01 และ 0.1 มิลลิเมตรตามลำดับ และคำนวณผลเฉลยเชิงตัวเลขทำให้ได้ว่าที่ระดับปริมาณน้ำฝนที่มากขึ้นจะส่งผลให้เกิดการระบาดของโรคเพิ่มขึ้น โดยจากการวิเคราะห์สัดส่วนประชากรคนที่ได้รับเชื้อและแสดงอาการพบว่าที่ประชากรคนเพียง 1 คนมีโอกาสที่จะได้รับเชื้อเพิ่มขึ้นโดยคิดเป็นร้อยละ 3.2 , 20.08 และ 42.53 ตามลำดับ ซึ่งทำให้ได้ข้อสรุปว่าปริมาณน้ำฝนในช่วงฤดูฝนสามารถส่งผลให้เกิดความชุกของการเกิดโรคเมลิออยโดสิสได้ จากผลของการวิเคราะห์ปัจจัยปริมาณน้ำฝนของบทความวิจัยในครั้งนี้จะเป็นประโยชน์ทำให้เกิดการเฝ้าระวังโรคเมลิออยโดสิสในช่วงฤดูฝน ซึ่งจะนำไปสู่การวางแผนเตรียมการและความคุมเพื่อลดการระบาดของโรคเมลิออยโดสิสในอนาคตต่อไป

อย่างไรก็ตามในการสร้างแบบจำลองทางคณิตศาสตร์การระบาดของโรคเมลิออยโดสิสจากสัตว์สู่คน โดยพิจารณาจากผลกระทบปัจจัยปริมาณน้ำฝน เนื่องจากการวิเคราะห์ปัจจัยปริมาณน้ำฝนที่ส่งผลกระทบต่อสัดส่วนประชากรคนที่ได้รับเชื้อและแสดงอาการเป็นการวิเคราะห์จากการคำนวณผลเฉลยเชิงตัวเลข ดังนั้นในบทความวิจัยครั้งต่อไปควรเลือกใช้วิธีการวิเคราะห์ความอ่อนไหวเพื่อให้ทราบว่าพารามิเตอร์ใดส่งผลกระทบต่อการระบาดของโรคกับสัดส่วนประชากรคนที่ได้รับเชื้อและแสดงอาการมากที่สุด และจากปริมาณน้ำฝนที่ใช้เป็นค่าที่เกิดจากการจำลองไม่ได้เกิดจากการวัดค่าจากข้อมูลจริงจึงอาจทำให้เกิดความคลาดเคลื่อนในการคำนวณประชากรคนที่มีโอกาสเกิดโรคเมลิออยโดสิสได้ ซึ่งในบทความวิจัยครั้งต่อไปสามารถนำข้อมูลปริมาณน้ำฝนจริงมาใช้ในการวิเคราะห์เพื่อให้การคำนวณหรือการทำนายจำนวนประชากรคนที่จะเกิดโรคเมลิออยโดสิสได้สอดคล้องมากขึ้น

เอกสารอ้างอิง

- Banerjee, S. (2014). *Mathematical Modeling: Models, Analysis and Applications*. London: Taylor & Francis.
- Boonpikun, A. & Chinvirasit, W. (2015). *Stability Analysis of an SVIR Epidemic Model with Nonlinear Incidence Rate*. pp. 300 – 307. The 34th National Graduate Research Conference, 27 March 2015, Faculty of Graduate School, Khon Kaen University.

- Division of Innovation and Research (DIR.). (2019). Melioidosis. *Research plan on disease prevention and health control*, 22- 23
- Edelstein-Keshet, L. (1988). Applications of Continuous Models to Population Dynamics. *Mathematical Models in Biology*.pp. 233-234. New York: Random House Publisher.
- Egilmez, H.I. Morozov, A.Y. Clokie, M.R.J. Shan, J. Letarov, A. & Galyov, E.E. (2018). Temperature – dependent virus lifecycle choices may reveal and predict facets of the biology of opportunistic pathogenic bacteria. *Scientific reports*, 8:9642. Doi: 10.1038/s41598-018-27716-3
- Jeenpun, A. (2013). *Epidemiology of Burkholderia pseudomallei in Songkhla Zoo* (M.S. thesis) Department of Veterinary Public Health, Faculty of Graduate School, Kasetsart University, Bangkok.
- Mahikul, W., Lisa J. White, Poovorawan, K., Soonthornworasiri, N., Sukontamarn, P., Chanthavilay, P., et al. (2019). Modelling population dynamics and seasonal movement to assess and predict the burden of melioidosis. *PLOS Neglected Tropical Diseases* 3(5) , e0007380. <https://doi.org/10.1371/journal.pntd.0007380>
- Mangkala, J. (2009). Diagnosis of Melioidosis in musculoskeletal system. *Thai Journal of Rheumatology*. 20(4), 137 – 166.
- Muangjaiphet, P.& Suggaravetsiri, P. (2019). Prevalence and factors associated with Melioidosis in Ubon Ratchathani province. *The Office of Disease Prevention and Control 7 Khon Kaen*, 26(2), 1-13.
- Yongzhen, P., Xuehui, J., Changguo, L. & Shujing, G. (2018). Dynamics of a model of toxoplasmosis disease in cat and human with varying size populations. *Elsevier B.V. on behalf of International Associations for Mathematics and Computers in Simulation (IMACS)* 144(2018), 52 - 59
- Suchatsoonthorn, C.& Sealim, S. (2019). Thailand’s national and health development status in the world and asia. *Journal of Health Science*, 28(4), 577 – 590
- Sumpuntharat, S., Ruangjiratain, S.& Moopayak, K. (2014). Factors associated with re-infection preventive behaviors among people infected with melioid in northeastern region. *Journal of Nursing Science*, 32(2), 14 – 22.
- Suwanchairob, O. (2015). Melioidosis. *Annual Epidemiological Surveillance Report* 201,82 – 85.
- Tanwisaid, K.& Surin, U. (2008). Monthly rainfall and severity of melioidosis in Nakhon Phanom, Northeastern Thailand. *Journal of Health Science*, 17(3), 363 – 375.
- Yosyingyong, P. & Viriyapong, R. (2017). Mathematical modeling the effects of vaccination on HBV infection in Thailand. *Burapha Science Journal*, 22, 400-410.

Research Article

ตัวแบบพยากรณ์อัตราเร็วลมรายวัน ที่ระดับความสูง 120 เมตร

อำเภอปากพนัง จังหวัดนครศรีธรรมราช

Daily Wind Speed Forecast Model at an Altitude of 120 Meters, Pak Phanang District, Nakhon Si Thammarat Province

วรangkนา เรียนสุทธิ^{1*}

Warangkhan Riansut^{1*}

¹สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยทักษิณ อำเภอป่าพะยอม จังหวัดพัทลุง 93210

¹Department of Mathematics and Statistics, Faculty of Science, Thaksin University, Pa Phayom, Phatthalung 93210, Thailand

*E-mail: warang27@gmail.com

Received: 10/01/2020; Revised: 29/04/2020; Accepted: 10/05/2020

บทคัดย่อ

วัตถุประสงค์ของการศึกษานี้คือเพื่อสร้างตัวแบบพยากรณ์อัตราเร็วลมรายวัน ที่ระดับความสูง 120 เมตร อำเภอปากพนัง จังหวัดนครศรีธรรมราช โดยอาศัยวิธีการทางสถิติทั้งหมด 4 วิธี ได้แก่ วิธีบ็อกซ์-เจนกินส์ วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของโฮลต์ วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของบราวน์ และวิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังที่มีแนวโน้มแบบแฉก อัตราเร็วลมเฉลี่ยต่อวันที่ใช้ศึกษาได้มาจากศูนย์วิจัยพลังงานและสิ่งแวดล้อม มหาวิทยาลัยทักษิณ ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2555 ถึงวันที่ 30 มิถุนายน พ.ศ. 2558 จำนวน 1,277 วัน แบ่งข้อมูลออกเป็น 2 ชุด ชุดที่ 1 ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2555 ถึงวันที่ 31 พฤษภาคม พ.ศ. 2558 จำนวน 1,248 วัน สำหรับการสร้างตัวแบบพยากรณ์ ชุดที่ 2 เป็นข้อมูลวันที่ 1 ถึง 30 มิถุนายน พ.ศ. 2558 จำนวน 30 วัน สำหรับการเปรียบเทียบความเหมาะสมของตัวแบบพยากรณ์ โดยใช้เกณฑ์รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ผลการศึกษา พบว่า วิธีที่มีความถูกต้องมากที่สุดคือ วิธีบ็อกซ์-เจนกินส์ (RMSE = 1.5556) โดยมีตัวแบบพยากรณ์ คือ

$$\hat{Y}_t = 0.82429Y_{t-1} + 0.69963Y_{t-2} - 0.80909Y_{t-3} + 0.28517Y_{t-4} + 0.22656e_{t-1} - 0.77254e_{t-2}$$

คำสำคัญ: อัตราเร็วลมเฉลี่ยต่อวัน, ตัวแบบพยากรณ์, บ็อกซ์-เจนกินส์, การทำให้เรียบแบบเลขชี้กำลัง

Abstract

The objective of this study was to construct a daily wind speed forecasting model at an altitude of 120 meters, Pak Phanang District, Nakhon Si Thammarat Province via the use of 4 statistical methods, namely, Box-Jenkins method, Holt's exponential smoothing method, Brown's exponential smoothing method, and damped trend exponential smoothing method. The average daily wind speed, which was gathered from the Research Center in Energy and Environment, Thaksin University during January 1, 2015 to June 30, 2015 of 1,277 days was divided into two datasets. The first dataset, which consisted of 1,248 days from January 1, 2015 to May 31, 2015 was used for constructing the forecasting models. The second dataset, which consisted of 30 days from June 1 – 30, 2015 was used for comparing the accuracy of the forecasts via the criterion of the lowest root mean square error (RMSE). The results indicated that the most accurate method was Box-Jenkins method (RMSE = 1.5556) which the forecasting model was

$$\hat{Y}_t = 0.82429Y_{t-1} + 0.69963Y_{t-2} - 0.80909Y_{t-3} + 0.28517Y_{t-4} + 0.22656e_{t-1} - 0.77254e_{t-2}.$$

Keywords: Average Wind Speed per Day, Forecasting Model, Box-Jenkins, Exponential Smoothing

บทนำ

พลังงานลมเป็นพลังงานตามธรรมชาติที่เกิดจากความแตกต่างของอุณหภูมิ ความกดดันของบรรยากาศ และแรงจากการหมุนของโลก ปัจจุบันมีการนำพลังงานลมมาใช้ประโยชน์กันมากขึ้น เช่น การนำมาผลิตไฟฟ้า เนื่องจากพลังงานลมมีอยู่ทั่วไป ไม่ต้องซื้อ และเป็นพลังงานที่สะอาดไม่ก่อให้เกิดอันตรายต่อสุขภาพ ซึ่งการผลิตกระแสไฟฟ้าจากพลังงานลมให้มีศักยภาพเพียงพอจำเป็นต้องอาศัยเครื่องจักรกลที่สามารถรับพลังงานจลน์จากการเคลื่อนที่ของลมเปลี่ยนเป็นพลังงานกลได้ เรียกเครื่องจักรกลชนิดนี้ว่า เทคโนโลยีกังหันลม อย่างไรก็ตาม การนำ เทคโนโลยีกังหันลมมาใช้ประโยชน์ในการผลิตกระแสไฟฟ้าเพื่อให้เกิดประสิทธิภาพสูงสุดจำเป็นต้องเข้าใจ พฤติกรรมของลม ลักษณะการแจกแจงทางสถิติของลม ความหนาแน่นของกำลังลมตามระดับความสูงของเสาตัวลม และลักษณะพื้นที่ของบริเวณที่จะนำพลังงานลมมาใช้ประโยชน์ รวมถึงอัตราเร็วลมซึ่งเป็นปัจจัยสำคัญที่จำเป็นต้องศึกษา ในบางช่วงเวลาอัตราเร็วลมอาจเกิดความไม่แน่นอน เป็นผลเนื่องมาจากสาเหตุหลายประการ เช่น ความแตกต่างของพื้นที่ (Waewsak et al., 2008) จากการที่อัตราเร็วลมมีคุณลักษณะที่มีความแปรปรวนสูง เพราะ พฤติกรรมของลมที่ไม่ต่อเนื่องและมีการเปลี่ยนแปลงสูง ทำให้การผลิตไฟฟ้าของกังหันลมผลิตไฟฟ้ามีลักษณะแปรปรวนตามพฤติกรรมของลม ส่งผลต่อปริมาณและคุณภาพไฟฟ้าที่ผลิตได้จากกังหันลมผลิตไฟฟ้า (Waewsak

et al., 2008; Bielecki, 2010) ด้วยเหตุผลดังกล่าว จึงเป็นที่น่าสนใจที่ควรมีการศึกษาการพยากรณ์ข้อมูลอัตราเร็วลม โดยใช้กระบวนการวิเคราะห์ข้อมูล การศึกษาครั้งนี้ผู้วิจัยจะให้ความสนใจกับการพยากรณ์อัตราเร็วลมรายวัน ที่ระดับความสูง 120 เมตร อำเภอปากพนัง จังหวัดนครศรีธรรมราช เนื่องจากการพยากรณ์อัตราเร็วลมรายวัน ที่ระดับความสูง 120 เมตรเหนือพื้นดิน เป็นสิ่งที่มีประโยชน์ เพราะเป็นระดับความสูงของกังหันลมผลิตไฟฟ้าขนาดใหญ่ในปัจจุบันที่มีหอคอยสูงระดับ 120 เมตร อีกทั้งที่ระดับความสูงเพิ่มขึ้น อัตราเร็วลมจะมีค่าเพิ่มขึ้นตาม จึงทำให้สามารถผลิตหน่วยไฟฟ้าได้มากยิ่งขึ้น

จากการทบทวนวรรณกรรมที่เกี่ยวข้อง พบว่า Keerativibool & Mahileh (2011) ได้พยากรณ์อัตราเร็วลม ราย 3 ชั่วโมง ที่ระดับความสูง 20 เมตร ตามแนวชายฝั่งของอำเภอจะนะ จังหวัดสงขลา ผลการวิจัยพบว่า วิธีการแยก ส่วนประกอบของอนุกรมเวลารูปแบบบวกให้ความแม่นยำในการพยากรณ์สูงกว่าวิธีบ็อกซ์-เจนกินส์ Keerativibool et al. (2011) ได้พยากรณ์อัตราเร็วลมราย 3 ชั่วโมง ที่ระดับความสูง 30 เมตร ตามแนวชายฝั่ง อำเภอจะนะ จังหวัด สงขลา ผลการวิจัยพบว่า วิธีบ็อกซ์-เจนกินส์เป็นวิธีที่ดีกว่าวิธีการแยกส่วนประกอบของอนุกรมเวลา Keerativibool and Mahileh (2013) ได้พยากรณ์อัตราเร็วลมรายชั่วโมง ที่ระดับความสูง 40 เมตร ตามแนวชายฝั่ง อำเภอท่าศาลา จังหวัดนครศรีธรรมราช ผลการวิจัยพบว่า วิธีบ็อกซ์-เจนกินส์ โดยใช้ตัวแบบ SARIMA(1, 1, 1)(0, 1, 1)₂₄ ไม่มีพจน์ ค่าคงตัว เป็นวิธีที่เหมาะสมกับข้อมูลชุดนี้มากที่สุด

จากการศึกษาที่ผ่านมา พบว่า มีการศึกษาการผลิตไฟฟ้าเชิงพาณิชย์ของภาคใต้ โดยการศึกษาพยากรณ์ อัตราเร็วลมที่ระดับความสูง 20 และ 30 เมตร อำเภอจะนะ จังหวัดสงขลา และที่ระดับความสูง 40 เมตร อำเภอท่า ศาลา จังหวัดนครศรีธรรมราช ซึ่งเป็นการศึกษาอัตราเร็วลมรายชั่วโมงหรือราย 3 ชั่วโมงเท่านั้น แต่ยังไม่เคยมี การศึกษาพยากรณ์อัตราเร็วลมรายวันและที่ระดับความสูง 120 เมตร อำเภอปากพนัง จังหวัดนครศรีธรรมราช ดังนั้น การศึกษาครั้งนี้ ผู้วิจัยจึงมีความสนใจที่จะศึกษาวิธีการสร้างตัวแบบพยากรณ์อัตราเร็วลมรายวัน ที่ระดับความสูง 120 เมตร อำเภอปากพนัง จังหวัดนครศรีธรรมราช โดยใช้วิธีการทางสถิติทั้งหมด 4 วิธี ได้แก่ วิธีบ็อกซ์-เจนกินส์ วิธีการทำให้เรียบแบบเลขชี้กำลังของโฮลด์ วิธีการทำให้เรียบแบบเลขชี้กำลังของบราวน์ และวิธีการทำให้เรียบแบบ เลขชี้กำลังด้วยแนวโน้มแบบแคม จากนั้นจึงคัดเลือกตัวแบบพยากรณ์ที่มีความเหมาะสมมากที่สุด 1 ตัวแบบ โดยใช้ เกณฑ์รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) ต่ำสุด เพื่อให้ได้ตัวแบบพยากรณ์ สำหรับใช้ในการพยากรณ์อัตราเร็วลมรายวัน ที่ระดับความสูง 120 เมตร อำเภอปากพนัง จังหวัดนครศรีธรรมราช แล้วนำผลการพยากรณ์อัตราเร็วลมที่ได้ไปใช้ประโยชน์ในการทำนายปริมาณหน่วยไฟฟ้าที่จะผลิตได้ในอนาคต ต่อไป

วิธีการทดลอง

การวิจัยครั้งนี้ผู้วิจัยดำเนินการสร้างตัวแบบพยากรณ์อัตราเร็วลมรายวัน (เมตร/วินาที) โดยใช้โปรแกรม SPSS รุ่น 17 ซึ่งข้อมูลที่ใช้ในการศึกษาได้มาจากศูนย์วิจัยพลังงานและสิ่งแวดล้อม มหาวิทยาลัยทักษิณ เป็นอัตราเร็ว

ลมรายชั่วโมง ที่ระดับความสูง 120 เมตรเหนือพื้นดิน อำเภอปากพนัง จังหวัดนครศรีธรรมราช ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2555 ถึงวันที่ 30 มิถุนายน พ.ศ. 2558 จำนวน 30,648 ค่า ซึ่งเป็นจำนวนข้อมูลที่มากเกินไป ทำให้รูปแบบการเปลี่ยนแปลงของอัตราเร็วลมในอดีตไม่คงที่ จึงยากแก่การสร้างตัวแบบพยากรณ์ ดังนั้นผู้วิจัยจึงนำอัตราเร็วลมรายชั่วโมงดังกล่าวมาคำนวณเป็นค่าเฉลี่ยต่อวัน จะได้อัตราเร็วลมรายวันตั้งแต่วันที่ 1 มกราคม พ.ศ. 2555 ถึงวันที่ 30 มิถุนายน พ.ศ. 2558 จำนวน 1,277 วัน จากนั้นแบ่งข้อมูลออกเป็น 2 ชุด ชุดที่ 1 คืออัตราเร็วลมรายวันตั้งแต่วันที่ 1 มกราคม พ.ศ. 2555 ถึงวันที่ 31 พฤษภาคม พ.ศ. 2558 จำนวน 1,248 วัน ใช้สำหรับการสร้างตัวแบบพยากรณ์ และชุดที่ 2 คืออัตราเร็วลมรายวันของวันที่ 1 ถึง 30 มิถุนายน พ.ศ. 2558 จำนวน 30 วัน ใช้สำหรับการเปรียบเทียบความเหมาะสมของตัวแบบพยากรณ์ด้วยเกณฑ์ RMSE ต่ำสุด

จากอนุกรมเวลาชุดที่ 1 ที่ต้องการศึกษา คือ อัตราเร็วลมรายวันตั้งแต่วันที่ 1 มกราคม พ.ศ. 2555 ถึงวันที่ 31 พฤษภาคม พ.ศ. 2558 จำนวน 1,248 วัน ผู้วิจัยจะดำเนินการตรวจสอบแนวโน้มและความแปรผันตามฤดูกาลโดยใช้การวิเคราะห์ความแปรปรวนทางเดียวโดยลำดับที่ของครัสคาล-วอลลิส (Kruskal-Wallis's One-Way Analysis of Variance by Rank) ซึ่งวิธีการนี้จะตรวจสอบแนวโน้มโดยการพิจารณาว่ามัธยฐานของอัตราเร็วลมในแต่ละเดือนหรือมัธยฐานของอัตราเร็วลม 12 กลุ่ม (12 เดือน) มีค่าแตกต่างกันหรือไม่ หากพบมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ จะหมายความว่าอนุกรมเวลามีแนวโน้ม ดังนั้นควรกำจัดแนวโน้มออกก่อนที่จะตรวจสอบความแปรผันตามฤดูกาล ซึ่งวิธีการกำจัดแนวโน้มมี 2 วิธี คือ ถ้าพิจารณาที่กราฟของอนุกรมเวลาเทียบกับเวลาแล้วพบว่าอนุกรมเวลามีการเคลื่อนไหวหรือมีการแกว่งตัวที่ค่อนข้างคงที่เมื่อเวลาเปลี่ยนแปลงไป กล่าวได้ว่า อนุกรมเวลา มีความเหมาะสมกับตัวแบบบวก ควรกำจัดแนวโน้มออกด้วยการลบ แต่ถ้าอนุกรมเวลามีการเคลื่อนไหวหรือมีการแกว่งตัวที่เพิ่มขึ้นหรือลดลงเมื่อเวลาเปลี่ยนแปลงไป กล่าวได้ว่า อนุกรมเวลา มีความเหมาะสมกับตัวแบบคูณ ควรกำจัดแนวโน้มออกด้วยการหาร สำหรับอนุกรมเวลาอัตราเร็วลมรายวันมีแนวโน้มและมีการเคลื่อนไหวหรือมีการแกว่งตัวที่เพิ่มขึ้นเมื่อเวลาเปลี่ยนแปลงไป ดังรูปที่ 1 ในผลและวิจารณ์ผลการทดลอง จึงมีความเหมาะสมกับตัวแบบคูณ ดังนั้นต้องกำจัดแนวโน้มออกด้วยการหาร เมื่อกำจัดแนวโน้มออกแล้วจึงตรวจสอบความแปรผันตามฤดูกาลโดยการพิจารณาว่ามัธยฐานของอัตราเร็วลมวันเดียวกันในแต่ละเดือนหรือมัธยฐานของอัตราเร็วลม 31 กลุ่ม (31 วัน) มีค่าแตกต่างกันหรือไม่ หากพบมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ จะหมายความว่าอนุกรมเวลา มีความแปรผันตามฤดูกาล (Riansut, 2018)

จากการสังเกตการเคลื่อนไหวของอัตราเร็วลมรายวัน ที่ระดับความสูง 120 เมตร อำเภอปากพนัง จังหวัดนครศรีธรรมราช ชุดที่ 1 ดังรูปที่ 1 กราฟฟังก์ชันสหสัมพันธ์ในตัวเอง (Autocorrelation Function: ACF) และกราฟฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (Partial Autocorrelation Function: PACF) และผลการทดสอบสมมุติฐานในผลและวิจารณ์ผลการทดลอง พบว่า อัตราเร็วลมมีแนวโน้มทั้งลดลงและเพิ่มขึ้น แต่ไม่มีความแปรผันตามฤดูกาล ดังนั้นวิธีการสร้างตัวแบบพยากรณ์ที่มีความเหมาะสมกับข้อมูลที่มีเพียงแนวโน้ม ได้แก่ วิธีบ็อกซ์-เจนกินส์ วิธีการทำให้เรียบแบบเลขชี้กำลังของโฮลต์ วิธีการทำให้เรียบแบบเลขชี้กำลังของบราวน์ และวิธีการทำให้เรียบแบบเลขชี้กำลัง

ด้วยแนวโน้มแบบเดิม วิธีการสร้างตัวแบบพยากรณ์ด้วยวิธีการทางสถิติทั้ง 4 วิธีดังกล่าวแสดงรายละเอียดดังหัวข้อที่ 1 สำหรับหัวข้อที่ 2 จะเป็นการเปรียบเทียบความเหมาะสมของตัวแบบพยากรณ์ โดยพิจารณาจากเกณฑ์ RMSE ต่ำสุด สัญลักษณ์ที่ใช้สำหรับวิธีการสร้างตัวแบบพยากรณ์ในหัวข้อที่ 1 แสดงดังนี้

Y_t แทนอัตราเร็วลมรายวัน ณ เวลา t

$\hat{Y}_{t+m}$ แทนค่าพยากรณ์อัตราเร็วลมรายวัน ณ เวลา $t + m$ โดยที่ m แทนจำนวนช่วงเวลาที่ต้องการพยากรณ์ไปข้างหน้า

a_t และ b_t แทนค่าประมาณระยะตัดแกน Y และความชันของแนวโน้ม ณ เวลา t ตามลำดับ

α , γ และ ϕ แทนค่าคงตัวการทำให้เรียบ โดยที่ $0 < \alpha < 1$, $0 < \gamma < 1$ และ $0 < \phi < 1$

t แทนช่วงเวลา ซึ่งมีค่าตั้งแต่ 1 ถึง n_t เมื่อ n_t แทนจำนวนข้อมูลชุดที่ 1 ($n_t = 1,248$)

1. วิธีการสร้างตัวแบบพยากรณ์

1.1 วิธีบ็อกซ์-เจนกินส์ (Box-Jenkins Method)

วิธีบ็อกซ์-เจนกินส์เป็นวิธีที่มีความนิยมสูง จึงมีการนำมาประยุกต์ใช้ในการพยากรณ์ด้านต่างๆ เป็นจำนวนมาก ขั้นตอนหนึ่งที่สำคัญของวิธีนี้ คือ การกำหนดตัวแบบ ซึ่งเราจำเป็นต้องทราบกราฟ ACF และ PACF ของอนุกรมเวลาที่คงที่ (Stationary) เพื่อใช้ในการกำหนดตัวแบบพยากรณ์ (Manmin, 2006) ตัวแบบของวิธีบ็อกซ์-เจนกินส์หรือตัวแบบ Seasonal Autoregressive Integrated Moving Average: SARIMA(p, d, q)(P, D, Q) แสดงดังสมการที่ (1) (Box et al., 1994) แต่ในกรณีที่มีข้อมูลมีเพียงแนวโน้มเท่านั้น ตัวแบบจะสามารถสรุปเหลือเพียง Autoregressive Integrated Moving Average: ARIMA(p, d, q)

$$\phi_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D Y_t = \delta + \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (1)$$

เมื่อ ε_t แทนความคลาดเคลื่อนของอัตราเร็วลมรายวัน

$\delta = \mu\phi_p(B)\Phi_p(B^s)$ แทนค่าคงตัว (Constant) โดยที่ μ แทนค่าเฉลี่ยของอนุกรมเวลาที่คงที่

$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ แทนตัวดำเนินการสหสัมพันธ์ในลำดับที่ p กรณีไม่มีฤดูกาล (Non-Seasonal Autoregressive Operator of Order p: AR(p))

$\Phi_p(B^s) = 1 - \phi_1 B^s - \phi_2 B^{2s} - \dots - \phi_p B^{ps}$ แทนตัวดำเนินการสหสัมพันธ์ในลำดับที่ P กรณีมีฤดูกาล (Seasonal Autoregressive Operator of Order P: SAR(P))

$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ แทนตัวดำเนินการเฉลี่ยเคลื่อนที่อันดับที่ q กรณีไม่มีฤดูกาล (Non-Seasonal Moving Average Operator of Order q: MA(q))

$\Theta_Q(B^s) = 1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_Q B^{Qs}$ แทนตัวดำเนินการเฉลี่ยเคลื่อนที่อันดับที่ Q กรณีมีฤดูกาล (Seasonal Moving Average Operator of Order Q: SMA(Q))

d และ D แทนลำดับที่ของการหาผลต่างและผลต่างฤดูกาล ตามลำดับ

B แทนตัวดำเนินการถอยหลัง (Backward Operator) โดยที่ $B^s Y_t = Y_{t-s}$

s แทนจำนวนคาบของฤดูกาล

ขั้นตอนการสร้างตัวแบบพยากรณ์โดยวิธีบ็อกซ์-เจนกินส์ แสดงรายละเอียดดังนี้ (Riansut, 2019)

1) พิจารณาอนุกรมเวลาว่าคงที่หรือไม่ นั่นคือ พิจารณาว่าอนุกรมเวลาที่ค่าสังเกต Y_t มีคุณสมบัติทางสถิติคือ ค่าเฉลี่ย ความแปรปรวน และฟังก์ชันความน่าจะเป็นของค่าสังเกต ณ เวลาต่างๆ กัน มีค่าคงที่ กล่าวคือ ไม่เปลี่ยนแปลงตามเวลาที่เปลี่ยนไป (Ket-iam; 2005) โดยพิจารณาจากกราฟของอนุกรมเวลาที่เทียบกับเวลา (Y_t , t) กราฟ ACF และ PACF หากพบว่าอนุกรมเวลาไม่คงที่ (Non-Stationary) ต้องแปลงอนุกรมเวลาให้คงที่ก่อนที่จะทำขั้นตอนต่อไป เช่น การแปลงข้อมูลด้วยการหาผลต่างหรือผลต่างฤดูกาล (Difference or Seasonal Difference) การแปลงข้อมูลด้วยลอการิทึมสามัญหรือลอการิทึมธรรมชาติ (Common Logarithm or Natural Logarithm) การแปลงข้อมูลด้วยเลขยกกำลัง เช่น ยกกำลัง 0.5 (Square Root Transformation) หรือยกกำลัง 2 (Square Transformation) เป็นต้น (Bowerman & O'Connell, 1993)

2) กำหนดตัวแบบพยากรณ์ที่เป็นไปได้จากกราฟ ACF และ PACF ของอนุกรมเวลาที่คงที่ นั่นคือ กำหนดค่า p, q, P และ Q พร้อมทั้งประมาณค่าพารามิเตอร์ของตัวแบบ

3) คัดพารามิเตอร์ที่ไม่มีนัยสำคัญออกจากตัวแบบพยากรณ์ครั้งละ 1 ตัว จากนั้นจึงกำหนดตัวแบบพยากรณ์และประมาณค่าพารามิเตอร์ใหม่จนกว่าจะได้ตัวแบบพยากรณ์ที่ประกอบด้วยพารามิเตอร์ที่มีนัยสำคัญทั้งหมด

4) คัดเลือกตัวแบบที่มีความคลาดเคลื่อนจากการพยากรณ์ผ่านข้อสมมุติ (Assumption) ได้แก่ ความคลาดเคลื่อนมีการแจกแจงปรกติ (ตรวจสอบโดยใช้การทดสอบคอลโมโกรอฟ-สมิรโนฟ: Kolmogorov-Smirnov's Test) ความคลาดเคลื่อนเป็นอิสระกัน (ตรวจสอบโดยใช้การทดสอบรันส์: Runs Test) ค่าเฉลี่ยของความคลาดเคลื่อนมีค่าเท่ากับศูนย์ (ตรวจสอบโดยใช้การทดสอบที: t-Test) และความแปรปรวนของความคลาดเคลื่อนจากวันเดียวกันในแต่ละเดือนมีค่าเท่ากัน (ตรวจสอบโดยใช้การทดสอบของเลวินภายใต้การใช้มัธยฐาน: Levene's Test based on Median)

5) พยากรณ์อนุกรมเวลา โดยใช้ตัวแบบพยากรณ์ที่เหมาะสมที่สุดจากขั้นตอนที่ 4

1.2 วิธีการทำให้เรียบแบบเลขชี้กำลังของโฮลต์ (Holt's Exponential Smoothing Method)

วิธีการทำให้เรียบแบบเลขชี้กำลังของโฮลต์มีความเหมาะสมกับข้อมูลที่มีแนวโน้มที่เป็นเส้นตรง โดยไม่มีความแปรผันตามฤดูกาล ตัวแบบพยากรณ์แสดงดังสมการที่ (2) (Ket-iam, 2005) เมื่อได้ตัวแบบพยากรณ์แล้วจะดำเนินการตรวจสอบข้อสมมุติของความคลาดเคลื่อนจากการพยากรณ์เช่นเดียวกับวิธีบ็อกซ์-เจนกินส์

$$\hat{Y}_{t+m} = a_t + b_t(m) \quad (2)$$

เมื่อ $a_t = \alpha Y_t + (1-\alpha)(a_{t-1} + b_{t-1})$, $b_t = \gamma(a_t - a_{t-1}) + (1-\gamma)b_{t-1}$

1.3 วิธีการทำให้เรียบแบบเลขชี้กำลังของบราวน์ (Brown's Exponential Smoothing Method)

วิธีการทำให้เรียบแบบเลขชี้กำลังของบราวน์มีความเหมาะสมกับข้อมูลที่มีแนวโน้มเป็นเส้นตรงและไม่มี
ความแปรผันตามฤดูกาลเช่นเดียวกับการทำให้เรียบแบบเลขชี้กำลังของโฮลต์ แต่มีการกำหนดให้ค่าคงตัวการทำให้
เรียบของค่าระดับ (α) และค่าคงตัวการทำให้เรียบของค่าความชัน (γ) เท่ากัน ตัวแบบพยากรณ์แสดงดังสมการที่
(3) (Ket-iam, 2005; Manmin, 2006) เมื่อได้ตัวแบบพยากรณ์แล้วจะดำเนินการตรวจสอบข้อสมมุติของความ
คลาดเคลื่อนจากการพยากรณ์เช่นเดียวกับวิธีบ็อกซ์-เจนกินส์

$$\hat{Y}_{t+m} = a_t + b_t[(m-1)+1/\alpha] \quad (3)$$

เมื่อ $a_t = \alpha Y_t + (1-\alpha)a_{t-1}$, $b_t = \alpha(a_t - a_{t-1}) + (1-\alpha)b_{t-1}$

1.4 วิธีการทำให้เรียบแบบเลขชี้กำลังด้วยแนวโน้มแบบแดม (Damped Trend Exponential Smoothing Method)

วิธีการทำให้เรียบแบบเลขชี้กำลังด้วยแนวโน้มแบบแดมมีความเหมาะสมกับข้อมูลที่มีแนวโน้มเป็น
เส้นตรงและไม่มีค่าความแปรผันตามฤดูกาลเช่นเดียวกับวิธีการทำให้เรียบแบบเลขชี้กำลังของโฮลต์ และวิธีการทำให้
เรียบแบบเลขชี้กำลังของบราวน์ แต่อัตราการเปลี่ยนแปลงของแนวโน้มจะช้ากว่าของวิธีการทำให้เรียบแบบเลขชี้
กำลังของโฮลต์ รวมถึงความชันจะมีค่าลดลงตามเวลา ตัวแบบพยากรณ์แสดงดังสมการที่ (4) (Manmin, 2006)
เมื่อได้ตัวแบบพยากรณ์แล้วจะดำเนินการตรวจสอบข้อสมมุติของความคลาดเคลื่อนจากการพยากรณ์เช่นเดียวกับ
วิธีบ็อกซ์-เจนกินส์

$$\hat{Y}_{t+m} = a_t + b_t \sum_{i=1}^m \phi^i \quad (4)$$

เมื่อ $a_t = \alpha Y_t + (1-\alpha)(a_{t-1} + \phi b_{t-1})$, $b_t = \gamma(a_t - a_{t-1}) + (1-\gamma)\phi b_{t-1}$

2. การเปรียบเทียบความเหมาะสมของตัวแบบพยากรณ์

การเปรียบเทียบความเหมาะสมของตัวแบบพยากรณ์อัตราเร็วลมรายวัน ที่ระดับความสูง 120 เมตร อำเภอ
ปากพนัง จังหวัดนครศรีธรรมราช ชุดที่ 2 หรืออัตราเร็วลมรายวันของวันที่ 1 ถึง 30 มิถุนายน พ.ศ. 2558 จำนวน 30
วัน กับค่าพยากรณ์จากวิธีการทางสถิติทั้ง 4 วิธี ได้แก่ วิธีบ็อกซ์-เจนกินส์ วิธีการทำให้เรียบแบบเลขชี้กำลังของโฮลต์
วิธีการทำให้เรียบแบบเลขชี้กำลังของบราวน์ และวิธีการทำให้เรียบแบบเลขชี้กำลังด้วยแนวโน้มแบบแดม โดยการ
คำนวณค่า RMSE ซึ่งถ้าพบว่าตัวแบบพยากรณ์ใดให้ค่า RMSE ต่ำสุด จะจัดเป็นตัวแบบที่มีความเหมาะสมกับข้อมูล
ชุดนี้มากที่สุด เนื่องจากให้ค่าพยากรณ์ที่มีความแตกต่างกับข้อมูลจริงน้อยที่สุด เกณฑ์ RMSE แสดงดังสมการที่ (5)
(Ket-iam, 2005)

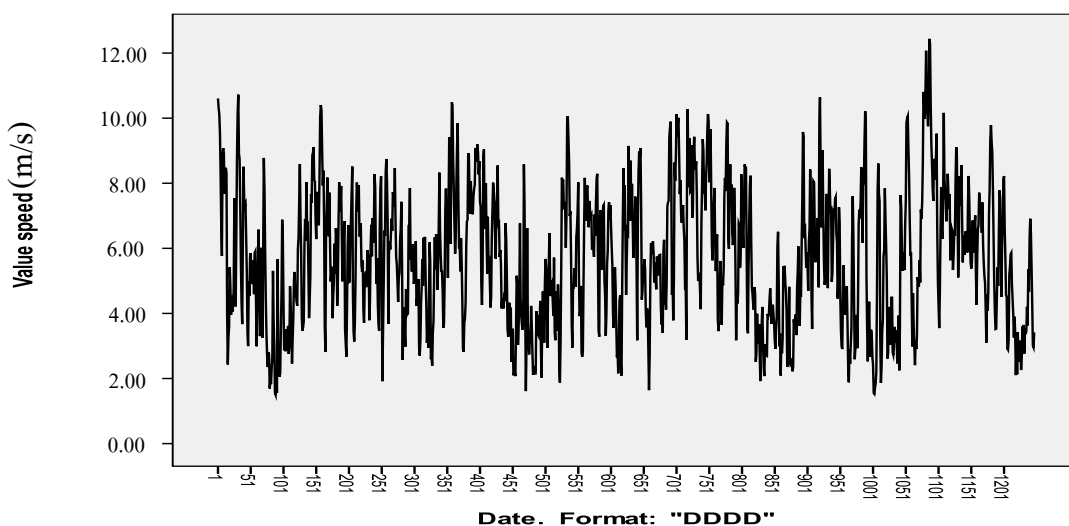
$$RMSE = \sqrt{\frac{1}{n_2} \sum_{t=1}^{n_2} e_t^2} \quad (5)$$

เมื่อ $e_t = Y_t - \hat{Y}_t$ แทนความคลาดเคลื่อนของอัตราเร็วลมรายวันที่ได้จากการพยากรณ์ ณ เวลา t

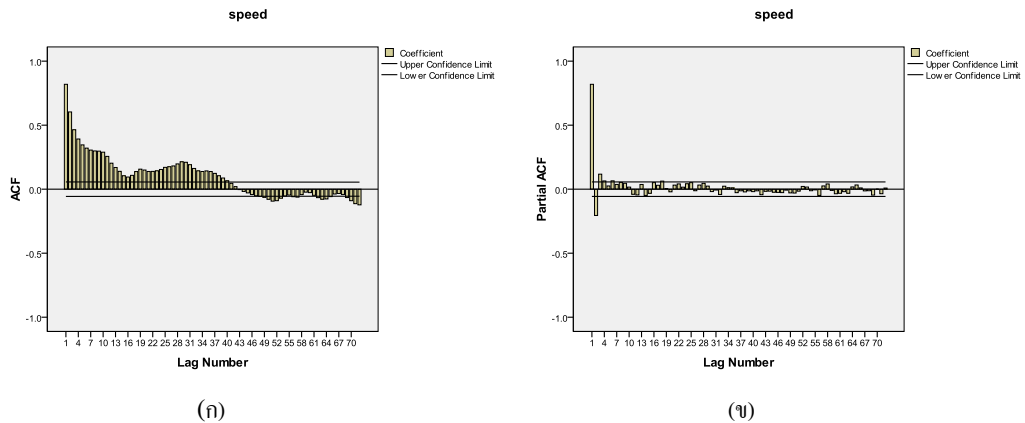
Y_t และ $\hat{Y}_t$ แทนอัตราเร็วลมรายวัน และค่าพยากรณ์อัตราเร็วลมรายวัน ณ เวลา t
 t แทนช่วงเวลา ซึ่งมีค่าตั้งแต่ 1 ถึง n_2 โดยที่ n_2 แทนจำนวนข้อมูลชุดที่ 2 ($n_2 = 30$)

ผลและวิจารณ์ผลการทดลอง

จากการตรวจสอบการเคลื่อนไหวของอัตราเร็วลมรายวัน ที่ระดับความสูง 120 เมตร อำเภอปากพ่อง จังหวัดนครศรีธรรมราช ชุดที่ 1 ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2555 ถึงวันที่ 31 พฤษภาคม พ.ศ. 2558 จำนวน 1,248 วัน ดังรูปที่ 1 พบว่า อัตราเร็วลมมีแนวโน้มทั้งเพิ่มขึ้นและลดลง และจากกราฟ ACF ดังรูปที่ 2 (ก) พบว่า สหสัมพันธ์ในตัวเองมีค่าลดลงอย่างช้าๆ (Die Down Slowly) และกราฟ PACF ดังรูปที่ 2 (ข) พบว่า ช่วงเวลาที่ 1 (Lag 1) มีค่าสหสัมพันธ์ในตัวเองบางส่วนสูงเกินจากขอบเขตที่กำหนดอย่างชัดเจน นั้นหมายความว่า อนุกรมเวลาชุดนี้มี ส่วนประกอบของแนวโน้ม ผลที่ได้มีความสอดคล้องกับที่สรุปไว้ในรูปที่ 1 และจากการทดสอบสมมติฐานเพื่อ ตรวจสอบว่าอัตราเร็วลมมีแนวโน้มและความแปรผันตามฤดูกาลหรือไม่ พบว่า ข้อมูลชุดนี้มีแนวโน้ม (Kruskal-Wallis $\chi^2 = 281.212$, p-value < 0.0001) แต่ไม่มีความแปรผันตามฤดูกาล (Kruskal-Wallis $\chi^2 = 10.839$, p-value = 0.999)



รูปที่ 1 ลักษณะการเคลื่อนไหวของอัตราเร็วลมรายวัน ที่ระดับความสูง 120 เมตร
ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2555 ถึงวันที่ 31 พฤษภาคม พ.ศ. 2558 จำนวน 1,248 วัน



รูปที่ 2 ACF และ PACF ของอัตราเร็วลมรายวัน ที่ระดับความสูง 120 เมตร

1. ผลการสร้างตัวแบบพยากรณ์

1.1 ผลของวิธีบ็อกซ์-เจนกินส์

เนื่องจากข้อมูลชุดนี้มีแนวโน้มหรืออนุกรมเวลายังไม่คงที่ ผู้วิจัยจึงกำจัดแนวโน้มออกเพื่อให้อนุกรมเวลาคงที่ โดยการหาผลต่างลำดับที่ 1 ($d = 1$) ได้กราฟ ACF และ PACF หลังจากการแปลงข้อมูลแสดงดังรูปที่ 3 ซึ่งพบว่า อนุกรมเวลามีลักษณะคงที่ จึงกำหนดตัวแบบพยากรณ์ที่เป็นไปได้เริ่มต้น คือ ตัวแบบ ARIMA(3, 1, 3) พร้อมกับประมาณค่าพารามิเตอร์ โดยตัวแบบพยากรณ์ที่มีพารามิเตอร์ทุกตัวมีนัยสำคัญที่ระดับ 0.05 คือ ตัวแบบ ARIMA(3, 1, 2) ไม่มีพจน์ค่าคงตัว ซึ่งจากสมการที่ (1) สามารถเขียนเป็นตัวแบบพยากรณ์ของวิธีบ็อกซ์-เจนกินส์ ได้ดังนี้

$$\begin{aligned} (1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B)Y_t &= (1 - \theta_1 B - \theta_2 B^2)\varepsilon_t \\ (1 - B - \phi_1 B + \phi_1 B^2 - \phi_2 B^2 + \phi_2 B^3 - \phi_3 B^3 + \phi_3 B^4)Y_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \\ Y_t &= (1 + \phi_1)Y_{t-1} - (\phi_1 - \phi_2)Y_{t-2} - (\phi_2 - \phi_3)Y_{t-3} - \phi_3 Y_{t-4} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \end{aligned}$$

เมื่อแทนค่าประมาณพารามิเตอร์ จะได้ตัวแบบพยากรณ์แสดงดังนี้

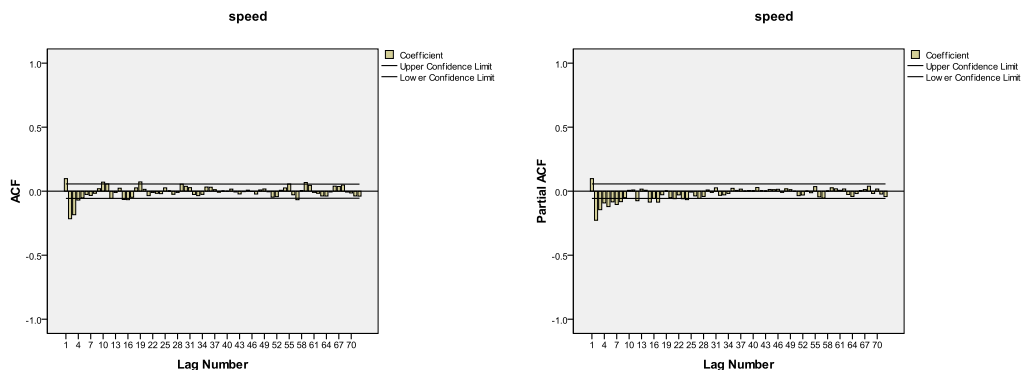
$$\hat{Y}_t = 0.82429Y_{t-1} + 0.69963Y_{t-2} - 0.80909Y_{t-3} + 0.28517Y_{t-4} + 0.22656\varepsilon_{t-1} - 0.77254\varepsilon_{t-2} \quad (6)$$

เมื่อ $\hat{Y}_t$ แทนค่าพยากรณ์อัตราเร็วลมรายวัน ณ เวลา t

Y_{t-j} แทนอัตราเร็วลมรายวัน ณ เวลา $t-j$

ε_{t-j} แทนความคลาดเคลื่อนของอัตราเร็วลมรายวันที่ได้จากการพยากรณ์ ณ เวลา $t-j$

ผลการตรวจสอบข้อสมมุติของความคลาดเคลื่อนที่ได้จากการพยากรณ์ ที่ระดับนัยสำคัญ 0.05 พบว่า มีการแจกแจงปกติ (Kolmogorov-Smirnov $Z = 0.988$, $p\text{-value} = 0.283$) เป็นอิสระกัน (Runs Test $Z = 0.624$, $p\text{-value} = 0.533$) ค่าเฉลี่ยมีค่าเท่ากับศูนย์ ($t = -0.313$, $p\text{-value} = 0.754$) และความแปรปรวนจากวันเดียวกันในแต่ละเดือนมีค่าเท่ากัน (Levene Statistic = 1.139, $p\text{-value} = 0.277$) ดังนั้นตัวแบบพยากรณ์ในสมการที่ (6) มีความเหมาะสม



รูปที่ 3 ACF และ PACF ของอัตราเร็วลมรายวัน ที่ระดับความสูง 120 เมตร
เมื่อแปลงข้อมูลด้วยการหาผลต่างลำดับที่ 1

1.2 ผลของวิธีการทำให้เรียบแบบเลขชี้กำลังของโฮลต์

ตัวแบบพยากรณ์ของวิธีการทำให้เรียบแบบเลขชี้กำลังของโฮลต์แสดงดังนี้

$$\hat{Y}_{t+m} = 3.41630 - 0.00472(m) \quad (7)$$

เมื่อ $\hat{Y}_{t+m}$ แทนค่าพยากรณ์อัตราเร็วลมรายวัน ณ เวลา $t + m$ โดยที่ $m = 1$ แทนวันที่ 1 มิถุนายน พ.ศ. 2558

ผลการตรวจสอบข้อสมมุติของความคลาดเคลื่อนที่ได้จากการพยากรณ์ ที่ระดับนัยสำคัญ 0.05 พบว่า มีการแจกแจงปรกติ (Kolmogorov-Smirnov $Z = 0.923$, $p\text{-value} = 0.362$) เป็นอิสระกัน (Runs Test $Z = -1.105$, $p\text{-value} = 0.269$) ค่าเฉลี่ยมีค่าเท่ากับศูนย์ ($t = -0.022$, $p\text{-value} = 0.983$) และความแปรปรวนจากวันเดียวกันในแต่ละเดือนมีค่าเท่ากัน (Levene Statistic = 1.453, $p\text{-value} = 0.055$) ดังนั้นตัวแบบพยากรณ์ในสมการที่ (7) มีความเหมาะสม

1.3 ผลของวิธีการทำให้เรียบแบบเลขชี้กำลังของบราวน์

ตัวแบบพยากรณ์ของวิธีการทำให้เรียบแบบเลขชี้กำลังของบราวน์แสดงดังนี้

$$\hat{Y}_{t+m} = 3.41055 - 0.30014[(m-1) + 1/0.54028] \quad (8)$$

เมื่อ $\hat{Y}_{t+m}$ แทนค่าพยากรณ์อัตราเร็วลมรายวัน ณ เวลา $t + m$ โดยที่ $m = 1$ แทนวันที่ 1 มิถุนายน พ.ศ. 2558

ผลการตรวจสอบข้อสมมุติของความคลาดเคลื่อนที่ได้จากการพยากรณ์ ที่ระดับนัยสำคัญ 0.05 พบว่า มีการแจกแจงปรกติ (Kolmogorov-Smirnov $Z = 1.026$, $p\text{-value} = 0.243$) และค่าเฉลี่ยมีค่าเท่ากับศูนย์ ($t = 0.011$, $p\text{-value} = 0.991$) แต่ความคลาดเคลื่อนไม่เป็นอิสระกัน (Runs Test $Z = -5.071$, $p\text{-value} < 0.0001$) และความแปรปรวนจากวันเดียวกันในแต่ละเดือนมีค่าไม่เท่ากันที่ระดับนัยสำคัญ 0.05 (Levene Statistic = 1.5, $p\text{-value} = 0.041$) ดังนั้นตัวแบบพยากรณ์ในสมการที่ (8) ไม่มีความเหมาะสม

1.4 ผลของวิธีการทำให้เรียบแบบเลขชี้กำลังด้วยแนวโน้มแบบแตร

ตัวแบบพยากรณ์ของวิธีการทำให้เรียบแบบเลขชี้กำลังด้วยแนวโน้มแบบแตรแสดงดังนี้

$$\hat{Y}_{t+m} = 3.41633 + 0.45908 \sum_{i=1}^m (0.09982)^i \quad (9)$$

เมื่อ $\hat{Y}_{t+m}$ แทนค่าพยากรณ์อัตราเร็วลมรายวัน ณ เวลา $t + m$ โดยที่ $m = 1$ แทนวันที่ 1 มิถุนายน พ.ศ. 2558

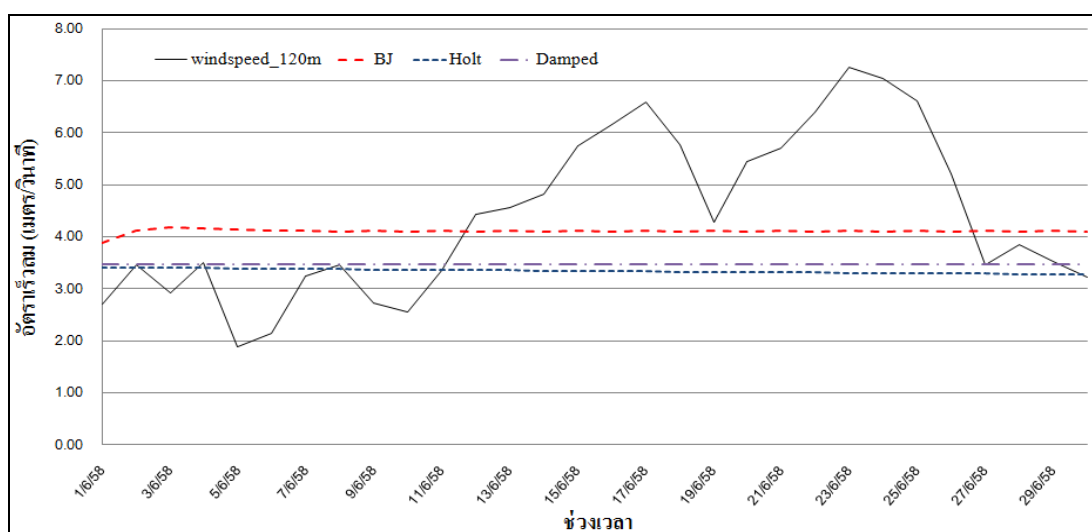
ผลการตรวจสอบข้อสมมุติของความคลาดเคลื่อนที่ได้จากการพยากรณ์ ที่ระดับนัยสำคัญ 0.05 พบว่า มีการแจกแจงปกติ (Kolmogorov-Smirnov $Z = 1.022$, $p\text{-value} = 0.247$) เป็นอิสระกัน (Runs Test $Z = 0.482$, $p\text{-value} = 0.630$) ค่าเฉลี่ยมีค่าเท่ากับศูนย์ ($t = -0.147$, $p\text{-value} = 0.883$) และความแปรปรวนจากวันเดียวกันในแต่ละเดือนมีค่าเท่ากัน (Levene Statistic = 1.355, $p\text{-value} = 0.097$) ดังนั้นตัวแบบพยากรณ์ในสมการที่ (9) มีความเหมาะสม

2. ผลการเปรียบเทียบความเหมาะสมของตัวแบบพยากรณ์

จากการตรวจสอบข้อสมมุติของความคลาดเคลื่อนที่ได้จากการพยากรณ์ ที่ระดับนัยสำคัญ 0.05 พบว่า ตัวแบบพยากรณ์ของวิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของบราวน์ในสมการที่ (8) ไม่มีความเหมาะสม ดังนั้นจึงเปรียบเทียบความเหมาะสมของตัวแบบพยากรณ์จากวิธีบ็อกซ์-เจนกินส์ในสมการที่ (6) วิธีการทำให้เรียบแบบเลขชี้กำลังของโสด์ในสมการที่ (7) และวิธีการทำให้เรียบแบบเลขชี้กำลังด้วยแนวโน้มแบบแฉกในสมการที่ (9) เท่านั้น โดยทำการพยากรณ์ข้อมูลชุดที่ 2 หรืออัตราเร็วลมรายวันของวันที่ 1 ถึง 30 มิถุนายน พ.ศ. 2558 จำนวน 30 วัน เพื่อคำนวณค่า RMSE ได้ผลแสดงดังตารางที่ 1 และรูปที่ 4 ซึ่งพบว่า วิธีบ็อกซ์-เจนกินส์มีความเหมาะสมมากที่สุด เนื่องจากมีค่า RMSE ต่ำสุด รองลงมาคือ วิธีการทำให้เรียบแบบเลขชี้กำลังด้วยแนวโน้มแบบแฉก และวิธีการทำให้เรียบแบบเลขชี้กำลังของโสด์ ตามลำดับ ผลการศึกษาครั้งนี้มีความสอดคล้องกับการศึกษาของ Keerativibool et al. (2011) ที่พบว่า วิธีบ็อกซ์-เจนกินส์เป็นวิธีที่ดีในการพยากรณ์อัตราเร็วลมราย 3 ชั่วโมง ที่ระดับความสูง 30 เมตร ตามแนวชายฝั่ง อำเภอนะนะ จังหวัดสงขลา และสอดคล้องกับการศึกษาของ Keerativibool & Mahileh (2013) ที่พบว่า วิธีบ็อกซ์-เจนกินส์เป็นวิธีที่มีความเหมาะสมมากที่สุดกับอัตราเร็วลมรายชั่วโมง ที่ระดับความสูง 40 เมตร ตามแนวชายฝั่ง อำเภอนาตาล จังหวัดนครราชสีมา แต่มีความขัดแย้งกับการศึกษาของ Keerativibool and Mahileh (2011) ที่พบว่า วิธีการแยกส่วนประกอบของอนุกรมเวลาในรูปแบบบวกให้ความแม่นยำในการพยากรณ์สูงกว่าวิธีบ็อกซ์-เจนกินส์ในการพยากรณ์อัตราเร็วลมราย 3 ชั่วโมง ที่ระดับความสูง 20 เมตร ตามแนวชายฝั่งของอำเภอนะนะ จังหวัดสงขลา อาจเนื่องมาจากลักษณะข้อมูล ความแตกต่างของช่วงเวลาในการเก็บรวบรวมข้อมูล ระดับความสูงที่วัดอัตราเร็วลม พื้นที่ที่ศึกษา และวิธีการสร้างตัวแบบพยากรณ์ที่นำมาศึกษาเปรียบเทียบ อีกทั้งการศึกษานี้ยังพบว่า อนุกรมเวลาอัตราเร็วลมรายวันไม่มีความแปรผันตามฤดูกาลขัดแย้งกับการศึกษาของ Keerativibool & Mahileh (2011), Keerativibool et al. (2011) และ Keerativibool & Mahileh (2013) ที่ศึกษาอัตราเร็วลมรายชั่วโมงและราย 3 ชั่วโมง ซึ่งจะพบความแปรผันตามฤดูกาล โดยสาเหตุที่อัตราเร็วลมรายวันไม่มีความแปรผันตามฤดูกาล อาจเนื่องมาจากการปรับอัตราเร็วลมที่เก็บราย 10 นาทีมาเฉลี่ยให้เป็นรายวัน จากนั้นจึงตรวจสอบความแปรผันตามฤดูกาลโดยการพิจารณามัธยฐานของอัตราเร็วลมวันเดียวกันในแต่ละเดือน ซึ่งพบว่า มัธยฐานของอัตราเร็วลมวันเดียวกันในแต่ละเดือนไม่ได้แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

ตารางที่ 1 RMSE ของข้อมูลชุดที่ 2

วิธีการพยากรณ์	RMSE
บ็อกซ์-เจนกินส์	1.5556
โซลต์	1.8791
แคม	1.7929



รูปที่ 4 การเปรียบเทียบอัตราเร็วลมรายวัน ที่ระดับความสูง 120 เมตร และค่าพยากรณ์

การศึกษาค้นคว้าพบว่า วิธีบ็อกซ์-เจนกินส์มีความเหมาะสมมากที่สุดกับอัตราเร็วลมรายวัน ที่ระดับความสูง 120 เมตร อำเภอปากพนัง จังหวัดนครศรีธรรมราช อาจเนื่องมาจากการสร้างตัวแบบพยากรณ์ของวิธีนี้จะพิจารณาลักษณะของข้อมูลว่ามีสหสัมพันธ์กันอย่างไร รวมถึงมีการพิจารณาข้อมูลและความคลาดเคลื่อนในอดีตเพื่อนำมาสร้างเป็นตัวแบบพยากรณ์ที่เหมาะสม โดยวิธีการพยากรณ์นี้สามารถใช้ได้กับข้อมูลที่มีลักษณะการเคลื่อนไหวทุกรูปแบบ อีกเหตุผลหนึ่งที่ผู้วิจัยคาดว่าวิธีบ็อกซ์-เจนกินส์มีความเหมาะสมมากกว่าวิธีการทำให้เรียบแบบเลขชี้กำลังของโซลต์ วิธีการทำให้เรียบแบบเลขชี้กำลังของบราวน์ และวิธีการทำให้เรียบแบบเลขชี้กำลังด้วยแนวโน้มแบบแคม อาจเนื่องมาจากตัวแบบพยากรณ์ที่ได้จากวิธีการทำให้เรียบแบบเลขชี้กำลังของโซลต์ และวิธีการทำให้เรียบแบบเลขชี้กำลังของบราวน์มีค่าความชัน (b) เป็นลบ ซึ่งหมายถึงผลที่ได้จากการพยากรณ์อัตราเร็วลมจะเป็นไปในทิศทางลดลงเท่านั้น ขณะที่วิธีการทำให้เรียบแบบเลขชี้กำลังด้วยแนวโน้มแบบแคมมีค่าความชันเป็นบวก ทำให้ผลที่ได้จากการพยากรณ์อัตราเร็วลมจะเป็นไปในทิศทางเพิ่มขึ้น แต่อัตราเร็วลมชุดที่ 1 ที่นำมาศึกษามีแนวโน้มทั้งลดลงและเพิ่มขึ้น

สรุปผลการทดลอง

การวิจัยครั้งนี้ได้สร้างตัวแบบพยากรณ์อัตราเร็วลมรายวัน ที่ระดับความสูง 120 เมตร อำเภอปากพ่อง จังหวัดนครศรีธรรมราช ผลการศึกษาพบว่า วิธีบ็อกซ์-เจนกินส์มีความเหมาะสมมากที่สุด รองลงมาคือ วิธีการทำให้เรียบแบบเลขชี้กำลังด้วยแนวโน้มแบบแฉก และวิธีการทำให้เรียบแบบเลขชี้กำลังของโฮลต์ ตามลำดับ ขณะที่วิธีการทำให้เรียบแบบเลขชี้กำลังของบราวน์ไม่มีความเหมาะสม เนื่องจากความคลาดเคลื่อนไม่ผ่านข้อสมมุติ ซึ่งตัวแบบพยากรณ์ของวิธีบ็อกซ์-เจนกินส์แสดงดังนี้

$$\hat{Y}_t = 0.82429Y_{t-1} + 0.69963Y_{t-2} - 0.80909Y_{t-3} + 0.28517Y_{t-4} + 0.22656e_{t-1} - 0.77254e_{t-2}$$

เมื่อ $\hat{Y}_t$ แทนค่าพยากรณ์อัตราเร็วลมรายวัน ณ เวลา t

Y_{t-j} แทนอัตราเร็วลมรายวัน ณ เวลา $t-j$

e_{t-j} แทนความคลาดเคลื่อนของอัตราเร็วลมรายวันที่ได้จากการพยากรณ์ ณ เวลา $t-j$

การศึกษาครั้งต่อไปควรมีการศึกษาการสร้างตัวแบบพยากรณ์อัตราเร็วลมรายวันด้วยวิธีการพยากรณ์อื่นๆ เช่น วิธีการพยากรณ์ในลักษณะของตัวแบบพลศาสตร์ของไหลเชิงคำนวณเข้ามาประยุกต์ใช้กับการจำลองแผนที่ลม ซึ่งสามารถแสดงผลการจำลองในพื้นที่ซับซ้อนได้ใกล้เคียงกับลมที่เกิดขึ้นจริง (Thongyai & Assawamartbunlue, 2017) หรือวิธีโครงข่ายประสาทเทียม (Artificial Neural Networks) รวมทั้งควรเพิ่มเติมข้อมูลอัตราเร็วลมให้เป็นปัจจุบัน เพื่อปรับปรุงตัวแบบพยากรณ์ที่ได้ให้มีความเหมาะสมและมีค่าพยากรณ์ที่มีความคลาดเคลื่อนน้อยที่สุด อีกทั้งเพื่อประโยชน์ในด้านการวางแผนและการพัฒนาเทคโนโลยีกังหันลมภายในประเทศ ควรสร้างตัวแบบพยากรณ์อัตราเร็วลมที่ระดับความสูงอื่นๆ เช่น 50 และ 100 เมตร เป็นต้น ถึงแม้ว่าที่ระดับความสูงต่ำๆ จะทำให้ได้อัตราเร็วลมที่ต่ำ และผลิตหน่วยไฟฟ้าได้น้อย แต่อย่างไรก็ตาม การสร้างกังหันลมผลิตไฟฟ้าที่ระดับความสูงต่ำๆ จะมีค่าใช้จ่ายน้อยตามไปด้วย

การศึกษาครั้งนี้พบว่า วิธีบ็อกซ์-เจนกินส์เป็นวิธีที่ให้ตัวแบบพยากรณ์เหมาะสมมากที่สุดในการพยากรณ์อัตราเร็วลมรายวัน ที่ระดับความสูง 120 เมตร อำเภอปากพ่อง จังหวัดนครศรีธรรมราช จึงสามารถนำผลการพยากรณ์อัตราเร็วลมรายวันจากตัวแบบดังกล่าวมาใช้ประโยชน์ในการทำนายปริมาณหน่วยไฟฟ้าที่จะผลิตได้ในอนาคตต่อไป ซึ่งพลังงานลมเป็นรูปแบบพลังงานอย่างหนึ่งที่น่าสนใจและมีศักยภาพที่จะใช้เป็นแหล่งพลังงานทดแทนสำหรับการผลิตพลังงานไฟฟ้า ข้อดีของพลังงานลมที่ดีกว่าพลังงานทดแทนชนิดอื่นๆ คือ พลังงานลมไม่จำเป็นต้องใช้เชื้อเพลิงในการผลิตพลังงานจึงไม่ต้องการจัดหาเชื้อเพลิง หรือควบคุมกากเชื้อเพลิงที่อาจก่อให้เกิดมลพิษต่อสภาวะแวดล้อมเหมือนกับพลังงานชีวมวล นอกจากนี้พลังงานลมยังมีโอกาสเกิดขึ้นได้ทั้งกลางวันและกลางคืน เมื่อเปรียบเทียบกับพลังงานแสงอาทิตย์ ที่สามารถผลิตพลังงานได้เฉพาะกลางวัน อีกทั้งในการนำพลังงานลมมาใช้ในการผลิตพลังงานไฟฟ้าไม่จำเป็นต้องใช้พื้นที่มากเมื่อเทียบกับการผลิตพลังงานไฟฟ้าด้วยพลังงานน้ำ แต่อย่างไรก็ตาม พลังงานลมอาจมีข้อจำกัดเกี่ยวกับปริมาณของลมซึ่งอาจขึ้นอยู่กับลักษณะ

ภูมิศาสตร์และสภาพอากาศที่ต้องเอื้อต่อการเกิดลม จากการสำรวจแล้วมีเพียงบางพื้นที่ในประเทศไทยเท่านั้นที่มีความเร็วลมเหมาะสมในการนำพลังงานลมมาใช้ประโยชน์ได้อย่างคุ้มค่า (Promdee, 2016)

กิตติกรรมประกาศ

ขอขอบคุณศูนย์วิจัยพลังงานและสิ่งแวดล้อม มหาวิทยาลัยทักษิณ ที่เอื้อเฟื้อข้อมูลอัตราเร็วลม สำหรับการวิจัยครั้งนี้

เอกสารอ้างอิง

- Bielecki, M. F. (2010). *Statistical characterization of errors in wind power forecasting* (M.S. thesis) Department of Engineering, Faculty of Engineering, Northern Arizona University, U.S.A.
- Bowerman, B. L. & O'Connell, R. T. (1993). *Forecasting and Time Series: An Applied Approach* (3rd ed.). California: Duxbury Press.
- Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control* (3rd ed.). New Jersey: Prentice Hall.
- Keerativibool, W. & Mahileh, J. (2011). Forecasting model of wind speed along the coast of Songkhla province. *Journal of Energy Research*, 8(3), 63-72. (in Thai)
- Keerativibool, W. & Mahileh, J. (2013). Forecasting model of wind speed along the coast of Tha Sala district, Nakhon Si Thammarat province. *KKU Research Journal*, 18(1), 32-50. (in Thai)
- Keerativibool, W., Waewsak, J. & Kanjnasamranwong, P. (2011). *Short-term forecast of wind speed at Chana district, Songkhla province Thailand*. pp. 343-350. In The 2011 International Conference on Alternative Energy in Developing Countries and Emerging Economies, 25-28 May 2011, Hat Yai, Thailand.
- Ket-iam, S. (2005). *Forecasting Technique* (2nd ed.). Songkhla, Thailand: Thaksin University. (in Thai)
- Manmin, M. (2006). *Time Series and Forecasting*. Bangkok, Thailand: Foreprinting. (in Thai)
- Promdee, Ch. (2016). *Electric power generation of savonius wind turbine with double wind tunnels* (M.S. thesis) Department of Electrical and Computer Engineer, Faculty of Engineer, Mahasarakham University, Bangkok. (in Thai)
- Riansut, W. (2018). Comparison of tangerine prices forecast model by exponential smoothing methods. *Thai Journal of Science and Technology*, 7(Supplement Issue 5), 460-470. (in Thai)
- Riansut, W. (2019). A comparison of forecasting models of live swine price. *RMUTSV Research Journal*, 11(2), 349-365. (in Thai)

- Thongyai N. & Assawamartbunlue, K., (2017). *Wind atlas simulation to find wind potential area to generate electric power of the Chanthaburi and Trat provinces, Thailand.* pp. 737-745. In The 31st Conference of Mechanical Engineering Network of Thailand, 4-7 July 2017, Faculty of Engineering, Srinakharinwirot University, Bangkok. (in Thai)
- Waewsak, J., Kongruang, C., Tirawanichakul, Y., Tirawanichakul, S., Matan, N., Promphat, C. & Noo-Thong, A. (2008). *The feasibility study of wind farm power plants along the coastal lines of southern Thailand.* Bangkok, Thailand: Office of the National Research Council of Thailand. (in Thai)

Research Article

การแพร่กระจายและแนวโน้มการเปลี่ยนแปลงเชิงพื้นที่และเวลา ของอุณหภูมิผิวน้ำทะเลในอ่าวไทยจากข้อมูลภาพถ่ายดาวเทียม

The spatial and temporal variations and trends of sea surface temperature in the Gulf of Thailand from satellite imageries

ธวัชชัย นาอุดม^{1*} อานุกุล บุณนประทีปรัตน์² และ ศิราพร ทองอุดม²

Tawatchai Na-u-dom^{1*} Anukul Buranapratheprat² and Siraporn Tong-u-dom²

¹ คณะวิทยาศาสตร์และสังคมศาสตร์ มหาวิทยาลัยบูรพา วิทยาเขตสระแก้ว

² ภาควิชาวาริชศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา

¹ Faculty of Science and Social Sciences, Burapha University, Sa Kaeo Campus

² Department of Aquatic Sciences, Faculty of Science, Burapha University

*E-mail: tawatchai.na@buu.ac.th

Received: 20/01/2020; Revised: 17/05/2020; Accepted: 30/05/2020

บทคัดย่อ

การเพิ่มขึ้นของอุณหภูมิที่ผิวน้ำทะเลได้ส่งผลกระทบต่อระบบนิเวศและทรัพยากรทางทะเลในปัจจุบันอย่างชัดเจน งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อศึกษาการแพร่กระจายและแนวโน้มการเปลี่ยนแปลงเชิงพื้นที่และเวลาของอุณหภูมิที่ผิวน้ำทะเลในอ่าวไทยจากดาวเทียม Aqua MODIS ตั้งแต่เดือนมกราคม พ.ศ. 2546 ถึง เดือนธันวาคม พ.ศ. 2560 (15 ปี) ซึ่งพบว่าการแพร่กระจายเชิงพื้นที่และเวลาของอุณหภูมิผิวน้ำทะเลเฉลี่ยรายปีในอ่าวไทยนั้นมีค่าใกล้เคียงกันในแต่ละปี (ประมาณ 29 ถึง 31 องศาเซลเซียส) และเพิ่มสูงขึ้น 1 ถึง 2 องศาเซลเซียสในช่วงปีเอลนีโญ โดยเฉพาะอย่างยิ่งบริเวณอ่าวไทยตอนบนและชายฝั่งบริเวณอ่าวไทยตอนล่าง นอกจากนี้อุณหภูมิผิวน้ำทะเลในอ่าวไทยมีแนวโน้มเพิ่มสูงขึ้นตามเวลา (ประมาณ 0.02 ถึง 0.06 องศาเซลเซียสต่อปี) โดยอัตราการเปลี่ยนแปลงจะแตกต่างกันไปในแต่ละพื้นที่ โดยเฉพาะอย่างยิ่งในบริเวณอ่าวไทยตอนบนและชายฝั่งทะเลบริเวณอ่าวไทยตอนล่างที่มีแนวโน้มการเปลี่ยนแปลงเชิงเวลาสูงกว่าบริเวณอื่นๆ

คำสำคัญ: อ่าวไทย, อุณหภูมิผิวน้ำทะเล, การเปลี่ยนแปลงเชิงพื้นที่และเวลา, Aqua MODIS

Abstract

Rising sea temperature have been obviously affected marine ecosystems and resources. This research aims to investigate the spatio – temporal distribution and trends of sea surface temperature in the Gulf of Thailand by using sea surface temperature data retrieved from Aqua MODIS from January 2003 to December 2017 (15 years). The results illustrated that the annual average sea surface temperature was normally between 29 to 31 °C and increased by approximately 1 to 2 °C during the El Niño condition, especially in the northern and southern Gulf of Thailand. The increasing trend was between 0.02 and 0.06 °C/year and its variation depended on sub-regions. High temporal variations were found in the upper Gulf of Thailand and along the coast of the lower gulf.

Keywords: Gulf of Thailand, Spatio – temporal distribution and trends, Sea surface temperature, Aqua MODIS

บทนำ

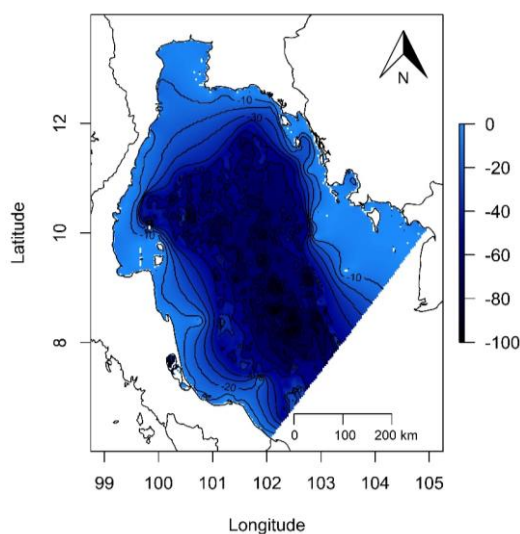
อ่าวไทยเป็นอ่าวกึ่งปิด มีพิกัดทางภูมิศาสตร์ที่ละติจูด 6 ถึง 14 องศาเหนือ และ ลองจิจูดที่ 99 ถึง 105 องศาตะวันออก มีความลึกเฉลี่ยประมาณ 45 เมตร โดยกลางอ่าวมีลักษณะเป็นแอ่งกระทะลึกประมาณ 80 เมตร มีแผ่นดินโอบล้อม 3 ด้านคือด้านเหนือ ด้านตะวันออกและตะวันตก โดยทางตอนใต้คือปากอ่าวมีความกว้างประมาณ 380 กิโลเมตรเชื่อมต่อกับทะเลจีนใต้ (South China Sea) (รูปที่ 1) สภาพภูมิอากาศของอ่าวไทยได้รับอิทธิพลจากลมมรสุมตะวันออกเฉียงเหนือและตะวันตกเฉียงใต้ ซึ่งส่งผลต่อการไหลเวียนของกระแสน้ำและระยะเวลาพำนักของมวลน้ำในอ่าวไทย นอกจากนี้ยังเป็นที่ยอมรับว่าท่าจากแม่น้ำสายสำคัญหลายสาย เช่น แม่น้ำเจ้าพระยา แม่น้ำท่าจีน แม่น้ำบางปะกง แม่น้ำตาปี เป็นต้น (Buranapratheprat & Bunpaong, 1998; Leenawarat et al., 2019)

การเพิ่มขึ้นของอุณหภูมิน้ำทะเลอันเนื่องมาจากการเปลี่ยนแปลงสภาพภูมิอากาศจากการกระทำของมนุษย์เป็นปัญหาที่มีความสำคัญในระดับโลกในปัจจุบัน (Yu et al., 2019; Ca, 2017; Park & Choi, 2017) สำหรับพื้นที่อ่าวไทยนั้น พบว่าอุณหภูมิของน้ำทะเลมีการเพิ่มขึ้นอย่างต่อเนื่องในช่วง 30 ปีที่ผ่านมา (Koad et al., 2012) ซึ่งการเปลี่ยนแปลงของอุณหภูมิที่ผิวน้ำทะเล (Sea surface temperature) เป็นปัจจัยสำคัญที่ส่งผลกระทบต่อระบบนิเวศและทรัพยากรทางทะเลโดย Siswanto & Tanaka (2014) และ Ackaboot et al. (2015) พบว่าประชากรแพลงก์ตอนพืชซึ่งเป็นผู้ผลิตเบื้องต้นที่สำคัญในระบบนิเวศทางทะเลมีความสัมพันธ์เชิงบวกกับการไหลของกระแสน้ำและอุณหภูมิปรากฏการณ์น้ำเปลี่ยนสี (Phytoplankton bloom) ที่เกิดจากการเพิ่มจำนวนขึ้นอย่างรวดเร็วของประชากรแพลงก์ตอนพืชในอ่าวไทยนั้นถือได้ว่าเป็นปัญหาที่มีความสำคัญต่อทรัพยากรทางทะเลและชายฝั่ง เนื่องจากก่อให้เกิดความ

เสียหายต่อการเพาะเลี้ยงสัตว์น้ำและทรัพยากรสัตว์น้ำชายฝั่งโดยเฉพาะอย่างยิ่งสัตว์น้ำหน้าดิน, การเกิดการฟอกขาวในหอยมือเสือ (Giant clams bleaching) (Junchompoo et al., 2013) รวมถึงการเกิดปรากฏการณ์ปะการังฟอกขาวในหลายพื้นที่ของอ่าวไทย (Chavanich et al., 2009; Hoeksema & Matthews, 2011; Sutthacheep et al., 2013)

โดยศึกษาการเปลี่ยนแปลงเชิงพื้นที่และเวลาของอุณหภูมิผิวน้ำทะเลในอดีตที่ผ่านมาเน้นการเก็บตัวอย่างจากพื้นที่ศึกษาในระดับภูมิภาค (Local scale) เช่น อ่าวไทยตอนบน (Na-u-dom et al., 2013) อันเนื่องมาจากข้อจำกัดด้านเวลาและงบประมาณวิจัย ทำให้ขาดความต่อเนื่องของข้อมูลทั้งในเชิงพื้นที่และเวลา ข้อมูลจากการรับรู้จากระยะไกล (Remote sensing) จึงได้ถูกนำใช้อย่างแพร่หลายเพื่อลดข้อจำกัดดังกล่าว (Boyer et al., 2009; Xu et al., 2013; Siswanto & Tanaka, 2014; Devi et al., 2015; Kratzer et al., 2014; Siswanto et al., 2016; Suwanlertcharoen & Prukpitikul, 2018) ซึ่งจะช่วยให้ได้ความต่อเนื่องของข้อมูลและทราบถึงการเปลี่ยนแปลงสภาพแวดล้อมและผลกระทบจากการเปลี่ยนแปลงสภาพภูมิอากาศในอ่าวไทยได้ชัดเจนยิ่งขึ้น

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาการแพร่กระจายและแนวโน้มการเปลี่ยนแปลงเชิงพื้นที่และเวลาของอุณหภูมิที่ผิวน้ำทะเลในอ่าวไทยซึ่งสามารถนำไปใช้เป็นข้อมูลพื้นฐานในการจัดการด้านทรัพยากรธรรมชาติและสิ่งแวดล้อมทางทะเลภายใต้สภาวะการเปลี่ยนแปลงสภาพภูมิอากาศเป็นพื้นฐานในการเตรียมความพร้อม การประเมินความเสี่ยงและผลกระทบเชิงพื้นที่ ซึ่งเป็นผลจากการเปลี่ยนแปลงสภาพภูมิอากาศในอ่าวไทย รวมถึงนำไปเป็นข้อมูลพื้นฐานในการกำหนดนโยบายด้านการจัดการทางด้านนิเวศวิทยาทางทะเลและชายฝั่งแบบบูรณาการในอ่าวไทยได้อย่างมีประสิทธิภาพต่อไปในอนาคต



รูปที่ 1 ขอบเขตและความลึกของอ่าวไทย โดยแถบสีและเส้นคอนทัวร์แสดงถึงความลึกน้ำ (เมตร)

วิธีการวิเคราะห์

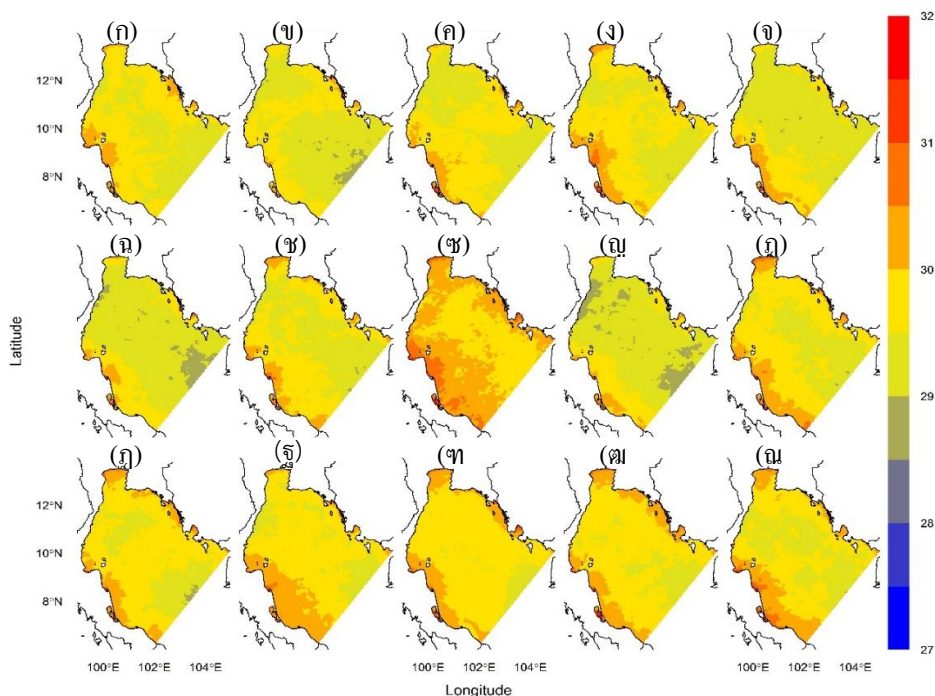
การศึกษาครั้งนี้ใช้ข้อมูลอุณหภูมิที่ผิวน้ำทะเลระดับ 3 (Level 3) จากดาวเทียมอะควา (Aqua) ซึ่งตรวจวัดโดยเซ็นเซอร์ โมดิส (Moderate resolution imaging spectroradiometer; MODIS) มีความละเอียดเชิงพื้นที่ (Spatial resolution) ประมาณ 4 ตารางกิโลเมตร ซึ่งสามารถเข้าถึงได้จาก <https://oceancolor.gsfc.nasa.gov/13> จากนั้นจึงดาวน์โหลดข้อมูลรายเดือนตั้งแต่เดือนมกราคม พ.ศ. 2546 ถึง เดือนธันวาคม พ.ศ. 2560 (15 ปี) เพื่อใช้ในการศึกษาการการแพร่กระจายและแนวโน้มเปลี่ยนแปลงเชิงพื้นที่และเวลาของอุณหภูมิที่ผิวน้ำทะเล โดยค่าความคลาดเคลื่อน (Bias) และส่วนเบี่ยงเบนมาตรฐาน (Standard deviation) ของข้อมูลอุณหภูมิที่ผิวน้ำทะเลที่ได้จากเซ็นเซอร์ โมดิส เทียบกับข้อมูลตรวจวัดจริงในบริเวณทะเลจีนใต้ อยู่ในช่วง -0.19 ถึง -0.34 และ 0.58 ถึง 0.68 องศาเซลเซียส ตามลำดับ (Qin et al., 2014)

การศึกษาการเปลี่ยนแปลงรายฤดูกาล จะแยกเป็น 4 ช่วงได้แก่ช่วงฤดูมรสุมตะวันออกเฉียงเหนือ (เดือนพฤศจิกายนถึงเดือนมกราคม) ช่วงการเปลี่ยนฤดูมรสุมจากตะวันออกเฉียงเหนือเป็นตะวันตกเฉียงใต้ (เดือนกุมภาพันธ์ถึงเดือนเมษายน) ช่วงฤดูมรสุมตะวันตกเฉียงใต้ (เดือนพฤษภาคมถึงเดือนสิงหาคม) และช่วงการเปลี่ยนฤดูมรสุมจากตะวันตกเฉียงใต้เป็นตะวันออกเฉียงเหนือ (เดือนกันยายนถึงเดือนตุลาคม) เพื่อศึกษาการแพร่กระจายและแนวโน้มการเปลี่ยนแปลงทั้งรายปีและรายฤดูกาลโดยวิธีการหาค่าเฉลี่ยและการวิเคราะห์การถดถอยเชิงเส้น (Linear regression) (Fang et al., 2006; Al-Rashidi et al., 2009) ตามลำดับ โดยพิจารณาการแนวโน้มการเปลี่ยนแปลงที่มีนัยสำคัญทางสถิติที่ระดับ 0.01 และ 0.05 ตามลำดับ ด้วยโปรแกรม R รุ่น 3.5.0 (R Core Team, 2018) และแพ็คเกจด้านภูมิสารสนเทศในการวิเคราะห์ข้อมูล (Na-u-dom, 2018; Pebesma et al., 2018; Dunnington, 2017; Hijmans, 2017)

ผลการวิเคราะห์

1. การแพร่กระจายเชิงพื้นที่และเวลาของอุณหภูมิผิวน้ำทะเลรายปี

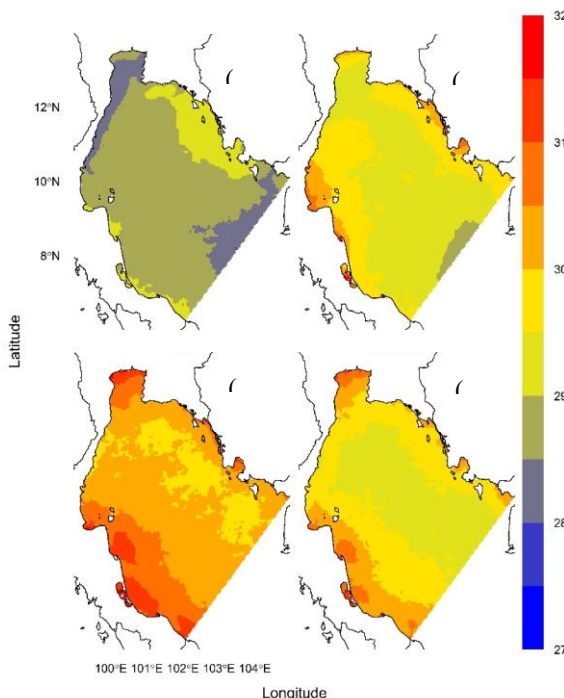
ผลการศึกษาการแพร่กระจายของอุณหภูมิที่ผิวน้ำทะเลรายปี ตั้งแต่ปี พ.ศ. 2546 ถึง 2560 ดังแสดงในรูปที่ 2 พบว่าการแพร่กระจายของอุณหภูมิกำลังใกล้เคียงกันในแต่ละปี (ประมาณ 29 ถึง 31 องศาเซลเซียส) โดยมวลน้ำที่อยู่ในบริเวณปากอ่าวไทยจะมีอุณหภูมิต่ำกว่ามวลน้ำภายในอ่าวและบริเวณชายฝั่ง โดยเฉพาะอย่างยิ่งบริเวณอ่าวไทยตอนล่าง นอกจากนี้ยังพบว่าในปี พ.ศ. 2550 (รูปที่ 2(จ)), 2551 (รูปที่ 2(ฉ)) และ 2554 (รูปที่ 2(ญ)) เป็นช่วงที่อุณหภูมิผิวน้ำทะเลมีค่าต่ำกว่าช่วงอื่น ๆ ทั้งบริเวณปากอ่าวและในอ่าว ยกเว้นบริเวณตอนในของอ่าวไทยตอนบน และชายฝั่งอ่าวไทยตอนล่างตั้งแต่จังหวัดสุราษฎร์ธานีถึงจังหวัดนครศรีธรรมราชที่ค่าอุณหภูมิผิวน้ำทะเลสูงกว่าพื้นที่อื่น ๆ ในช่วงเวลาเดียวกัน อย่างไรก็ตามในช่วงพ.ศ. 2553 (รูปที่ 2(ซ)), 2557 (รูปที่ 2(ฐ)) และ 2558 (รูปที่ 2(ฑ)) มวลน้ำอ่าวไทยมีอุณหภูมิสูงกว่าช่วงเวลาอื่น ๆ ประมาณ 1 ถึง 2 องศาเซลเซียส



รูปที่ 2 การแพร่กระจายเชิงพื้นที่และเวลาของอุณหภูมิผิวน้ำทะเลรายปีตั้งแต่พ.ศ. (ก) 2546, (ข) 2547, (ค) 2548, (ง) 2549, (จ) 2550, (ฉ) 2551, (ช) 2552, (ซ) 2553, (ญ) 2554, (ฎ) 2555, (ฉ) 2556, (ช) 2557, (ญ) 2559 และ (ฎ) 2560 แถบสีแสดงถึงระดับของอุณหภูมิผิวน้ำทะเล (องศาเซลเซียส)

2. การแพร่กระจายเชิงพื้นที่อุณหภูมิผิวน้ำทะเลเฉลี่ยรายฤดูกาล

พิจารณาการแพร่กระจายของอุณหภูมิผิวน้ำทะเลเฉลี่ยรายฤดูกาล (รูปที่ 3) พบว่าอุณหภูมิผิวน้ำทะเลมีค่าต่ำสุด (ประมาณ 27 ถึง 29 องศาเซลเซียส) ในช่วงฤดูมรสุมตะวันออกเฉียงเหนือ (รูปที่ 3(ก)) โดยเฉพาะอย่างยิ่งบริเวณอ่าวไทยตอนบนถึงชายฝั่งทะเลบริเวณจังหวัดประจวบคีรีขันธ์และปากอ่าวไทย จากนั้นอุณหภูมิจะค่อย ๆ เพิ่มขึ้นเมื่อเข้าสู่ช่วงการเปลี่ยนฤดูมรสุมจากตะวันออกเฉียงเหนือเป็นตะวันตกเฉียงใต้ (ประมาณ 29 ถึง 30 องศาเซลเซียส) (รูปที่ 3(ข)) โดยเฉพาะอย่างยิ่งบริเวณชายฝั่ง และมีค่าสูงสุดในช่วงฤดูมรสุมตะวันตกเฉียงใต้ (มากกว่า 30 องศาเซลเซียส) (รูปที่ 3(ค)) โดยเฉพาะบริเวณอ่าวไทยตอนบนและชายฝั่งทะเลอ่าวไทยตอนล่าง อย่างไรก็ตาม อุณหภูมิที่ผิวน้ำทะเลจะลดลงเมื่อเข้าสู่ช่วงการเปลี่ยนฤดูมรสุมจากตะวันตกเฉียงใต้เป็นตะวันออกเฉียงเหนือ (รูปที่ 3(ง)) แต่ในอ่าวไทยตอนบนและชายฝั่งทะเลอ่าวไทยตอนล่างก็ยังมีอุณหภูมิสูงกว่าเมื่อเทียบกับบริเวณโดยรอบ รวมถึงรูปแบบการแพร่กระจายของอุณหภูมิที่ผิวน้ำทะเลก็ยังคล้ายคลึงกับรูปแบบการแพร่กระจายในช่วงฤดูมรสุมตะวันตกเฉียงใต้แต่มีอุณหภูมิที่ต่ำกว่า

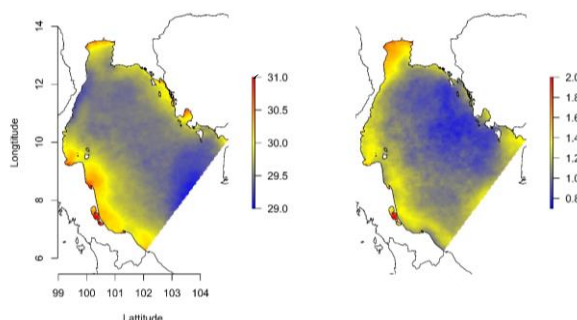


รูปที่ 3 การแพร่กระจายเชิงพื้นที่ของอุณหภูมิผิวน้ำทะเลเฉลี่ยใน (ก) ฤดูมรสุมตะวันออกเฉียงเหนือ (ข) ช่วงการเปลี่ยนฤดูมรสุมจากตะวันออกเฉียงเหนือเป็นตะวันตกเฉียงใต้ (ค) ฤดูมรสุมตะวันตกเฉียงใต้ และ (ง) ช่วงการเปลี่ยนฤดูมรสุมจากตะวันตกเฉียงใต้เป็นตะวันออกเฉียงเหนือ โดยแถบสีแสดงถึงระดับของอุณหภูมิผิวน้ำทะเล (องศาเซลเซียส)

3. การแพร่กระจายเชิงพื้นที่ของอุณหภูมิผิวน้ำทะเลเฉลี่ยรายปีตั้งแต่เดือนมกราคม พ.ศ. 2546 ถึง เดือนธันวาคม พ.ศ. 2560

อุณหภูมิผิวน้ำทะเลเฉลี่ยรายปีและส่วนเบี่ยงเบนมาตรฐาน (รูปที่ 4) นั้นมีรูปแบบการแพร่กระจายเชิงพื้นที่อย่างชัดเจนโดยมีค่าอยู่ในช่วงระหว่าง 29 ถึง 31 องศาเซลเซียส โดยค่าเฉลี่ยของอุณหภูมิผิวน้ำทะเลบริเวณชายฝั่งทะเล มีค่าประมาณ 30 ถึง 31 องศาเซลเซียส โดยเฉพาะอย่างยิ่งบริเวณอ่าวไทยตอนบนและชายฝั่งทะเลของอ่าวไทยตอนล่าง ซึ่งแตกต่างจากมวลน้ำบริเวณกลางอ่าวที่มีอุณหภูมิต่ำกว่า (ประมาณ 29 องศาเซลเซียส) (รูปที่ 4 ก))

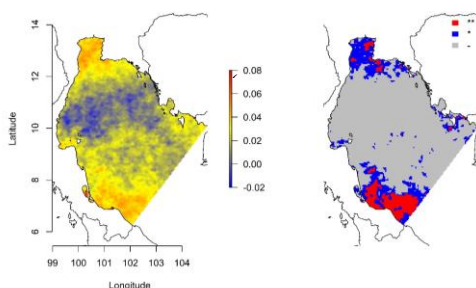
ส่วนเบี่ยงเบนมาตรฐาน (ความแปรปรวน) ของอุณหภูมิผิวน้ำทะเลในแต่ละปี (รูปที่ 4(ข)) ในอ่าวไทยมีค่าตั้งแต่ 0.75 ถึง 2.00 องศาเซลเซียส โดยในบริเวณชายฝั่งมีค่าความแปรปรวนจากค่าอุณหภูมิผิวน้ำทะเลเฉลี่ยสูงกว่ามวลน้ำในบริเวณกลางอ่าว โดยเฉพาะอย่างยิ่งบริเวณอ่าวไทยตอนบนและบริเวณทะเลสาบสงขลา



รูปที่ 4. (ก) การแพร่กระจายเชิงพื้นที่ของอุณหภูมิผิวน้ำทะเลเฉลี่ยรายปี และ (ข) ส่วนเบี่ยงเบนมาตรฐานของอุณหภูมิผิวน้ำทะเลเชิงพื้นที่ โดยแถบสีแสดงถึงระดับของอุณหภูมิผิวน้ำทะเลและค่าส่วนเบี่ยงเบนมาตรฐาน (องศาเซลเซียส) ตามลำดับ

4. การแพร่กระจายเชิงพื้นที่ของแนวโน้มการเปลี่ยนแปลงอุณหภูมิผิวน้ำทะเลเฉลี่ยรายปีตั้งแต่เดือนมกราคม พ.ศ. 2546 ถึง เดือนธันวาคม พ.ศ. 2560

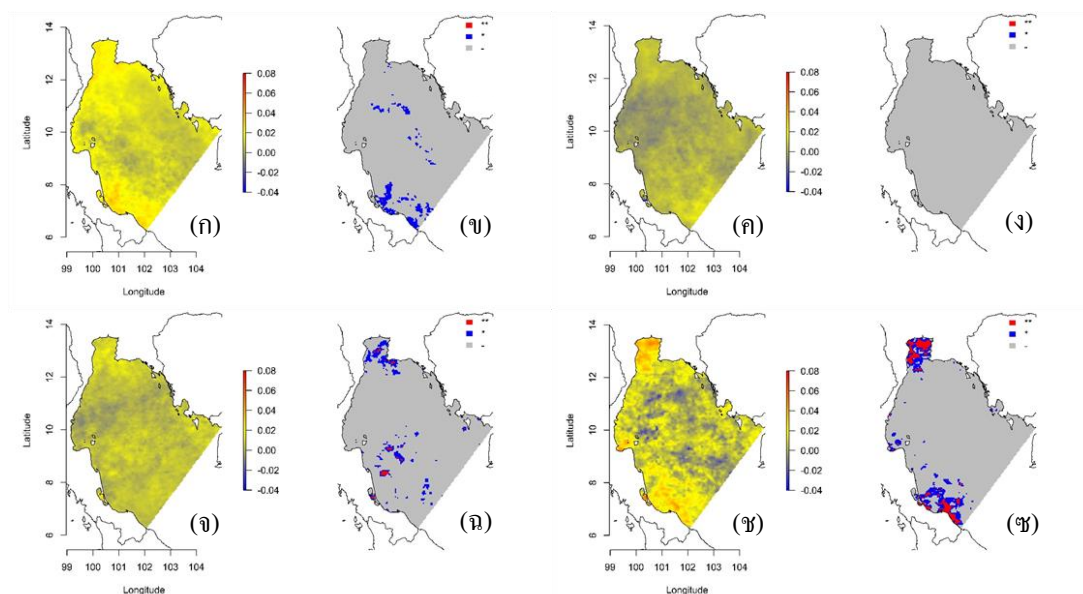
จากผลการศึกษาการแพร่กระจายเชิงพื้นที่ของแนวโน้มการเปลี่ยนแปลงอุณหภูมิผิวน้ำทะเลเฉลี่ยรายปี (รูปที่ 4) พบว่าอุณหภูมิที่ผิวน้ำทะเลบริเวณอ่าวไทยตอนบนและชายฝั่งทะเลอ่าวไทยตอนล่างเพิ่มสูงขึ้นอย่างมีนัยสำคัญและอย่างมีนัยสำคัญยิ่งทางสถิติ ($0.01 < p\text{-value} < 0.05$ และ $p\text{-value} < 0.01$ ตามลำดับ) โดยเฉพาะอย่างยิ่งในบริเวณจังหวัดชลบุรี ระยองและตั้งแต่บริเวณจังหวัดสงขลาลงไปที่มีมวลน้ำในบางบริเวณมีอุณหภูมิสูงขึ้นอย่างมีนัยสำคัญทางสถิติ (ประมาณ 0.06 ถึง 0.08 องศาเซลเซียสต่อปี; $0.01 < p\text{-value} < 0.05$) ดังรูปที่ 4(ก) และ (ข) ตามลำดับ อย่างไรก็ตามมวลน้ำบริเวณกลางอ่าวนั้นมีอุณหภูมิที่ผิวน้ำทะเลลดลงเล็กน้อย ซึ่งการเปลี่ยนแปลงดังกล่าวนี้ไม่มีนัยสำคัญทางสถิติ ($p\text{-value} > 0.05$)



รูปที่ 5. (ก) การแพร่กระจายเชิงพื้นที่ของแนวโน้มการเปลี่ยนแปลงอุณหภูมิผิวน้ำทะเลเฉลี่ยรายปี โดยแถบสีแสดงถึงระดับของแนวโน้มการเปลี่ยนแปลง (องศาเซลเซียสต่อปี) และ (ข) ระดับนัยสำคัญ ($p\text{-value}$) โดยสีแดง (**) แสดงถึงบริเวณที่มีการเปลี่ยนแปลงที่มีนัยสำคัญยิ่งทางสถิติ ($p\text{-value} < 0.01$) สีน้ำเงิน (*) แสดงถึงบริเวณที่มีการเปลี่ยนแปลงที่มีนัยสำคัญทางสถิติ ($0.01 < p\text{-value} < 0.05$) และสีเทา (-) แสดงถึงบริเวณที่มีการเปลี่ยนแปลงอย่างไม่มีนัยสำคัญทางสถิติ ($p\text{-value} > 0.05$)

5. การแพร่กระจายเชิงพื้นที่ของแนวโน้มการเปลี่ยนแปลงอุณหภูมิผิวน้ำทะเลเฉลี่ยรายฤดูกาล

จากการศึกษาการแพร่กระจายเชิงพื้นที่ของแนวโน้มการเปลี่ยนแปลงอุณหภูมิผิวน้ำทะเลเฉลี่ยรายฤดูกาล (รูปที่ 4) แสดงให้เห็นว่าในช่วงฤดูมรสุมตะวันออกเฉียงเหนือ ช่วงเปลี่ยนฤดูมรสุมจากตะวันออกเฉียงเหนือเป็นตะวันตกเฉียงใต้และฤดูมรสุมตะวันตกเฉียงใต้ (รูปที่ 4 (ก) (ค) และ (จ)) แนวโน้มการเปลี่ยนแปลงเชิงพื้นที่ของอุณหภูมิที่ผิวน้ำทะเลโดยส่วนมากในช่วงเวลาดังกล่าวมีการเปลี่ยนแปลง (ประมาณ ± 0.02 องศาเซลเซียสต่อปี) อย่างไม่มีนัยสำคัญทางสถิติ ($p\text{-value} > 0.05$) (รูปที่ 4 (ข) (ง) และ (ฉ)) อย่างไรก็ตามเมื่อเข้าสู่ช่วงการเปลี่ยนฤดูมรสุมจากตะวันตกเฉียงใต้เป็นตะวันออกเฉียงเหนือ อุณหภูมิของผิวน้ำทะเลมีแนวโน้มเพิ่มสูงขึ้นอย่างมีนัยสำคัญและอย่างมีนัยสำคัญยิ่งทางสถิติ ($0.01 < p\text{-value} < 0.05$ และ $p\text{-value} < 0.01$ ตามลำดับ) (รูปที่ 4(ซ) และ (ซ)) โดยเฉพะอย่างยิ่งในบริเวณอ่าวไทยตอนบนและชายฝั่งอ่าวไทยตอนล่างตั้งแต่จังหวัดสงขลาลงไป โดยมีแนวโน้มการเพิ่มขึ้นของอุณหภูมิที่ผิวน้ำทะเลประมาณ 0.04 ถึง 0.07 องศาเซลเซียสต่อปี



รูปที่ 6 การแพร่กระจายเชิงพื้นที่ของแนวโน้มการเปลี่ยนแปลงอุณหภูมิผิวน้ำทะเลเฉลี่ยรายฤดูกาล ใน (ก) ฤดูมรสุมตะวันออกเฉียงเหนือ, (ค) ช่วงการเปลี่ยนฤดูมรสุมจากตะวันออกเฉียงเหนือเป็นตะวันตกเฉียงใต้, (จ) ฤดูมรสุมตะวันตกเฉียงใต้ และ (ซ) ช่วงการเปลี่ยนฤดูมรสุมจากตะวันตกเฉียงใต้เป็นตะวันออกเฉียงเหนือ ตามลำดับ โดยแถบสีแสดงถึงระดับของแนวโน้มการเปลี่ยนแปลง (องศาเซลเซียสต่อปี) และ (ข ง ฉ และ ซ) ระดับนัยสำคัญ ($p\text{-value}$) ในแต่ละฤดูกาล โดยสีแดง (**) แสดงถึงบริเวณที่มีการเปลี่ยนแปลงที่มีนัยสำคัญยิ่งทางสถิติ ($p\text{-value} < 0.01$) สีน้ำเงิน (*) แสดงถึงบริเวณที่มีการเปลี่ยนแปลงที่มีนัยสำคัญทางสถิติ ($0.01 < p\text{-value} < 0.05$) และสีเทา (-) แสดงถึงบริเวณที่มีการเปลี่ยนแปลงอย่างไม่มีนัยสำคัญทางสถิติ ($p\text{-value} > 0.05$)

วิจารณ์ผลการทดลอง

จากการศึกษาการแพร่กระจายเชิงพื้นที่ของอุณหภูมิผิวน้ำทะเลเฉลี่ยรายปี (รูปที่ 2) เห็นได้ว่าปี พ.ศ. 2553 (รูปที่ 2(ข)) อุณหภูมิผิวน้ำทะเลมีค่าสูงเมื่อเทียบกับช่วงเวลาอื่น ๆ ประมาณ 1 ถึง 2 องศาเซลเซียส โดยเฉพาะอย่างยิ่งบริเวณอ่าวไทยตอนบนและชายฝั่งทะเลบริเวณอ่าวไทยตอนล่าง สะท้อนให้เห็นถึงความสัมพันธ์กับปรากฏการณ์เอลนีโญ (El Niño) ในช่วงเดือนมิถุนายน พ.ศ. 2552 ถึง เดือนเมษายน 2553 (Lorpittayakan, 2016) ซึ่งสถานการณ์การเพิ่มขึ้นของอุณหภูมิผิวน้ำทะเลสามารถส่งผลกระทบต่อระบบนิเวศทางทะเลโดย Junchompoo et al. (2013) และ Prukpittikul et al. (2013) ได้สรุปถึงสถานการณ์ของปะการังและหอยมือเสือบริเวณอ่าวไทย ในช่วงปี พ.ศ. 2553 พบว่าอยู่ในภาวะวิกฤต โดยหอยมือเสือบริเวณเกาะมันใน จังหวัดระยอง ที่เริ่มพบการฟอกขาวประมาณร้อยละ 90 ในเดือนพฤษภาคม พ.ศ. 2553 (Junchompoo et al., 2013) และการปรากฏการณ์ปะการังฟอกขาวในอ่าวไทยโดยเริ่มพบการฟอกขาวตั้งแต่ปลายเดือนมีนาคม พ.ศ. 2553 เกิดซึ่งมีสาเหตุมาจากอุณหภูมิผิวน้ำทะเลที่สูงขึ้น (มากกว่า 31 องศาเซลเซียส) ต่อเนื่องกันเป็นเวลานาน (8 ถึง 10 เดือน) รวมถึงปัจจัยทางธรรมชาติอื่นๆ เช่นความเข้มของแสงแดดที่เป็นจุดวิกฤติที่กระตุ้นให้เกิดการฟอกขาวของปะการัง (Prukpittikul et al., 2013)

ผลการศึกษาการแพร่กระจายเชิงพื้นที่ของอุณหภูมิผิวน้ำทะเลในอ่าวไทยเฉลี่ยรายฤดูกาลพบว่ามีค่าสูงที่สุด (ประมาณ 31 ถึง 32 องศาเซลเซียส) ในช่วงฤดูมรสุมตะวันตกเฉียงใต้ ซึ่งเป็นผลมาจากการผสมผสานของมวลน้ำจากพลังงานคลื่น (Wave mixing energy) ในฤดูมรสุมตะวันตกเฉียงใต้ที่มากกว่าในช่วงฤดูมรสุมตะวันออกเฉียงเหนือ ส่งผลให้อุณหภูมิที่ผิวน้ำทะเลอ่าวไทยในฤดูมรสุมตะวันตกเฉียงใต้สูงกว่าช่วงเวลาอื่น (Pithayamaythakul & Sojisuporn, 2016) โดย Luadnakrob & Buranapratheprat (2012) ในช่วงฤดูมรสุมตะวันตกเฉียงใต้นั้นค่าฟลักซ์ความร้อนรวมทั้งที่ผิวน้ำทะเลอ่าวไทย (Total heat flux) และค่าฟลักซ์ความร้อนจากการนำและการพา (Sensible heat flux) บริเวณทะเลฝั่งตะวันตกของอ่าวไทยนั้นมีค่าสูงกว่าชายฝั่งทะเลฝั่งตะวันออกนั้นมีค่าสูงกว่าบริเวณอื่น ๆ ในฤดูมรสุมตะวันตกเฉียงใต้และมีทิศทางเป็นลบซึ่งแสดงถึงการส่งผ่านความร้อนจากผิวน้ำทะเลสู่บรรยากาศ ซึ่งให้เห็นว่าบริเวณชายฝั่งทะเลตะวันตกของอ่าวไทยที่มีอุณหภูมิผิวน้ำสูงกว่าอุณหภูมิอากาศและอุณหภูมิที่ผิวน้ำ ณ บริเวณอื่น ๆ ในช่วงเวลาเดียวกัน อย่างไรก็ตามหากพิจารณาฟลักซ์ความร้อนรวมทั้งที่ผิวน้ำทะเลจาก Luadnakrob & Buranapratheprat (2012) มีค่าสูงที่สุดในช่วงเปลี่ยนฤดูมรสุมจากมรสุมตะวันออกเฉียงเหนือเป็นมรสุมตะวันตกเฉียงใต้แทนที่จะเป็นฤดูมรสุมตะวันตกเฉียงใต้ ซึ่งผลที่ได้นี้อาจชี้ให้เห็นถึงความเกี่ยวข้องกับระยะเวลาที่เกิดปฏิสัมพันธ์ (Time lag) ระหว่างฟลักซ์ความร้อนรวมและการเพิ่มขึ้นของอุณหภูมิที่ผิวน้ำทะเล ซึ่งเป็นประเด็นที่น่าสนใจและควรศึกษาเพิ่มเติมในอนาคต

นอกจากนี้ผลการศึกษาชี้ให้เห็นถึงอุณหภูมิผิวน้ำทะเลบริเวณอ่าวไทยตอนบนในช่วงฤดูมรสุมตะวันตกเฉียงใต้ ซึ่งมีค่าสูงกว่าบริเวณอ่าวไทยตอนล่างอย่างเห็นได้ชัด ซึ่งอาจสอดคล้องกับการลดลงของปริมาณน้ำท่า (Prasanchum et al., 2013; Vinayachandran et al., 2015) โดยเฉพาะอย่างยิ่งในกลุ่มน้ำเจ้าพระยา (Kim et al., 2005) ซึ่งเป็นผลมาจากการเปลี่ยนแปลงการใช้ที่ดิน (Land use change) และผลจากการเปลี่ยนแปลงสภาพภูมิอากาศ

(Kim et al., 2005; Prasanchum et al., 2013) พบว่ากับอิทธิพลของค่าพลาสมาความร้อนรวมที่ผิวหนังน้ำทะเลและค่าพลาสมาความร้อนจากการนำและการพาในช่วงฤดูมรสุมตะวันตกเฉียงใต้ (Luadnakrob & Buranapratheprat., 2012) จึงทำให้ในช่วงเวลาดังกล่าวมีอุณหภูมิผิวหนังน้ำทะเลสูงกว่าในช่วงเวลาอื่น ๆ

จากผลการศึกษาแนวโน้มการเปลี่ยนแปลงของอุณหภูมิผิวหนังน้ำทะเลพบว่ามีผลสอดคล้องกับการศึกษาของ Koad et al. (2012) โดยเฉพาะอย่างยิ่งในบริเวณจังหวัดสุราษฎร์ธานีและประจวบคีรีขันธ์ นอกจากนี้แนวโน้มการเปลี่ยนแปลงของอุณหภูมิผิวหนังน้ำทะเลเฉลี่ยทั้งรายปีและรายฤดูกาลที่มีแนวโน้มสูงขึ้นนั้นยังสอดคล้องกับการแนวโน้มการเปลี่ยนแปลงของอุณหภูมิผิวหนังน้ำทะเลในบริเวณใกล้เคียง เช่น ชายฝั่งทะเลของประเทศเวียดนามและทะเลจีนใต้ (Yu et al., 2019; Ca, 2017; Park & Choi, 2017)

สำหรับการแนวโน้มการเปลี่ยนแปลงของอุณหภูมิผิวหนังน้ำทะเล ในช่วงการเปลี่ยนฤดูมรสุมตะวันตกเฉียงใต้เป็นตะวันออกเฉียงเหนือที่มีแนวโน้มการเพิ่มขึ้นสูงสุดนั้น อาจเป็นผลมาจากอุณหภูมิผิวหนังน้ำทะเลที่สูงในช่วงมรสุมตะวันตกเฉียงใต้มีระยะเวลายาวนานมากขึ้นจนเลยเข้ามาสู่ช่วงเปลี่ยนฤดูมรสุม จึงทำให้อุณหภูมิผิวหนังน้ำทะเลในช่วงเปลี่ยนฤดูมรสุมมีการเปลี่ยนแปลงและมีค่าเพิ่มขึ้นมากกว่าในช่วงฤดูกาลอื่น ซึ่งจะเห็นได้จากรูปแบบการเชิงพื้นที่ของบริเวณที่มีแนวโน้มอุณหภูมิผิวหนังน้ำทะเลสูงขึ้นอย่างมีนัยสำคัญทางสถิติ ($p\text{-value} < 0.01$ และ $0.01 < p\text{-value} < 0.05$) นั้นสอดคล้องกับการแพร่กระจายเชิงพื้นที่ของบริเวณที่มีค่าอุณหภูมิผิวหนังน้ำทะเลเฉลี่ยสูงในช่วงการเปลี่ยนฤดูมรสุมตะวันตกเฉียงใต้เป็นตะวันออกเฉียงเหนือ

สำหรับประยุกต์ใช้เทคนิคการรับรู้จากระยะไกลในการศึกษารังนี้ชี้ให้เห็นถึงประสิทธิภาพของเทคนิคดังกล่าวในการประเมินสถานการณ์การเปลี่ยนแปลงด้านสิ่งแวดล้อมทางทะเลที่มีผลมาจากการเปลี่ยนแปลงสภาพภูมิอากาศที่มีความต่อเนื่องทั้งในเชิงพื้นที่และเวลา ซึ่งเป็นข้อมูลที่สำคัญในการประเมิน การเตรียมความพร้อม การจัดการ รวมถึงระบุพื้นที่ที่อาจได้รับผลกระทบจากการเปลี่ยนแปลงสภาพภูมิอากาศ ซึ่งควรมีการถึงการต่อยอดและบูรณาการข้อมูลระหว่างหน่วยงานที่เกี่ยวข้องทั้งข้อมูลตรวจวัดและข้อมูลที่ได้จากดาวเทียมรวมเพื่อให้ได้การติดตามและการประเมินผลกระทบสถานการณ์การเปลี่ยนแปลงด้านสิ่งแวดล้อมทางทะเลที่ได้รับผลจากการเปลี่ยนแปลงสภาพภูมิอากาศมีประสิทธิภาพมากยิ่งขึ้น ซึ่งจะนำไปสู่การระบุพื้นที่และการประเมินความเสี่ยงรวมถึงผลกระทบต่อระบบนิเวศทางทะเลที่อาจเกิดขึ้นจากเปลี่ยนแปลงสภาพภูมิอากาศในอ่าวไทยรวมถึงนำไปเป็นข้อมูลพื้นฐานในการกำหนดนโยบายด้านการจัดการทางด้านนิเวศวิทยาทางทะเลและชายฝั่งแบบในอ่าวไทยได้ต่อไปในอนาคต

สรุปผลการทดลอง

จากผลการศึกษาการแพร่กระจายและแนวโน้มการเปลี่ยนแปลงเชิงพื้นที่และเวลาของอุณหภูมิผิวหนังน้ำทะเลในอ่าวไทยนั้นสรุปได้ว่า การแพร่กระจายเชิงพื้นที่และเวลาของอุณหภูมิผิวหนังน้ำทะเลเฉลี่ยรายปีในอ่าวไทยมีค่าใกล้เคียงกัน (ประมาณ 29 ถึง 31 องศาเซลเซียส) แต่อย่างไรก็ตามอุณหภูมิผิวหนังน้ำทะเลในปี พ.ศ. 2553 นั้นมีค่าสูงที่สุดอัน

เนื่องมาจากปรากฏการณ์เอลนีโญ นอกจากนี้หากพิจารณาการแพร่กระจายของอุณหภูมิผิวน้ำทะเลเฉลี่ยราย ฤดูกาลพบว่าอุณหภูมิผิวน้ำทะเลฤดูมรสุมตะวันตกเฉียงใต้มีค่าสูงที่สุด โดยเฉพาะอย่างยิ่งในบริเวณชายฝั่งทะเลอ่าวไทยฝั่งตะวันตกและอ่าวไทยตอนบน นอกจากนี้ แนวโน้มการเปลี่ยนแปลงอุณหภูมิผิวน้ำทะเลชี้ให้เห็นว่าอุณหภูมิผิวน้ำทะเลในอ่าวไทยโดยทั่วไปมีแนวโน้มเพิ่มสูงขึ้น (ประมาณ 0.02 ถึง 0.06 องศาเซลเซียสต่อปี) โดยอัตราการเปลี่ยนแปลงนั้นจะแตกต่างกันไปในแต่ละพื้นที่ โดยเฉพาะอย่างยิ่งในบริเวณอ่าวไทยตอนบนและชายฝั่งทะเลตั้งแต่จังหวัดสงขลาลงไปนั้น มีการเปลี่ยนแปลงอย่างมีนัยสำคัญทางสถิติ ($p\text{-value} < 0.01$ และ $0.01 < p\text{-value} < 0.05$)

กิตติกรรมประกาศ

ขอขอบคุณคณะวิทยาศาสตร์และสังคมศาสตร์ มหาวิทยาลัยบูรพา วิทยาเขตสระแก้ว สำหรับสิ่งอำนวยความสะดวกในการทำวิจัยและผู้ประเมินบทความ (Reviewers) ที่ได้กรุณาให้คำแนะนำและข้อคิดเห็นที่เป็นประโยชน์ต่อการปรับปรุงบทความวิจัยฉบับนี้

เอกสารอ้างอิง

- Ackaboot, J., Daraneesrisuk, J., Kongtong, D., Trerattawittaya, T., Waewmayara, W., Jeasombunpanit, S., & Choosungwal, A. (2015). Monitoring plankton in the upper Gulf of Thailand through remote sensing. In Proceeding of The 8th Thai student symposium on geography and geoinformatics. (pp. 1-10).
- Al-Rashidi, T. B., El-Gamily, H. I., Amos, C. L., & Rakha, K. A. (2009). Sea surface temperature trends in Kuwait Bay, *Arabian Gulf. Natural Hazards*, 50(1), 73–82.
- Boyer, J. N., Kelble, C. R., Ortner, P. B., & Rudnick, D. T. (2009). Phytoplankton bloom status: Chlorophyll a biomass as an indicator of water quality condition in the southern estuaries of Florida, USA. *Ecological Indicators*, 9(6), 56–67.
- Buranapratheprat, A., & Bunapong, M. (1998). A two dimensional hydrodynamic model for the Gulf of Thailand. In Proceedings of the IOC/WESTPAC Fourth International Scientific Symposium. (pp. 469-478).
- Ca, V.T. (2017). A climate change assessment via trend estimation of certain climate parameters with in situ measurement at the coasts and islands of Viet Nam. *Climate*, 5(2), 36.
- Chavanich, S., Viyakarn, V., Loyjiw, T., Pattaratamrong, P., & Chankong, A. (2009). Mass bleaching of soft coral, sarcophyton spp. in Thailand and the role of temperature and salinity stress. *ICES Journal of Marine Science*, 66(7), 1515–1519.

- Devi, G. K., Ganasri, B. P., & Dwarakish, G. S. (2015). Applications of Remote Sensing in Satellite Oceanography: A Review. *Aquatic Procedia*, 4, 579–584.
- Dunnington, D. (2017). Package “prettymapr”. Retrived June 18, 2018, from <https://cran.r-project.org/web/packages/prettymapr/prettymapr.pdf>
- Fang, G., Chen, H., Wei, Z., Wang, Y., Wang, X., & Li, C. (2006). Trends and interannual variability of the South China Sea surface winds, surface height, and surface temperature in the recent decade. *Journal of Geophysical Research: Oceans*, 111, 1-16.
- Hijmans, R. J. (2017). Introduction to the “raster” package. Retrived March 16, 2018, from <https://cran.r-project.org/web/packages/raster/raster.pdf>
- Hoeksema, B. W., & Matthews, J. L. (2011). Contrasting bleaching patterns in mushroom coral assemblages at Koh Tao, Gulf of Thailand. *Coral Reefs*, 30(1), 95.
- Junchompoo, C., Sinrapasan, N., Penpian, C., & Patsorn, P. (2013). Changing seawater temperature effects on giant clams bleaching, Mannai Island, Rayong province, Thailand. In Proceeding of the design symposium on conservation of ecosystem. (pp 71–76).
- Kim, W., Kanae, S., Agata, Y., & Oki, T. (2005). Simulation of potential impacts of land use/cover changes on surface wate fluxes in the Chaophraya river basin, Thailand. *Journal of geophysical research*, 110, 2-10.
- Koad, P., Jaroensutasinee, M., & Jaroensutasinee, K. (2012). Sea surface temperature trends in the gulf of Thailand and the Andaman Sea. In 2012 OCEANS - Yeosu. (pp. 1–8).
- Kratzer, S., Therese Harvey, E., & Philipson, P. (2014). The use of ocean color remote sensing in integrated coastal zone management-A case study from Himmerfjärden, Sweden. *Marine Policy*, 43, 29–39.
- Leenawarat, D., Tong-u-dom, S., & Buranapratheprat., A. (2019). A hydrodynamic model for the study of the seasonal variation in the residence time of water mass in the Inner Gulf of Thailand. *Burapha Science Journal*, 24(1), 1-16.
- Lorpittayakorn, P. (2016). Relationship between sea surface temperature in the inner Gulf of Thailand and Rainfall in Eastern Thailand during ENSO period. *Thai Science and Technology Journal*, 24, 400 - 417.
- Luadnakrob., P. & Buranapratheprat., A. (2012). Annual surface heat flux in the Gulf of Thailand. *Burapha Science Journal*, 17(1), 77-86.
- Na-u-dom, T. (2018). Application of R Programming for the Geoinformatic Display. *The Journal of Applied Science*, 17(2), 118-128.

- Na-u-dom, T., Buranapratheprat, A., Homhual, K., & Intracharoen, P. (2013). Temporal and spatial variations of water qualities in the upper Gulf of Thailand during two seasons in 2009. *Burapha Science Journal*, 18(2013), 32–42.
- Park, Y.G., & Chai, A. (2017). Long-term changes of South China Sea sea surface temperatures in winter and summer. *Continental Shelf Research*, 143, 185-193.
- Pebesma, E., Bivand, R., Rowlingson, B., Gómez-Rubio, V., Hijmans, R., Summer, M. et al. (2018). Package “sp”. Retrived March 16, 2018, from <https://cran.r-project.org/web/packages/sp/sp.pdf>
- Pithayamaythakul., P. & Sojisuporn., P. (2016). Water stratification under wave influence in the Gulf of Thailand. *Bullein of Earth Sciences of Thailand*, 7, 33-41.
- Prasanchum, H., Kangrang, A., Hormwichian, R., & Compliew, S. (2013). Impact of climate and rapid landuse change of runoff wualities in Lower-Lampao river basin. In Proceeding of The 5th International conference on Science, Technology and Innovation for Sustainable Weill-Being (STISWB V). pp (1-6).
- Prukpitikul, S., Buakeaw, V., Kesdesh, W., Kongprom., A., & Kaewpoo, N. (2013). The assessment of coral situation in Thailand by using geo-informatics technology. *RMUTSV Research Journal*, 5(1), 61-77.
- Qin, H., Chen, G., Wang, W., Wang, D., & Zeng, L. (2014). Validation and application of MODIS-derived SST in the South China Sea. *International Journal of Remote Sensing*, 35, 4315-4328.
- R Core Team. (2018). R: A language and environment for statistical computing. Retrived May 16, 2018, from <http://www.r-project.org/>
- Siswanto, E., Honda, M. C., Matsumoto, K., Sasai, Y., Fujiki, T., Sasaoka, K., & Saino, T. (2016). Sixteen-year phytoplankton biomass trends in the northwestern Pacific Ocean observed by the SeaWiFS and MODIS ocean color sensors. *Journal of Oceanography*, 72(3), 479–489.
- Siswanto, E., & Tanaka, K. (2014). Phytoplankton biomass dynamics in the Strait of Malacca within the period of the seawifs full mission: Seasonal cycles, interannual variations and decadal-scale trends. *Remote Sensing*, 6(4), 2718–2742.
- Sutthacheep, M., Yucharoen, M., Klinthong, W., Pengsakun, S., Sangmanee, K., & Yeemin, T. (2013). Impacts of the 1998 and 2010 mass coral bleaching events on the Western Gulf of Thailand. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 96, 25–31.
- Suwanlertcharoen, T., & Prukpitikul, S. (2018). Detection of Phytoplankton Blooms in the Upper Gulf of Thailand Using Sentinel-3A OLCI Imagery. *Environment and Natural Resources Journal*, 16(1), 9–20.

- Vinayachandran, P.N., Jahfer, S., & Nanjundiah, R.S. (2015). Impact of river runoff into the ocean on Indian Summer Monsoon. *Environmental Research Letters*. 10, 1-10.
- Xu, Y., Ishizaka, J., Yamaguchi, H., Siswanto, E., & Wang, S. (2013). Relationships of interannual variability in SST and phytoplankton blooms with giantjellyfish (*Nemopilema nomurai*) outbreaks in the Yellow Sea and East China Sea. *Journal of Oceanography*, 69(5), 511–526.
- Yu, Y., Zhang, H.R., Jin, J., & Wang, T. (2019). Trends of sea surface temperature and sea surface temperature fronts in the South China Sea during 2003-2017. *Acta Oceanologica Sinica*, 38(4), 106-115.

Research Article

การใช้ประโยชน์ของเสียจากระบบบำบัดน้ำเสียโรงงานแป้งมัน สำปะหลังดัดแปร มาผลิตพลังงานทดแทนโดยระบบหมักไร้อากาศ แบบหนึ่งขั้นตอน

The application of wastes from modified tapioca starch plant to produce the renewable energy using one stage anaerobic digestion system

ปิยะวดี ศรีวิชัย^{1*} และ ชุมพร รอดสีดา²

Piyavadee Srivichai^{1*} and Chumaporn Rodsrida²

¹สาขาวิชานามัยสิ่งแวดล้อม คณะแพทยศาสตร์ มหาวิทยาลัยพะเยา อำเภอเมือง จังหวัดพะเยา 56000

²สาขาวิชาอาชีวอนามัยและความปลอดภัย คณะแพทยศาสตร์ มหาวิทยาลัยพะเยา อำเภอเมือง จังหวัดพะเยา 56000

¹Program in Environmental Health, Faculty of Medicine, University of Phayao, Mueang, Phayao 56000, Thailand

²Program in Occupational Health and Safety, Faculty of Medicine, University of Phayao, Mueang, Phayao 56000, Thailand

*E-mail: u4108047@hotmail.com

Received: 24/03/2020; Revised: 21/04/2020; Accepted: 12/05/2020

บทคัดย่อ

ระบบบำบัดน้ำเสียของโรงงานแป้งมันสำปะหลังดัดแปรผลิตของเสียหลัก 2 ประเภท ประเภทแรกคือเศษตะกอนแป้งจากบ่อดักตะกอนเบื้องต้นและอีกประเภทคือตะกอนชีวภาพจากบ่อดักตะกอนเร่ง มีความเหมาะสมที่จะนำมาใช้เป็นวัสดุหมักโดยระบบหมักไร้อากาศเพื่อผลิตก๊าซชีวภาพ โดยผลการศึกษาพบว่าเศษตะกอนแป้งมีศักยภาพสูงในการผลิตก๊าซชีวภาพและก๊าซมีเทนสูงมากเท่ากับ 382 และ 225 ลิตรต่อกิโลกรัมของแข็งระเหยที่เดิม (L/kg TVS_{added}) ตามลำดับ (ร้อยละมีเทนเท่ากับ 59) แต่อย่างไรก็ตามการหมักเศษตะกอนแป้งเพียงอย่างเดียวระบบเสี่ยงต่อการล้มเหลว เนื่องจากการสะสมตัวของกรดอินทรีย์ระเหยที่ทำให้พีเอชของระบบลดลงอย่างรุนแรง ซึ่งการนำเศษตะกอนแป้งมาหมักเพียงชนิดเดียว จึงจำเป็นต้องมีการปรับพีเอชระหว่างการหมักย่อย อีกทั้งควรมีการพิจารณาความคุ้มค่าและเหมาะสมเมื่อมีการเติมสารเคมีเพื่อปรับพีเอชในระบบหมักย่อย ในขณะที่การหมักร่วมของ

เศษตะกอนแป้งกับตะกอนชีวภาพที่อัตราส่วน 0.5:0.5 ช่วยปรับเสถียรภาพของระบบให้ดีขึ้น แต่ได้ประสิทธิภาพในการผลิตก๊าซชีวภาพและก๊าซมีเทนค่อนข้างต่ำกว่าเพียง 196 และ 127 L/kg TVS_{added} ตามลำดับ (ร้อยละมีเทนเท่ากับ 65) แสดงให้เห็นว่าการเติมตะกอนชีวภาพมีประโยชน์ในการปรับเสถียรภาพของระบบหมักย่อยให้ดีขึ้น ดังนั้นจึงควรมีการศึกษาเพิ่มประสิทธิภาพในการผลิตก๊าซชีวภาพและก๊าซมีเทนให้สูงขึ้นต่อไป

คำสำคัญ: แป้งมันสำปะหลังคัดแปร, ก๊าซชีวภาพ, เศษตะกอนแป้ง, ตะกอนชีวภาพ

Abstract

The wastewater treatment system of modified tapioca starch plant produces two major kinds of waste. One is where waste starch sludge (WSS) from a primary sedimentation pond and the other is waste activated sludge (WAS) from an activated sludge pond. They are suitable to use as substrates in anaerobic digestion for biogas production. The results of these studies found that biogas and methane production with WSS at 382 and 225 L/kg TVS_{added}, respectively (methane content of 59%), had high potential. However, the single WSS digestion risked to fail because of the volatile fatty acid accumulation, leading to the dramatic drop in pH. The single WSS digestion may need pH adjustment during the digestion process. Also, it must be considered in an economic feasibility and appropriateness when adding chemicals to adjust the pH in the digestion system. As the co-digestion of WSS:WAS at 0.5:0.5 improved the system stability, but it obtained lower biogas and methane production efficiencies of 196 and 127 L/kg TVS_{added}, respectively (methane content of 65%). These indicated that the addition of WAS had the advantage to enhance the stability of anaerobic digestion. Therefore, there should be further studied to increase the biogas and methane production efficiencies.

Keywords: modified tapioca starch, biogas, waste starch sludge, waste activated sludge

บทนำ

อุตสาหกรรมแป้งมันสำปะหลังเป็นอุตสาหกรรมหลักของประเทศไทย ที่สร้างรายได้เป็นจำนวนมากให้กับประเทศไทย โดยเฉพาะอย่างยิ่งแป้งมันสำปะหลังคัดแปร ประเทศไทยเป็นผู้ส่งออกรายใหญ่ที่สุดของโลก โดยในปี พ.ศ.2561 ทำการส่งออกสูงถึงประมาณปีละ 1.03 ล้านตันต่อปี คิดเป็นมูลค่าการส่งออก 24,335 ล้านบาท (Thai Tapioca Starch Association, 2018) โดยแป้งมันสำปะหลังคัดแปรถูกนำไปใช้ในหลายอุตสาหกรรม อาทิเช่น อาหาร กระดาษ สิ่งทอ กาว เป็นต้น เนื่องจากมีคุณสมบัติพิเศษหลายอย่าง เช่น ช่วยปรับปรุงเรื่องของความเหนียว การละลายในน้ำเย็น การทนความร้อน และทนกรดและเป็นที่ยอมรับกันว่ากระบวนการผลิตจะมีน้ำเสียเกิดขึ้นเป็นจำนวนมากประมาณ 10-20 ลูกบาศก์เมตรต่อตันแป้งที่ผลิต (Thailand Tapioca Starch, 2018) โดยส่วนใหญ่

น้ำเสียจะถูกส่งเข้าระบบบำบัดน้ำเสียของทางโรงงาน และระบบที่ได้รับความนิยมใช้สำหรับโรงงานที่ตั้งอยู่ในเขตชุมชน ซึ่งมีพื้นที่จำกัดคือ ระบบบำบัดน้ำเสียแบบตะกอนเร่ง (Activated sludge wastewater treatment system) (Wang et al., 2006) จากระบบบำบัดน้ำเสียดังกล่าวจะก่อให้เกิดของเสียหลักๆ สองชนิดด้วยกันคือ เศษตะกอนแป้ง (Waste starch sludge) จากบ่อตกตะกอนเบื้องต้น และตะกอนชีวภาพ (Waste activated sludge) บ่อเติมอากาศของระบบตะกอนเร่ง

ปริมาณของเศษตะกอนแป้งที่เกิดขึ้นจะขึ้นอยู่กับคุณภาพของเทคโนโลยี และการบริหารจัดการของแต่ละโรงงาน ซึ่งโดยทั่วไปจะมีปริมาณแป้งที่สูญเสียในรูปของเศษตะกอนแป้งอยู่ที่ประมาณ 1 เปอร์เซ็นต์ของแป้งที่ผลิตได้ (Yono et al., 2018) ซึ่งถ้าคิดจากปริมาณผลิตรวมประมาณ 1.5 ล้านตันต่อปี (National Food Institute, 2011) จะมีปริมาณสูงถึงปีละ 150,000 ตัน ซึ่งการจัดการโดยส่วนใหญ่ถูกขายในราคาถูก (The Thai Tapioca Products Factory Association, 2017) ให้กับพ่อค้าในพื้นที่เพื่อนำไปใช้ในการผลิตอาหารสัตว์ ซึ่งไม่มีความคุ้มค่าทางเศรษฐศาสตร์ มีปัญหาพื้นที่ที่จัดเก็บ และกลิ่นเหม็นรบกวนระหว่างรอการเก็บขน โดยเฉพาะในฤดูฝน การนำเศษตะกอนแป้งมาใช้ในการผลิตก๊าซชีวภาพจึงเป็นทางเลือกที่น่าสนใจเนื่องจากทำในระบบปิด ช่วยแก้ปัญหาเรื่องของกลิ่น และสามารถใช้เป็นแหล่งพลังงานทดแทนให้กับทางโรงงานได้ แต่อย่างไรก็ตามจากคุณสมบัติของเศษตะกอนแป้งที่มีอัตราส่วนคาร์บอนต่อไนโตรเจนที่สูง และมีองค์ประกอบหลักเป็นแป้ง ที่ย่อยได้ค่อนข้างง่าย อาจส่งผลให้ระบบหมักแบบไม่ใช้ออกซิเจนที่ใช้ในการผลิตก๊าซชีวภาพล้มเหลวได้เนื่องจาก การสะสมของกรดอินทรีย์ระเหยปริมาณมากที่จะเกิดขึ้นอย่างรวดเร็วจากการย่อยของแป้ง ส่งผลให้พีเอชของระบบลดลงอย่างรวดเร็ว ทำให้ระบบขาดเสถียรภาพและล้มเหลวได้ นอกจากนี้จากการที่เมทาโนเจนที่ค่อนข้างไวต่อการเปลี่ยนแปลงพีเอช ทำให้หยุดสร้างก๊าซมีเทน และตายในที่สุด (Nijaguma, 2006) มีหลายงานวิจัยที่แสดงให้เห็นว่าการย่อยวัสดุที่มีแป้งเป็นองค์ประกอบหลักสูงระบบจะมีความเสี่ยงดังกล่าว (Chen et al., 2015) ด้วยเหตุนี้การหมักร่วมกับวัสดุอื่นๆ จึงเป็นทางเลือกที่เหมาะสม ซึ่งพบว่าของเสียจากระบบบำบัดน้ำเสียอีกชนิดหนึ่งคือ ตะกอนชีวภาพจากบ่อเติมอากาศ มีความเหมาะสมในการนำมาทดลองหมักร่วม เนื่องจากมีอัตราส่วนคาร์บอนต่อไนโตรเจนที่ค่อนข้างต่ำโดยทั่วไปอยู่ระหว่าง 6-9 (Liu et al., 2016) ซึ่งน่าจะเป็นประโยชน์ช่วยในการแก้ปัญหาดังกล่าวข้างต้น เนื่องจากตะกอนชีวภาพส่วนใหญ่ประกอบด้วยเซลล์ของแบคทีเรียเป็นหลัก (Tomasik, 2017) ซึ่งย่อยสลายได้ยากกว่าแป้ง รวมถึงการนำมาหมักร่วมจะช่วยปรับสมดุลของปริมาณธาตุอาหารให้ดีขึ้นกว่าการหมักเศษแป้งเพียงอย่างเดียวที่มีอัตราส่วนคาร์บอนต่อไนโตรเจนที่สูงมาก อีกทั้งตะกอนชีวภาพที่เกิดมีปริมาณสูงถึงประมาณ 0.8 กิโลกรัมแห้งต่อกิโลกรัมบีโอดีที่ถูกกำจัด (Gray, 2017) โดยงานวิจัยเกี่ยวกับการผลิตก๊าซชีวภาพจากอุตสาหกรรมแป้งมันสำปะหลังส่วนใหญ่ เป็นการศึกษาในอุตสาหกรรมแป้งมันสำปะหลังดิบ ที่ใช้เปลือกหั่ว กากที่เหลือจากการสกัดแป้ง หรือน้ำเสียจากกระบวนการผลิต (Kullavanijaya & Thongduang, 2012; Ben & Michael, 2018; Izah, 2019) มาใช้ในการผลิตก๊าซชีวภาพ ซึ่งยังขาดการศึกษาในของเสียจากโรงงานระบบบำบัดน้ำเสีย

งานวิจัยนี้จึงสนใจในการนำของเสียจากระบบบำบัดน้ำเสียของโรงงานแปรงมันสำปะหลังคัดแปรรูปคือ เศษตะกอนแป้ง และตะกอนชีวภาพ มาทำการหมักร่วมโดยระบบหมักไร้อากาศแบบหนึ่งขั้นตอน เพื่อหาศักยภาพในการผลิตก๊าซชีวภาพ รวมถึงวิเคราะห์ประสิทธิภาพ และเสถียรภาพของระบบหมักร่วมว่ามีความเหมาะสมในการนำไปใช้ประโยชน์ในการผลิตพลังงานทดแทนหรือไม่ต่อไป

วิธีการทดลอง

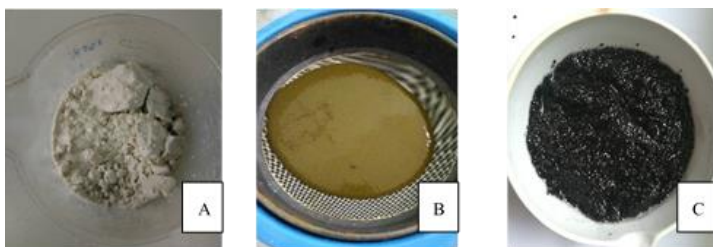
1. วัสดุหมักและหัวเชื้อที่ใช้ในการทดลอง

1.1 เศษตะกอนแป้ง (Waste starch sludge: WSS) เก็บมาจากบ่อตกตะกอนเบื้องต้นที่รองรับน้ำเสียจากการล้างกระบวนการผลิตและถังปฏิกรณ์ชีวภาพ ใส่ถุงพลาสติกและเก็บรักษาไว้ที่อุณหภูมิ 4 องศาเซลเซียส (°C) ไว้จนกว่าจะใช้ทดลอง โดยปกติเศษตะกอนแป้งจะถูกส่งไปรีดน้ำออก และขายเพื่อนำไปผลิตอาหารสัตว์

1.2 ตะกอนชีวภาพจากระบบตะกอนเร่ง (Waste activated biosludge: WAS) ทำการเก็บจากบ่อเติมอากาศของระบบตะกอนเร่งแบบเอสบีอาร์ (Sequence batch reacto: SBR) ภายหลังจากมีการสูบน้ำใสออก โดยนำมาผ่านการกรองด้วยตะแกรงขนาด 30 mesh และเก็บใส่ถุงพลาสติกและเก็บรักษาไว้ที่อุณหภูมิ 4°C เพื่อเตรียมไว้ใช้ในการทดลอง

ทั้งเศษตะกอนแป้ง และตะกอนชีวภาพนำมาจากระบบบำบัดน้ำเสียของโรงงานผลิตแปรงมันสำปะหลังคัดแปรรูปที่ตั้งอยู่ในจังหวัดระยอง

1.3 หัวเชื้อที่ใช้ในการทดลอง (Inoculum) นำมาจากระบบบำบัดน้ำเสียแบบยูเอสบี (Up flow anaerobic sludge blanket: UASB) ของโรงงานผลิตเครื่องดื่มน้ำ ในจังหวัดนนทบุรี ทำการเตรียมความเข้มข้นของหัวเชื้อเท่ากับ 2.5% ของของแข็งระเหย (Total volatile solids: TVS) วัสดุหมักและหัวเชื้อที่ใช้ในการทดลองมีการวิเคราะห์ลักษณะสมบัติดังนี้ คือ ความชื้น พีเอช (pH) อัตราส่วนคาร์บอนต่อไนโตรเจน (Carbon to nitrogen ratio: C/N) และเปอร์เซ็นต์ของของแข็งระเหย (%TVS)



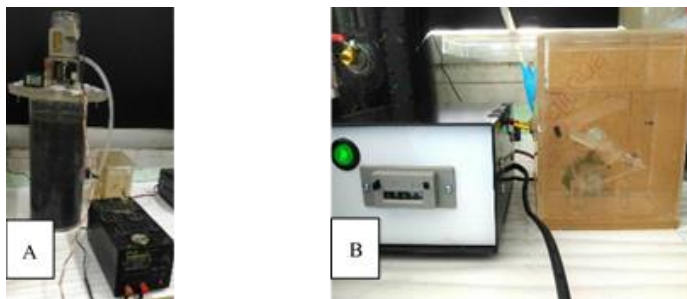
รูปที่ 1. วัสดุหมักและหัวเชื้อที่ A) WSS B) WAS C) Inoculum

2. ระบบหมักไร้อากาศแบบหนึ่งขั้นตอนโดยถังปฏิกรณ์กวนผสมแบบต่อเนื่อง (Continuous stirred tank reactor: CSTR)

ถังที่ใช้ในการหมักเป็นถังกลมขนาด 6 ลิตร ดังรูปที่ 2A ซึ่งมีปริมาตรในการหมักเท่ากับ 5 ลิตร ปริมาตรของวัสดุหมักต่อหัวเชื้อ เท่ากับ 60:40 โดยปริมาตร (3 ลิตร:2 ลิตร) และกำหนดความเข้มข้นของวัสดุหมัก และหัวเชื้อที่ใช้ในการหมัก เท่ากับ 1%TVS และ 2.5%TVS ตามลำดับ กำหนดอัตราส่วนของวัสดุหมักสองชนิดคือ เศษตะกอนแป้งต่อตะกอนชีวภาพ WSS:WAS เท่ากับ 1:0 (เศษตะกอนแป้งอย่างเดียว), 0.7:0.3, 0.5:0.5, 0.3:0.7 และ 0:1 (ตะกอนชีวภาพอย่างเดียว) ทำการหมักที่อุณหภูมิห้องระหว่าง 25-30°C โดยมีการกวนผสมด้วยอัตรา 15 นาทีต่อชั่วโมง ทำการหมักเป็นเวลา 15 วัน โดยทำการเก็บตัวอย่างจากจุดเก็บตัวอย่างที่อยู่บริเวณกึ่งกลางถังเพื่อวิเคราะห์ พีเอช ความเข้มข้นกรดอินทรีย์ระเหย (Volatile fatty acids: VFA) ค่าความเป็นด่าง (Alkalinity: ALK) และมีการวัดอุณหภูมิภายในถัง และปริมาตรก๊าซชีวภาพที่เกิดขึ้นในแต่ละวันด้วย Gas counter (รูปที่ 2B) รวมถึงมีการเก็บตัวอย่างก๊าซชีวภาพเพื่อวิเคราะห์หาค่าร้อยละของมีเทนโดยต่อท่อด้านบนเข้ากับถุงเก็บก๊าซ (Tedlar bag) ขนาด 2 ลิตร และทำการวิเคราะห์ค่าร้อยละมีเทนด้วยเครื่องก๊าซโครมาโตกราฟี วิเคราะห์ทั้งหมดทำตามวิธี Standard methods for the examination of water and wastewater (American Public Health Association, 2017) รายละเอียดในการวิเคราะห์แต่ละพารามิเตอร์ดังแสดงในตารางที่ 1

ตารางที่ 1. วิธีการวิเคราะห์ในแต่ละพารามิเตอร์

พารามิเตอร์	วิธีวิเคราะห์
ความชื้น	- Dried 103-105°C (Standard 2540B)
พีเอช (pH)	- pH meter (Mettler Toledo, Seven Easy, Switzerland)
ของแข็งระเหย (TVS)	- Dried 550°C (Standard 2540E)
อัตราส่วนของคาร์บอนต่อไนโตรเจน (C/N)	- Total organic analyzer และ Kjeldahl method
กรดอินทรีย์ระเหย (VFA)	- Titration method (Standard 5560B)
ความเป็นด่าง (ALK)	- Titration method (Standard 2320B)
ปริมาตรก๊าซชีวภาพ	- Gas counter
ร้อยละของก๊าซมีเทน	- Gass Chromatography (Model Shimadzu GC-2014, Shimadzu Corp., Japan) with thermal conductivity detector - Molecular sieve packed column (MS-5A, stainless steel, Outside diameter 3 mm. Inside diameter 2.1 mm, Length 2 m.(Restek Corp., USA), Temperature of injector, column and detector at 150, 80, and 200°C, respectively, Argon as carrier gas with flow rate 30 ml/m



รูปที่ 2. ระบบหมัก ไร้อากาศแบบหนึ่งขั้นตอน A) ถัง CSTR B) Gas counter

3. การคำนวณประสิทธิภาพในการผลิตก๊าซชีวภาพ และประสิทธิภาพในการกำจัดของแข็งระเหย

$$\text{ประสิทธิภาพการในผลิตก๊าซชีวภาพ (L/kg TVS}_{\text{added}}) = \frac{\text{Cumulative Biogas}}{\text{TVS}_{\text{added}}} \quad (1)$$

โดย

Cumulative Biogas คือ ปริมาตรก๊าซชีวภาพที่สามารถผลิตได้สะสม (L)

TVS_{added} คือ น้ำหนักของแข็งระเหยที่เดิมเข้าสู่ถังหมัก ไร้อากาศ (kg)

$$\text{ประสิทธิภาพในการกำจัดของแข็งระเหย (\%)} = \frac{(\text{TVS}_{\text{initial}} - \text{TVS}_{\text{final}}) \times 100}{\text{TVS}_{\text{initial}}} \quad (2)$$

โดย

TVS_{initial} คือ ความเข้มข้นของแข็งระเหยเริ่มต้นการหมักย่อย ไร้อากาศ (mg/l)

TVS_{final} คือ ความเข้มข้นของแข็งระเหยสิ้นสุดการหมักย่อย ไร้อากาศ (mg/l)

ผลและวิจารณ์ผลการทดลอง

1. ลักษณะสมบัติของวัสดุหมัก

จากผลการวิเคราะห์ลักษณะสมบัติของวัสดุหมักและหัวเชื้อที่ใช้ในการหมัก ดังตารางที่ 2 พบว่า เศษตะกอนแป้งมีค่าพีเอช 4.48 ซึ่งมีความเป็นกรดสูง และมีอัตราส่วนคาร์บอนต่อไนโตรเจนที่ค่อนข้างสูงมาก เท่ากับ

242 โดยจากการศึกษาของ Promthong et al., (2006) พบว่าโดยทั่วไปองค์ประกอบหลักของแป้งโดยทั่วไปจะมีองค์ประกอบหลักเป็น อะไมโลเพกตินสูงถึง 83 เปอร์เซ็นต์ ที่ย่อยได้ค่อนข้างง่ายและรวดเร็ว ซึ่งจากองค์ประกอบและคุณสมบัติ เศษตะกอนแป้งจึงไม่มีความเหมาะสมที่จะนำมาใช้ในการหมักย่อยเพียงชนิดเดียว โดยเศษตะกอนแป้งจะถูกย่อยได้โดยง่ายและรวดเร็วเป็นกรดอินทรีย์ระเหย ทำให้มีการสะสมระบบโดยที่เมทาโนเจนไม่สามารถนำไปใช้ในการผลิตมีเทนได้ทัน จากการสะสมดังกล่าวส่งผลให้พีเอชของระบบลดลงอาจส่งผลให้ระบบขาดเสถียรภาพ และล้มเหลวในที่สุดเนื่องจากเมทาโนเจนค่อนข้างไวต่อการเปลี่ยนแปลงพีเอช เกิดการหยุดสร้าง และตายในที่สุดเมื่อพิจารณาลักษณะสมบัติของตะกอนชีวภาพ พบว่าพีเอชเป็นกรดเพียงเล็กน้อย เท่ากับ 6.50 เหมาะสำหรับระบบการหมักแบบไร้อากาศ ซึ่งช่วยให้การปรับพีเอชให้มีความเหมาะสมมากขึ้นจากการนำมาหมักร่วมกับเศษตะกอนแป้งที่มีความเป็นกรด และองค์ประกอบหลักส่วนใหญ่ที่เป็นเซลล์ของแบคทีเรียที่มีโครงสร้างค่อนข้างซับซ้อนย่อยได้ยากกว่าเศษตะกอนแป้ง จะทำให้อัตราการย่อยวัสดุหมักกลายเป็นกรดอินทรีย์ระเหยเกิดขึ้นช้าลง ทำให้เมทาโนเจนสามารถนำกรดอินทรีย์ระเหยไปใช้ได้ทัน ระบบจึงมีสมดุล เนื่องจากการไม่มีการสะสมจนเกิดปัญหาข้างต้น อีกทั้งอัตราส่วนของคาร์บอนต่อไนโตรเจนที่ค่อนข้างต่ำเพียง 4.27 จึงจะเป็นประโยชน์ช่วยในการปรับปริมาณสารอาหารให้เหมาะสม สำหรับการหมักย่อยแบบไร้อากาศมากขึ้นดีกว่าการหมักย่อยโดยใช้วัสดุหมักเป็นเศษตะกอนแป้ง หรือตะกอนชีวภาพ เพียงอย่างเดียว ซึ่งมีค่าสูง และต่ำเกินไป ตามลำดับ

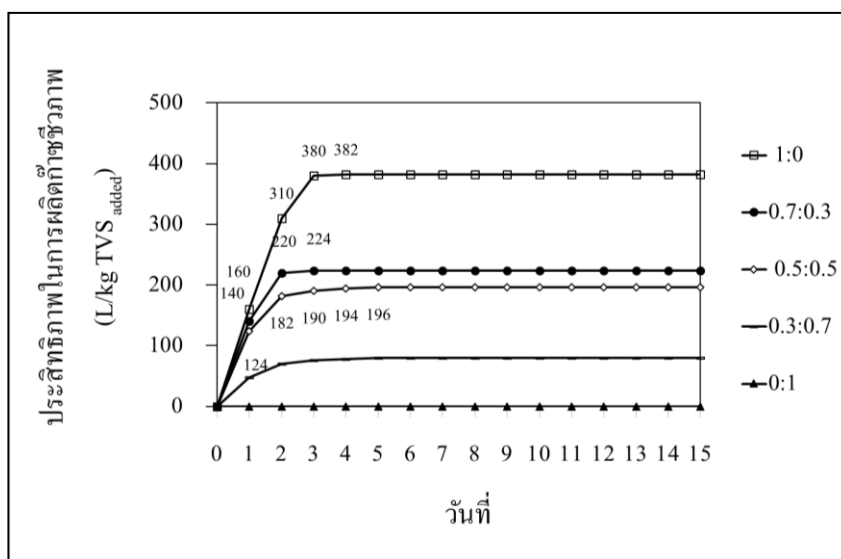
ตารางที่ 2. ลักษณะสมบัติของ WSS และ WAS

พารามิเตอร์	WSS	WAS
ความชื้น (%)	39.09±0.62	92.44±1.01
pH	4.48±0.01	6.50±0.03
TVS (%)	31.26±0.03	5.92±0.04
อัตราส่วน C/N	242	4.27

2. ประสิทธิภาพในการผลิตก๊าซชีวภาพและมีเทน

ประสิทธิภาพในการผลิตก๊าซชีวภาพในแต่ละอัตราส่วนการหมักย่อย พบว่าอัตราส่วนที่ใช้แป้งเพียงอย่างเดียว อัตราส่วนเท่ากับ 1:0 สามารถผลิตก๊าซชีวภาพได้สูงสุดเท่ากับ 382 ลิตรต่อกิโลกรัมของแข็งระเหยเริ่มต้น (L/kg TVS_{added}) ภายในวันที่ 4 ของการหมัก รองลงมาเป็นอัตราส่วนที่มีการหมักร่วมระหว่างเศษตะกอนแป้งและตะกอนชีวภาพ ที่อัตราส่วน 0.7:0.3 และ 1:1 มีประสิทธิภาพในการผลิตก๊าซชีวภาพเท่ากับ 224 และ 196 L/kg TVS_{added} ตามลำดับ ในขณะที่อัตราส่วนที่มีการใช้ตะกอนชีวภาพเพียงอย่างเดียวพบไม่มีก๊าซชีวภาพเกิดขึ้นจาก

ระบบหมัก ดังแสดงในรูปที่ 3 อาจเนื่องมาจากอัตราส่วน C/N ของวัสดุหมักที่ค่อนข้างต่ำมากเท่ากับ 4.27 ซึ่งห่างจากอัตราส่วนที่แนะนำว่าเหมาะสมสำหรับการผลิตก๊าซชีวภาพ ที่เท่ากับ 20-30 (Dioha et al., 2013) จากองค์ประกอบหลักของก๊าซมีเทนเป็นคาร์บอนและไฮโดรเจน โดยในอัตราส่วน C/N ที่ต่ำมาก อาจทำให้ขาดแหล่งคาร์บอนในการผลิตก๊าซมีเทน ขณะที่อัตราส่วนการหมักรวม 1:0, 0.7:0.3 และ 0.5:0.5 มีอัตราส่วน C/N เท่ากับ 242, 79 และ 42 ตามลำดับ ซึ่งสูงกว่าระดับที่แนะนำ แสดงให้เห็นว่าอัตราส่วน C/N ที่สูงระบบก็ยังคงสามารถผลิตก๊าซชีวภาพได้ โดยอัตราส่วน C/N ที่สูงเนื่องจากขาดไนโตรเจน แต่มีผลทางลบต่อการสร้างพลังงาน และโปรตีนซึ่งเป็นโครงสร้างเซลล์ของจุลินทรีย์ (Deublein & Steinhaue, 2011) โดยระบบยังผลิตก๊าซชีวภาพได้



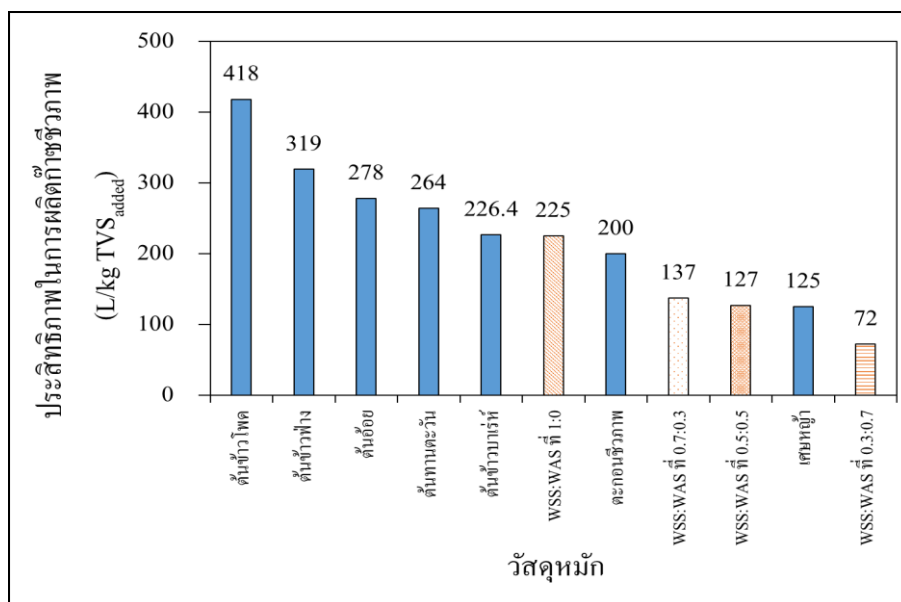
รูปที่ 3. ประสิทธิภาพในการผลิตก๊าซชีวภาพของการหมักร่วมของ WSS:WAS ในอัตราส่วนที่แตกต่างกัน

ผลการวิเคราะห์ร้อยละของก๊าซมีเทน ดังตารางที่ 3 ของทุกอัตราส่วนการหมัก พบว่า ค่อนข้างสูงมีความเหมาะสม และคุ้มค่าในการนำไปใช้เป็นแหล่งพลังงานทดแทน ซึ่งแนะนำว่าองค์ประกอบของก๊าซชีวภาพ ควรมีก๊าซมีเทนไม่น้อยกว่าร้อยละ 50 (Ahmmad & Haque, 2014) เมื่อทำการเปรียบเทียบประสิทธิภาพในการผลิตก๊าซมีเทนจากการหมักแบบไม่ใช้ออกซิเจนจากชีวมวลอื่นๆ (รูปที่ 4) จากการทบทวนวรรณกรรมของ (Cesaro & Belgiomo, 2015) โดยมีประสิทธิภาพในการผลิตก๊าซมีเทนอยู่ระหว่าง 80-418 L/kg TVS_{added} โดยอัตราส่วน 1:0 มีประสิทธิภาพในการผลิตก๊าซมีเทน 225 L/kg TVS_{added} ใกล้เคียงกับ ดินข้าวบาเร่ (226.4 L/kg TVS_{added}) ดินฮ้อย (278 L/kg TVS_{added}) ดินทานตะวัน (231-297 L/kg TVS_{added}) ส่วนอัตราส่วน 0.7:0.3 และ 0.5:0.5 มีประสิทธิภาพใน

การผลิตก๊าซมีเทน 137 และ 127 L/kg TVS_{added} ตามลำดับ ใกล้เคียงกับ เศษหญ้า (125 L/kg TVS_{added}) ซึ่งพบว่าทั้งสามอัตราส่วนมีประสิทธิภาพปานกลาง โดยน้อยกว่าชีวมวลที่มีประสิทธิภาพสูง อาทิเช่น ต้นข้าวโพด (418 L/kg TVS_{added}) ต้นข้าวฟ่าง (286–319 L/kg TVS_{added}) อีกทั้งพบว่าทั้งสองอัตราส่วนประสิทธิภาพในการผลิตก๊าซมีเทนใกล้เคียงกับการศึกษาของ (Karlsson et al., 2011) ซึ่งนำตะกอนชีวภาพจากระบบตะกอนเร่งของโรงงานกระดาษ 6 แห่ง ของสวีเดน โดยทำการทดลองในขวดทดลองบีเอ็มพี มีประสิทธิภาพระหว่าง 100-200 L/kg TVS_{added} โดยจากประสิทธิภาพที่ได้ควรมีการศึกษาเพื่อปรับปรุงเพิ่มประสิทธิภาพในการผลิตก๊าซมีเทนต่อไป

ตารางที่ 3. ร้อยละของมีเทนและประสิทธิภาพในการผลิตก๊าซมีเทน

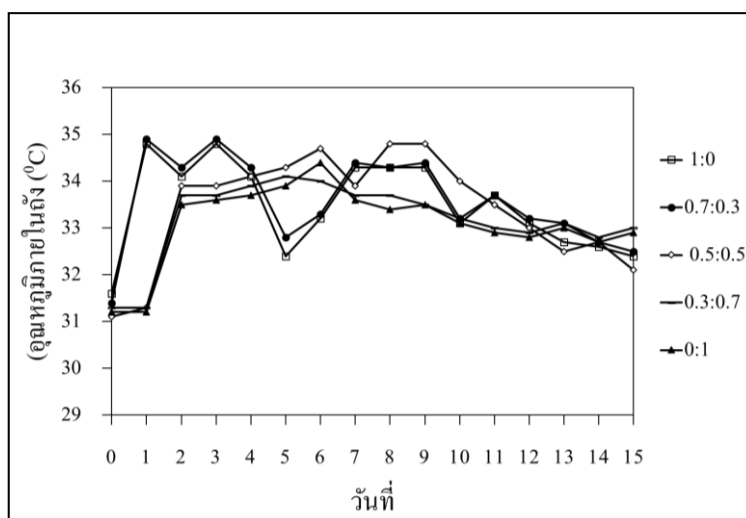
อัตราส่วนการหมักร่วมของ	ร้อยละของมีเทน	ประสิทธิภาพในการผลิตก๊าซมีเทน
WSS:WAS	(%)	(L/kg TVS _{added})
1:0	59	225
0.7:0.3	61	137
0.5:0.5	65	127
0.3:0.7	58	72



รูปที่ 4. การเปรียบเทียบประสิทธิภาพในการผลิตก๊าซมีเทนกับวัสดุหมักชนิดอื่น

3. สภาพการทำงานและเสถียรภาพของระบบ CSTR

อุณหภูมิภายในถัง CSTR ของแต่ละอัตราส่วนการหมัก แสดงดังรูปที่ 5 พบว่าอุณหภูมิอยู่ระหว่าง 30-35°C ซึ่งอุณหภูมิภายในถังจะมีการเพิ่มอย่างรวดเร็วภายในวันที่ 1-3 ของการหมัก ในอัตราส่วนการหมัก 1:0 และ 0.7:0.3 มีอุณหภูมิสูงสุดเท่ากับ 34.80°C และ 34.92°C ในวันที่ 1 และ 3 ของการหมัก และพบว่าอัตราส่วนการหมัก 0.5:0.5 มีอุณหภูมิสูงจนถึงวันที่ 6 ของการหมักแสดงให้เห็นว่าระบบยังมีการหมักย่อยและผลิตก๊าซเกิดขึ้น ในขณะที่อัตราส่วน 0:1 ที่มีแต่ตะกอนชีวภาพ จะมีอุณหภูมิเพิ่มน้อยกว่าอัตราส่วนอื่น ซึ่งอัตราส่วนนี้ไม่มีการผลิตก๊าซชีวภาพเกิดขึ้นอุณหภูมิที่สูงขึ้นจึงต่ำกว่าอัตราส่วนอื่น และการย่อยแบบไม่ใช้ออกาศเกิดไม่ถึงกระบวนการสร้างก๊าซมีเทน อาจอยู่แค่กระบวนการไฮโดรไลซิส (Hydrolysis) เท่านั้น เนื่องจากเมื่อพิจารณาค่าพีเอชยังค่อนข้างคงที่ และอัตราส่วน VFA/ALK ก่อนข้างต่ำ

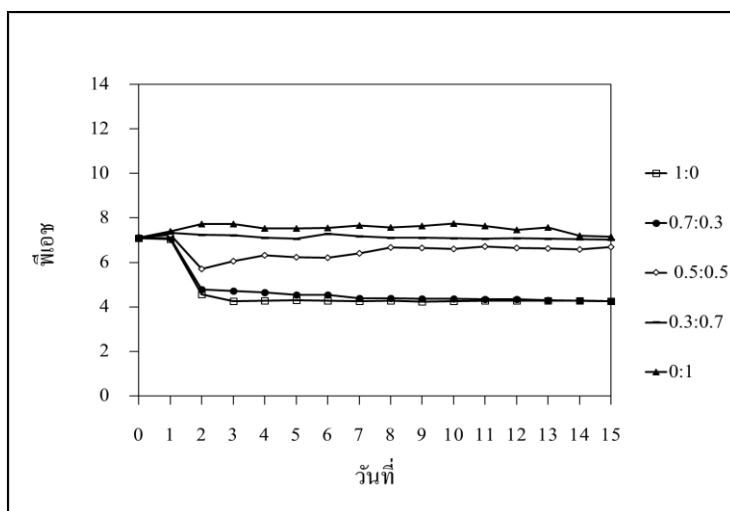


รูปที่ 5. อุณหภูมิภายในถัง CSTR ของการหมักร่วมของ WSS:WAS ในอัตราส่วนที่แตกต่างกัน

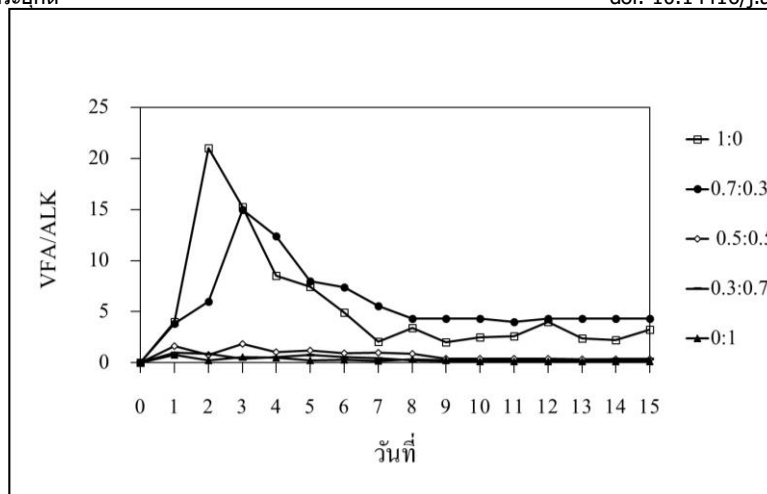
เมื่อพิจารณาสภาพการทำงาน of ระบบหมักย่อย พบว่าทั้งสามอัตราส่วนการหมัก (1:0, 0.7:0.3 และ 0.5:0.5) ค่าพีเอชภายในถัง CSTR ลดลงอย่างรวดเร็วภายใน 3 วันของการหมักย่อย โดยในวันที่ 2 ของการหมักลดลงมากที่สุด ซึ่งลดลงจนต่ำกว่าระดับที่เหมาะสมสำหรับการย่อยแบบไม่ใช้ออกาศสภาวะมีโซฟิลิก (pH ระหว่าง 6.5-8.0) (Seadi et al., 2008) โดยเฉพาะอัตราส่วน 1:0 ที่มีการเติมเศษตะกอนแป้งเพียงอย่างเดียว มีค่าพีเอชต่ำมากอยู่ระหว่าง 4.25-4.56 อาจเนื่องมาจากเศษตะกอนแป้งที่ย่อยง่ายทำให้มีการสะสมของกรดอินทรีย์ระเหย ดังแสดงในรูปที่ 6 สอดคล้องกับอัตราส่วนของ VFA/ALK ซึ่งพบว่ามีค่าสูงถึง 21 ในวันที่ 2 ของการหมัก และหลังจากนั้นจึงลดลง

โดยหลังจากวันที่ 4 ที่หยุดผลิตก๊าซชีวภาพอัตราส่วนของ VFA/ALK ก็ยังคงสูงเป็นอย่างมาก อยู่ระหว่าง 2.08-4.0 ห่างจากอัตราส่วนที่แนะนำเป็นอย่างมากเช่นกัน (VFA/ALK เหมาะสมไม่ควรเกิน 0.4) (Bakili, 2014) เช่นเดียวกันกับอัตราส่วนการหมัก 0.7:0.3 ที่พบว่าค่าพีเอชภายในถัง CSTR ลดลงอย่างรวดเร็วภายใน 3 วันของการหมัก และอัตราส่วนของ VFA/ALK ที่สูงถึง 15 โดยหลังจากที่ระบบหยุดผลิตก๊าซชีวภาพแล้ว ระบบยังคงมีค่าพีเอชที่ต่ำกว่า 4.5 และอัตราส่วน VFA/ALK ที่สูงเท่ากับ 4

ในขณะที่อัตราส่วนการหมัก 0.5:0.5 พบว่าหลังจากวันที่ 5 ที่ระบบหยุดผลิตก๊าซชีวภาพ มีค่าพีเอชกลับมาค่อนข้างเป็นกลาง อยู่ระหว่าง 6.21-6.70 และอัตราส่วน VFA/ALK ของระบบเท่ากับ 0.38 ในวันที่ 12 ของการหมัก ดังรูปที่ 7 จากผลการวิเคราะห์พีเอชและอัตราส่วน VFA/ALK แสดงให้เห็นว่าทั้งสองอัตราส่วน 1:0 และ 0.7:0.3 มีความไม่มีเสถียรภาพเกิดขึ้น โดยเป็นอัตราส่วนที่มีการเติมแอมโมเนียมากกว่าตะกอนชีวภาพ มีความเสี่ยงต่อการล้มเหลวมากกว่าอัตราส่วน 0.5:0.5 ดังนั้นการนำเศษตะกอนแอมโมเนียมาใช้เป็นวัสดุหลักในการผลิตก๊าซชีวภาพควรมีการปรับพีเอชระหว่างการหมัก และพบว่าการเติมตะกอนชีวภาพเพื่อหมักร่วมช่วยให้ระบบมีเสถียรภาพดีขึ้น ซึ่งถ้ามีการเติมสารเคมีเพื่อใช้ในการปรับพีเอช จะใช้ปริมาณน้อยกว่าอัตราส่วนที่ใช้แอมโมเนียอย่างเดียว



รูปที่ 6. ค่าพีเอชภายในถัง CSTR ของการหมักร่วมของ WSS:WAS ในอัตราส่วนที่ต่างกัน

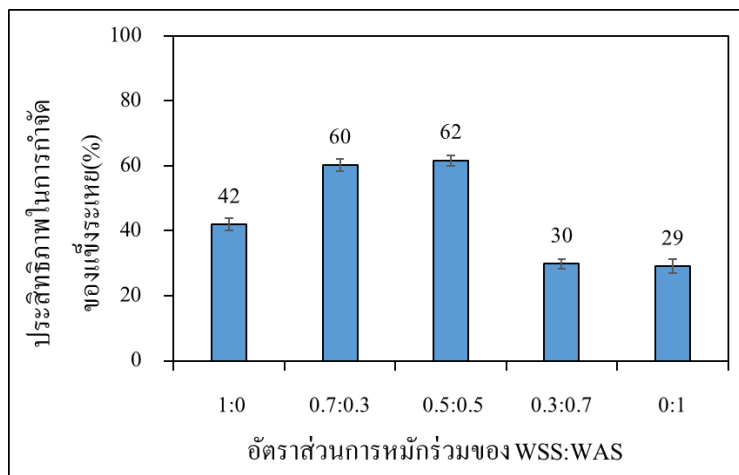


รูปที่ 7. อัตราส่วนของ VFA/ALK ภายในถัง CSTR ของการหมักร่วมของ WSS:WAS ในอัตราส่วนที่แตกต่างกัน

4. ประสิทธิภาพในการกำจัดของแข็งระเหยของระบบ CSTR

รูปที่ 8 แสดงประสิทธิภาพในการกำจัดของแข็งระเหย (Total volatile solid removal efficiency) พบว่าอัตราส่วน 0.5:0.5 มีประสิทธิภาพในการกำจัดสูงสุด เท่ากับ 62 เปอร์เซ็นต์ รองลงมาคือที่อัตราส่วน 0.7:0.3 และ 1:0 เท่ากับ 60 และ 42 เปอร์เซ็นต์ ตามลำดับ ซึ่งตรงกันข้ามกับประสิทธิภาพในการผลิตก๊าซมีเทน จากผลดังกล่าวพบว่าของแข็งระเหยที่ถูกกำจัดจากอัตราส่วน 1:0 มีศักยภาพในการผลิตก๊าซชีวภาพได้ดีกว่าอัตราส่วน 0.7:0.3 และ 0.5:0.5 โดยเมื่อคำนวณประสิทธิภาพการผลิตก๊าซมีเทนต่อน้ำหนักของของแข็งระเหยที่ถูกกำจัด ($L/kg TVS_{removed}$) สำหรับอัตราส่วน 1:0 0.7:0.3 และ 0.5:0.5 ได้เท่ากับ 545, 228 และ 205 $L/kg TVS_{removed}$ ตามลำดับ

จากผลการทดลองแสดงให้เห็นอย่างชัดว่าอัตราส่วน 1:0 มีประสิทธิภาพในการผลิตก๊าซมีเทนสูงเท่ากับ 2.67 เท่า ของอัตราส่วน 0.5:0.5 เนื่องจากเศษตะกอนแบ่งเป็นตัวหลักที่ส่งผลต่อประสิทธิภาพในการผลิตก๊าซมีเทน แต่อย่างไรก็ตามเมื่อพิจารณาสภาวะการทำงานของระบบหมักก็ พบว่ามีความเสถียรค่อนข้างต่ำ ระบบเสี่ยงต่อการล้มเหลว ดังนั้นจึงจำเป็นต้องมีการปรับปรุงสภาวะการทำงานของระบบก่อน โดยอาจมีการปรับพีเอชระหว่างการหมักย่อย ในขณะที่การนำการหมักร่วมระหว่างเศษตะกอนแบ่งกับตะกอนชีวภาพที่ 0.5:0.5 ไปใช้นั้น พบว่าระบบมีเสถียรภาพดีขึ้นแต่ประสิทธิภาพในการผลิตก๊าซชีวภาพยังต่ำอยู่ จึงควรมีการศึกษาเพิ่มประสิทธิภาพในการผลิตก๊าซให้สูงขึ้น อาทิเช่น การปรับอุณหภูมิในการหมักย่อย การกวนผสม การปรับสภาพ (pretreatment) วัสดุหมักเบื้องต้นด้วยวิธีต่างๆ เป็นต้น



รูปที่ 8. ประสิทธิภาพในการกำจัดของแ่งระเหย ของการหมักร่วมของ WSS:WAS ในอัตราส่วนที่แตกต่างกัน

สรุปผลการทดลอง

เศษตะกอนแ่งจากโรงงานแ่งมันสำปะหลังดัดแปรเป็นของเสียที่มีศักยภาพสูงในการผลิตก๊าซมีเทน แต่การนำใช้ในระบบหมักไร้อากาศเพียงชนิดเดียว พบว่ามีปัญหาเรื่องของการเดินระบบขาดเสถียรภาพมีการสะสมตัวของกรดอินทรีย์ระเหยส่งผลให้พีเอชของระบบลดลงอย่างรวดเร็ว ทำให้ระบบหยุดการผลิตก๊าซชีวภาพและล้มเหลวได้ ซึ่งการนำมาหมักร่วมกับตะกอนชีวภาพด้วยอัตราส่วน 0.5:0.5 ช่วยให้ระบบมีเสถียรภาพของระบบดีขึ้น แต่ประสิทธิภาพในการผลิตก๊าซมีเทนยังค่อนข้างต่ำมากเมื่อเทียบกับการหมักเศษตะกอนแ่งเพียงอย่างเดียว ซึ่งควรมีการศึกษาหาวิธีเพื่อเพิ่มประสิทธิภาพในการผลิตก๊าซมีเทนต่อไป

กิตติกรรมประกาศ

ขอขอบคุณคุณศิริชัย สุวรรณละมัยที่คอยช่วยเหลือ และสนับสนุนให้งานวิจัยนี้สำเร็จลุล่วงไปได้ด้วยดี และคณะแพทยศาสตร์ มหาวิทยาลัยพะเยาที่เอื้อเฟื้อสถานที่ในการทำวิจัยครั้งนี้

เอกสารอ้างอิง

Ahmmad, R.M. & Haque, S. (2014). Providing electricity by digester types on biogas productions from municipal solid waste in Dhaka city, Bangladesh. *International Journal of Energy, Information and Communications*, 5(3), 13-22.

- American Public Health Association. (2017). *Standard methods for the examination of water and wastewater*. (23rd ed). Washington DC: American Public Health Association (APHA).
- Bakili, S.S. (2014). Anaerobic Digestion of Banana Winery Effluent for Biogas Production. *International Journal of Environmental Protection and Policy*, 2(5), 168-173.
- Ben, M.G. & Michael, E.I. (2018). Biogas Production from Farm Waste (Cassava Peels and SwineDung): Co-Digestion and Prospect on Economic Growth. *Research & Reviews: Journal of Ecology and Environmental Sciences*, 6(2), 58-63.
- Cesaro, A. & Belgiorno, V. (2015). Combined Biogas and Bioethanol Production: Opportunities and Challenges for Industrial Application. *Energies*, 8(8), 8121-8144.
- Chen, S., Zhang, J. & Wang, X. (2015). Effects of alkalinity sources on the stability of anaerobic digestion from food waste. *Waste Management Research*, 33(11), 1033-1040.
- Deublein, D. & Steinhauser, A. (2011). *Biogas from Waste and Renewable Resources: An Introduction* (2th ed.). Germany: John Wiley & Sons.
- Dioha, I.J., Ikeme, C.H., Nafi'u, T., Soba, N.I. & Yusuf M.B.S. (2013). Effect of carbon to nitrogen ratio on biogas production. *International Research Journal of Natural Sciences*, 1(3), 1-10.
- Gray, N.F. (2017). *Water Science and Technology. An Introduction* (4th ed.). Taylor & Francis Incorporated.
- Karlsson, A., Truong, X.B., Gustavsson, J., Svensson, B.H., Nilsson, F. & Ejlertsson J. (2011). Anaerobic treatment of activated sludge from Swedish pulp and paper mills--biogas production potential and limitations. *Environmental Technology*, 32(13-14), 1559-1571.
- Kullavanijaya, P. & Thongduang, P. (2012). Enhanced Biogas Production in Anaerobic Digestion of Cassava Wastewater Through Supplementation of Biodiesel Waste as Co-Substrate. *International Journal of Renewable Energy Research*, 2(3), 510-515.
- Liu, C., Li, H. & Zhang, Y. (2016). Improve biogas production from low-organic-content sludge through high-solids anaerobic co-digestion with food waste. *Bioresource Technology*, 219, 252-260.
- Izah, S.C. (2019). *Cassava Mill Effluents Recycling through Bioenergy Production: A Review*. Nigeria: Niger Delta University.
- Nijaguna, B.T. (2006). *Biogas Technology*. new Delhi, India: K.K Gupta for New Age International (P) Ltd.

- Promthong, S., Kanto, U., Tirawattanawanich C., Tongyai, S., Isariyodom, S. & Markvichitr K. (2006). *Comparison of nutrient compositions and carbohydrate fractions of corn, cassava chip and cassava pellet ingredients*. Animals. In Proceedings of the 44th Kasetsart University Annual Conference, Kasetsart University, Bangkok. (in Thai)
- Seadi, T A., Rutz, D., Prassl, H., Köttner, M., Finsterwalder, T., Volk, S. et al. (2008). *BiogasHandbook (T. A. Seadi Ed)*, University of Southern Denmark Esbjerg, Denmark.
- Thai Tapioca Starch Association. (2018). *Export Tapioca Products*. Retrieved March 3, 2020, from http://www.thaitapiocastarch.org/en/information/statistics/export_tapioca_products. (in Thai)
- Thailand Tapioca Starch. (2018). *Pollution/Waste and Environmental Management for Cassava Starch Industry*. Retrieved March 2, 2020, from <http://www.thailandtapiocastarch.net/technology-detail/10/3/>. (in Thai)
- The Thai Tapioca Products Factory Association. (2017). *The price of tapioca by-Product in Thailand*. Retrieved March 1, 2020, from <http://thaitapioca.org/>. (in Thai)
- Tomasik, B. (2017). *Microorganisms created by wastewater-treatment systems*. Retrieved March 5, 2020, from https://reducing-suffering.org/microorganisms-wastewater-treatment/#Microorganism_composition.
- Wang, L.K., Hung, Y.T., Lo, H.H., Yapijakis, C. (2006). *Waste treatment in the food processing industry* (1st ed.). New York: Taylor & Francis.
- Yono, B., Primaloka, A.D., Ardhanari, L., Matin, H.H.A., Sumardiono, S. (2018). Study of Biogas Production from Cassava Industrial Waste by Anaerobic Process. *MATEC Web of Conferences*. 1-6.

Faculty of Applied Science

King Mongkut's University of Technology North Bangkok

1518 Pracharat 1 Road, Wongsawang, Bangsue, Bangkok 10800 : Thailand

