Geographia Technica

Geographia Technica

Technical Geography

an International Journal for the Progress of Scientific Geography

2023 – No. 1

Cluj University Press

Contents

Geographia Technica

DEVELOPING A FLOOD FORECASTING SYSTEM WITH MACHINE LEARNING AND APPLYING TO GEOGRAPHIC INFORMATION SYSTEM

Jirayu PUNGCHING[1] 🆔 *and Sitang PILAILAR[1]* 🆔

ABSTRACT:

Floods are natural disasters that can damage lives, property, and the economy. Therefore, it is necessary to have a reliable and accurate flood forecasting system to provide early warning in time. Although several Mathematical models have been developed and used to forecast floods continuously for decades, most require up-to-date and specific physical data, including a high experience user, to provide and interpret the result. It is an obstacle for use in remote areas with incomplete information and a lack of specialists. This study, therefore, developed a real-time flood forecasting system with Machine Learning by applying a 2-variable sliding window technique to restructure the data, which can solve the problem of data limitation. Thung Song District Nakhon Si Thammarat Province was selected to test this newly developed model. By importing the water level data of two water level observed stations, SWR025 at the upstream and NKO001 at Thung Song Municipality, into five machine learning algorithms (Linear Regression, Support Vector Machine, K-Nearest Neighbor, Decision Tree, and Random Forest) for forecasting the water level every 30 minutes for the next 5 hours. Their performance was compared by the MSE, MAE, and R^2, which ranged from 0.006-0.013, 0.044-0.063, and 0.518-0.750, respectively. The Random Forest was the most efficient algorithm for the 3-hour forecast with an efficiency value of MSE 0.006, MAE 0.044, and R^2 0.75. The developed ML flood forecasting model was validated by the flood data in November 2021 and showed good agreement. Then, the extent of the inundation area was evaluated by the mathematical model. Next, the water depth and surface elevation were transformed and applied to GIS. Finally, the flood risk areas on Google Maps under that specific rainfall are promptly notified to the people three hours before the flood occurs.

Key-words: *Flood Forecasting, Flood Maps, Machine Learning, Linear Regression, Support Vector Machine, K-Nearest Neighbor, Decision Tree, Random Forest, Thung Song Municipality*

1. INTRODUCTION

Floods are threatening natural disasters that cause damage to life, property, and economic security at both local and national levels. There has been a tendency to occur more frequently, more violently, and become increasingly difficult to predict due to many factors such as changing rain behavior from the past, land-use change, encroachment on people's waterways, and shallowness of natural waterways, etc. Urbanization significantly changes hydrological behavior by changing the infiltration rate, baseflow, and lag time and influencing flow patterns (Rafiei et al., 2016). It emphasizes the need for a reliable flood forecasting system to provide timely early warning for citizens and government agencies to take preventative or evacuation measures to alleviate the damage. Additionally, a real-time flood inundation map that indicates the extent of flood risk areas corresponding to the amount of precipitation upstream helps deal with potential disasters.

Although flood prediction mathematical models have been constantly developed and applied for hazard assessment and extreme event management for decades, most physical models require accurate physical data of the basin for model setting up (Chang et al., 2020). The recorded events over a long period are also necessary. Obviously, the more extended and accurate data results in more precise forecasting closer to reality, but it requires substantial computational effort. Furthermore, changes in the physical data of the area inevitably affect the accuracy of forecasts over the expectations of system users. There are also quite specific data import restrictions. In addition, the area risk factors cannot be readily increased or decreased, and aspects of expertise and user experience can significantly

―――――――――――――――
[1]Department of Water Resource Engineering, Faculty of Engineering, Kasetsart University,
Bangkok 10900, Thailand; jirayu.pun@ku.th, fengstpl@ku.ac.th

influence the forecasting system's accuracy. Likewise, several studies suggested a gap in the short-term prediction capability of physical models (Castabile and Macchione, 2015). For example, Van de Honert and McAneney, 2011 published the failure in flood prediction in Queensland, Australia. Shrestha et al., 2013 also reported the systematic error caused by the unreliable numerical weather prediction model. The inappropriate short-term prediction that needs significant improvement has encouraged the usage of advanced data-driven models, e.g., machine learning (ML). Machine Learning Neural Networks model how the brain performs a particular task or function of interest (Haykin, 2008). The use of a Neural Network is computation through the process of "learning" that is adaptive in the machine, known as "Machine Learning (ML)." The computing algorithms can improve automatically through experience and by using information for training (Learning) to find reasonable approximate solutions to complex problems. Researchers have moved from physical-based flood forecasting methods to ML techniques over the last two decades (Chang et al.,2018). Chang et al., 2014 conducted the prediction of flood volume using hybrid Som and dynamic neural networks. Lohani et al., 2014 and Badrzadeh et al., 2015 applied ML for runoff prediction in real-time flood forecasting. Granata et al., 2016 examined the runoff in an urban area and compared the performances between the support vector regression and the EPA's stormwater management model. The study confirmed the usefulness of SVR for urban flood prediction.

After estimating the runoff and flood volume according to various return periods of rainfall, the predicted inundation map has been determined. In Kemaman river, West Malaysia, Chang et al., 2018 developed the flood inundation forecast model by combining two artificial neural networks, Self Organizing Map (SOM) and Recurrent nonlinear autoregressive with exogenous inputs (RNARX). The forecast inundation maps were then generated ahead of the event. The accuracy of inundation maps RMSE was found to range from 0.08 to 0.68, and R^2 ranged from 0.94 to 1. A few seconds carried out three to twelve hours ahead of inundation maps. Chang et al., 2020 predicted the inundation maps to build a real-time warning system for people in the risk area. He developed the effective real-time pluvial flood forecasting AI platform in Taoyuan City, Taiwan, with 6,000 sets of color-classified rainfall hyetograph maps. Three hundred thousand simulated flooding maps were used for learning with Inception V3 Convolutional Neural Network (CNN). The method's accuracy is shown by comparing AI-generating map images and simulation model images in RMSE ranging from 0.02 to 0.44. This AI platform can predict pluvial floods one-hour ahead; the total execution time is less than 6 minutes. Kim, 2020 provided the expected inundation area in Gangnam, Korea, due to simultaneous rainfall. The Probabilistic Neural Network (PNN) was used to perform the return period for observed rainfall, and the Support Vector Machine (SVM) and Self-Organizing Mapping (SOM) were used to predict the flood volume and inundation maps.

Although the above studies successfully applied machine learning techniques for flood forecasting and inundation map-generating, the rainfall data, including the rainfall return period statistic, has been crucial. Unfortunately, flood forecasting has still been problematic because of insufficient rainfall data, especially in remote urban areas. Therefore, this study attempts to apply the sliding window techniques for a machine learning-based model to cope with the future flood situation under data scarcity from a limited number of stations. Five algorithms of ML were selected and applied with the proposed sliding window techniques. The most suitable algorithms that resulted in the best agreement of water level at the concerned station were then applied under the specific return period-rainfall. Finally, the example of an application with GIS was conducted. Thus, the inundation map of the study area, Thung Song Municipality, Nakhon Si Thammarat Province, was generated to support flood surveillance and disaster alleviation.

2. STUDY AREA AND METHODOLOGY

2.1 Study Area

Thung Song Municipality in Nakhon Si Thammarat Province locates in the southern part of Thailand. It covers an area of 802.977 sq. km. The geographical characteristic is the high land in the northeast and low land on the Middle side, where the elevation ranges from +19.00 MSL to +1,255.00

MSL. The main river that flows through Thung Song Municipality is Khlong Thalao, as shown in **Fig. 1**, with a steep upstream slope of approximately 0.002. Thus, the runoff from the upstream, namely Namtok Yong, Khlong Thaloan, and Khlong Purk, flow through the city quickly. Consequently, it causes flash floods almost yearly, such as the large February 2006.
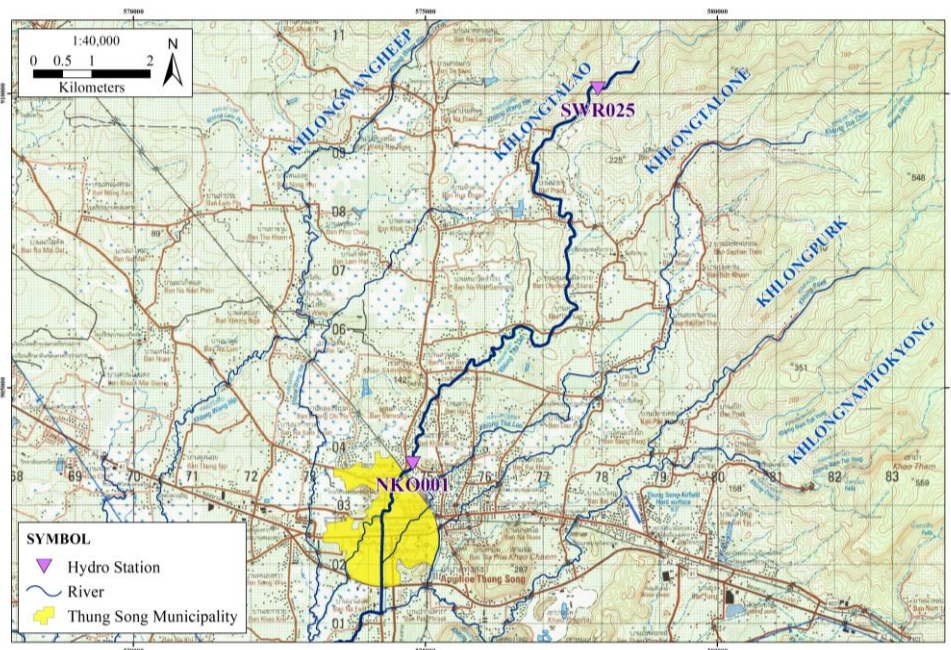
Fig. 1. Thung Song Municipality; Study Area and Route of Runoff Flow through the City.

On February 13-14, 2006, Thung Song accumulated 160 mm. of rainfall, estimated to have a volume of 3.80 million cum, causing the flow in Khlong Thalao of 151 cum/s. As a result, a flash flood occurred at 0200 a.m. February 14 without warning. The water level rose along the river to +51.42 MSL at Sapan Talad Kaset Bridge, +50.47 MSL at Asia Road Bridge, and +51.80 MSL at the Fire Station.

2.2 Flood Event for the Study

To generate the inundation map under a particular flood, this study reviewed data on flood events from 2020 to 2022. The severe flooding was recorded in December 2020, as they appear flood stains on the wall of Thung Song Municipality Office, measuring 1.20 meters, as shown in **Fig. 2**.
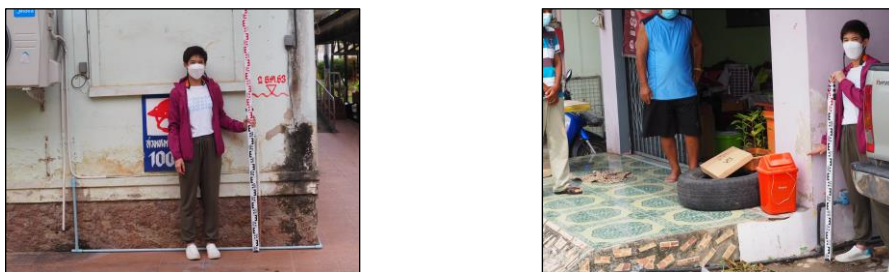
Fig.2. Flood Stains on the wall of Thung Song Municipality Office
due to Flood Event in December 2020.

The water level at the municipality, station NKO001, increased from +52.000 MSL at 0700 p.m. December 1 to the peak of +55.80 MSL at 0200 p.m. December 2, 2020, as shown in **Fig. 3.**
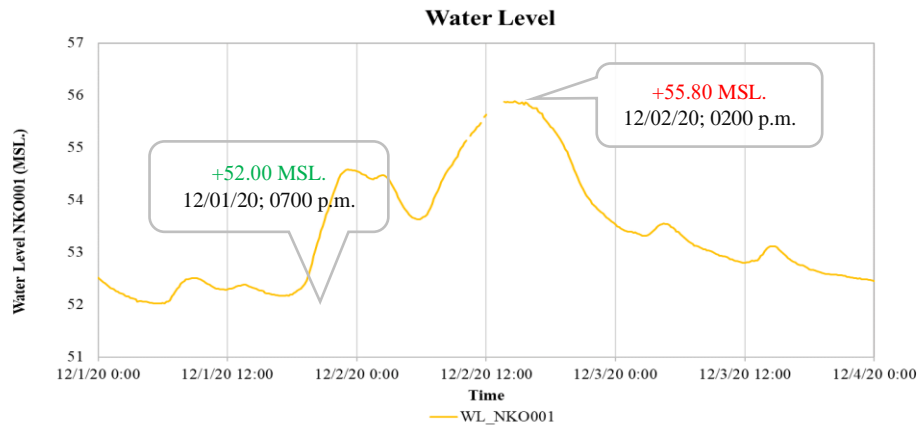
Fig. 3. Water Level of Flood Event in December 2020 at Thung Song Municipality, Station NKO001.

Another flood event occurred in November 2021, where the flood depth at Thung Song Municipality was about 30 cm, as shown in **Fig. 4.**

Fig. 4. Flood Event in November 2021 at Thung Song Municipality (Thairath, 2021).

The flood peak reached within five hours, from 51.33 MSL to 53.26 MSL, as shown in **Fig.5.** However, under the water situation, Thung Song Municipality has a surveillance and warning practice by monitoring the water level at the upstream station. As soon as the officials notice the water level at Station SWR025, as shown in **Fig.1**, rises to the watchful level, they will inform people to lift things high, arrange sandbags and avoid traveling in the path that may be flooded. Generally, the water travel time from Station SRW025 to Station NKO001 at the municipality, which is 10 km apart, is about 3 hours; thus, the people have enough time to deal with the flood situation.
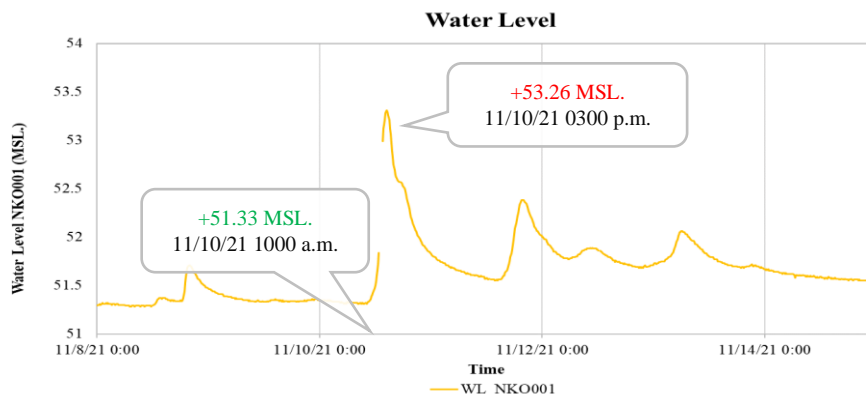
Fig. 5. Water Level of Flood Event in November 2021 at Thung Song Municipality, Station (NKO001).

2.3 Methodology

This study attempts to develop the ML flood forecasting model with the best algorithm. Then, the inundation map, which is practically helpful in flood warning and protection, can be generated. Thus, the works have been divided into six steps, as shown in **Fig. 6.**

Collect Data
- Rainfall Data
- Water Level Data
- Discharge Data

↓

Screen and Set up Data for Sliding Window Technique

↓

Set up Machine Learning Flood Forecasting Model with Sliding Window and 5 Algorithms

↓

Compare Efficiency of 5 Algorithms Machine Learning Flood Forecasting Model
- MSE
- MAE
- R-squared

↓

Generate Inundation Map by Apply the Best Algorithm Machine Learning Flood Forecasting Model with GIS

↓

Compare the Generate Inundation Map with Actual Inundation Area

Fig. 6. Steps of Work.

2.3.1. Data Collection and Handling with Sliding Window Techniques

To forecast the water level downstream where the city is located, the developed model has to learn the flow pattern and relationship between the upstream and downstream water levels. Thus, supervised learning for Machine Learning is considered, and the available data has been explored.

Along Khlong Thalao, which flows directly through Thung Song Municipality, two water level stations belong to Hydro-Informatics Institute (HII). The data throughout 2021 have been selected for ML algorithms testing and learning. Nevertheless, for every 10 minutes of data, 61,566 timesteps, it was found that 3,152 data at SWR025 and 1,045 data at NKO001 were missing. The information in that period was, therefore, cut off.

Considering the consistency of water level data at the two stations 10 km apart, they have a good agreement, as shown in **Fig. 7** and **Fig. 8**. The travel time is about three hours; thus, the time lag could be observed.
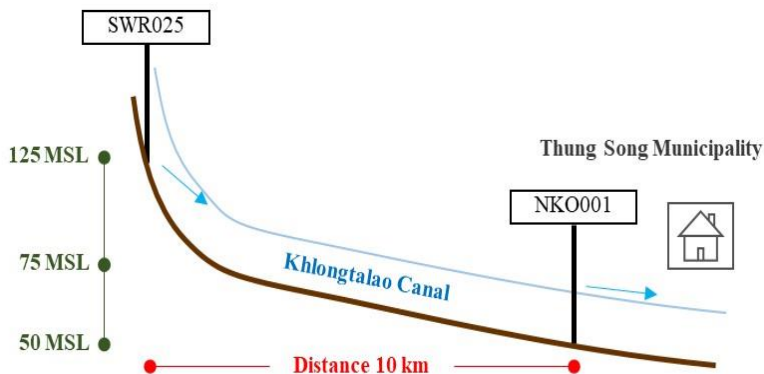
Fig.7 Longitudinal Profile of SWR025 and NKO001 Water Level Stations.
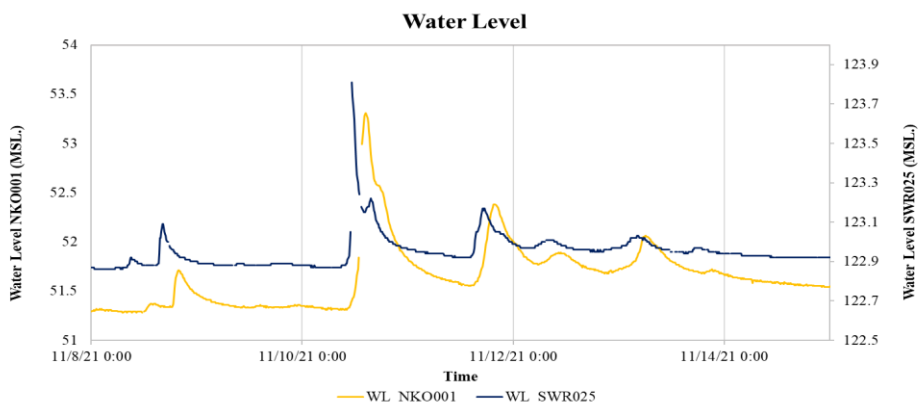
Fig. 8. Water Levels at SRW025 and NKO001 stations during the Flood Event in November 2021.

Time series forecasting can be framed as a supervised problem. The use of previous time steps to predict the next time step is called the sliding window method. Brownlee, 2020 introduced the sliding window technique in reframing data for predicting future values with regression problems. This technique has advantages in predicting multiple time-step ahead of one output variable. The function of regression in the algorithm learns the mapping function from the input variables (X) to predict the output variable (y) as $y = f(X)$. Sequential data analysis is used to restructure the data for supervision for predicting the output of variable (y) in the present time of variable input (X). The input and output sequence data can be used to directly supervise the machine learning algorithm without restruct data $(y_t = f(X_t))$, where t is time. To predict variable (y) multiple steps ahead, the sequence data have to be restructed for the learning algorithm by using previous timesteps for supervised data $(y_{t-n} = f(X_t))$, where n is the number of timesteps (Dietterich, 2002). Variable (y) is shifted to many steps, such as $y_t, y_{t-0.5}, y_{t-1}, …, y_{t-5}$ in the data frame, as shown in **Fig. 9.**

This study applied the sliding window for time series forecasting for multi-variable learning. Two water-level stations' data were restructed. The water level at the upper stream station was the variable (X) to predict the water level at Thung Song Municipality, which was the variable (y), multiple steps ahead, as shown in **Fig. 10.**
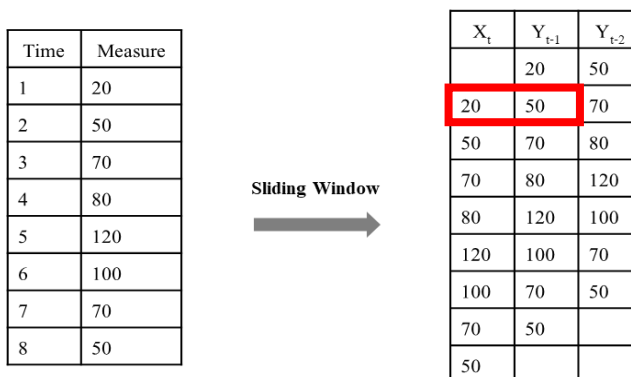
Fig. 9. Sliding Window for Time Series Forecasting Data Frame.
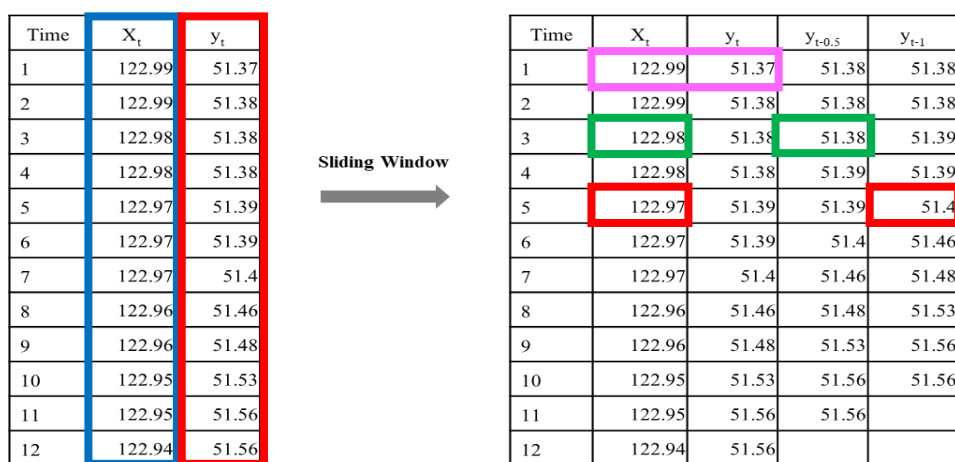
Fig. 10. Sliding Window with Two Variables in this Study.

The water level at SWR025 (X_t) was used to supervise the water level at NKO001, which was restructured to arrange a sequence for prediction every 30 minutes, and five hours ($y_t, y_{t-0.5}, y_{t-1}, ..., y_{t-5}$) ahead. Plotting the water level of SWR025 with sliding window multi-step forecasting water level of NKO001 is shown in **Fig.11**. After arranging the sequence for prediction, every forecasting timestep of the data frame has to start and end at the same period, and the row of missing data must be removed. The total timestep of the data frame for machine learning is 51,109 timesteps. Data have been divided into 60%, 20%, and 20% for training, validation, and testing. Then the data was input into five machine-learning regression algorithms.

2.3.2. Machine Learning Flood Forecasting Model Setting up

Mosavi et al., 2018 defined ML as a field of artificial intelligence (AI) used to induce regularities and patterns, providing more straightforward implementation with low computation cost, as well as fast training, validation, testing, and evaluation, with high performance compared to physical models, and relatively less complexity. Furthermore, several ML algorithms, e.g., artificial neural network (ANNs), neuro-fuzzy, support vector machine (SVM), and support vector regression (SVR), were reported as practical flood forecasts. However, the capability of forecasting depends on their training, whereby the system learns the target task based on past data. Géron, 2019 classified ML systems according to the amount and type of supervision they get during the training into four categories: supervised learning, unsupervised learning, semisupervised learning, and reinforcement learning.
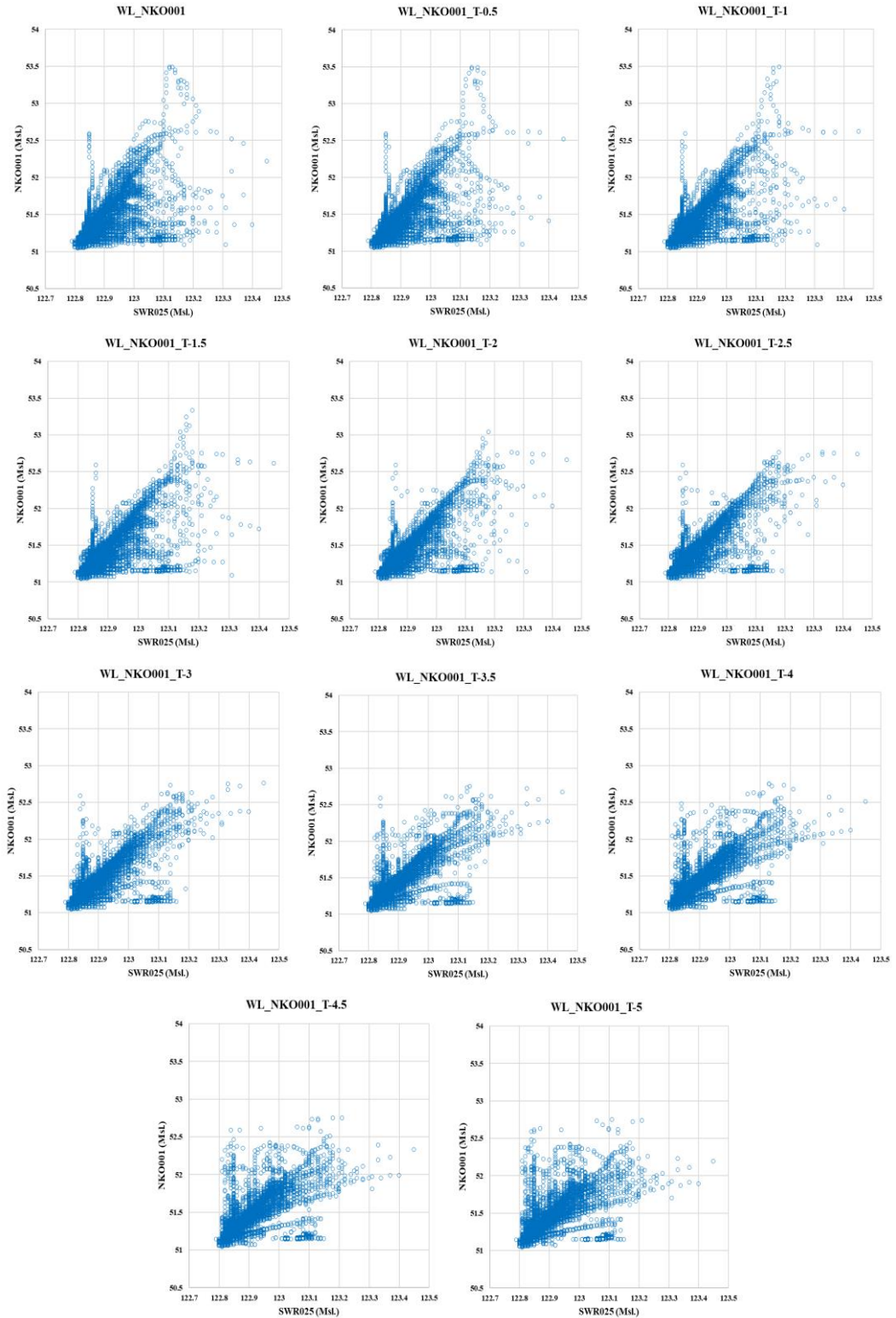
Fig. 11. Distribution of Water Level data of SWR025 and NKO001 after arranging the sequence for prediction.

In this study, five supervised learning regression algorithms, which are Linear Regression (LR), Support Vector Machines (SVMs), K-Nearest Neighbors (KNN), Decision Trees (DT), and Random Forest model (RF), have been selected. The detail of each algorithm is as briefly described.

1) The Linear regression model (LR) predicts by simply computing a weighted sum of the input features plus a constant called the *bias term*, as Linear regression model prediction Equation (1) shown in **Fig. 12** (Géron, 2019).

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n \tag{1}$$

$\hat{y}$ is the predicted value,

n is the number of features,

x_1 is the number of features, and

θ_j is the jth model parameter (including the *bias term* θ_0 and the feature weights $\theta_1, \theta_2, \ldots, \theta_n$)
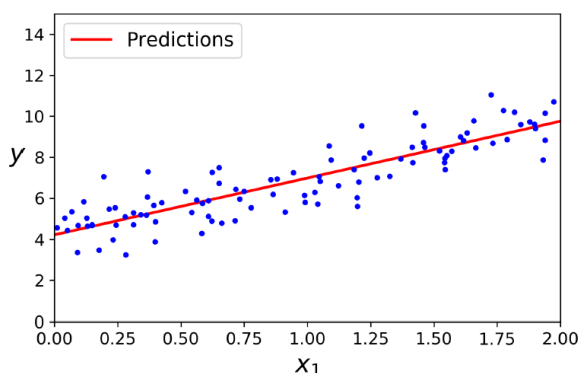
Fig. 12. Linear Regression and Model Prediction (Géron, 2019: p.118)

2) The Support Vector Machine model (SVM) finds hyperplanes by the maximum linear margin of support vectors. Suppose data cannot classify as linear. Then, kernel function handling with nonlinear datasets adds more features, such as polynomial, sigmoid and radial basis functions (RBF). **Fig. 13** shows two linear SVM Regression models train on random linear data, one with a large margin ($\epsilon = 1.5$) and the other with a small margin ($\epsilon = 0.5$) (Géron, 2019).
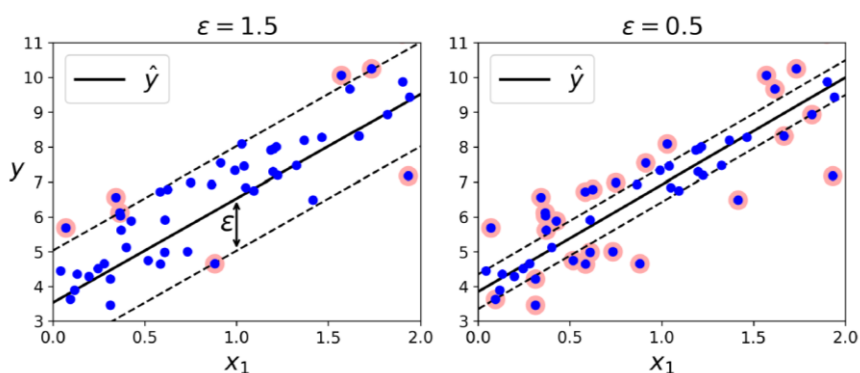
Fig. 13. SVM Regression (Géron, 2019: p.165).

3) K-Nearest Neighbors model (KNN) predicts the category of test samples according to training sample *k,* which is the closest neighbor to the test sample and inserts it into a category with the greatest

probability. Near or far distances to neighboring points can be calculated using the Euclidean distance equation (2) (Andrian, 2019). **Fig. 14** shows a sample of new instance in the K-Nearest Neighbors model (Géron, 2019).

$$D(a,b) = \sqrt{\sum_{k=1}^{d}(a_k - b_k)^2} \qquad (2)$$

D is the distance between points
a is the known point
b is the unknown point
d is the dimension of the point being measured
k is the value of neighboring data measured
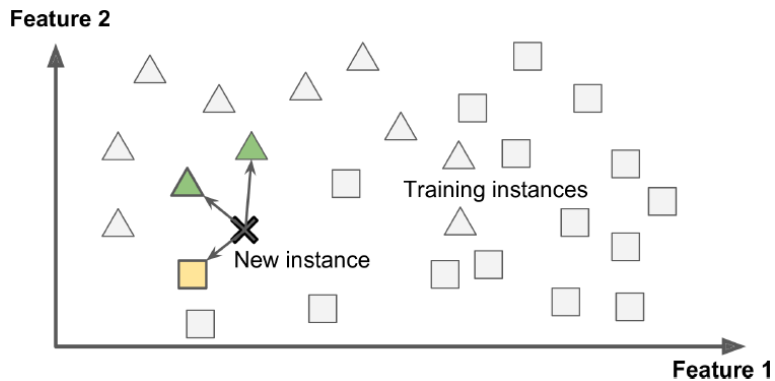
Fig. 14. K-Nearest Neighbors Model (KNN), (Géron, 2019: p.19).

4) The Decision Tree model (DT) expresses data with a tree-like graph based on the rules or conditional statements of the variables, subdivides them into similar data types by separation rules, and continues this classification until the final classification criteria are satisfied. In a DT, through a binary recursive partitioning process, split variables and split points that minimize the mean squared error (MSE) are identified in each step. In addition, "pruning" is performed to determine the tree size that minimizes the MSE using cross-validation to prevent overfitting. A Decision Tree (DT) conceptual diagram is shown in **Fig. 15** (Lee et al., 2020).
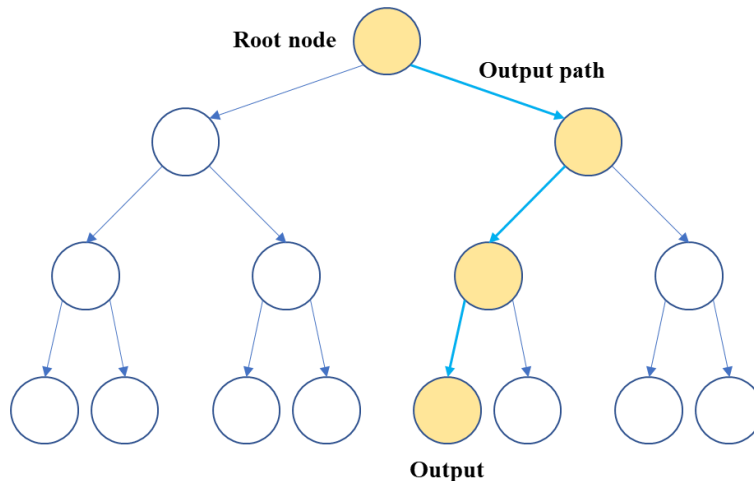
Fig. 15. Conceptual Diagram of a Decision Tree (DT); modified from: (Lee et al., 2020: p.4).

5) The Random Forest model (RF) is an ensemble of Decision Trees, generally trained via the bagging method. Random Forest increases the number of Decision Trees and creates multiple training data sets from one data set. It has improved predictive power as a result of creating multiple DTs through multiple learnings and then combining multiple DTs. The Conceptual diagram of a Random Forest (RF) is shown in **Fig. 16** (Lee et al., 2020).
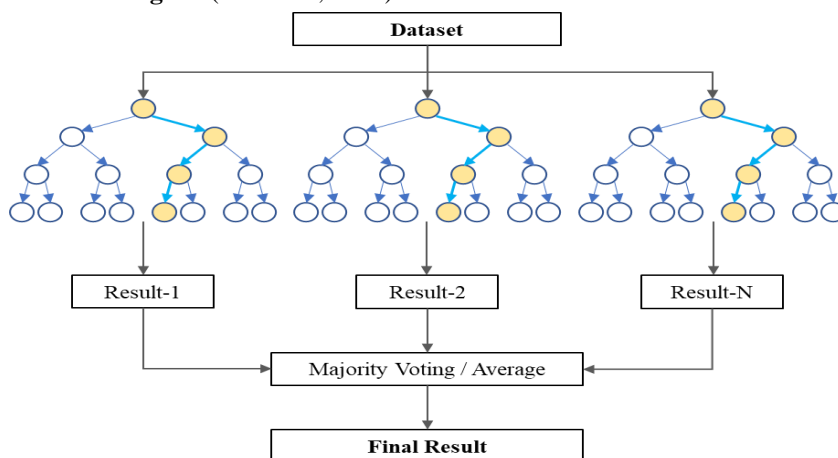
Fig. 16. Conceptual Diagram of a Random Forest (RF); modified from: (Tbico, 2021).

Regression is a modeling method that learns relationships within provide and predicts using this relationship. The kind of model that learns and predicts continuous data is called a regression problem. Regression was typically found in the following scenario (Haykin, 2008). One of the random variables is considered to be of particular interest; that random variable is referred to as a dependent variable or *response*.

The remaining random variables are called independent variables or *regressors.* Their role is to explain or predict the response's statistical behavior and the response's dependence on the regressors. It includes an additive *error* term to account for uncertainties in how this dependence is formulated. The error term is called the expectational or *expectational error*, which is used interchangeably.

2.3.3. Five Algorithms Forecasting Efficiency Comparison

As previously illustrated how various algorithms function; LR, SVM, KNN, DT, and RF. LR often solves regression problems by examining relationships between two or more variables. It supports only linear solutions, while SVM, KNN, DT, and RF support both linear and non-linear solutions. Therefore, when there is a large number of features with fewer data-set (with low noise), LR may outperform DT/RF. In this study, five algorithm performances of forecasting have been evaluated by three parameters: MSE, MAE, and R-squared. Mean Square Error (MSE) is an average squared deviation of the predicted value, Mean Absolute Error (MAE) indicates how the predicted values are far from the measured values, and the R-squared is the strength of the relationship between the predicted values and actual values (Yafouz et al., 2021), as equations shown below:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{3}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{4}$$

$$R - squared = 1 - \frac{(y_i - \hat{y}_i)^2}{(y_i - \bar{y}_i)^2} \tag{5}$$

where N is the number of observations, y is the vector of the actual values, and $\hat{y}$ is the vector of the predicted values (Surakhi et al., 2021).

2.3.4. Inundation Map Generating by Applying ML Flood Forecasting with GIS

Once the best algorithm for the ML Flood Forecasting model has been obtained, the model is input with water level data at the upstream station, SWR025, to forecast the water level rise at municipal sites. Next, the water level overflows of the bank will be analyzed for the extent of the flood area by Mike 11HD, Mike21FM, and Mike Flood mathematical models.

MIKE 21 Flow Model (Mike21FM) is a modeling system for 2D free-surface flows. It is applicable to the simulation of hydraulic and environmental phenomena in lakes, estuaries, bays, coastal areas, and seas (DHI, 2017b). It requires information on bathymetry. Thus, a Digital Elevation Model (DEM) resolution of 20 m^2 covering all river networks was used to generate a bathymetry file in this study, as shown in **Fig.17.** Mike Flood is coupling Mike11HD and Mike21FM to simulate flood maps in hourly steps. Its linkages set up to Lateral Link allow a string of MIKE21FM cells/elements to be laterally linked to a given reach in MIKE11HD, either a branch section or an entire branch. Flow through the lateral link is calculated using a structural equation. This link is particularly useful for simulating overflow from a river channel onto a flood plain, as the concept shown in **Fig.18** (DHI, 2017c). Further, the extent of the inundation area is illustrated with GIS techniques.
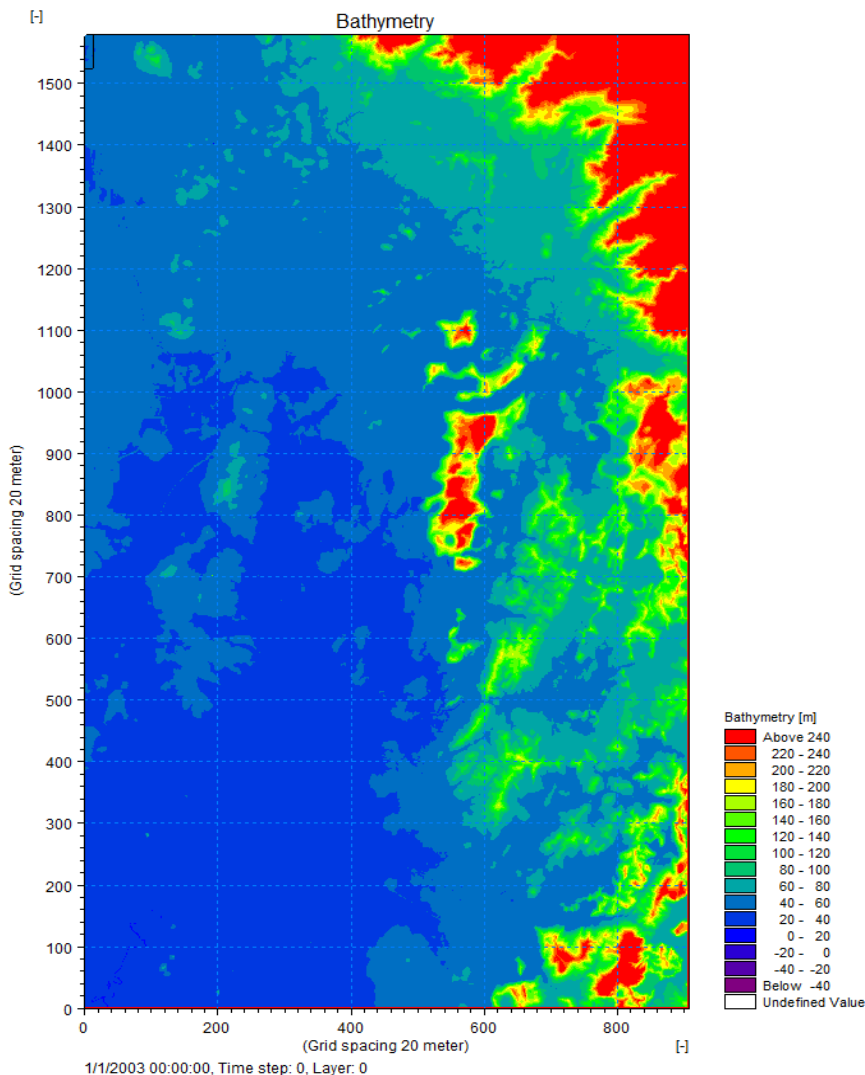
Fig. 17. Bathymetry of the Study Area.
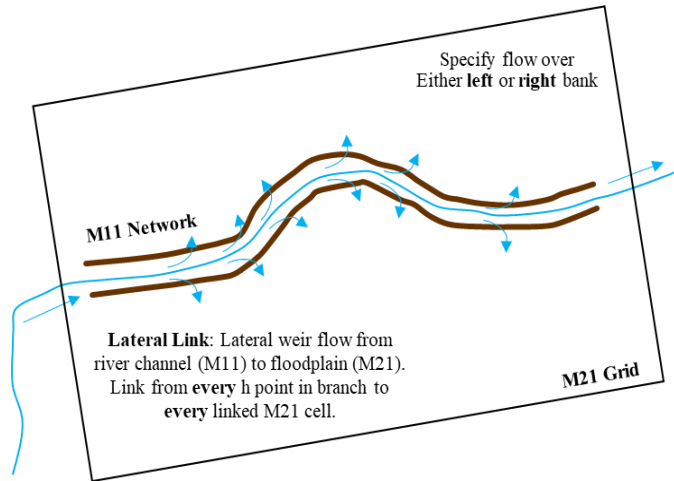
Fig. 18. Application of Mike 11 and Mike 21 on Flood Area Analysis: modified from: (DHI, 2017c: p.17).

3. RESULTS AND DISCUSSIONS

3.1. Data Scattering Analyze by Sliding Window Technique

To figure out the scattering pattern of the water level data at SWR025 and NKO001, the data at these two stations at present and multiple future steps have been analyzed and viewed by scatter plot, as shown in **Fig. 19**. The relationship shows a linear trend with many noises. The noise of the data was largely found below the 45° line from the Timesteps 0-3 hours, and a large amount of noise above the 45 ° line in the Timesteps 3.5-5 hours, while the noise at the Timesteps of 3 hours is comparatively less than the other Timesteps. It indicates that the rise and fall of the water level at NKO001 at 3 hours is close to the current SWR025 water level.
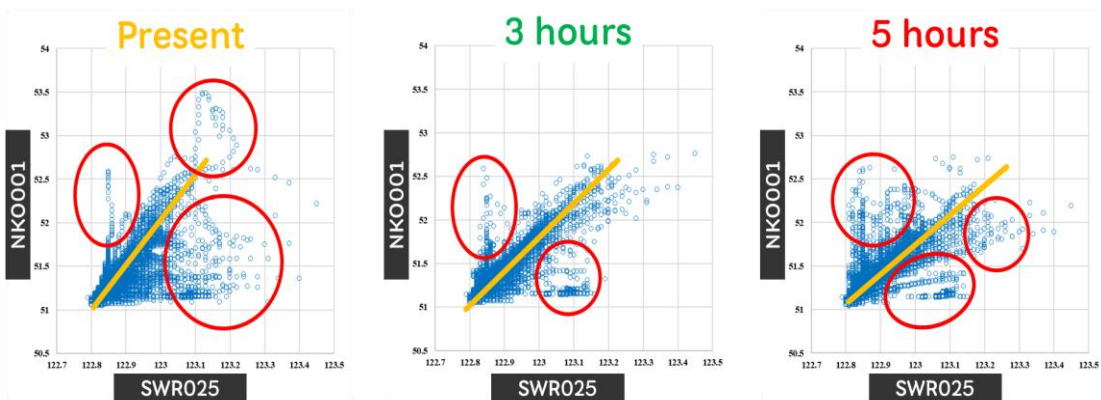
Fig. 19. Water Level Data Scattering Patterns in Two Stations-Scattering Analyzes.

To finish the sequence for prediction, every forecasting timestep of the data frame has to start and end at the same period, and the row of missing data must be removed. The total timestep of the data frame for machine learning is 51,109 timesteps. Later, the data were divided into 60%, 20%, and 20% for training, validation, and testing.

3.2. Machine Learning Forecasting Models Performance

Since the appropriate time step forecasting was decided according to the scattering pattern of the water level data, five algorithms have been supervised to learn with water level data of SWR001 and NKO001 with multi-step forecasting. In addition, MSE, MAE, and R^2 compared the algorithms' performance.

The performance of five algorithms on multi-step water level forecasting has been evaluated, and the results are presented in **Table 1**. The Random Forest algorithm performs best among other algorithms by MSE 0.006, MAE 0.044, and R^2 0.75. Further, most algorithms show that the best forecast time is 3 hours ahead, except for K-Nearest Neighbors, whose best forecast time is 2.5 hours. It is consistent with the flow time from station SWR025 to station NKO001, which takes 3 hours.

Table 1.

Performance Comparison among Five Algorithms on Multi-step Water Level Forecasting.

Algorithm	Model Performance	Forecasting time (hr.)										
		+0	+0.5	+1.0	+1.5	+2.0	+2.5	+3.0	+3.5	+4.0	+4.5	+5.0
Linear Regression	MSE	0.013	0.012	0.012	0.011	0.01	0.01	**0.01**	0.01	0.01	0.01	0.011
	MAE	0.063	0.061	0.06	0.058	0.056	0.055	**0.054**	0.054	0.054	0.055	0.056
	R^2	0.518	0.54	0.56	0.581	0.594	0.603	**0.604**	0.598	0.584	0.567	0.548
Support Vector Machine	MSE	0.01	0.009	0.009	0.008	0.008	0.007	**0.007**	0.007	0.007	0.007	0.008
	MAE	0.052	0.05	0.053	0.052	0.052	0.052	**0.051**	0.051	0.051	0.052	0.053
	R^2	0.641	0.66	0.67	0.693	0.7	0.706	**0.708**	0.704	0.697	0.685	0.671
K-Nearest Neighbors	MSE	0.01	0.009	0.008	0.008	0.007	**0.007**	0.007	0.007	0.007	0.007	0.008
	MAE	0.054	0.052	0.051	0.05	0.049	**0.047**	0.047	0.046	0.046	0.047	0.049
	R^2	0.628	0.657	0.676	0.697	0.713	**0.724**	0.722	0.719	0.708	0.694	0.676
Decision Tree	MSE	0.009	0.008	0.007	0.007	0.007	0.006	**0.006**	0.006	0.006	0.007	0.007
	MAE	0.051	0.049	0.048	0.047	0.046	0.045	**0.044**	0.044	0.045	0.045	0.046
	R^2	0.672	0.697	0.718	0.734	0.738	0.746	**0.749**	0.744	0.733	0.719	0.704
Random Forest	MSE	0.009	0.008	0.007	0.007	0.006	0.006	**0.006**	0.006	0.006	0.007	0.007
	MAE	0.051	0.049	0.048	0.047	0.046	0.045	**0.044**	0.044	0.045	0.045	0.046
	R^2	0.672	0.697	0.719	0.736	0.741	0.748	**0.75**	0.744	0.734	0.719	0.704

Additionally, the forecasted water level at the NKO001 station from Random Forest Algorithm shows good agreement with the observed water level, as shown in **Fig. 20.** Thus, the developed ML model's accuracy of 3 hours ahead forecasting is validated.
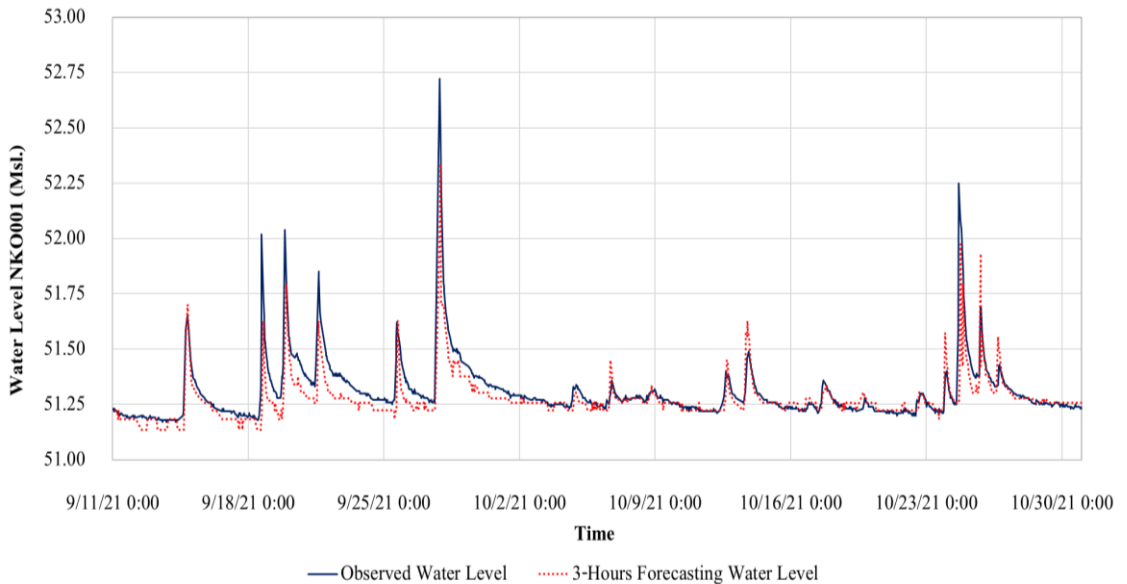
Fig. 20. Comparison between the Observed Water Level and the 3-hours Forecasting at Station NKO001; flood events in 2021.

Then, based on the water level at the upstream station, SWR025, the forecasting of the water level at the surveillance station, NKO001, can be conducted.

3.3. Dataset of Flood Maps

Since the Random Forest algorithm showed the best performance in ML flood forecasting, as previously described in 3.2, it was then applied to another set of water levels that caused the flood in June 2022. The comparison between the forecasted water level and the observed at the downstream station, NKO001, is shown in **Fig. 21.** The forecasting accuracy was then reconfirmed with the MSE 0.009, MAE 0.056, and R^2 0.75.
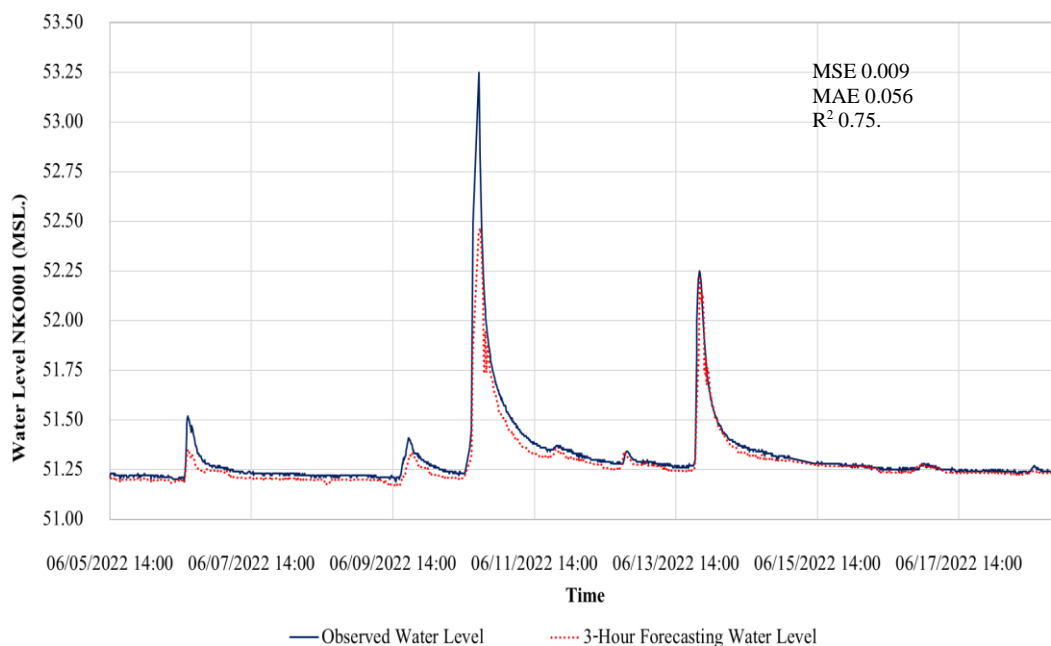
Fig. 21. Comparison between the Observed Water Level and the 3-hours Forecasting at Station NKO001, flood events in June 2022.

As an example of applying the results of flood forecasting to create flood maps, further, the extent of the flood area was evaluated by generating a flood map with the Mike Flood, coupling 1D and 2D models where the Mike11HD models the river and Mike21FM models the floodplain with lateral links.

The simulation consumed the amounts of input data and time processing the flood maps; the water depth and surface elevation are obtained in the ".dfs2" file type. It is shortly transformed into a ".KML" file type by Mike to Google Earth; to store the flood maps with a resolution of 20x20 m. in the database. It can be updated in real-time according to the reported water level upstream frequency. Thus, the people in flood-prone areas can properly take flood alleviation action in time.

The generated flood map is overlayed with the actual inundation area obtained from the municipality, as shown in **Fig. 22.** The agreement means that whenever the water level is reported at the upstream station, the inundation area can be evaluated. Then, the flood warning procedures can proceed based on the generated flood map.
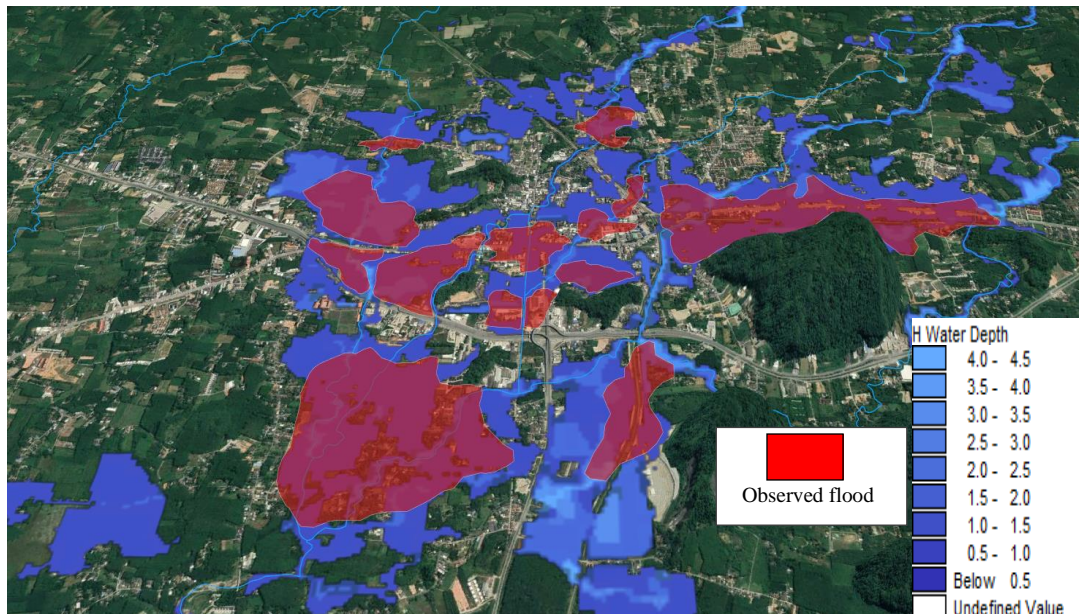
Fig. 22. Flood Map of Thung Song Municipality due to the peak flow on June 10, 2022.

3.4. Flood Forecasting System

Finally, the Machine Learning Forecasting Model and the Flood Maps Database have been gathered as the Flood Forecasting System. Since the real-time water level data of SWR025 and NKO001 are input into the Flood Forecasting System, the Machine Learning Forecasting Model forecasts the water level +0, +0.5, +1.0, … ,+5.0 hours ahead. The forecasted result will be matched with the flood map from the database at the same water elevation at NKO001. Whether the flood maps indicate the risk area in Thung Song Municipality, the system will send an alarm to the Thung Song Municipality officers and related agencies.

The procedure for surveillance and alerting people at risk of flooding is shown in **Fig. 23.**
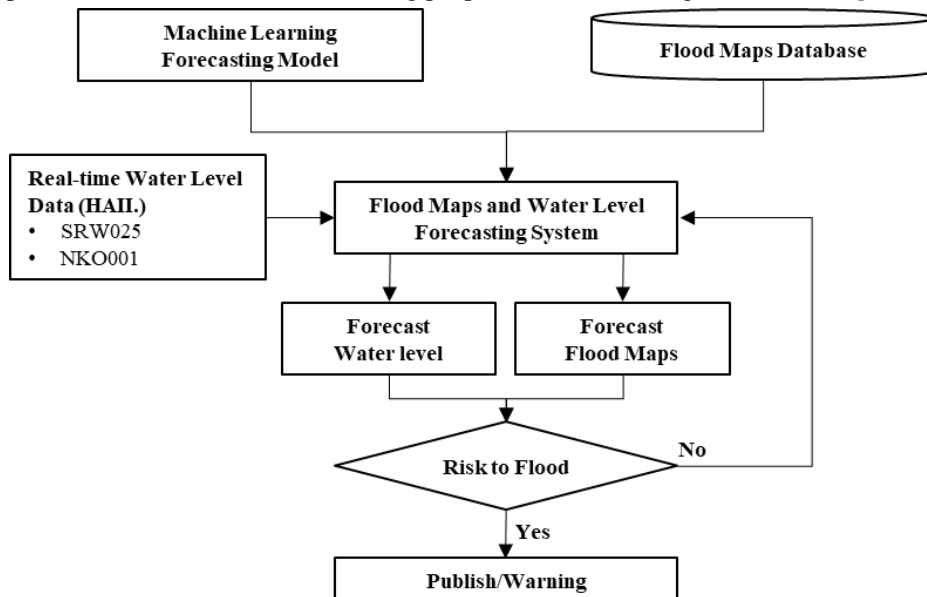
Fig. 23. Overall Procedure of Flood Forecasting and Surveillance.

4. CONCLUSION

The short-term flood warning information in remote areas can be problematic due to insufficient rainfall data and the limitation of water level observed stations. In addition, most numerical flood forecasting models require detailed physical data and are highly computational and time-consuming. Thus, the related agencies may not be able to warn people in risk areas accurately and timely.

This study developed a machine learning-based model to cope with the future flood situation in Thung Song Municipality in Nakhon Si Thammarat under data scarcity. Using the water level relationship between the upper station, SWR025, and the downstream station, NKO001, a 2-variable sliding window technique has been applied. First, the data has been arranged and restructed; the water level at the upstream station was the variable input to predict the water level downstream. After the water level data scattering has been analyzed, the best prediction sequence can be decided. Then, five Machine Learning algorithms which are Linear Regression, Support Vector Machine, K-Nearest Neighbor, Decision Tree, and Random Forest, were supervised to learn with water data with multi-step forecasting. Their performance was compared based on the MSE, MAE, and R^2. As a result, the Radom Forest performed the best by MSE 0.006, MAE 0.044, and R^2 0.75, under 3 hours ahead of forecasting.

Consequently, the developed ML flood forecasting model has been validated by inputting the water level at the upstream station, which is 10 km from the municipality. The 3-hour forecasted water level at Thung Song Municipality station showed good agreement with the observed flood in November 2021. Furthermore, the "Flood Forecasting System" can be conducted in the next step as the extent of the flood area has been evaluated by generating a flood map with the Mike Flood mathematical model. The water depth and surface elevation are shortly transformed by Mike to Google Earth. As a result, the 3-hour forecasted inundation area under that specific rainfall, which can be real-time updated, is helpful for flood warning and disaster alleviated procedures.

In summary, even in remote areas with limited water level measurement stations and incomplete data, Flood warnings can also be notified by a developed machine learning-based model. The introduced 2-variable sliding window technique and the algorithm must be selected to suit the data set and then applied with the Mike Flood mathematical model. Further, the extent of flood-prone areas can be shown in conjunction with the GIS system. The information under specific rainfall is promptly notified in advance.

A C K N O W L E D G E M E N T

R E F E R E N C E S

Andrian, R., Naufal, M.A., Hermanto, B., Junaidi, A., and Lumbanraja, F.R. (2019). k-Nearest Neighbor (k-NN) Classification for Recognition of the Batik Lampung Motifs. *Journal of Physics*. 1338(2019) 012061. https://doi:10.1088/1742-6596/1338/1/012061

Brownlee, J. (2020). *Introduction to Time Series Forecasting with Python, How to Prepare Data and Develop Models to Predict the Future*. Machine Learning Mastery.

Chang, L.C., Amin, M.Z.M., Yang, S.N., and Chang, F.J. (2018). Building ANN-Based Regional Multi-Step-Ahead Flood Inundation Forecast Models. *Water*. 10. 1283. https://doi:10.3390/w10091283

Chang, D.L., Yang, S.H., Hsieh, S.L., Wang, H.J., and Yeh, H.C. (2020). Artificial Intelligence

Methodologies Applied to Prompt Pluvial Flood Estimation and Prediction. *Water*. 12. 3552. https://doi:10.3390/w12123552

DHI. (2017a). *MIKE 11 A modeling system for rivers and channels*, User Guide. Denmark.

DHI. (2017b). *MIKE 21 Flow model FM, hydrodynamic module*, User Guide. Denmark.

DHI. (2017c). *MIKE FLOOD, 1D-2D Modelling*, User Manual. Denmark.

Dietterich, T.G. (2002). Machine Learning for Sequential Data: A Review. *LNCS*. 2396.

Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.).* Penguin Books. O'Reilly Media.

Haykin, S. (2008). *Neural Networks and Learning Machines (3rd ed)*. McMaster University. Pearson Prentice Hall.

Hydro-Informatics Institute (HII.), (2021, November 25). Flood Forecasting System. http://www.thaiwater.net/floodforecast

Kim, H. I., and Han, K.Y. (2020). Inundation Map Prediction with Rainfall Return Period and Machine Learning. *Water*. 12. 1552. https://doi:10.3390/w12061552

Lee, J.Y., Choi, C., Kang, D., Kim, B.S., and Kim, T.W. (2020). Estimating Design Floods at Ungauged Watersheds in South Korea Using Machine Learning Models. *Water*. 12. 302. https://doi:10.3390/w12113022

Rafiei Emam, A., Mishra, B., Kumar, P., Masago, Y., & Fukushi, K. (2016). Impact Assessment of Climate and Land-Use Changes on Flooding Behavior in the Upper Ciliwung River, Jakarta, Indonesia. *Water*, 8(12). doi:10.3390/w8120559

Surakhi, O., Zaidan, M.A., Fung, P.L., Motlagh, N.H., Serhan, S., Khanafseh, M.A., Ghoniem, R.M., and Hussien, T. (2021). Time-Lag Selection for Time-Series Forecasting Using Neural Network and Heuristic Algorithm. MDPI. *Electronics*. 10. 2518. https://doi:10.3390/electronics10202518

Tibco. (2021). *What is a Random Forest?* [Online] Available from: https://www.tibco.com/reference-center/what-is-a-random-forest [Accessed April 20, 2022]

Thairath. (2021). *Flooding on November 29, 2021, in Nakhon Si Thammarat Province*. [Online] Available from: https://www.thairath.co.th/news/local/south/2253247 [Accessed April 25, 2022]

Yafouz, A., Ahmed, A.N., Zaini, N., Sherif, M., Sefelnasr, A., and Shafie, A.E. (2021). Hybrid deep learning model for ozone concentration prediction: comprehensive evaluation and comparison with various machine and deep learning algorithms. *Engineering Applications of Computational Fluid Mechanics*. 15. 1. (902-933) https://doi.org/10.1080/19942060.2021.1926328

INVESTIGATION OF SOIL EROSION IN AGRO-TOURISM AREA: GUIDELINE FOR ENVIRONMENTAL CONSERVATION PLANNING

Ni Made TRIGUNASIH [1] and Moh SAIFULLOH [2]*

ABSTRACT:

Most agro-tourism areas are located in the upstream area of the watershed, in the highlands, steep slopes, and the air temperature is peaceful and calm. One of them is in the tourism area in the Penet upstream watershed, Bali, Indonesia. Agro-tourism, in addition to providing economic benefits to agricultural products, also from tourist visits to the panorama and landscape of agricultural land. Land use that is not wise and not following conservation rules can cause erosion. This study aims to investigate soil erosion rate and plans for environmental-based conservation. The method used to quantify erosion uses Universal Soil Loss Equation (USLE) with a geographic information system (GIS) approach. This research integrates field surveys and soil analysis in the laboratory. The result showed that the erosion rate in agro-tourism areas ranges from 0.32 t ha^{-1} yr^{-1} (very light) to 1,535.34 t ha^{-1} yr^{-1} (very heavy). The agro-tourism areas affected by erosion with heavy to very heavy categories are the villages of Antapan, Bangli, Apuan, Angseri, and Candikuning. Conservation actions can be taken, including improving plant management factors with dense vegetation, increasing land management actions by constructing bench terraces, and planting parallel to contour lines. It is recommended for further researchers to model erosion by integrating remote sensing data and geographic information systems in order to complement data that cannot be obtained in the field or the laboratory.

Key-words: *Erosion, USLE, Agro-tourism, Conservation, Geographic Information System (GIS)*

1. INTRODUCTION

The island of Bali has been known by many tourists worldwide as a famous tourist destination (Rideng et al., 2020). Natural tourism objects such as beaches, lakes, mountains, cultural tourism objects, and agro-tourism are all found on the island of Bali (Utama, 2020). Many researchers have studied the agro-tourism sector regarding marketing, development concepts, and economic benefits . Agro-tourism is an activity that seeks to develop the natural resources of an area with the potential for agriculture to be used as a tourist area. Plantation areas, specific vegetable-producing centers, and rural areas have great potential to become agro-tourism objects (Mahanani et al., 2021; Marin, 2015; Rosardi et al., 2021). The contained potential must be viewed in terms of the natural environment, geographical location, types of products, agricultural commodities produced, facilities, and infrastructure (Satriawan et al., 2015). In this study, we examine from a different perspective, namely, the impact of agro-tourism activities in the upstream area of the river on the environment. Apart from lowland rice, horticultural crops such as vegetables are one of the commodities cultivated in this upstream of agrotourism area. Characteristics of horticultural cultivation are intensive soil cultivation. Intensive tillage can cause the stability of soil aggregates to be disturbed. The upstream area of this agro-tourism area has an average annual rainfall of 3,110 mm/year. Cultivation of horticultural crops in areas with steep slopes, intensive tillage, and high rainfall can cause erosion.

Another problem with agricultural activities on steep slopes is that they are prone to landslides (Diara et al., 2022; Trigunasih & Saifulloh, 2022). Therefore, using sloping land to meet human

---

[1]* *Soil Sciences and Environment, Faculty of Agriculture Udayana University, 80234 Denpasar, Indonesia,*
  Corresponding author *, *trigunasih@unud.ac.id*

[2] *Spatial Data Infrastructure Development Center Udayana University, 80234 Denpasar, Indonesia,*
  moh9saifulloh@gmail.com

interests requires the application of soil and water conservation. Soil and water conservation are efforts to maintain and increase soil productivity, quantity, and water quality. Applying soil and water conservation techniques is the primary strategy in efforts to preserve and utilize the environment and natural resources (Narendra et al., 2021; Nugroho et al., 2022). Soil and water conservation is beneficial in mitigating climate change and land degradation in the long term. Therefore, it is essential always to pay attention to conservation rules so that the land is not susceptibility to erosion.

Based on the abovementioned problems, the authors assume it is necessary to research erosion mapping and determine appropriate conservation plans in agro-tourism areas. This study aims to quantify the amount of erosion and its spatial distribution and to plan environmental-based conservation measures.

2. STUDY AREA

This research was conducted in the upstream area of agro-tourism on the island of Bali (**Fig 1a**). Hydrologically this area is included in the upstream Penet and Beratan watersheds. The slope of the slope is dominated by steep, with altitudes ranging from 334 - 2142 masl (**Fig 1b**). This area is located in Baturiti District, Tabanan Regency, Bali Province. Several villages use the land for agro-tourism activities, such as Candikuning, Angseri, Baturiti, and Antapan villages. The main tourist attraction is Lake Batur, with superior commodities on the surrounding agricultural land: strawberries, peppers, carrots, etc. Geographically, is located at 08° 14'30"-08° 38'07" latitude and 114° 59'00"-115° 02'57" east longitude.
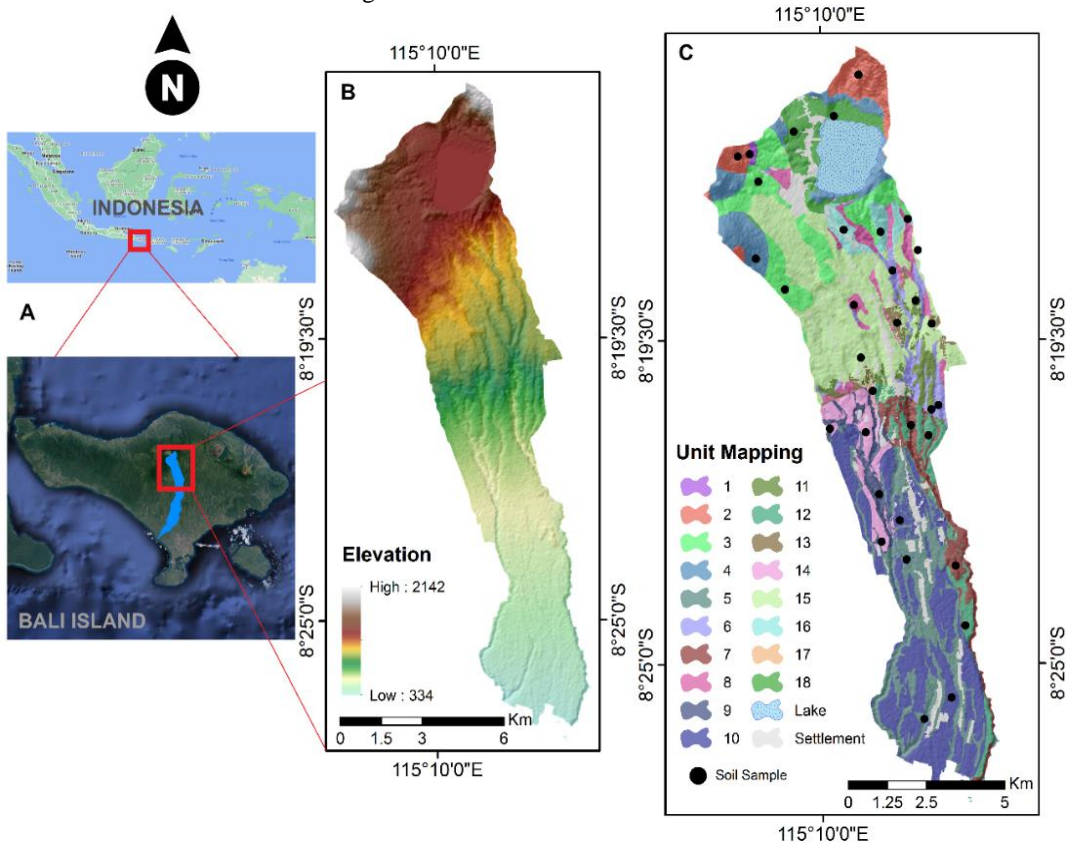
Fig. 1. The research location is viewed from a global scale in Indonesia and the island of Bali (a),
a regional scale and shows the altitude of the research area (b),
mapping units and the location of soil samples in the field (c).

3. DATA AND METHODS

3.1. Tools and Materials

The materials needed for our reserch include (1) chemical compounds (Calgon 5%, H_2O_2 30%, H_2O, $K_2Cr_2O_7$, concentrated H_2SO_4, concentrated H_3PO_4, DPA, $FeSO_4$); (2) soil samples obtained from field observations; (3) rainfall data for the last ten years from Meteorological, Climatological, and Geophysical Agency (BMKG); (4) soil type maps ; (5) land use maps, and slope class maps derived from Alos Palsar Digital Elevation Model (DEM), and Google Earth Satellite Imagery. The tools used include (1) Belgy drill; (2) sample ring; (3) field knife; (4) Abney levels; (5) ArcGIS 10.8 software; (6) laboratory analysis equipment, such as permeability determination apparatus, mug or glass beaker, pipette, tin, Erlenmeyer, burette, Petri dish, texture sieve, measuring cup, oven, and scale.

3.2 Data Analysis

The method used in this research is a field survey method by taking soil samples using the purposive sampling technique and then the soil samples analysis at the Laboratory. The stages of research implementation are (1) preparation (library study, determination of homogeneous land units, and preliminary survey); (2) field survey and soil sampling; (3) soil analysis in the Laboratory; (4) erosion calculations; (5) Tolerable erosion (EDP) calculation; (6) soil and water conservation planning; and (7) erosion mapping. This research mapping unit is based on thematic map overlays (soil type, land use, and slope). Eighteen mapping units were obtained, which were then used to take soil samples in the field. Soil sampling was carried out by purposive sampling and compositely, referring to the mapping unit that had been made previously. The map of the mapping unit and the soil sampling point is presented in **Fig. 1c**.Two parameters are observed, namely those observed in the field and those observed in the Laboratory. The parameters observed in the field include:

1. the length of the slope (L) measured using a meter,
2. the slope (S) measured or observed with the Abney level,
3. visually observed soil structure based on the shape and type of soil structure,
4. adequate depth measured with a Belgy Drill,
5. visually observed type and density of vegetation (C), and
6. visually observed land management (P).

Meanwhile, the parameters observed in the Laboratory include

1. soil texture using the pipette method,
2. organic materials using the Walkey and Black method,
3. permeability using the De Booth method based on Darcy's law, and
4. soil volume weight using the ring sample method.

Parameter criteria through field observations and laboratory analysis adopted from previous researchers, which were carried out in the Galunggung Watershed, Bali Province (Trigunasih et al., 2018). We adopted the previous researchers because our area's biophysical characteristics are similar to those of previous researchers. The difference is in the research area and the recommended conservation output, that this research is specific for conserving the upstream watershed area for ago-tourism activities.

a. Calculation of soil Erosion

The method used for calculating erosion is the USLE method proposed by (Wischmeier & Smith, 1978) with the following Eq. 1

$$A = R \times K \times LS \times C \times P \tag{1}$$

where A is the weight of soil lost (t ha^{-1} yr^{-1}), R is the rain erosivity factor (kj ha^{-1} cm^{-1}), K is the soil erodibility factor (t kj^{-1}), LS is the factor of slope length and slope, C is the factor of vegetation cover and plant management, and P is the factor of land management.

b. Rainfall Erosivity (R)

Rain erosivity (R) is the ability of rainwater to cause erosion which is calculated using the Bols methode (Wischmeier & Smith, 1978) with the Eq. 2

$$EI_{30} = 6.119(RAIN)^{1.21} \times DAYS^{-0.47} \times MAXP^{0.53} \tag{2}$$

where EI_{30} is the monthly rainfall erosivity, RAIN is the average monthly rainfall (cm), DAYS is the average rainy day in one month, and MAXP is the maximum rainy day in the month concerned (cm).

c. Soil Erodibility (K)

Soil erodibility (K) is the ease with which the soil erodes, which can be calculated using the following (Wischmeier & Smith, 1978) Eq. 3

$$100K = 1.292 [2.1 M^{1.14} (10^{-4}) (12-a) + 3.25 (b-2) + 2.5 (c-3)] \tag{3}$$

where K is the soil erodibility value. M is the percentage of very fine sand fraction (0.1 – 0.05 mm diameter) and dust fraction (0.05 – 0.002 mm diameter) $\times$ (100 – clay fraction percentage), *a* is the percentage of organic matter, *b* is the soil structure code, and *c* is the soil profile, permeability class.

d. Slope Length (LS)

Length and slope (LS) is the ratio between the amount of erosion from a plot of land with a certain slope length and steepness, which can be calculated using the (Renard et al., 1991) Eq. 4

$$LS = \sqrt{x(0{,}0138 + 0{,}00965s + 0{,}00138s^2)} \tag{4}$$

where x is the length of the slope (m) and s is the steepness of the slope (%).

e. Ground Cover Vegetation and Plant Management and Land Management (CP)

Ground cover vegetation and plant management (C) is the ratio between the amounts of erosion from planted soils with certain management to the amount of soil erosion that is not planted and processed cleanly, whose value is obtained by observing visually at the field survey stage.

Land management (P) is the ratio of the amount of soil erosion with a certain conservation measure to the amount of erosion from soil that is processed according to the direction of the slope, whose value is also obtained by direct observation at the field survey stage. The C and P values are then calculated by multiplying the two factors to get the combined CP value.

f. Permissible Erosion (Edp)

Setting the maximum allowable erosion rate (Edp) is essential because it is impossible to reduce the erosion rate to zero on agricultural land, especially in sloping areas. The Edp value was obtained using the (Renard et al., 1991) Eq. 5

$$Edp (t\ ha^{-1}\ yr^{-1}) = Edp (mm\ yr^{-1}.) \times Bulk\ Density \times 10 (t\ ha^{-1}\ yr^{-1}.) \tag{5}$$

Where Edp is the allowable erosion rate (t ha^{-1} yr^{-1}.), the equivalent depth is the effective depth $\times$ value of the depth factor (mm), the useful life of the soil is 400 years, and Bulk Density is the weight of the soil volume.

4. RESULTS AND DISCUSSIONS

4.1. Parameters Causing Erosion

The rain erosivity values obtained each month are then added together to get the annual erosivity value. The results show that the annual rainfall erosivity at Pos Luwus is 2,414.18 t ha^{-1} cm^{-1}; the annual rainfall erosivity at Baturiti Post is 2,443.00 t ha^{-1} cm^{-1}, and rain erosivity at Candikuning Post has a value of 2,807.23 t ha^{-1} cm^{-1} (**Fig 2a**). Soil erodibility in the study area ranges from 0.08-0.51 (very low to high). Very low soil erodibility is found in Unit 2, with a value of 0.08. Low soil erodibility was found in Units 3, 7, 12, 13, 17 and 18. Medium soil erodibility was found in Units 4, 5, 8, 9, 10, and 11. Slightly high soil erodibility was found in Unit 14. High soil erodibility was found in Units 1, 6, 15, and 16 (**Fig 2b**). Calculating the length and slope of the slopes produce LS values between 0.39 to 6.06. The lowest LS value of 0.39 is found in Unit 5, with a length and slope of 50 m and 3%, respectively. While the highest LS value, which is 6.06, is found in Unit 2, with a slope length of 3.5 m and a slope of 45% (**Fig 2c**). The CP value obtained starts from the lowest, which is 0.001, and the highest is 0.28. The lowest CP values were found in Units 1, 2, 3, and 4 with forest land use and no land management or conservation action. In comparison, the CP with the highest value can be found in Units 15, 16, 17, and 18 with the use of dry land and land management actions in the form of traditional terraces (**Fig 2d**). Statistical data regarding the parameters that cause erosion are presented in **Table 1**.
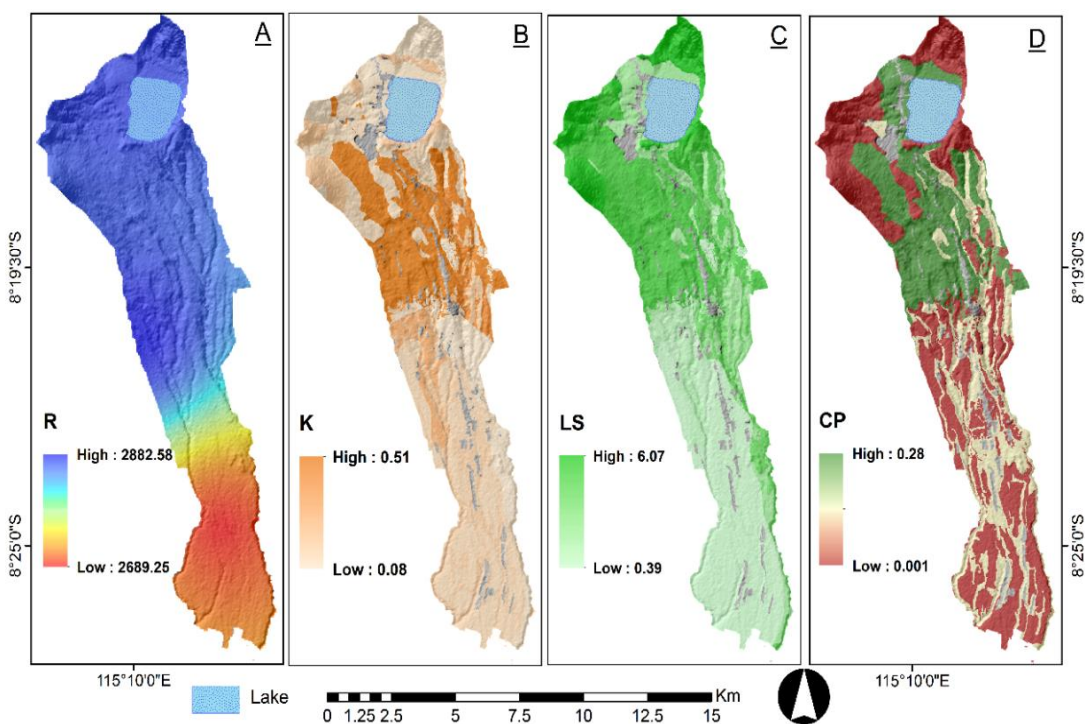
Fig 2. The thematic map that we use to calculate the amount of erosion i.e. Erosivity (R) is derived from data on rainfall (a), erodibility (K) (b), slope length (LS) (c), and land management (CP) (d).

4.2. Spatial Distribution of Soil Erosion and Conservation Effort

The amount of erosion (A) calculation shows that the study area experienced erosion from 0.32 to 1535.34 t ha^{-1} yr^{-1}. Based on the classification of erosion levels according to (Morgan et al., 1984), the amount of erosion in the study area is very light to very heavy. Very light erosion was

found in Units 1, 2, 3, 4, 10, 11, 12, 13, and 14 (4,848.31 ha). Mild erosion was found in Unit 5 (971.88 ha). Moderate erosion was found in Units 7, 8, 9, and 17 (989.61 ha). Heavy erosion occurred in Units 6 and 18 (599.70 ha). Very heavy erosion was found in Units 15 and 16 (2,516.77 ha) (**Fig 3a**).

Table 1.

Statistical data on the parameters that cause erosion calculation results and conservation efforts.

Land Unit	Landuse	Area (ha)	R	K	LS	CP	A (t ha^{-1} yr^{-1})	Categories	EDP	Conservation
1	Primary dry land forest	266.26	2443	0.47	5.76	0.001	6.68	Very Light	28.2	Maintenance effort
2	Primary dryland forest	722.13	2443	0.08	6.07	0.001	1.25	Very Light	24.03	Maintenance effort
3	Secondary dryland forest	897.32	2443	0.21	4.89	0.001	2.49	Very Light	33.3	Maintenance effort
4	Secondary dryland forest	661.25	2443	0.27	5.48	0.001	3.63	Very Light	35.1	Maintenance effort
5	Mixed Dryland Farming	971.88	2414.18	0.25	0.39	0.08	19.04	Light	27.79	Maintenance effort
6	Mixed Dryland Farming	243.31	2443	0.46	3.22	0.08	290.78	Heavy	24.75	Mixed dryland farming (high density) and bench terraces of good construction
7	Mixed Dryland Farming	358.58	2414.18	0.19	3.79	0.08	139.5	Moderate	25.65	Mixed dryland farming (high density) and bench terraces of good construction
8	Mixed Dryland Farming	280.58	2443	0.22	1.8	0.08	76.09	Moderate	19.6	Mixed dryland farming (medium-high density) and bench terraces of medium-fine construction
9	Mixed Dryland Farming	255.31	2443	0.27	1.4	0.08	74.85	Moderate	24.7	Mixed dryland farming (medium-high density) and bench terraces of medium-fine construction
10	Ricefield	1652.35	2414.18	0.22	0.4	0.0015	0.32	Very Light	17.1	Maintenance effort
11	Ricefield	169.81	2443	0.26	1.79	0.0015	1.71	Very Light	14.55	Maintenance effort
12	Ricefield	305.26	2414.18	0.19	2.01	0.0015	1.38	Very Light	19.95	Maintenance effort
13	Ricefield	103.12	2443	0.19	0.82	0.0015	0.58	Very Light	16.5	Maintenance effort
14	Ricefield	337.07	2807.23	0.34	1.01	0.0015	1.43	Very Light	18.53	Maintenance effort
15	Dryland Farming	2208.66	2443	0.49	4.61	0.28	1535.3	Very heavy	22.27	Peanuts with straw mulch 4 t h^{-1} and bench terraces of good construction
16	Dryland Farming	308.11	2443	0.51	4.32	0.28	1505.6	Very heavy	22.73	Peanuts with straw mulch 4 t h^{-1} and bench terraces of good construction
17	Dryland Farming	95.14	2443	0.14	1.54	0.28	147.36	Moderate	23.75	Dryland farming and terrace benches are of good construction
18	Dryland Farming	356.39	2443	0.2	2.1	0.28	289.56	Heavy	26.33	Mixed dryland farming (high density) and bench terraces of medium construction

Source: Researcher Analysis in 2022.

Very light erosion was spread in the villages of Angseri, Antapan, Apuan, Bangli, Batunya, Baturiti, Candikuning, Mekarsari, Perean, Perean Kangin, and Perean Tengah (Units 1, 2, 3, 4, 10, 11, 12, 13, and 14). The K factor that causes very mild to mild erosion in Units 2, 3, 5, 12, and 13 indicates that the soil in the area is not easily eroded. Meanwhile, the CP factor causes very mild erosion in Units 1, 2, 3, 4, 10, 11, 12, 13, and 14. Units 1, 2, 3, and 4 have land use in the form of forests with a lot of litter. Forests have a high density between plants to reduce the impact of rainwater because it is blocked by plant canopies that touch each other. In addition, a lot of litter also plays a role in increasing infiltration so that it can reduce the surface runoff velocity. Units 10, 11, 12, 13, and 14 have land use in the form of rice fields with bench terraces of medium construction.

Irrigated rice has a dense plant density to reduce runoff and block the soil surface from the impact of rainwater. In addition, the fibrous roots in rice can control erosion because the fine threads in the fibrous roots can bind the primary soil grains so that the stability of soil aggregates becomes stable (Trigunasih et al., 2018; Trigunasih et al., 2017). Using rice fields with terraces also contributes to reducing the CP value. Bench terraces reduce the length and slope of the slope to minimize the amount and speed of runoff (Gao et al., 2020; Lowe et al., 2021; Ma et al., 2019).

Light erosion was spread in the villages of Angseri, Antapan, Apuan, Bangli, Baturiti, Luwus, Mekarsari, Perean, Perean Kangin, and Perean Tengah (Unit 5). The area has the lowest LS value with a slope classified as sloping. The gentle slope can increase infiltration, reducing the amount and speed of runoff. Erosion is spreading in the villages of Angseri, Antapan, Apuan, Bangli, Batunya, Baturiti, Candikuning, Luwus, Mekarsari, and Perean Kangin (Units 7, 8, 9, and 17). The slope in this area is classified as slightly sloping to sloping, allowing surface runoff. Units 7, 8, 9 (**Fig 3a**) have medium density mixed dryland farming with traditional terraces.
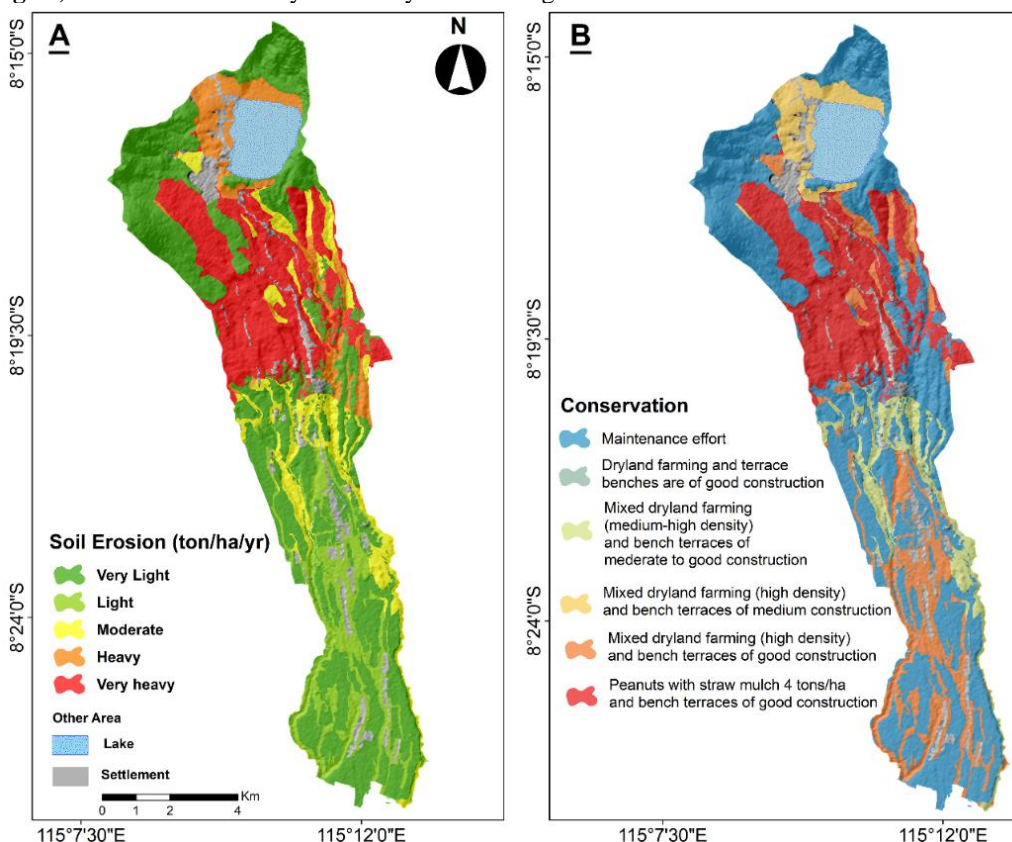
Fig. 3. Map of spatial distribution of erosion magnitude in agro-tourism areas (a), spatial distribution of environmental-based conservation action plans (b).

Raindrops will easily destroy soil particles because the plant density is less dense, so the plant canopy does not completely cover the soil. In addition, the ability of traditional terraces will not be as good as that of bench terraces of moderate to good quality in reducing the amount and speed of runoff (Chen et al., 2022; Deng et al., 2021; Khelifa et al., 2017).

Heavy erosion was spread in the villages of Antapan, Bangli, Batunya, Baturiti, and Candikuning (Units 6 and 18). The K factor is the cause of heavy erosion in Unit 6 with a high category, indicating that the soil in the area is easily eroded. The slope in this area is classified as slightly sloping to slightly steep so that it can cause surface runoff. Land Unit 6 has a land use of medium density mixed dryland farming of medium density and dry fields can increase the erosion value because it has a density between less dense plants. So that the soil is easily destroyed by the destructive power of rain and traditional terraces that do not have as good quality as bench terraces of medium and good construction.

Very heavy erosion was spread in the villages of Angseri, Antapan, Apuan, Bangli, Batunya, Baturiti, Candikuning, Luwus, and Mekarsari (Units 15 and 16). The high value of soil erodibility (K) indicates that the soil in the area is easily eroded. The LS value in the unit is high which can increase the amount and speed of runoff. CP with the use of dry land can increase the erosion value because it has a loose plant density, causing soil aggregates to be easily destroyed when it rains with high intensity and traditional terraces do not have the ability as well as bench terraces in reducing the amount and speed of runoff (Gao et al., 2020; Khelifa et al., 2017).

Recent research conducted by (Jemai et al., 2021), estimating soil erosion using GIS and remote sensing in South-Eastern Tunisia. The average soil loss rate is about 0.2 t ha^{-1} yr^{-1}. The maximum soil loss in the study area is 17 t ha^{-1} yr^{-1} in the Matmata mountain area. Further research by (Pham et al., 2018) in Sap Watershed, Central Vietnam. The results showed that forest land experienced the most soil erosion at around 19 t ha^{-1} year^{-1}, plantation forests at 7 t ha^{-1} yr^{-1}, and other agricultural lands at 3.70 and 1.45 t ha^{-1} yr^{-1} for annual crops and lowland rice.

Comparing the two previous researchers in different regions of the country shows that our findings regarding the erosion rate are drastically different, namely for the maximum erosion rate reaching 1,535.3 t ha^{-1} yr^{-1} on dry land farming land use. This condition is caused by the characteristics of the regions in the different research areas, causing different results for each parameter built using the USLE model. We assume that the findings of our study should be evaluated by conducting further research with a different approach.

This research is not based on raster/pixel but the based data from vector/polygon. The uncertainty and weakness are that the resulting erosion model is general. Other researchers analyzed the erosive factor with remote sensing data such as the Global Precipitation Measurement Mission (GPM) and Tropical Rainfall Measuring Mission (TRMM) (Chen et al., 2021; Das et al., 2022). Land use and vegetation density through remote sensing spectral index approach (e.g., Normalized Difference Vegetation Index) (Ayalew et al., 2020), and use of detailed scale soil type to analyze erodibility (Ayalew et al., 2020). The effects of land use change must also be considered in quantifying erosion. The study investigates the impact of land use change on the rate of soil erosion, and informs that the rate of soil erosion is increasing along with changes in forest land to crop with lowest vegetation density and bare land (Eskandari Damaneh et al., 2022; Mariye et al., 2022; Sourn et al., 2022).

The conservation actions are determined after comparing the amount of erosion that occurs (A) with the tolerable erosion (EDP). If the value of (A) is greater than (EDP) then conservation action is required. Meanwhile, if the value of A is less than Edp, then there is no need to take conservation measures and only implement maintenance efforts. Units with an A value lower than the Edp value are Units 1, 2, 3, and 4 with forest land use, Unit 5 with mixed dryland farming, and Units 10, 11, 12, 13, and 14 with paddy field use. The unit does not require soil and water conservation measures and requires only maintenance efforts. Meanwhile, Units 6, 7, 8, and 9 with mixed dryland farming and 15, 16, 17, and 18 with dry land use require soil and water conservation measures because these eight units have A values greater than EDP. Conservation data based on allowable erosion conditions are presented in **Table 1**, and their spatial distribution is shown in **Fig. 3b.**

5. CONCLUSIONS

The erosion rate in agro-tourism areas ranges from 0.32 t ha^{-1} yr^{-1} (very light) to 1,535.34 t ha^{-1} yr^{-1} (very heavy). The spatial distribution of erosion rates in the heavy to very heavy categories, respectively, is 599.70 and 2,516.77 ha. The agro-tourism areas affected by the erosion rate in this category are the villages of Antapan, Bangli, Apuan, Angseri, and Candikuning. Conservation actions can be taken, including improving plant management factors with dense vegetation, increasing land management actions by constructing bench terraces, and planting parallel to contour lines. It is recommended for further researchers to model erosion by integrating remote sensing data and geographic information systems in order to complement data that cannot be obtained in the field or the laboratory.

R E F E R E N C E S

Ayalew, D. A., Deumlich, D., Šarapatka, B., & Doktor, D. (2020). Quantifying the sensitivity of NDVI-based C Factor Estimation and Potential Soil Erosion Prediction using Spaceborne Earth Observation Data. Remote Sensing, 12(7). https://doi.org/10.3390/rs12071136

Chen, Y., Xu, M., Wang, Z., Gao, P., & Lai, C. (2021). Applicability of two satellite-based precipitation products for assessing rainfall erosivity in China. Science of the Total Environment, 757. https://doi.org/10.1016/j.scitotenv.2020.143975

Chen, Y., Jian, S., Wu, Q., Hu, C., & Soomro, S. e. hyde. (2022). Identify runoff generation patterns of check dams and terraces and the effects on runoff: a case study. Acta Geophysica, 70(2). https://doi.org/10.1007/s11600-022-00728-4

Das, S., Jain, M. K., & Gupta, V. (2022). A step towards mapping rainfall erosivity for India using high-resolution GPM satellite rainfall products. Catena, 212. https://doi.org/10.1016/j.catena.2022.106067

Deng, C., Zhang, G., Liu, Y., Nie, X., Li, Z., Liu, J., & Zhu, D. (2021). Advantages and disadvantages of terracing: A comprehensive review. In *International Soil and Water Conservation Research* (Vol. 9, Issue 3). https://doi.org/10.1016/j.iswcr.2021.03.002

Diara, I. W., Suyarto, R., & Saifulloh, M. (2022). Spatial Distribution Of Landslide Susceptibility In New Road Construction Mengwitani-Singaraja, Bali-Indonesia: Based On Geospatial Data. *GEOMATE Journal*, *23*(96), 95–103. https://doi.org/https://doi.org/10.21660/2022.96.3320

Eskandari Damaneh, H., Khosravi, H., Habashi, K., Eskandari Damaneh, H., & Tiefenbacher, J. P. (2022). The impact of land use and land cover changes on soil erosion in western Iran. Natural Hazards, 110(3). https://doi.org/10.1007/s11069-021-05032-w

Gao, J., Bai, Y., Cui, H., & Zhang, Y. (2020). The effect of different crops and slopes on runoff and soil erosion. *Water Practice and Technology*, *15*(3). https://doi.org/10.2166/wpt.2020.061

Jemai, S., Kallel, A., Agoubi, B., & Abida, H. (2021). Soil Erosion Estimation in Arid Area by USLE Model Applying GIS and RS: Case of Oued El Hamma Catchment, South-Eastern Tunisia. Journal of the Indian Society of Remote Sensing, 49(6). https://doi.org/10.1007/s12524-021-01320-x

Khelifa, W. Ben, Hermassi, T., Strohmeier, S., Zucca, C., Ziadat, F., Boufaroua, M., & Habaieb, H. (2017). Parameterization of the Effect of Bench Terraces on Runoff and Sediment Yield by Swat Modeling in a Small Semi-arid Watershed in Northern Tunisia. *Land Degradation and Development*, *28*(5). https://doi.org/10.1002/ldr.2685

Lowe, M. A., McGrath, G., & Leopold, M. (2021). The Impact of Soil Water Repellency and Slope upon Runoff and Erosion. *Soil and Tillage Research*, *205*. https://doi.org/10.1016/j.still.2020.104756

Ma, B., Liu, G., Ma, F., Li, Z., & Wu, F. (2019). Effects of crop-slope interaction on slope runoff and erosion in the Loess Plateau. *Acta Agriculturae Scandinavica Section B: Soil and Plant Science*, *69*(1). https://doi.org/10.1080/09064710.2018.1488988

Mahanani, R. S., Hidayat, T., Wardatic, I., Galushasti, A., & Wiyono, L. C. (2021). Local economic development strategies to increase economic growth in agrotourism areas. *Tourism(Poland)*, *31*(2). https://doi.org/10.18778/0867-5856.31.2.07

Marin, D. (2015). Study on the Economic Impact of Tourism and of Agrotourism on Local Communities. *Research Journal of Agricultural Science*, *47*(4).

Mariye, M., Mariyo, M., Changming, Y., Teffera, Z. L., & Weldegebrial, B. (2022). Effects of land use and land cover change on soil erosion potential in Berhe district: a case study of Legedadi watershed, Ethiopia. International Journal of River Basin Management, 20(1). https://doi.org/10.1080/15715124.2020.1767636

Morgan, R. P. C., Morgan, D. D. V., & Finney, H. J. (1984). A predictive model for the assessment of soil erosion risk. *Journal of Agricultural Engineering Research*, *30*(C). https://doi.org/10.1016/S0021-8634(84)80025-6

Narendra, B. H., Siregar, C. A., Dharmawan, I. W. S., Sukmana, A., Pratiwi, Pramono, I. B., Basuki, T. M., Nugroho, H. Y. S. H., Supangat, A. B., Purwanto, Setiawan, O., Nandini, R., Ulya, N. A., Arifanti, V. B., & Yuwati, T. W. (2021). A review on sustainability of watershed management in Indonesia. In *Sustainability (Switzerland)* (Vol. 13, Issue 19). https://doi.org/10.3390/su131911125

Nugroho, H. Y. S. H., Basuki, T. M., Pramono, I. B., Savitri, E., Purwanto, Indrawati, D. R., Wahyuningrum, N., Adi, R. N., Indrajaya, Y., Supangat, A. B., Putra, P. B., Auliyani, D., Priyanto, E., Yuwati, T. W., Pratiwi, Narendra, B. H., Sukmana, A., Handayani, W., Setiawan, O., & Nandini, R. (2022). Forty Years of Soil and Water Conservation Policy, Implementation, Research and Development in Indonesia: A Review. In *Sustainability (Switzerland)* (Vol. 14, Issue 5). https://doi.org/10.3390/su14052972

Pham, T. G., Degener, J., & Kappas, M. (2018). Integrated universal soil loss equation (USLE) and Geographical Information System (GIS) for soil erosion estimation in A Sap basin: Central Vietnam. International Soil and Water Conservation Research, 6(2). https://doi.org/10.1016/j.iswcr.2018.01.001

Renard, K. G., Foster, G. R., Weesies, G. A., & Porter, J. P. (1991). RUSLE: revised universal soil loss equation. *Journal of Soil & Water Conservation*, *46*(1).

Rideng, I. W., Budiartha, I. N. P., & Sukandia, I. N. (2020). The development of bali tourism through cultural and local wisdom of customary village. *International Journal of Entrepreneurship*, *24*(5).

Rosardi, R. G., Prajanti, S. D. W., Atmaja, H. T., & Juhadi. (2021). Sustainable tourism model in pagilaran tea plantation agrotourism, in Indonesia. *International Journal of Sustainable Development and Planning*, *16*(5). https://doi.org/10.18280/ijsdp.160519

Satriawan, I. K., Pujaastawa, I. B. G., & Sarjana, I. M. (2015). Development of small-scale agro-tourism in the Province of Bali, Indonesia. *Advances in Environmental Biology*, *9*(21).

Sourn, T., Pok, S., Chou, P., Nut, N., Theng, D., & Vara Prasad, P. V. (2022). Assessment of Land Use and Land Cover Changes on Soil Erosion Using Remote Sensing, GIS and RUSLE Model: A Case Study of Battambang Province, Cambodia. Sustainability (Switzerland), 14(7). https://doi.org/10.3390/su14074066

Trigunasih & Saifulloh. (2022). Spatial Distribution of Landslide Potential and Soil Fertility: A Case Study in Baturiti District, Tabanan, Bali, Indonesia. *Journal of Hunan University Natural Sciences*, *49*(2). https://doi.org/https://doi.org/10.55463/issn.1674-2974.49.2.23

Trigunasih, N. M., Kusmawati, T., & Yuli Lestari, N. W. (2018). Erosion Prediction Analysis and Landuse Planning in Gunggung Watershed, Bali, Indonesia. *IOP Conference Series: Earth and Environmental Science*, *123*(1). https://doi.org/10.1088/1755-1315/123/1/012025

Trigunasih, Ni Made, Merit, I. N., Wiyanti, W., Narka, I. W., & Dibia, I. N. (2017). Evaluation Of Land Suitability For Increasing Productivity In Degraded Unda Watershed, District Of Karangasem, Bali. *International Journal of Biosciences and Biotechnology*, *5*(1). https://doi.org/10.24843/ijbb.2017.v05.i01.p03

Utama, I. G. B. R. (2020). Tourism Visitor Center Flowchart As Recommendation for Bali Tourism Destination. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3669739

Wischmeier, W., & Smith, D. (1978). Predicting rainfall erosion losses: a guide to conservation planning. In *U.S. Department of Agriculture Handbook No. 537*. https://doi.org/10.1029/TR039i002p00285

THERMAL REGIME OF THE NORTHWESTERN PART OF THE BLACK SEA

Viktor VYSHNEVSKYI [1] ID, Alexander MATYGIN [2] ID and Viktor KOMORIN [3*] ID*

ABSTRACT:
The spatio-temporal features of the surface water thermal regime of the northwestern part of the Black Sea, its bays and estuaries have been studied. Based on the long-term observations, the increase in the air and water temperatures was identified. Since 1894 the mean annual air temperature in Odesa city has increased by at least 2.5 °C. Also significant, but somewhat smaller (about 2.0 °C), is the increase in the water temperature. The July and August water temperature has the biggest increase. Water temperatures generally are higher than the air temperature, but this ratio is varying in value and sign for different seasons. In January–February, when it is the coldest, the water temperature is higher than the air temperature. In spring and summer, the air heats up faster than the water and it is warmer than water. In autumn and December, the sea water temperature usually is significantly higher than the air temperature. In spring and summer, the water temperatures in the bays and estuaries are higher than the water temperatures in the sea. In autumn and winter, it is the other way around. The dominance of the northwest and the north winds in the studied region is often induces the noticeable decrease in the water temperature near the coast. It was shown that even light winds have the impact on the spatial distribution of sea water temperature.

Key-words: *Air temperature, Water temperature, Black Sea, Wind; Remote Sensing.*

1. INTRODUCTION

The Black Sea is a unique sea, eighty-seven percent of which is fulfilled the hydrogen sulfide. The most environmentally sensitive area of the Black Sea for climate change and anthropogenic influence is northwestern part of the Black Sea shelf. This was indicated with assessment data of the ecosystem of the Black Sea to MSFD descriptors (Directive 2008/56 / EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive) presented in (Slobodnik, et al. 2020a, 2020b, 2020c). As it was shown in study (Komorin, 2021) the water temperature is the one of the main indicators of marine water state that influence to the sustainability of the marine shelf ecosystem. There are quite a few papers devoted to the thermal regime of the Black Sea. In particular, the study (Ilyin, et al., 2012) shows that within the coastal zone the mean annual water temperature is significantly higher than the air temperature. This excess in the northwestern part of the Black Sea is about 1 °C, which is less than near the coast of the Caucasus, where this excess is about 2 °C. This is due to the Main Black Sea Current, in which water moves counter-clockwise. Therefore, the relatively low water temperature near the northwestern shores is due to the movement of the sea water from the north to the south.

The study (Ilyin, et al., 2012) contains information on the mean long-term water temperature at several observation points located on the sea shore. Thus, the mean annual water temperatures for

---

[1] *National Aviation University, Liubomyra Huzara Ave., 1, Kyiv, 03058, Ukraine, vishnev.v@gmail.com*

[2] *Hydrometeorological Center of the Black and Azov Seas, Franch Ave., 89, Odesa, 65000, Ukraine, acm32alex@gmail.com*

[3] *Research Institution "Ukrainian Scientific Centre of Ecology of the Sea", Franch Ave., 89, Odesa, 65000, Ukraine, vkomorin@gmail.com*

*Corresponding authors

the period 1915–2011 (Odesa Port) were 11.1 °C, and 12.1 °C (period of 1951–2011, in the village Prymorske). The long-term mean water temperatures for August (the warmest month for both points) are 20.7 °C and 22.5 °C accordingly. The above-mentioned study showed the existence of noticeable increase of water temperature in the Black Sea, with the average rate of 0.08 °C per decade during period 1923–2011. Considerable research of the water temperature in the northwestern part of the Black Sea was carried out in the study (Popov, et al., 2016), where images and maps of the characteristic temperatures of the sea water surface layer for each month of year were presented. According to those data, the lowest water temperature usually observed in February and the highest – in August. The studies (Ilyin, et al., 2012; Bolshakov, Matygin, 2017) highlighted unusually high and low water temperatures in Odesa. Thus, on July 17, 1976, the water temperature there was only 7.2 °C, and on August 13, 1983 it was only 8.0 °C. The very high water temperature (31.2 °C) was recorded on August 11, 2010.

There are many publications (Adrian, et al., 2009; Czernecki, Ptak, 2018; Lieberherr, Wunderle, 2018; Ptak, et al., 2020; Shaltout, Omstedt, 2014; Vyshnevskyi, Shevchuk, 2021; Woolway, et al., 2017) where the facts of the increase in the water temperature had been detected.

It should be noted, that the observation at coastal zone does not represent the spatio-temporal patterns for the whole region of the open sea water temperature. This is difficult to carry out even with the use of the research fleet data. The issue could be resolved with the using of remote sensing, in particular the using of data from the thermal bands of Landsat 8 and Landsat 9 satellites. However, until recently, such data for the studied region of the Black Sea were very limited, mainly due to the small number of the good quality (without clouds) satellite images. According to the authors, the appearance of a sufficiently large number of them makes it possible to study not only seasonal features of surface water temperature, but also features caused by even a light wind.

Therefore, the main purpose of this research is the identification of the spatio-temporal patterns in the recent distribution of the north-western region of the Black Sea water temperature and its long-term trends.

2. STUDY AREA

The studied area of the Black Sea, as the title of the article implies, is located in its northwestern part (**Fig. 1**). Observations of the climate and water temperature are carried out at the number of meteorological and hydrological stations. The majority of the analyzed information are based on the observations at Odesa and Vylkove meteorological stations, which are located, respectively, in the northern and southern parts of the studied region.
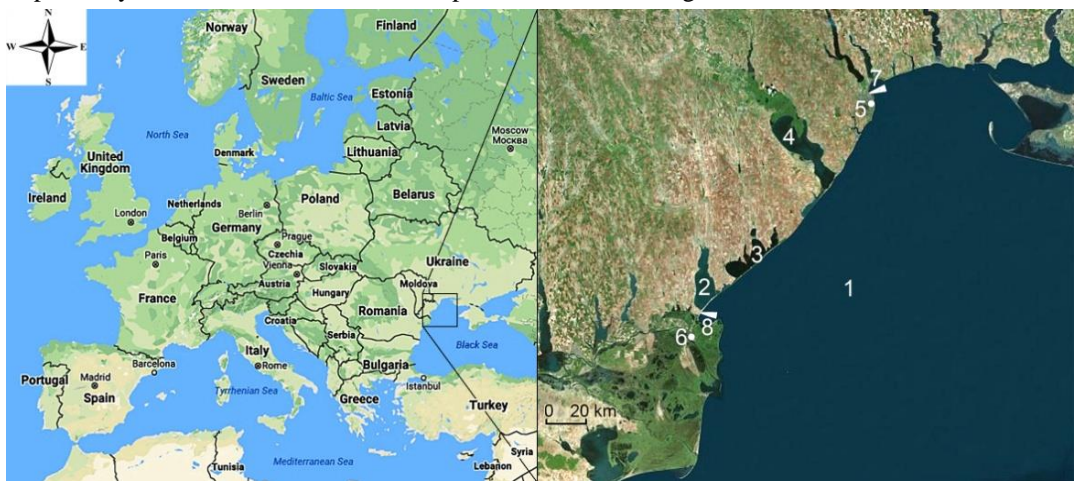
Fig. 1. Location of the studied part of the Black Sea and hydrometeorological observation points:
1 – Black Sea, 2 – lake Sasyk, 3 – Tuzliv estuaries, 4 – Dniester estuary, 5–6 – Odesa and Vylkove meteorological stations, 7–8 – Odesa and Prymorske hydrological stations.

Odesa meteorological station is located near the center of the city only 150 m from the sea. The observations started there in 1894, and since then the meteorological station was never moved. Vylkove meteorological station is located on the bank of the Danube River, 13 km from the sea (**Fig. 1**).

3. DATA AND METHODS

The study of air temperature and its long-term changes for Odesa meteorological station was carried out for the period 1894–2021, and for Vylkove meteorological station for the period of 1961–2021. The water temperature in Odesa is measured near the city centre in the port, where the depth is quite significant. At the same time, at Prymorske station, the water temperature is measured on the sea beach, where the depth is small. As will be shown below, this has some effect on the results, primarily on the intra-annual temperature distribution.

Changes in the mean annual water temperature at Odesa meteorological station were identified for the period of 1915–2021, and the mean monthly water temperature for the period of 1961–2021. Data on water temperature at Prymorske station were analysed for the period of 2016–2021.

In addition to the data from the regular monitoring, in our research were used the remote sensing data. The main attention was paid to the images from Landsat 8 and Landsat 9 satellites, which are available on the website https://earthexplorer.usgs.gov. The images of series LC08_L1TP_180028 and LC09_L1TP_180028 cover most of the studied region. The time of the earth surveying of the studied area is 08:44 GMT, which is corresponds to about the noon of the local time. The spatial resolution of B10 thermal band of these satellite images is 100 m, the revisit time is 16 days.

Although Landsat 8 satellite has been in the orbit since 2013, there are not many high-quality images, especially in the cold season. The largest part of images is completely or partially overcast. There are even fewer high-quality images from Landsat 9 satellite, which was launched in September 2021. In research we had used images that are not overcast and those are representative for the different seasons. Satellite images were processed using ArcMap 10 program.

Water surface temperature was calculated applying formulas recommended by NASA:

for Landsat 8:

$t = (1321.08/(Ln((774.89/(("LC8\_B10.TIF" * 0.0003342)+0.1))+1)))-273.15,$

for Landsat 9 satellite:

$t = (1329.2405/(Ln((799.0284/(("LC9\_L1TP\_B10.TIF" * 0.00038)+0.1))+1)))-273.15.$

The reliability of these regressions was confirmed by many studies (Schaeffer et al, 2018; Vyshnevskyi, Shevchuk, 2018).

The territory, which does not belong to water area, was identified using the calculation of Normalized Difference Pond Index (NDPI). This index is calculated by the equation

$NDPI = (B6 - B3) / (B6 + B3),$

in which B3 and B6 are the meanings of corresponded values of the satellites Landsat bands. The territory, which is not the water area, was presented in grey color for better visualization. The territory that did not get into the pictures is shown in white.

4. RESULTS AND DISCUSSIONS

4.1. Air temperature and its long-term changes

The considerable duration of observations at Odesa and Vylkove meteorological stations makes it possible to characterize the air temperature in the studied region and its long-term changes. As in many regions of the world, the air temperature here has risen significantly. During the period 1894–2021, the air temperature in Odesa increased by more than 2.5 °C, the most in the last 30 years.

At the same time, in case of approximation by a linear trend, this increase would be noticeably less – about 1.8 °C. The changes of air temperature during 1894–2021 according to the linear trend are 0.15 °C per decade (**Fig. 2**).
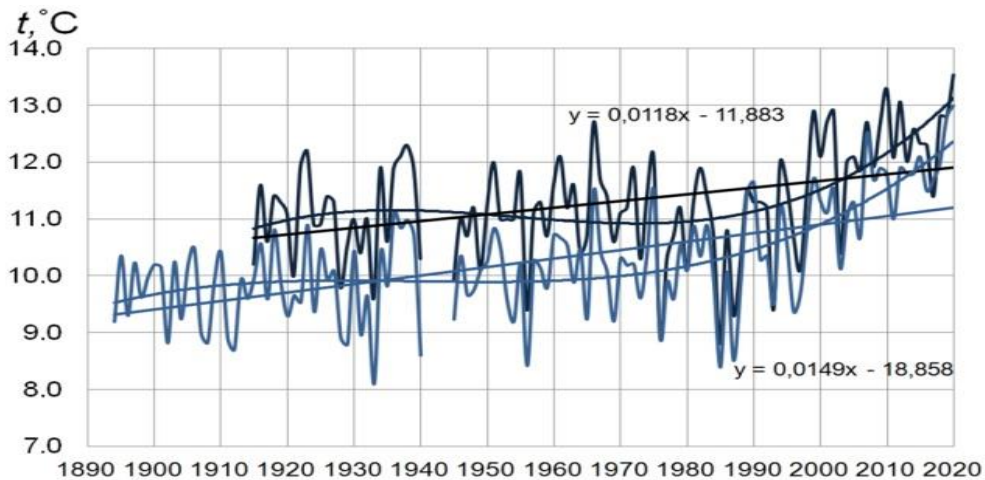
Fig. 2. Long-term changes in the average annual of air (light blue) and water temperature (deep blue) in Odesa.

Data processing for the 30-years period 1961–1990 shows the mean annual air temperature in Odesa is 10.2 °C and the corresponding value for the period 1991–2020 is 11.3 °C. The air temperature in Vylkovo during these periods is 11.1 °C and 12.2 °C respectively. As it can be seen in both cases, the difference in air temperature is 1.1 °C. The year 2020 turned out to be the warmest in the entire history of the observations, when the mean annual air temperature in Odesa reached 13.0 °C, and in Vylkove – 13.6 °C.

During the last 30 years, the highest mean air temperature in Odesa (23.4 °C) is observed in July, and the lowest (minus 0.4 °C) in February. The highest air temperature in Vylkovo is also observed in July (24.1 °C), the lowest one – in January (plus 0.4 °C). Compared to the previous 30-year period (1961–1990), the air temperature increased in all months of the year, but the most in July–August.

4.2. Wind regime

To some extent, the water temperature in the northwestern part of the Black Sea depends on the speed and direction of the wind. It is well known that wind can cause the changes in water level, which are accompanied by significant changes in temperatures. The cases of the low water temperatures in Odesa which were mentioned in the Introduction were caused by the northwest wind. In general, the northwest and the north winds dominate in the studied part of the Black Sea. The predominant wind directions in Odesa are from the northwest, which occurs in 17.5 % of the observations. Such recurrence for the July–August period increases to 22–23 %. The prevailing wind in Vylkovo is from the north. The south wind at this station has somewhat lower recurrence (**Fig. 3**).

According to the available data the wind speed in the studied region is low. Its mean long-term value at Odesa station is 2.9 m/sec. The highest wind speed is observed in November–December (3.4–3.5 m/sec), the lowest one – in July–August (2.3 m/sec). According to the chronological observations, there is the decreasing trend of wind speed.
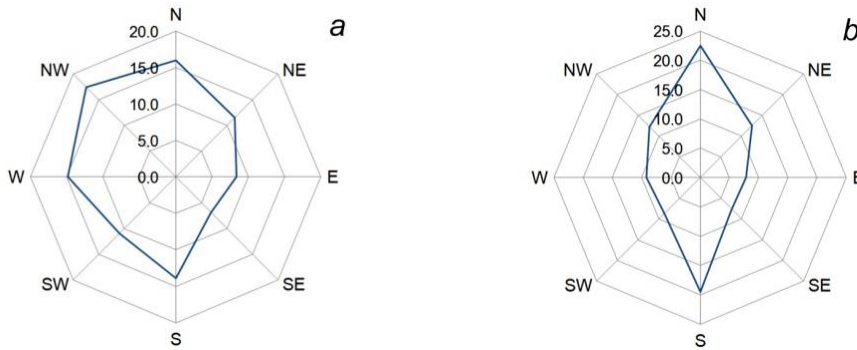
Fig. 3. Mean annual recurrence of wind at Odesa (*a*)
and Vylkove (*b*) meteorological stations during 1991–2020.

4.3. Water temperature and its long-term changes

A comparison of the air and water temperature data (see **Fig. 2**) shows that the water temperature is generally higher than the air temperature. This excess is explained by the fact that the main absorption of the Sun's energy occurs in the upper layer of water, where, in fact, the temperature is measured. In addition, the fact that seawater has a certain minimum temperature, while the air temperature does not have such limit, must be considered. For the northwestern part of the Black Sea, which has a salinity approximately half that of the World Ocean, this minimum is about minus 1 °C.

During the observation period 1915–2021, the mean annual water temperature in the studied part of the Black Sea increased significantly – by approximately 2.0 °C. According to the linear trend, the increase is less – about 1.1 °C (see **Fig. 2**). However, this increase is still greater than that obtained in the study (Ilyin, et al, 2012) for the period 1923–2011.

In 1991–2021, the mean annual water temperature in Odesa was 12.0 °C, in 2001–2021 – 12.4 °C. In 2020, which turned out to be the warmest, the average annual water temperature in Odesa reached 13.4 °C.

Analysis of **Fig. 2** shows that mean annual water temperature generally increases more slowly than the air temperature. This can be explained by the large volume of water in the Black Sea and the inertia of its heating. The difference between the mean annual air and water temperature in the middle of the 20th century was about 1.1 °C, nowadays it is 0.7–0.8 °C. A somewhat greater increase in air temperature compared to water temperature was recorded for the Kakhovske reservoir located relatively not far from the Black Sea (Vyshnevskyi, Shevchuk, 2021). However, when the volume of water is small, the increase in air and water temperature is practically similar.

During the year, the ratio between air and water temperature does not remain constant. In January–February, when it is the coldest, the water temperature is higher than the air temperature. In spring and summer, the air temperature is higher than the water temperature. In autumn and December, the water temperature becomes significantly higher than the air temperature (**Fig. 4a**).

Different rates (and signs) of the differences in the values of monthly averaged air and water temperatures causes, that the correlation between these parameters varies throughout the year. However, the correlation between the air and water temperatures is generally strong. For the mean annual values in Odesa and period 1961–2020, the correlation coefficient is $r = 0.893$, for the mean values of the period of May to October relevant value of $r = 0.864$, for the mean values of summer period – $r = 0.856$.

Comparison of the monthly means of water temperatures averaged for the two 30 years periods (1961–1990) and (1991–2020) clearly displays the increase in water temperature, throughout a year, with little bigger increases in July and August (**Fig. 4b**).
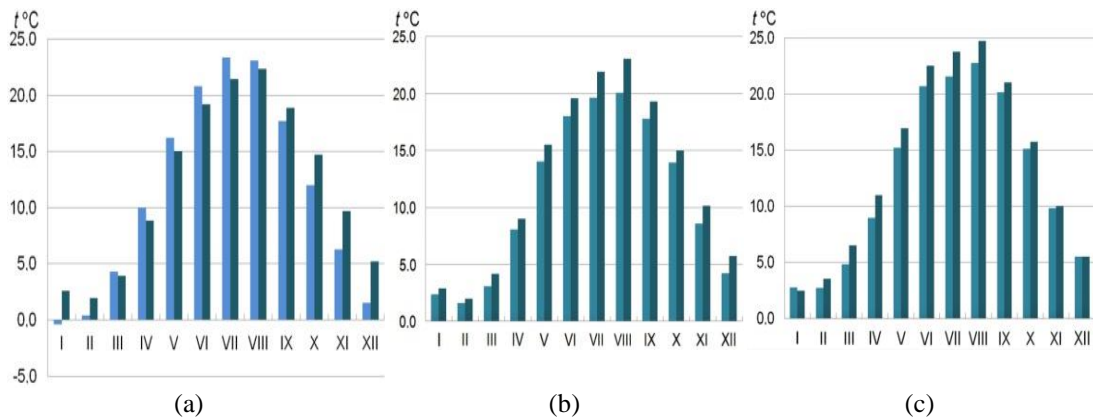
Fig. 4. Year course of temperature: a – monthly means of temperature air (left columns) and water (right columns) in Odesa averaged for the period of 1991–2020; b – water temperature by months in Odesa: left columns – 1961–1990, right columns – 1991–2020; c – water temperature by months in 2016–2021: left columns – in Odesa, right columns – in Prymorske.

Comparison of the averaged values of water temperatures monthly means in Prymorske and in Odesa detected the differences in these values. In spring and summer, the shallow Zhebryanivska Bay, in which Prymorske station is located, warms up faster and as result the water at this station is higher than in Odesa port. In January, the water temperature in this bay is lower than in Odesa (**Fig. 4c**). As can be seen in **Fig. 4c**, in recent years the water temperature in August at Prymorske station approached the mark of 25 °C (the values typical for resorts in Bulgaria).

There is a close correlation ($r = 0.989$) between the data on mean annual water temperature in Odesa and Prymorske, which has the form

$t_{Primor} = 1.0598 * t_{Odesa} + 0.396$ (°C).

According to this regression, the mean water temperature at Prymorske station in 1991–2021 is 13.1 °C, which is higher than in Odesa by 1.1 °C. This is close to the difference in air temperature.

4.4. Spatial features of the water temperature

The given data sufficiently fully represents the temporal patterns of water temperature at Odesa and Prymorske stations. However, the spatial patterns of the water temperature remain unclear, primarily in the open sea and estuaries. These gaps can be filled with the use of remote sensing, namely with data from thermal bands of Landsat 8 and Landsat 9 satellites.

Corresponding data shows that at the beginning of the year the water temperature in the Black Sea is low, especially in the bays and estuaries. It is the highest in the open part of the sea, where there is greater depth, heat reserves and better mixing. At this time shallow bays of the sea can be covered with ice (**Fig. 5a**). Similar conditions are observed in February and early spring, with the difference that in March the shallow bays of the sea already begin to warm up (**Fig. 5b**).

A comparison of images obtained in January and March shows that the sea water temperature may be even lower in the latter case. First of all, it depends on the air temperature observed in a certain month or season. Thus, the weather in January and February 2014 in the studied region was cool. On the other hand, the weather in December 2021 was warmer than normal and this affected the water temperature not only this but also the next month. In general, according to the long-term observations, the water temperature in March is a little bit higher than in January (see **Fig. 4b**).

Meanwhile, the water temperature is affected by specific weather conditions. This, in particular, applies to the conditions of 11.03.2014. Since the night before a rather strong north wind was observed in the studied region. At Odesa and Vylkove meteorological stations, its speed was 6–8 m/sec. This wind caused the decrease of water temperature near the shore.
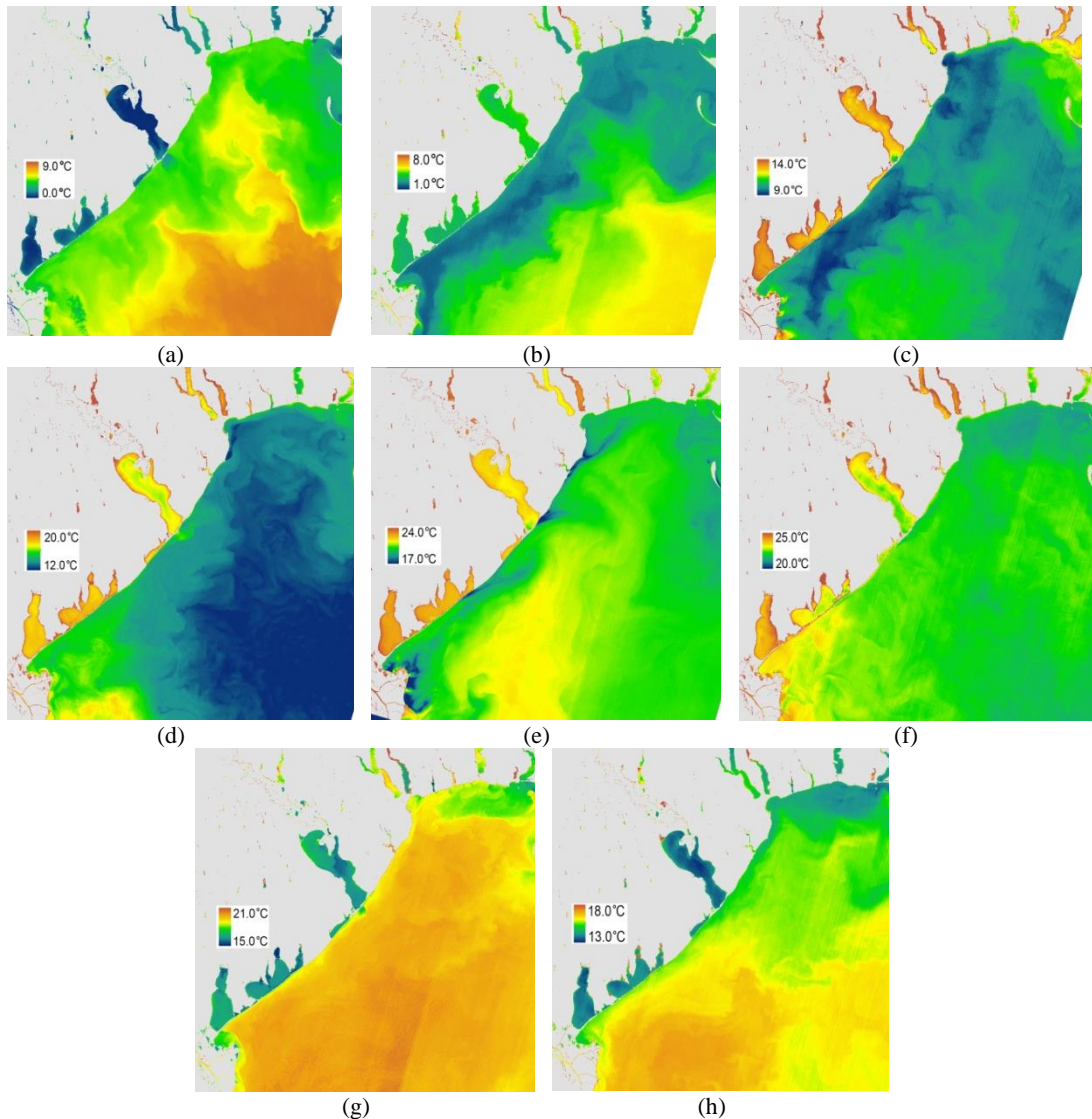
Fig. 5. Spatial features of water temperature in the northwestern part of the Black Sea according to data from Landsat 8 and Landsat 9 satellites: a) on January 4, 2022; b) on March 11, 2014; c) on April 28, 2020; d) on May 12, 2022; e) on July 20, 2015; f) on August 18, 2020; g) on September 19, 2014; h) on October 16, 2018.

It can be added that on March 11, 2014, the mean daily air temperature in Odesa was 7.9 °C, in Vylkove – 6.4 °C. In turn, the mean daily water temperature that day in Odesa was 4.0 °C, in Prymorske – 3.2 °C. As can be seen from these data and those ones which are shown on **Fig. 4b**, there is a close relationship between the temperature measured and identified with use of the satellite image.

In April, when it warms up noticeably, the highest temperature is observed in shallow bays and water bodies isolated from the sea. This, in particular, is typical for the lake Sasyk, Dniester estuary and other water bodies (**Fig. 5c**).

Analysis of **Fig. 5c** allows us to assume that on April 28, 2020, a wind from the south or southwest was probably observed. This, in particular, is evidenced by the appearance of cool water in the southern part of the Dniester estuary, as well as the relatively high temperature in its northern and northeastern parts. Indeed, in the first half of this day, up to the time of the satellite survey, a

south wind with a speed of 3–4 m/sec was observed at the Odesa meteorological station. At Vylkove station, the wind was of approximately the same strength from the south and southwest. On April 28, 2020, the mean daily air temperature in Odesa was 11.7 °C, in Vylkove it was 13.4 °C. The water temperature in Odesa was 9.4 °C, in Prymorske – 13.2 °C. The latest data show that water temperature at Prymorske station in spring can be significantly higher than the water temperature at Odesa station, which is located further north and has a greater depth. We should add that at the end of April there is usually a peak of spring flood of the Danube River. The consequence of this is a slight increase in water temperature near its mouth, because the water masses of the Danube River warm up faster than that of the Black Sea.

In comparison to the April values, the May air and water temperatures rise significantly. First of all, this applies to shallow water bodies, which warm up faster than the Black Sea. At this time, the water temperature in shallow water bodies can rise to 17–18 °C. Simultaneously the sea remains significantly colder. The characteristic conditions, which confirms these features, were observed on May 12, 2022 (**Fig. 5d**). In the first half of this day a south wind with a speed of 3–4 m/sec was recorded in the studied region. The mean daily air temperature at Odesa station was 16.8 °C.

Analysis of **Fig. 5d** shows that in May (this also applies to June and July) the spatial difference in water temperature can be very significant. This can be explained by the slow heating of the Black Sea water mass. In case of strong wind, the water temperature near the shore can sharply drop or rise by several degrees.

The condition recorded on the satellite image from July 20, 2015 seems interesting. This image shows the existence of several zones with a low water temperature near the shore (**Fig. 5e**). At the beginning of this day a south wind with a speed of 2–3 m/sec was observed in the studied region. At noon, the wind in Odesa was from the south-southeast, its speed increased to 5 m/sec. A similar wind was observed in Vylkovo, but in the afternoon its direction was from the south. The mean daily air temperature in Odesa on July 20, 2015 was 23.0 °C, in Vylkove – 23.6 °C. The water temperature in Odesa was 19.0 °C. As it could be seen on the image **Fig. 5e**, the even gentle south wind can cause the appearance of the relatively small upwelling zones at some parts of the coast, where the temperature is lower than in the surrounding area.

In August, when the water temperature in the sea is usually the highest, the difference between the water temperatures in the open sea and the bay's water temperatures somewhat decreases. This feature confirms the conditions observed on August 18, 2020 (**Fig. 5f**). At the beginning of this day a light wind of a variable direction was observed both at the Odesa and Vylkove meteorological stations. Later on in the day, from 6:00 a.m. to 12:00 p.m., the wind prevailing direction became the south-southeast, with a speed of 3–4 m/sec. The mean daily air temperature at Odesa meteorological station was 23.1 °C, at Vylkove meteorological station – 22.7 °C. The mean daily water temperature was as follows: Odesa – 21.9 °C, Prymorske – 25.1 °C. Similarly, to other cases, the comparison of measured water temperature and values, determined from the satellite image, shows a good correlation.

The decrease in the air temperature, which is usually observed in the second half of September, causes the fact that the water temperature in the sea becomes warmer than near the shore – primarily in its open part, which is deeper and located further south. This can be seen on the example observed on September 19, 2014 (**Fig. 5g**). In the first half of this day the light wind with a speed of 2–3 m/sec from the north-northeast was observed in the studied region. The mean daily air temperature in Odesa was 17.7 °C, in Vylkove – 16.1 °C. The water temperature was as follows: Odesa – 20.3 °C, Prymorske – 18.8 °C.

As the weather gets colder (October), the open deep waters of the studied region of the Black Sea cools down slower than its shallow waters. These conditions can be seen on the example conditions, observed on October 16, 2018 (**Fig. 5h**).

In the first half of the day on October 16, 2018, a light northeast wind was observed in the studied region. Its speed at Odesa and Vylkove meteorological stations was 2–3 m/sec. The mean daily air temperature in Odesa was 16.0 °C, in Vylkove – 14.7 °C. The water temperature was as follows: Odesa – 15.6 °C, Prymorske – 15.5 °C.

The given data make it possible to characterize the main spatio-temporal features of the water temperature in the northwestern part of the Black Sea. Observation data, as well as data from remote sensing, show that the lowest water temperature in the open part of the sea is observed in the second half of winter. With the onset of spring, the shallow bays of the sea begin to warm up, and therefore the water temperature here becomes higher than in open sea. It is important that spatial differences in water temperature in the first half of the year are very significant – up to 8–10 °C. Relatively higher water temperature in the bays of the sea is observed until the end of summer. In September, when the air temperature drops, water temperature there becomes lower than in the open sea. The subsequent decrease in air temperature is usually accompanied by a drop in water temperature in shallow bays and some slower cooling in the deep open part of the sea.

It could be added that water has a great capacity to hold energy. It takes a lot of energy to raise the temperature, but once the water temperature is raised, the heat energy is dissipating very slowly.

Even the light wind has a significant influence on the spatial sea water temperature distribution. The dominance of the northwest and the north winds in the studied part of the Black Sea often leads to a noticeable decrease in water temperature near the shore.

5. CONCLUSIONS

During the long-term observation period, the air and water temperature in the northwestern part of the Black Sea has increased significantly. Since 1894, the mean annual air temperature in Odesa has risen by at least 2.5 °C. Significant, but somewhat smaller (about 2 °C), is the increase in water temperature. During 1991–2021, the mean annual water temperature in Odesa was 12.0 °C, in Prymorske, which lies to the south near the mouth of the Danube River, is 13.1 °C. The highest annual water temperature during entire observation period was observed in 2020: in Odesa it was 13.5 °C, in Prymorske – 13.4 °C.

The water temperature rose the most in July and August. During 1991–2021, the water temperature in Odesa in these months averaged 21.5 °C and 22.4 °C, and in recent years (2016–2021) – 21.8 °C and 22.7 °C, respectively. The water temperature in Prymorske (2016–2021) in these months is even higher – 23.9 °C and 24.7 °C.

The mean annual water temperature is generally higher than the air temperature, but in individual seasons this ratio is different. In January–February, when it is the coldest, the water temperature is higher than the air temperature. In spring and summer, the air temperature is higher than the water temperature. In autumn and in December, the water temperature is significantly higher than the air temperature.

In spring and summer, the water temperature in the bays is higher than the sea temperature, in autumn and winter, the opposite is true.

Quite often, the water temperature is affected by the wind. The predominance of the northwest and the north winds in the studied region is often accompanied by a noticeable decrease in water temperature near the coast. The reliable satellite images show that even the light wind has the impact on the spatial distribution of sea water temperature.

REFERENCES

Adrian, R. et al. (2009). Lakes as sentinels of climate change. Limnol. Oceanogr. Nov; 54 (6): 2283–2297 https://aslopubs.onlinelibrary.wiley.com/doi/abs/10.4319/lo.2009.54.6_part_2.2283

Bolshakov, V.N., Matygin, A.S. (2017). Climate of Odesa (reference book for the curious). Odesa, 29 p. (in Russian).

Czernecki, B., Ptak, M. (2018). The impact of global warming on lake surface water temperature in Poland – the application of empirical-statistical downscaling, 1971–2100. Journal of Limnology. 77 (2). 340–348

Ilyin, Y.P., Repetin, L.N., Belokopytov, V.N. et al. (2012). Hydrometeorological conditions of the seas of Ukraine. T. 2. The Black sea. Sevastopol. 421 p. (in Russian).

Komorin, V. (2021). Assessment of the Black sea shelf ecosystem sustainability with mathematical simulation method. Geographia Technica, Vol. 16, Issue 2, 2021, pp. 19–28 DOI: 10.21163/GT_2021.162.02. http://technicalgeography.org/index.php/on-line-first/378-02_komorin

Lieberherr, G., Wunderle, S. (2018). Lake Surface Water Temperature Derived from 35 Years of AVHRR Sensor Data for European Lakes. Remote Sens. 10 (7), 990. 1–25. https://www.mdpi.com/2072-4292/10/7/990

Marszelewski, W., Pius, B. (2016). Long-term changes in temperature of river waters in the transitional zone of the temperate climate: A case study of Polish rivers. Hydrol. Sci. J. 61, 1430–1442. https://www.tandfonline.com/doi/full/10.1080/02626667.2015.1040800

Popov, Yu., Matygin, A.S., Matveev, A.V. et al. (2016). North-western part of the Black Sea: structure and climatic changes of oceanological fields. Odessa. 439 p. (in Russian).

Ptak, M., Sojka, M., Nowak, B. (2020). Effect of climate warming on a change in thermal and ice conditions in the largest lake in Poland – Lake Sniardwy. J. Hydrol. Hydromech., 68, 3, 260–270. http://www.uh.sav.sk/Portals/16/vc_articles/2020_68_3_Ptak_260.pdf

Shaltout, M., Omstedt, A. (2014). Recent sea surface temperature trends and future scenarios for the Mediterranean Sea. Oceanologia. 56. 3. 411–443.

Schaeffe, B.A., Iiame, J., Dwyer, J. et al. (2018). An initial validation of Landsat 5 and 7 derived surface water temperature for U.S. lakes, reservoirs, and estuaries. International Journal of Remote Sensing Volume 39. 7789-7805. https://www.tandfonline.com/doi/full/10.1080/01431161.2018.1471545

Slobodnik, J. et al. (2020a). National Pilot Monitoring Studies and Joint Open Sea Surveys in Georgia, Russian Federation and Ukraine, 2016: Final Scientific Report / J. Slobodnik, B. Alexandrov, V. Komorin, A. Mikaelyan, A. Guchmanidze, M. Arabidze, A. Korshenko, S. Moncheva. Dnipro: Seredniak T.K.

Slobodnik, J. et al. (2020b). 12-Months National Pilot Monitoring Studies in Georgia, Russian Federation and Ukraine, 2016-2017: Final Scientific Report/ J. Slobodnik, V. Medinets, B. Alexandrov, V. Komorin, A. Mikaelyan, A. Guchmanidze, M. Arabidze, A. Korshenko. Dnipro: Seredniak T.K.

Slobodnik, J. et al. (2020c). National Pilot Monitoring Studies and Joint Open Sea Surveys in Georgia, Russian Federation and Ukraine, 2017: Final Scientific Report / J. Slobodnik, B. Alexandrov, V. Komorin, A. Mikaelyan, A. Guchmanidze, M. Arabidze, A. Korshenko. Dnipro: Seredniak T.K.

Vyshnevskyi, V.I., Shevchuk, S.A. (2018). Use of remote sensing data for study of water bodies of Ukraine. 84 p. (in Ukrainian).

Vyshnevskyi, V., Shevchuk, S. (2021). Thermal regime of the Dnipro Reservoirs. J. Hydrol. Hydromech., 69, 3, 300–310. https://doi.org/10.2478/johh-2021-0016

Woolway, R.I., Dokulil, M.T, Marszelewski W. et al. (2017). Warming of Central European lakes and their response to the 1980s climatic regime shift. Climate change, 141: 759–773. https://link.springer.com/content/pdf/10.1007/s10584-017-1966-4.pdf

DROUGHT HAZARD ASSESSMENT USING ANOMALY DROUGHT INDEX AND GEOGRAPHIC INFORMATION SYSTEM IN THE CHI RIVER BASIN, THAILAND

Sarunphas IAMAMPAI[1] , *Jirawat KANASUT[1]* , *Banramee KANTAWONG[1]*
and Prem RANGSIWANICHPONG[1]*

ABSTRACT:
Drought is a natural disaster that causes problems in agriculture. Such events have been frequented in the Chi River basin, Thailand, in the last 10 years. Currently, drought assessment is conducted in several ways, including the use of product data from satellites or meteorological and hydrological data. In this study, we developed the anomaly drought index (ANDI) based on a combination of normalized difference vegetation index and soil water index product data and runoff station data using the entropy weight method. The ANDI was created and compared to historical data from the Emergency Events Database (EM-DAT) and drought-related data from 2011 to 2020. The ANDI shows a correlation with 0.80 with rice yield and 0.9 with reservoir (dry), and 0.95 with rainfall (wet). Furthermore, we classified ANDI's drought intensity into three categories: mild, moderate, and severe. Droughts were discovered to have occurred in 2013, 2015, 2016, 2019, and 2020. They began in wet seasons and materialized in the subsequent dry seasons. These results were consistent with the drought reports of EM-DAT and the National Hydroinformatics Data Center. Therefore, the ANDI is suitable for assessing drought in the Chi River basin in Thailand.

Key-words: Natural hazard, NDVI, Remote Sensing, Spatial analysis, Thailand.

1. INTRODUCTION

Drought is the one of natural disasters that occurs in several countries, and its impact can huge damage to environmental, energy, and human activities, especially agriculture (Wilhite et al., 2007). Generally, drought hazard occurred by a prolonged lack of precipitation or insufficient rainfall (Kallis, 2008; Dracup et al., 1980). Additionally, extended periods of precipitation deficit deplete surface water and groundwater reservoirs (Marcos-Garcia et al., 2017). Thailand frequently experiences droughts and is one of the most drought-affected countries in Southeast Asia (Pandey et al., 2007). Its agricultural sector is directly affected by droughts. Most of the Thai people are farmers and are often affected by drought. The damage from droughts in Thailand is concentrated in the northeastern part of the country, given the large area used for rice cultivation in this region (Prabnakorn et al., 2018). In 2004‑2005 and 2015‑2017, drought events in Thailand occurred with high damage to economic system approximately $220 million and $330 million, respectively (Wichitarapongsakun et al., 2016).

Several researchers have developed drought indices for analyzing drought hazards using meteorological and satellite data (Anderson et al., 2016; Cui et al., 2021). Drought indices are mostly analyzed based on precipitation data; they include the Palmer drought severity index (Palmer, 1965), standard precipitation index (SPI) (McKee et al., 1993), and standard precipitation evapotranspiration index (SPEI) (Vicente-Serrano et al., 2010). Despite all of these indices, the SPI is the most suitable for assessing drought in rain-fed areas, but not for irrigated areas (Teweldebirhan et al., 2019). The normalized difference vegetation index (NDVI) can represent quantified vegetation based on satellite data and connected to drought occurrences. It is calculated from red and infrared reflectances.

The NDVI is widely used to indicate vegetation health because it can detect chlorophyll well (Tucker, 1979; Amri et al., 2011; Rouse et al., 1974). The soil water index (SWI) is a drought index developed from a fusion of surface soil moisture (SSM) observations from the Sentinel-1 C-band SAR and Metop ASCAT sensors; it is calculated via recursive formulation. The SWI has eight characteristic T values for various soil depths (Bauer-Marschallinger et al., 2018; Albergel et al., 2008; Brocca et al., 2010). The global drought index (GDI) was developed based on a combination of the NDVI and SWI. The GDI varies between areas with different land uses (Zribi et al., 2021). The composite drought indicator is a drought index calculated by a weighted combination of the SPI, NDVI, evapotranspiration, and land surface temperature (Bijaber et al., 2018). These studies show that composite drought indices can better identify drought hazard areas than single indices.

In this study, we developed a new index (Anomaly Drought Index [ANDI]) using the Entropy Weight (EW) method for analyzing the drought hazard in the Chi River basin by combining runoff station data, SPI, NDVI, and SWI. In addition, drought reports from relevant agencies were used for comparison to verify the accuracy of the drought assessment results obtained from the ANDI.

2. STUDY AREA

The Chi River basin is a major watershed in northeast Thailand. It has an area of 49,131 km^2 (**Fig. 1**) and consists of the following provinces: Chaiyaphum, Khon Kaen, Nong Bua Lamphu, Udon Thani, Maha Sarakham, Nakhon Ratchasima, Loei, Phetchabun, Kalasin, Roi Et, Yasothon, Ubon Ratchathani, Sisaket, and Mukdahan.
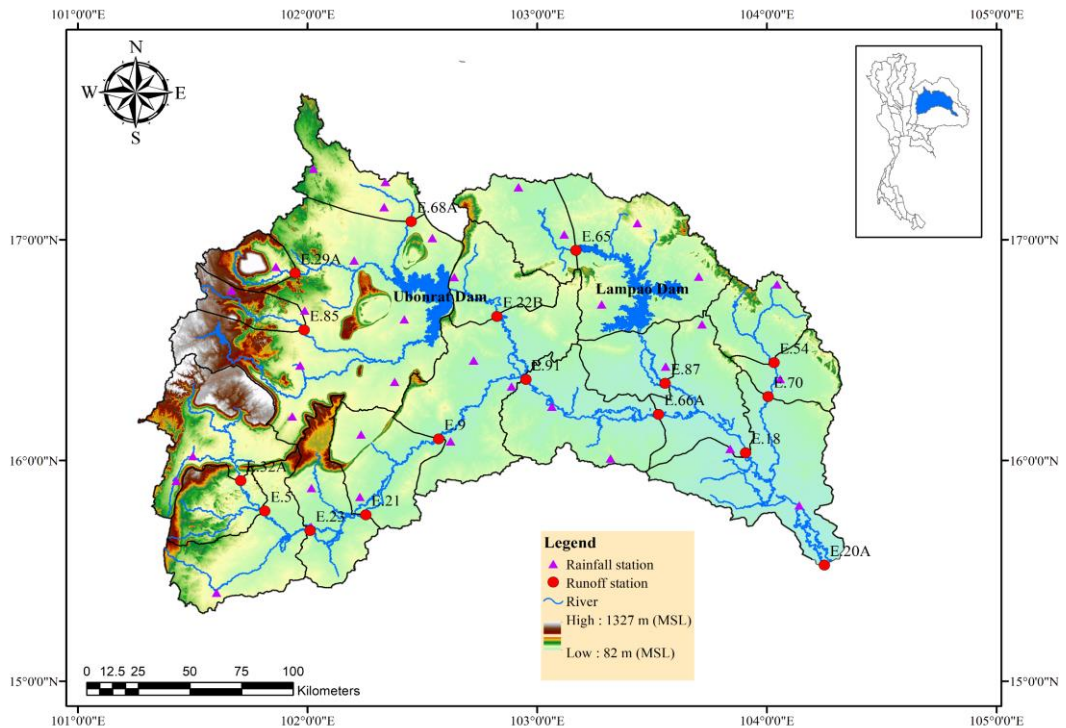
Fig. 1. Topography map of the Chi River basin.

The climate in the Chi River basin is influenced by northeast and southeast monsoons, which result in a dry season (November‑April) and a wet season (May‑October), respectively. Annual rainfall in the Chi River basin averages approximately 1,285 mm and more than 90% of rainfall occurs in the wet season (**Table 1**).

Monthly average rainfall in the Chi River basin.

month	1	2	3	4	5	6	7	8	9	10	11	12	sum
average rainfall (mm)	12	7	36	56	150	136	242	309	253	67	16	1	1285

There are two large, important reservoirs, namely, the Ubon Rat Reservoir and the Lampao Reservoir, which have storage capacities of 2,431 million and 1,980 million m^3, respectively. According to land use data produced by the Department of Land Development in 2016, agricultural land accounts for 70% of the total area, whereas forest land accounts for 19% (**Fig. 2**). In the dry season, normally farmers planted rice crop during December-January. For harvesting activities is usually between March and April.
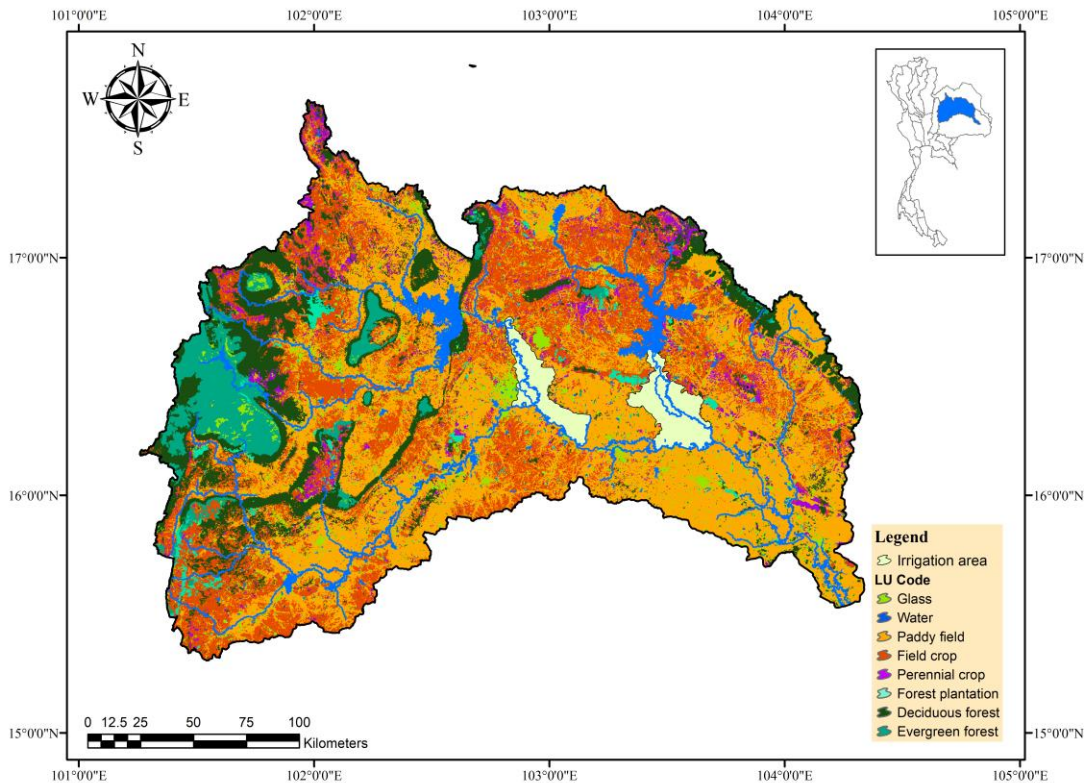
Fig. 2. Land used map of the Chi River basin.

3. DATA AND METHODS

In this study, we used data from 15 runoff stations operated by the Royal Irrigation Department (RID), observed monthly runoff data from 2011 to 2020 (10 years), data from two large reservoirs (Ubon Rat and Lampao dam), and rice yield during dry season data (collected by the RID). The monthly rainfall data were collected by 37 rain gauge stations of the Thai Meteorological Department during 2010 to 2020 for calculating SPI. Furthermore, we gathered land use data from the Land Development Department. We then generated a drought hazard map for the Chi River basin using remote sensing data and a geographic information system (GIS).

In 1993, the SPI was developed by Mckee et al., which it based only on monthly precipitation data. The SPI can be described using Eqs. (1)-(3) (Asadi Zarch, M.A. et al. 2014).

$$H(x_k) = q + (1 - q)G(x_k) \tag{1}$$

For 0 < H(x_k) < 0.5

$$SPI = -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right), \; t = \sqrt{ln\left(\frac{1}{H(x_k)^2}\right)} \tag{2}$$

For 0.5 < H(x_k) < 1.0

$$SPI = \left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right), \; t = \sqrt{ln\left(\frac{1}{(1 - H(x_k))^2}\right)} \tag{3}$$

the $G(x_k)$ is Gamma probability density function, q is the probability of zero precipitation and $H(x_k)$ is the cumulative probability. Where c_0 = 2.515517, c_1 = 0.802853, c_2 = 0.010328, d_1 = 1.432788, d_2 = 0.189269, and d_3 = 0.001308.

The SWI, which was included to represent the water content, is a product of the Copernicus Global Land Service (https://land.copernicus.eu/global/products/swi accessed on 1 March 2022). The SWI has a daily spatial resolution of 10 km^2 and a characteristic temporal length (T) parameter that includes 1, 5, 10, 15, 20, 40, 60, and 100 to reflect soil depth. (Albergel et al., 2008; Paulik et al., 2014). This index is highly correlated with soil moisture content, especially when the T length is 20 (Zribi et al., 2010). The SWI is calculated as Eqs. (4)-(6).

$$SWI(t_n) = \sum_i^n SSM(t_i) \, e^{-\frac{t_n - t_i}{T}} \Big/ \sum_i^n e^{-\frac{t_n - t_i}{T}} \tag{4}$$

$$SWI(t_n) = SWI(t_{n-1}) + K_n(SSM(t_n) - SWI(t_{n-1})) \tag{5}$$

$$K_n = K_{n-1} / (K_{n-1} + e^{-(\frac{t_n - t_{n-1}}{T})}) \tag{6}$$

the variables t_n and t_{n-1} are the current and previous dates, respectively, in Julian days. T is the parameter of characteristic time length. SSM is Surface Soil Moisture observations from Sentinel-1 C-band SAR and Metop ASCAT sensors. K_n is factor at time n. The initial value of SWI_0 and K_0 were set = 1.

The NDVI, an effective remote sensing indicator of green vegetation distribution, is derived by determining the difference in spectral reflectance value between the red and NIR bands (Rouse Jr et al., 1974). Additionally, the NDVI can indicate the growing conditions of plants (Bharathkumar et al., 2015; Gillespie et al., 2018) and is used for drought monitoring (Son et al., 2012; Nanzad et al., 2019). Generally, the NDVI value ranges between −1 and 1 and can be calculated as in Eq. (7). In this study, we used NDVI values from MODIS (MOD13A2). The MOD13A2 product is corrected for atmospheric conditions, has a resolution of 1 km^2, and is provided every 16 days (https://e4ftl01.cr.usgs.gov/MOLT/MOD13A2.006 (accessed on 1 March 2022).

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{7}$$

where *NIR* denotes the near-infrared band and *RED* is the red band. They have wavelengths of 859 and 649 nm, respectively.

3.1. Relationship of drought indices

This method calculates data values in a way that indicates the change in such values for the same month in each year. Thus, the obtained data are not affected by the season. The standard anomaly time series is calculated by Eq. (8). Previously, the usual anomaly approach was utilized to determine the NDVI and SWI (Amri et al., 2011; Amri et al., 2012). To analyze drought without the influence of seasonality, we produced standard anomalies for the NDVI, SWI, and runoff data; we renamed the indices by adding Sa before the original names, thus, Sa-NDVI, Sa-SWI, and Sa-R. Due of its significant relationship to drought episodes in the Chi River basin, a 3-month cumulative runoff was used in drought estimate for Sa-R. The results shown that accumulated runoff from the previous 3 months (Sa-R) is strongly related to the drought events in the Chi River basin.

The SPI was created to track precipitation shortages over time. These time scales describe the impacts of drought on various water supplies, which have two timings for detecting drought: 1. short durations (1-6 months) and 2. long timescales (12-36 months). In this study, the SPI's 6-month cumulative rainfall was estimated.

$$Sa_i(\text{x}) = \frac{X_i - X_{avg,i}}{\sigma_i} \tag{8}$$

where *X* denotes *SPI, SWI, NDVI*, and *Sa-R*. X_i is the data estimate for a given month *i*. $X_{avg,i}$ is the mean value of the data for month *i* generated from ten years of data. and σ_i is the standard deviation of the data during month *i* over the same 10-year period.

Additionally, we analyzed the relationship between the four indices (SPI, Sa-NDVI, Sa-SWI, and Sa-R) and the drought data (rice yield, rainfall, and water storage) by dividing the data into two seasonal groups (dry and wet seasons). Rice yield and water storage data were compared between dry season, and rainfall data was compared between wet season. The Pearson correlation coefficient is usually used to test relationships as Eq. (9).

$$R_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{9}$$

where *x* is the indices (*SPI, Sa-SWI, Sa-NDVI, and Sa-R*) and *y* is the drought data (yield, rainfall, and water storage).

3.2. ANDI

In this research, we developed the ANDI for assessing the drought hazard in the Chi River basin. This new index was developed as a combination of SPI, Sa-SWI, Sa-NDVI, and Sa-R obtained from the same period, and it was computed by Eq. (10). The Entropy Weight (EW) was used to evaluate the criterion weight based on the intensity of the available data, and it calculated the weights of the four indices (*a*, *b*, *c*, and *d*) (Shannon et al., 1948). A specified variable will vary more the higher the entropy. The composite drought index weight is often determined using this approach (Chen et al., 2020, Waseem et al., 2015). The EW is computed as in Eqs. (11)-(14).

$$ANDI = a(SPI) + b(Sa\text{-}NDVI) + c(Sa\text{-}SWI) + d(Sa\text{-}R) \tag{10}$$

$$P_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}} \tag{11}$$

$$E_j = -\frac{1}{ln(n)} \sum_{i=1}^{n} P_{ijt} \, ln(P_{ij}) \tag{12}$$

$$D_j = 1 - E_j \tag{13}$$

$$Ew_j = \frac{D_j}{\sum_{j=1}^{n} D_j} \tag{14}$$

where i is i^{th} year ($i = 1, 2, ... 10$); X_{ij} is value of j index at the ith year ($j = 1, 2, 3$, and 4 refer to SPI, Sa-SWI, Sa-NDVI, and Sa-R) and P_{ij} is the projection value that corresponds to the normalized value. E_j is entropy for j index; D_j is Degree of divergence for j index; Ew_j are the weights allocated to the variables.

The ANDI map, which was created using the SPI, Sa-NDVI, Sa-SWI, and Sa-R data sets, should aid in the efficiency of drought assessment in each sub-basin area. Because the NDVI and SWI data are held at spatial resolutions of 1 and 10 km^2, respectively, we aggregated the NDVI resolution to 10 km^2 by average value, which has the same resolution as SWI, and then computed it to the grid map of Sa-NDVI and Sa-SWI at 10 km^2 resolution. The SPI map was constructed by distributing rainfall station data to all grids with a resolution of 10 km^2 using the inverse distance squared technique (IDW), where rainfall station weights change with distance, as indicated in Eq (15). Then it was grid-calculated into the SPI grid map. We computed Sa-R from runoff station data at sub-basin outlets and used it to represent all grids in these sub-basins. The Sa-R grid map utilized in this study has 15 sub-basins with no reservoir area (See **Fig. 3** for the Sa-R map). Following that, the grid maps of Sa-NDVI, Sa-SWI, Sa-R, and SPI were weighted using the EW technique, which was derived using the grid as shown in **Fig.3.**

$$W_i = \frac{\frac{1}{d_i^2}}{\sum_{i=1}^{n}\left(\frac{1}{d_i^2}\right)} \tag{15}$$

where d_i is the distance between the runoff station and the considered grid location. W_i is the grid weighting of each rainfall station.

For effective drought assessment by ANDI. We determined three levels of severity for drought (mild, moderate, and severe) by comparing the ANDI values in the basin with quantities of rice yield and drought events recorded from EM-DAT. Drought conditions were represented by rice crops since the amount of rice cultivated between December and April of each year is determined by the amount of available water. However, droughts have the greatest impact on rice harvests.

3.3. Validation of ANDI via comparison of drought events recorded

The results of the ANDI drought map assessment in the Chi River basin were compared with the above drought levels. We divided the drought assessments into two seasons (dry and wet) and compared them with actual drought situations to analyze the relationship between them. We used drought reports for the Chi River basin from two sources, namely, the Centre for Research on the Epidemiology of Disasters (CRED) and the National Hydroinformatics Data Center (NHC). The CRED provides free access to the full Emergency Events Database (EM-DAT) for noncommercial purposes. EM-DAT is a global database containing details on the occurrences and effects of over 22,000 natural disasters, such as the number of people killed, injured, or affected and economic damage estimates. The NHC is maintaining a database for the management of water resources and natural disasters in Thailand. The NHC has been collecting records on water events for droughts and floods since 2005. **Fig. 4** shows a flowchart of the methodology in this study.
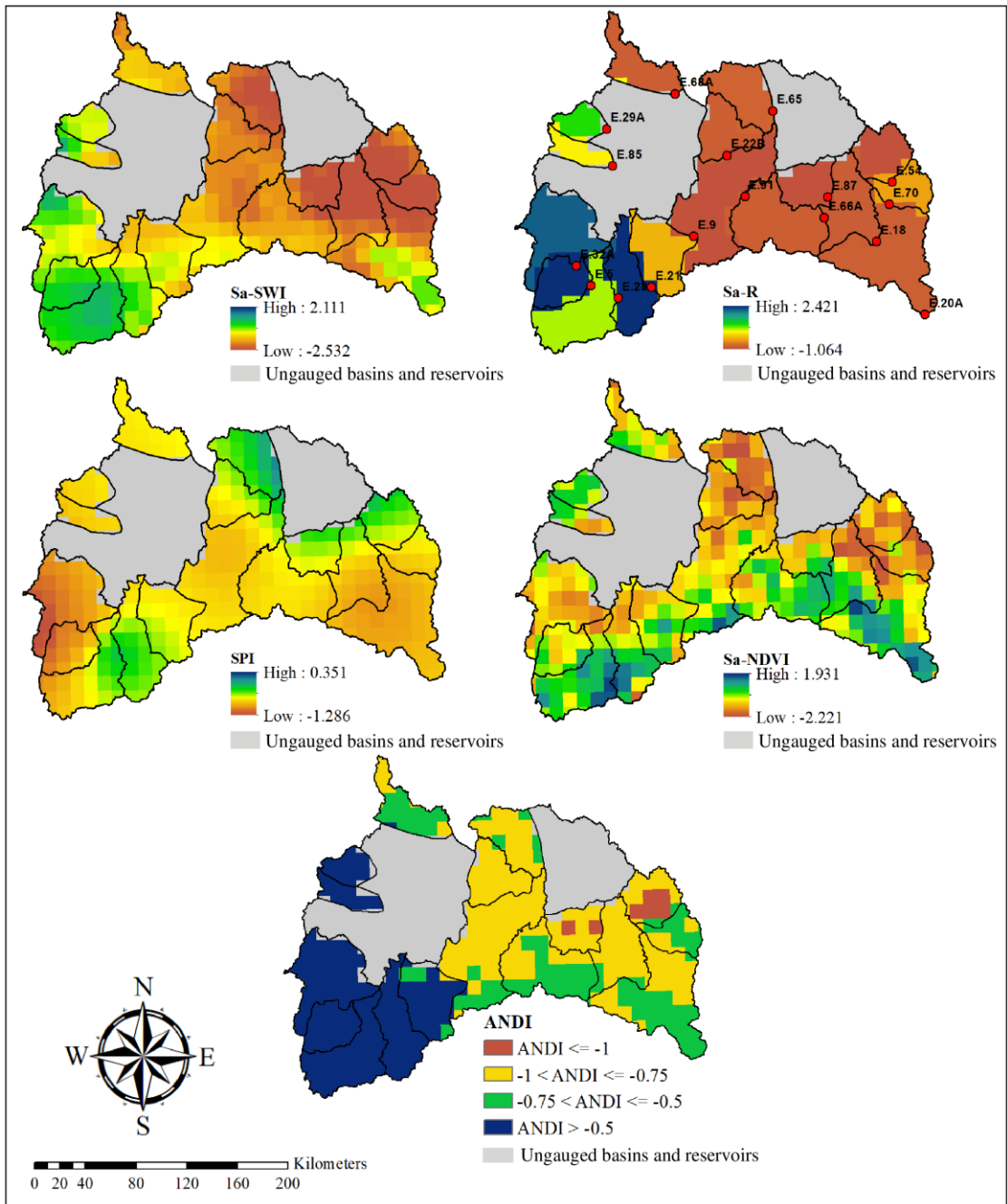
Fig. 3. Process of generating the ANDI map for assessing drought hazard (February 2013).
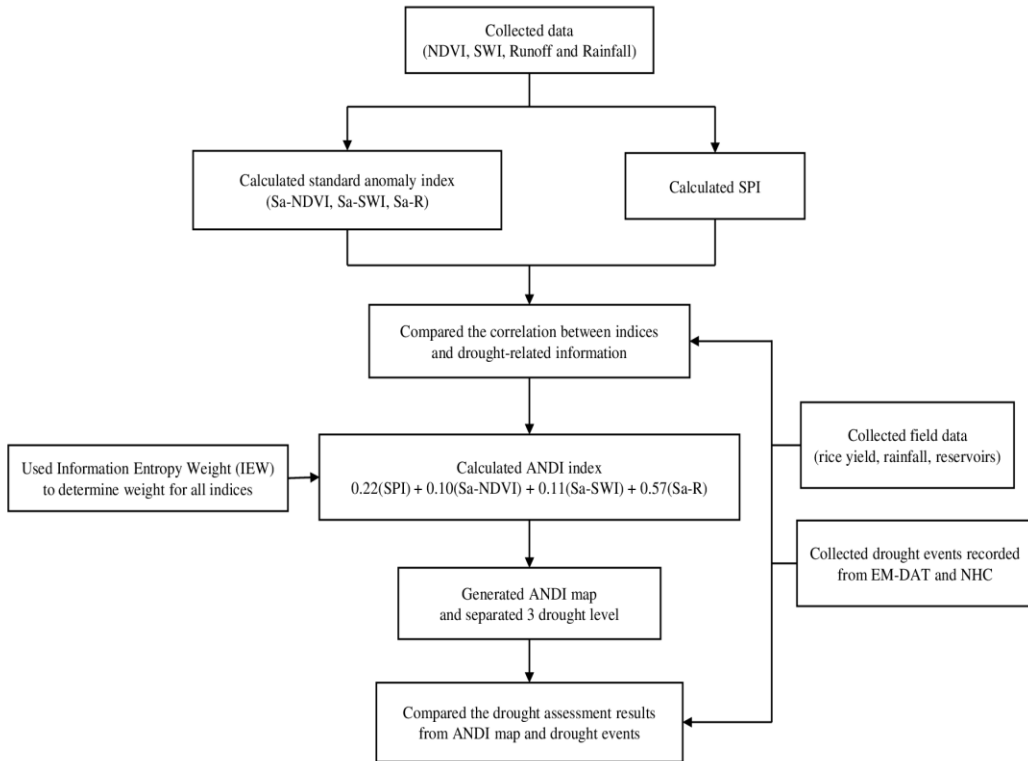
Fig. 4. Flowchart of the study methodology hazard.

4. RESULTS AND DISCUSSIONS

4.1. Relationship between drought indices and historical data

In the drought analysis for this study, **Fig. 5** shows the time series of the four data sets had distinct characteristics (SPI NDVI, SWI, and Runoff). The SPI demonstrates the change in cumulative rainfall at 6 months, and a negative value indicates that cumulative rainfall at 6 months is less than average. The remaining three data sets indicate that NDVI, SWI, and runoff are greater in the rainy season and lower in the dry season, with seasonality. Furthermore, each drought index reacts differently to drought occurrences. The NDVI is a useful tool for detecting changes in agricultural areas. However, it may not be suitable for non-agricultural locations. In December–April, the NDVI values in the irrigation area are clearly higher than those in the rain-fed area (**Fig. 6a**). This is because rice crops are cultivated in the irrigation area but not in the rain-fed area. **Fig. 6b** shows that the SWI values in the irrigated and rain-fed areas are nearly identical. Although the SWI accurately measures soil moisture content, the soil moisture that causes dryness in each place may differ, depending on the type of land use in that area. Additionally, the runoff data considered in this study directly shows the amount of water available to everyone, which accurately reflects the available quantity. However, it drops to zero during the dry season, making it difficult to quantify the severity of each year's drought (**Fig. 5d**).
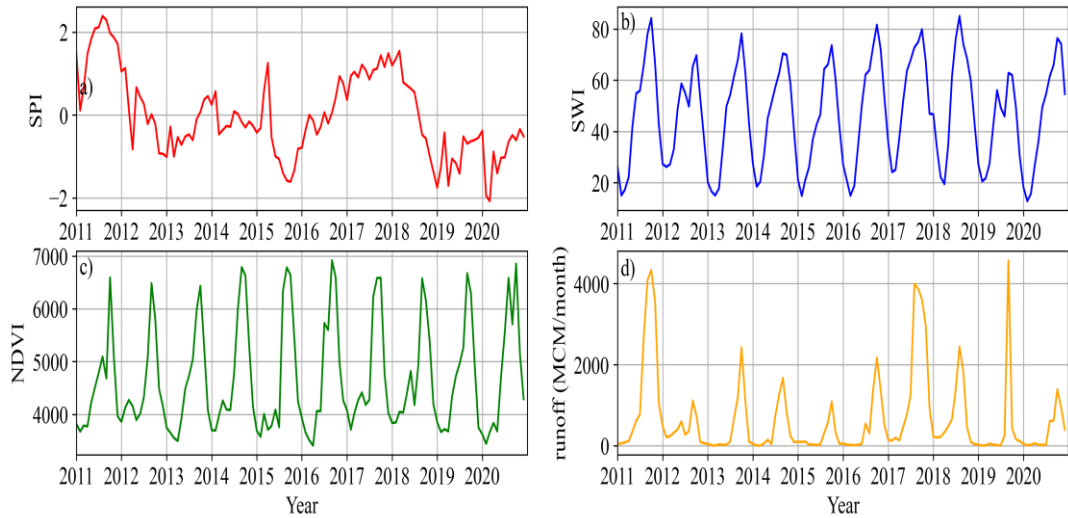
Fig. 5. SPI, NDVI, SWI, and monthly runoff (million cubic meters (MCM)/month) data time series from 2011–2020.
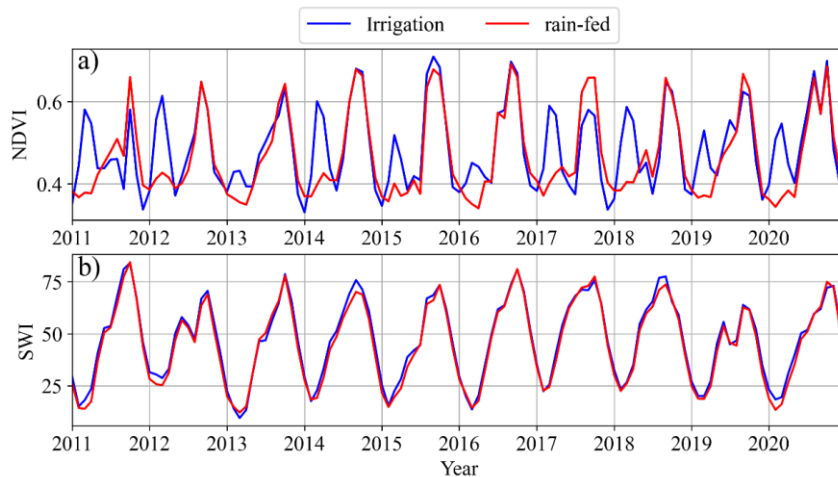
Fig. 6. SWI and NDVI time series in irrigation and rain-fed areas.

The results of the standard anomaly method (Sa-NDVI, Sa-SWI and Sa-R) are shown in **Fig. 7**, which can be separated into two main groups, namely, negative (the index is below the annual average for that month) and positive (the index is above the annual average for that month). These results indicate that if the standard anomaly value is negative during the dry season, then there is less soil moisture than typical. However, if it occurs during the rainy season, then rainfall is delayed or less than typical. The analysis of the seasonal correlation between the four indices and the measurement data (rice yield, reservoirs, and rainfall) reveals that, in dry seasons, Sa-R has the highest correlation with rice yields (0.86), followed by Sa-NDVI (0.67), SPI (0.61), and Sa-SWI (0.48), whereas SPI has the strong correlation with reservoir (0.88), followed by Sa-R (0.85), Sa-NDVI (0.77), and Sa-SWI (0.65). For the wet season, SPI is the most correlated with rainfall (0.96) and is followed by Sa-R (0.89), Sa-SWI (0.72), and Sa-NDVI (0.28) (**Table 2-3**). In addition, Sa-R stands out in its relationship to rice yield, while its relationship with reservoirs and rainfall is slightly less SPI. Although the Sa-R is more powerful than the others, it is a station data. When used as a representative of the sub-basin, it cannot see the difference between each pixel in sub-basins.
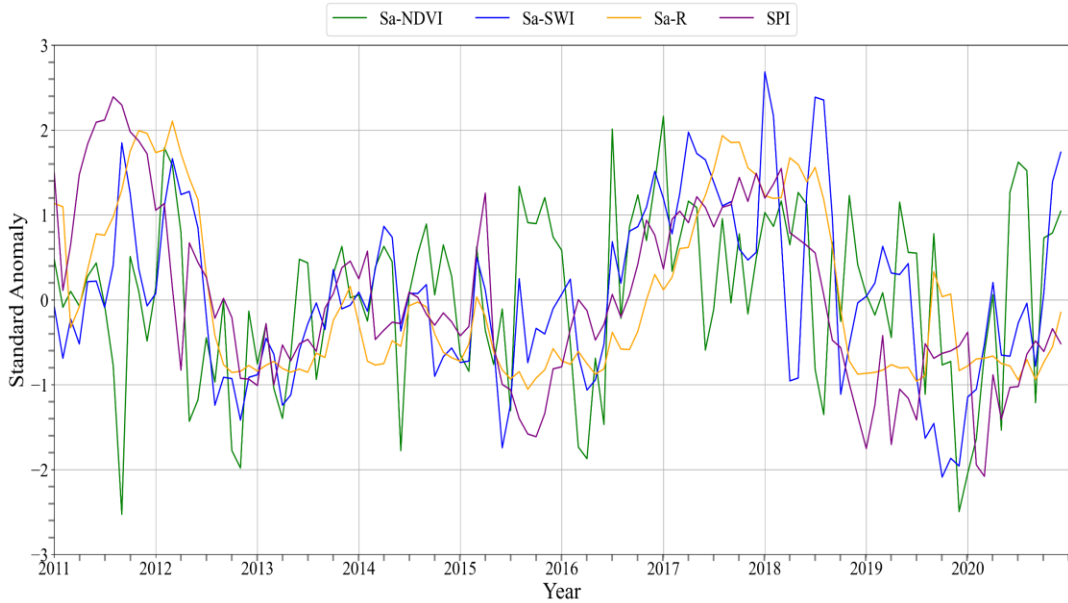
Fig. 7. SPI, Sa-NDVI, Sa-SWI, and Sa-R time series from 2011–2020.

Table 2.
Correlation coefficient between drought indices during before harvest,
rice yield and reservoir storage in the Chi River basin.

	SPI	Sa-NDVI	Sa-SWI	Sa-R	ANDI
Rice yield (dry)	0.61	0.67	0.48	0.86	0.80
Reservoir (dry)	0.88	0.77	0.65	0.85	0.90

Table 3.
Correlation coefficient between drought indices during
end of rainy, rainfall in the Chi River basin.

	SPI	Sa-NDVI	Sa-SWI	Sa-R	ANDI
Rainfall (wet)	0.96	0.28	0.72	0.89	0.95

The combining of Sa-R with spatial data indices. Combining Sa-R with spatial data indices (Sa-NDVI and Sa-SWI) will assist to overcome this challenge.

Afterward, we estimated the weights of indices that were used to generate ANDI via the EW method. The results of this approach are provided in (**Table 4-5**), which were calculated in February. The weights of SPI, Sa-NDVI, Sa-SWI, and Sa-R are 0.22, 0.10, 0.11, and 0.57, respectively. However, this weight is unique for each month. The ANDI has correlations with rice yield, reservoirs, and rainfall of 0.80, 0.90, and 0.95, respectively (**Table 2-3**). It has the highest correlation with reservoir, while the correlation with rice yield is larger than SPI, Sa-NDVI, and Sa-SWI, and the correlation with rainfall is greater than Sa-NDVI, Sa-SWI, and Sa-R. ANDI also outperformed Sa-R and SPI in terms of pixel-difference. Therefore, ANDI was selected to represent drought conditions because it is well related to the drought situation in the dry and wet seasons (**Fig. 8**).

Practical Example of Entropy Weight Calculation in February.

	February										
	Year	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
	SPI	0.11	0.15	0.09	0.13	0.08	0.08	0.15	0.17	0.04	0.01
P_{ij}	Sa-NDVI	0.09	0.18	0.08	0.07	0.06	0.06	0.08	0.13	0.10	0.14
	Sa-SWI	0.06	0.14	0.07	0.09	0.06	0.10	0.13	0.19	0.10	0.05
	Sa-R	0.20	0.27	0.03	0.03	0.05	0.03	0.13	0.21	0.02	0.03
	SPI	-2.25	-1.87	-2.45	-2.06	-2.47	-2.48	-1.93	-1.80	-3.21	-5.04
$ln(P_{ij})$	Sa-NDVI	-2.40	-1.71	-2.55	-2.67	-2.81	-2.88	-2.47	-2.01	-2.25	-1.95
	Sa-SWI	-2.74	-1.96	-2.60	-2.43	-2.77	-2.26	-2.06	-1.68	-2.28	-3.02
	Sa-R	-1.59	-1.32	-3.62	-3.46	-2.98	-3.58	-2.08	-1.54	-3.97	-3.38

**Entropy Weight for SPI, Sa-NDVI, Sa-SWI
and Sa-R for combining into ANDI.**

Indices	e_j	div_j	W_j
SPI	0.94	0.06	0.22
Sa-NDVI	0.97	0.03	0.10
Sa-SWI	0.97	0.03	0.11
Sa-R	0.83	0.17	0.57
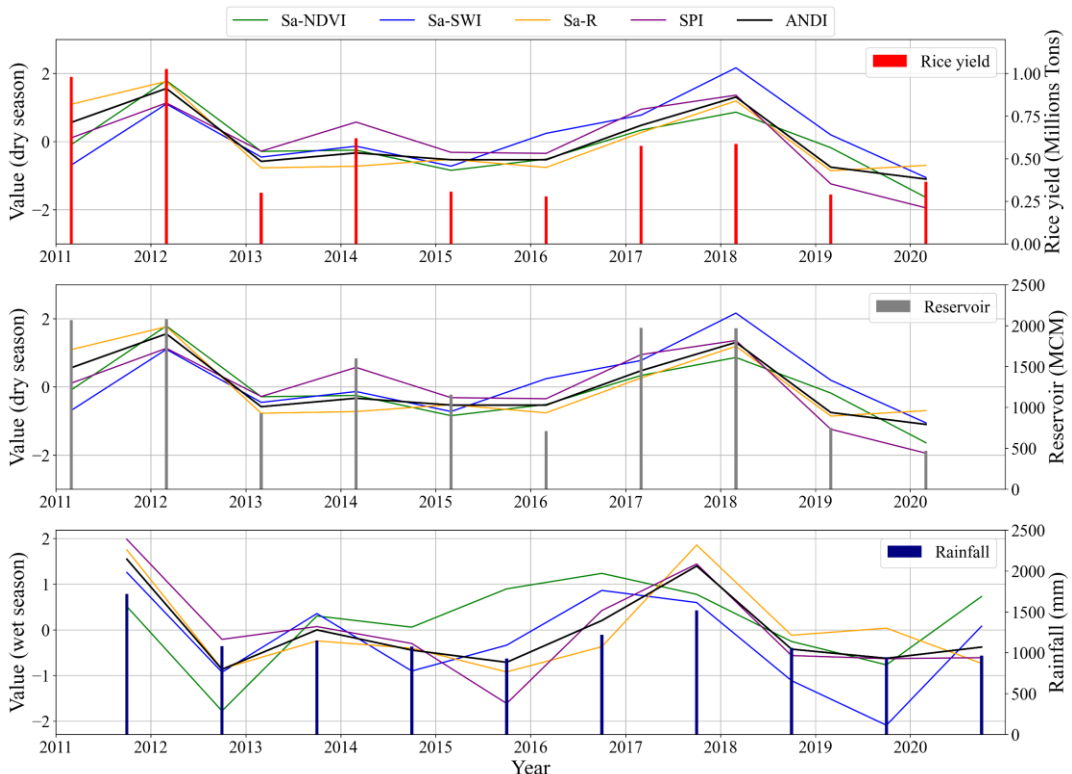
Fig. 8. Relationship between SPI, Sa-NDVI, Sa-SWI, and Sa-R and rice yield, reservoirs and rainfall.

4.2. Generation of ANDI drought hazard map

Remote sensing and GIS are used to produce the ANDI to investigate the spatial distribution of drought hazard areas in the basin with a 10 km^2 resolution. Most of the ANDI values in the 10 studied years (2011–2020) are between −1.5 and 1.5, but they are less than −0.5 when considering only the drought situation in the Chi River basin (2013, 2015, 2016, 2019, and 2020) (**Fig. 9**).
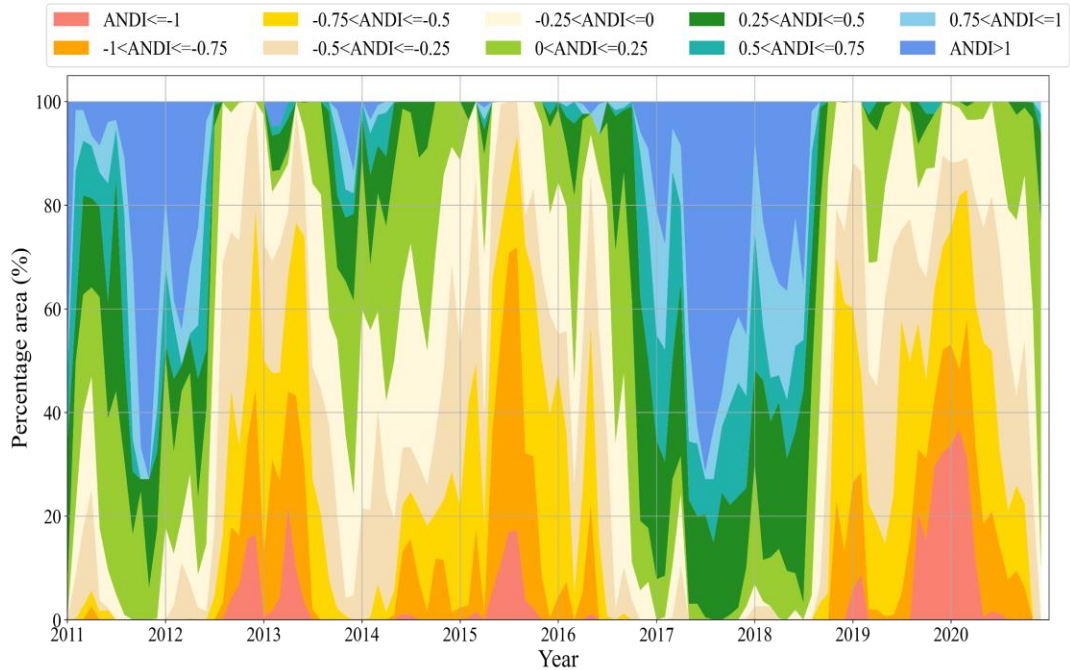
Fig. 9. Time series of monthly ANDI in the Chi River basin during 2011 – 2020.

According to data on rice yield, rainfall, and reservoir water quantities, drought occurred in 2013, 2015, 2016, 2019, and 2020. The initial ANDI, which was used to determine the occurrence of drought, was then examined.

In **Fig. 9**, the pink, orange, and yellow bars show the examined ANDI values of ANDI ≤−1, −1 < ANDI ≤ −0.75, and −0.75 < ANDI ≤ −0.5. Findings show that the ANDI of −0.25 is overvalued as a drought indicator because it sees areas in 2017 and 2018 that were not undergoing droughts. Moreover, the values of −0.75 and −1 are too small; they identify too little of the drought area in 2016. Thus, we set ANDI = −0.5 as the threshold for drought conditions in this study, and we established three drought levels for drought monitoring in the Chi River basin. **Fig. 9** shows the proportion of ANDI values dispersed in each range of ANDI values. Hence, we experimented with various random values to determine the appropriate value for the drought events (2013, 2015, 2016, 2019, and 2020). Then, we set the following ranges of drought levels: mild drought (−0.75 < ANDI ≤ −0.5), moderate drought (−1 < ANDI ≤ −0.75), and severe drought (ANDI ≤ −1).

The ANDI map was generated using the determined drought levels. There were three apparent droughts, namely, 2012 (wet) – 2013 (dry), 2014 (wet) –2016 (dry), and 2018 (wet) – 2020 (wet), all of which were followed by a wet season and then a dry season (**Fig. 10**). The drought event of 2019 – 2020 was the worst in 10 years, followed by those in 2012–2013 and 2014–2015; these severe droughts affected 37%, 22%, and 17%, respectively, of the basin.
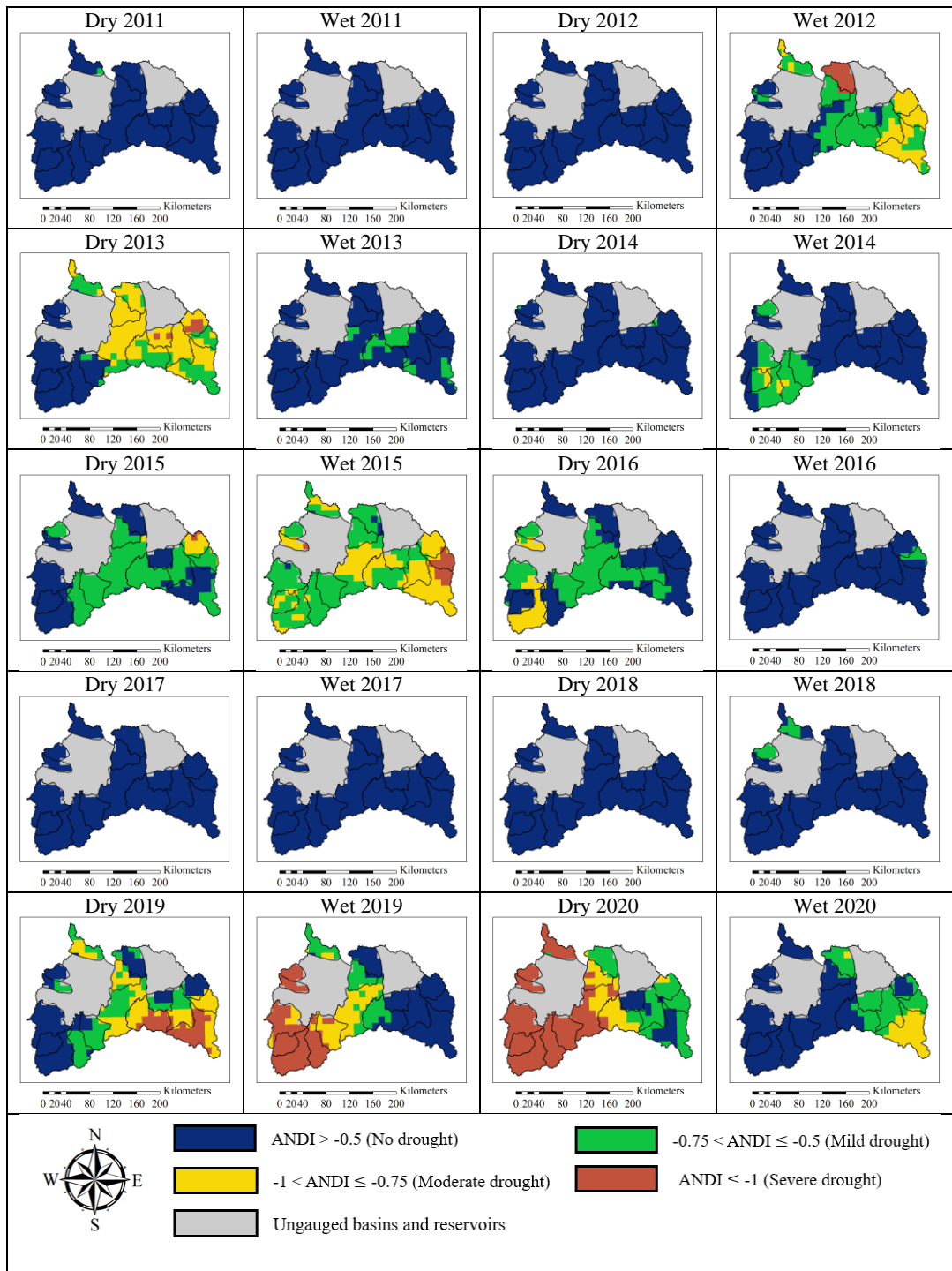
Fig.10. Drought hazard map in the Chi River basin based on the ANDI during 2011–2020.

4.3. Validation of ANDI map with reported drought events

Data on drought occurrences in the Chi River basin were collected from the georeferenced EM-DAT (https://public.emdat.be/data accessed on 1 March 2022) and the NHC (https://tiwrm.hii.or.th/v3/archive accessed on 1 March 2022). The EM-DAT contains three disaster reports of drought in the Chi River basin in 2011–2020. In 2012, the April–August drought incurred damages of $1.2 million. From January 2015 to May 2017 (2 years and 4 months), the long, intense drought incurred a total damage value of $3.3 billion. The last recorded drought occurred in July 2019 and lasted 6 months, but no damage was reported (**Table 6**). The NHC contains data about five drought events that occurred during 2011–2020 (2013, 2014, 2015, 2016, and 2020), which may have occurred during the dry season (November–April). According to a comparison between the ANDI map's drought assessment results and the drought situation reports from the CRED and NHC, the ANDI map's assessment is consistent with that of EM-DAT during 2015, 2016, and 2020 but not in 2017 and 2019. It is also consistent with drought data from the NHC for 2013, 2015, 2016, and 2020 but not for 2014 and 2019. A comparison of EM-DAT and NHC data shows corresponding drought events during dry seasons in 2015, 2016, and 2020. This is consistent with the drought analysis results from the ANDI map; inaccuracies are observed for 2019 (**Table 7**).

In Thailand, several research articles on drought monitoring have been published. In 2019, Zenkoji et al. analyzed drought in the upper Chao Phraya River basin using data on rainfall and dam inflows of large reservoirs (Bhumibol Dam and Sirikit Dam) during 1953–2015; their findings showed that droughts occurred in 2012–2015, especially in 2015, when the worst drought occurred (Zenkoji et al., 2019). Raksapatcharawong et al. (2020) studied droughts in 2012, 2014, 2015, 2018, and 2019 using information gathered from the Department of Agricultural Extension and concluded that northeast Thailand experienced the greatest impact. Jomsrekrayom et al. (2021) conducted a study on droughts in the northeast region of Thailand using VCI values from MODIS products; their results showed droughts during the wet seasons of 2011, 2012, 2013, 2017, 2018, and 2019, with those in 2013 and 2019 being severe ones. The drought assessment results from these previous studies show droughts in 2012, 2013, 2014, 2015, 2018, and 2019 in agreement with the ANDI drought assessment results.

Table 6.

Drought hazard events in Thailand from EM-DAT

Disaster No.	Disaster Type	Country	Start year	Start month	End year	End month	Total Damages (US$)
2012-9109-THA	Drought	Thailand	2012	4	2012	8	1,200,000
2015-9574-THA	Drought	Thailand	2015	1	2017	5	3,300,000,000
2019-9359-THA	Drought	Thailand	2019	7	2020	2	-

Table 7.

Comparison of drought hazard results of the ANDI with EM-DAT and NHC data.

Year	2011		2012		2013		2014		2015		2016		2017		2018		2019		2020	
Season	D	W	D	W	D	W	D	W	D	W	D	W	D	W	D	W	D	W	D	W
ANDI				✓	✓			✓	✓	✓	✓							✓	✓	✓	
EMDAT				✓					✓	✓	✓	✓	✓						✓	✓	
NHC				✓	✓	✓	✓	✓	✓	✓	✓								✓	✓	
Combined				✓					✓	✓	✓								✓	✓	

5. CONCLUSIONS

Drought analysis in the Chi River basin using the proposed ANDI, a composite index incorporating SPI, Sa-NDVI, Sa-SWI, and Sa-R, is conducted from a combination analysis of satellite data and station measurement data. The NDVI and SWI values are based on MODIS and ASCAT products, respectively, whereas SPI is based on precipitation and utilizes IDW to distribute into pixels. Runoff measurements are used to construct runoff maps. All four data points are connected to various drought scenarios. The SPI detects variations in precipitation, the NDVI distinguishes agricultural and non-agricultural places, the SWI shows soil moisture, and runoff indicates the quantity of water farmers may utilize at a given moment.

ANDI outperforms other indices in its association with the reservoir during the dry season, but it performs similarly to Sa-R and SPI in its link with rice yield during the rainy season. Furthermore, ANDI is more efficient than the Sa-R and SPI for estimating droughts in any season.

Drought occurs during the dry season and is driven mostly by water scarcity; during the wet season, it is influenced primarily by rainfall. The ANDI is analyzed and compared with drought event data from 2011 to 2020 to establish drought severity levels. During times of drought, most ANDI values are below -0.5. Therefore, the ANDI value of -0.5 is set as the threshold for entering a drought state. Then, the following drought levels are set: mild drought ($-0.75 <$ ANDI ≤ -0.5), moderate drought ($-1 <$ ANDI ≤ -0.75), and severe drought (ANDI ≤ -1). Drought analysis using the ANDI and the defined levels shows that each drought occurred during the wet season and arrived during the dry season because of the low rainfall during the wet season. This results in insufficient water supply in dams during dry seasons. Additionally, findings indicate a total of 5 years of drought (2013, 2015, 2016, 2019, and 2020). For 2015, 2016, and 2020, the ANDI drought assessment results agree with EM-DAT and NHC drought reports. Therefore, the ANDI can assess drought effectively (close to actual events) and is a good tool for monitoring and assessing future droughts.

ACKNOWLEDGEMENTS

R E F E R E N C E S

Albergel, C., Rudiger, C., Pellarin, T., Calvet, J.C., Fritz, N., Froissard, F., Suquia, D., Petitpa A., Piguet, B., Martin, E. (2008) From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations. *Hydrology and Earth System Sciences*. 12(6), 1323-1337.

Amri, R., Zribi, M., Lili-Chabaane, Z., Duchemin, B., Gruhier, C., Chehbouni, A. (2011) Analysis of vegetation behavior in a North African semi-arid region, using SPOT-VEGETATION NDVI data. *Remote sensing*. 3(12), 2568-2590.

Amri, R., Zribi, M., Lili-Chabaane, Z., Wagner, W., Hasenauer, S. (2012) Analysis of C-band scatterometer moisture estimations derived over a semiarid region. *IEEE transactions on geoscience and remote sensing*. 50(7), 2630-2638.

Anderson, M.C., Zolin, C.A., Sentelhas, P.C., Hain, C.R., Semmens, K., Yilmaz, M.T., Gao, F., Otkin, J.A., Tetrault, R. (2016) The Evaporative Stress Index as an indicator of agricultural drought in Brazil: An assessment based on crop yield impacts. *Remote Sensing of Environment*. 174, 82-99.

Bauer-Marschallinger, B., Paulik, C., Hochstöger, S., Mistelbauer, T., Modanesi, S., Ciabatta, L., Massari, C., Brocca, L., Wagner, W. (2018) Soil moisture from fusion of scatterometer and SAR: Closing the scale gap with temporal filtering. *Remote Sensing*. 10(7), p. 1030.

Bharathkumar, L., Mohammed-Aslam, M.A. (2015) Crop pattern mapping of Tumkur Taluk using NDVI technique: a remote sensing and GIS approach. *Aquatic Procedia*. 4, 1397-1404.

Bijaber, N., El Hadani, D., Saidi, M., Svoboda, M.D., Wardlow, B.D., Hain, C.R., Poulsen, C.C., Yessef, M., Rochdi, A. (2018) Developing a remotely sensed drought monitoring indicator for Morocco. *Geosciences*. 8(2), 55.

Brocca, L., Melone, F., Moramarco, T., Wagner, W., Hasenauer, S. (2010) ASCAT soil wetness index validation through in situ and modeled soil moisture data in central Italy. *Remote Sensing of Environment*. 114(11), 2745-2755.

Chen, S., Zhong, W., Pan, S., Xie, Q., Kim, T. W. (2020). Comprehensive drought assessment using a modified composite drought index: A case study in Hubei Province, China. *Water*, 12(2), 462.

Cui, A., Li, J., Zhou, Q., Zhu, R., Liu, H., Wu, G., Li, Q. (2021) Use of a multiscalar GRACE-based standardized terrestrial water storage index for assessing global hydrological droughts. *Journal of Hydrology*. 603, p. 126871.

Dracup, J.A., Lee, K.S., Paulson Jr, E.G. (1980) On the definition of droughts. *Water resources research*. 16(2), 297-302.

Gillespie, T.W., Ostermann-Kelm, S., Dong, C., Willis, K.S., Okin, G.S., MacDonald, G.M. (2018) Monitoring changes of NDVI in protected areas of southern California. *Ecological Indicators*. 88, 485-494.

Jomsrekrayom, N., Meena, P., & Laosuwan, T. (2021) Spatiotemporal analysis of vegetation drought variability in the middle of the northeast region of Thailand using terra/modis satellite data. *Geographia Technica*. 16, 70-81.

Kallis, G. (2008) Droughts. Annual review of environment and resources. 33, 85-118.

Marcos-Garcia, P., Lopez-Nicolas, A., Pulido-Velazquez, M. (2017) Combined use of relative drought indices to analyze climate change impact on meteorological and hydrological droughts in a Mediterranean basin. *Journal of Hydrology*. 554, 292-305.

McKee, T.B., Doesken, N.J., Kleist, J. (1993) The relationship of drought frequency and duration to time scales. In Proceedings of the 8th Conference on Applied Climatology, *California*, pp. 179-183.

Nanzad, L., Zhang, J., Tuvdendorj, B., Nabil, M., Zhang, S., Bai, Y. (2019) NDVI anomaly for drought monitoring and its correlation with climate factors over Mongolia from 2000 to 2016. *Journal of arid environments*. 164, 69-77.

Palmer, W.C. 1965. Meteorological drought. US Department of Commerce, Weather Bureau.

Pandey, S., Bhandari, H., Ding, S., Prapertchob, P., Sharan, R., Naik, D., Taunk, S.K., Sastri, A. (2007) Coping with drought in rice farming in Asia: insights from a cross-country comparative study. *Agricultural Economics*. 37, 213-224.

Paulik, C., Dorigo, W., Wagner, W., Kidd, R. (2014) Validation of the ASCAT Soil Water Index using in situ data from the International Soil Moisture Network. *International journal of applied earth observation and geoinformation*. 30, 1-8.

Prabnakorn, S., Maskey, S., Suryadi, F. X., de Fraiture, C. (2018) Rice yield in response to climate trends and drought index in the Mun River Basin, Thailand. *Science of the Total Environment*. 621, 108-119.

Raksapatcharawong, M., Veerakachen, W., Homma, K., Maki, M., Oki, K. (2020) Satellite-based drought impact assessment on rice yield in Thailand with SIMRIW− RS. *Remote Sensing*. 12(13), p. 2099.

Rouse Jr, J.W., Haas, R.H., Deering, D.W., Schell, J.A., Harlan, J.C. (1974) Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation (No. E75-10354).

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3), 379-423.

Son, N.T., Chen, C.F., Chen, C.R., Chang, L.Y., Minh, V.Q. (2012) Monitoring agricultural drought in the Lower Mekong Basin using MODIS NDVI and land surface temperature data. *International Journal of Applied Earth Observation and Geoinformation*. 18, 417-427.

Teweldebirhan Tsige, D., Uddameri, V., Forghanparast, F., Hernandez, E.A., Ekwaro-Osire, S. (2019) Comparison of meteorological-and agriculture-related drought indicators across Ethiopia. *Water*. 11(11), p. 2218.

Tucker, C.J. (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*. 8(2), 127-150.

Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I. (2010) A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate*. 23(7), 1696-1718.

Waseem, M., Ajmal, M., Kim, T. W. (2015). Development of a new composite drought index for multivariate drought assessment. *Journal of Hydrology*, 527, 30-37.

Wichitarapongsakun, P., Sarin, C., Klomjek, P., Chuenchooklin, S. (2016) Rainfall prediction and meteorological drought analysis in the Sakae Krang River basin of Thailand. *Agriculture and Natural Resources*. 50(6), 490-498.

Wilhite, D.A., Svoboda, M.D., Hayes, M.J. (2007) Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness. *Water resources management*. 21(5), 763-774.

Zarch, M. A. A., Sivakumar, B., Sharma, A. (2015). Droughts in a warming climate: A global assessment of Standardized precipitation index (SPI) and Reconnaissance drought index (RDI). *Journal of hydrology*, 526, 183-195.

Zenkoji, S., Tebakari, T., Dotani, K. (2019) Rainfall and reservoirs situation under the worst drought recorded in the Upper Chao Phraya River Basin, Thailand. *Journal of Japan Society of Civil Engineers, Ser. G (Environmental Research)*. 75(5), 115-124.

Zribi, M., Nativel, S., Le Page, M. (2021) Analysis of agronomic drought in a highly anthropogenic context based on satellite monitoring of vegetation and soil moisture. *Remote Sensing*. 13(14), p. 2698.

Zribi, M., Paris Anguela, T., Duchemin, B., Lili, Z., Wagner, W., Hasenauer, S., Chehbouni, A. (2010) Relationship between soil moisture and vegetation in the Kairouan plain region of Tunisia using low spatial resolution satellite data. *Water Resources Research*. 46(6), 823-835.

GIS-BASED ANALYTICAL HIERARCHY PROCESS MODELING
FOR FLOOD-PRONE AREA MAPPING IN VIETNAM

Huu Duy NGUYEN[1]* 🆔 *, Gheorghe ŞERBAN[2]** 🆔

ABSTRACT:
Flood is a more frequent natural hazard, having destroying effects on human life and property around the world. Water resource management strategies require understanding of the flood susceptibility. This study's objective is to develop an effective new approach to the construction of a flood susceptibility map (FSM) in the Nhat Le‑Kien Giang River watershed in Vietnam, based on the analytical hierarchy process (AHP) technique. AHP was used to calculate the weighting of each of nine conditioning factors to build a flood susceptibility map. The results showed that the AHP technique would be successful in constructing a FSM with an AUC value of 0.95. The area of high and very high flood susceptibility in the residential area increased from 95 km^2 in 2005 to 251 km² in 2020. While the area of high and very high flood susceptibility in the agricultural area increased from 239 km^2 in 2005 to 245 km² in 2020. The findings of this study can support decision-makers and planners working in flood management and the development of mitigation strategies. Although the approach in this study has been applied to construct a flood susceptibility map for one area in Vietnam, it can be applied both to other types of natural hazards and in other countries.

Key-words: Flood susceptibility, AHP, Nhat Le–Kien Giang River, Vietnam, flood management.

1. INTRODUCTION

Flood is now the most common and the most damaging of natural hazards globally, with occurrences having increased by around 40% over the past two decades (Hirabayashi et al. 2013, Prasad et al. 2021). According the United Nations Office for Disaster Risk Reduction (UNISDR), 150,061 floods occurred globally between 1995 and 2016, causing approximately 157,000 deaths and affecting 200 million people every year (Towfiqul Islam et al. 2021, UNISDR 2015). Asia is particularly exposed: nine of the top ten countries affected by floods are Asian (Pham et al. 2020). In addition, Vietnam is often affected by major floods and typhoons every year, causing significant damage to people, agriculture, and housing. Although the Vietnamese government has focused on structural measures like dikes, dams, and early warning systems, flood prediction tools are still limited (Luu et al. 2021). Climate change combined with socio-economic growth has had a significant effect on flood and is set to further increase the risk of flood in the future (Costache 2019, Nachappa et al. 2020). Therefore, the assessment of FS is important steps when preparing management and mitigation strategies regarding future emergencies.

Various methods have been carried out by the global scientific community to construct flood susceptibility maps. Various studies have employed the MIKE hydraulic model (Patro et al. 2009, Tansar et al. 2020) and the SWAT model (Narsimlu et al. 2015, Rajib et al. 2020), which have both proven effective in analyzing the effects of flood on a given territory. However, these models utilized detailed field data such as river cross-river, meteorological, and hydrological data series, so they have mainly been applied in smaller regions with good data quality. In addition, the remote sensing and geographic information systems (GIS) in FSM have contributed to a variety of studies in the field (Dewan et al. 2007, Kabenge et al. 2017). However, flood usually occurs quickly, following bad weather, so in many cases the sensors cannot accurately identify flood time and are often affected by

───────────────
[1] *Faculty of Geography, University of Science, Vietnam National University, Hanoi, Vietnam; nguyenhuuduy@hus.edu.vn*
[2] *Faculty of Geography, Babes-Bolyai University, Cluj-Napoca, Romania; gheorghe.serban@ubbcluj.ro*
* *correspondance: Huu Duy Nguyen (nguyenhuuduy@hus.edu.vn), Gheorghe Şerban (gheorghe.serban@ubbcluj.ro)*

cloud cover. All of this prevents adequate monitoring. Although remote sensing and GIS can show the spatial distribution of flood, they are limited in their ability to detect the driving forces behind the flood. These models require further development or replacement with more powerful methods, to reduce the restrictions of hydrological models (Hostache 2010).

The causes of flood depend on natural conditions, and anthropogenic activities (Bui et al. 2020). We have reviewed several studies that employed AI methods to predict FS. They can be categorized two approaches: machine learning methods and expert knowledge. Identifying the correlations between these factors enables screening for suitable flood prediction models. Machine learning approaches include support vector machine (Tehrany et al. 2014), bagging (Chen et al. 2019), dagging (Yariyan et al. 2020), random forest (Chen et al. 2020, Lee et al. 2017), adaboost (Bui et al. 2016, Hong et al. 2018), artificial neural network (Falah et al. 2019, Khoirunisa et al. 2021), and K-nearest neighbor (Abu El-Magd et al. 2021, Ren et al. 2019). The calculation principle of these methods is very complex and involves a significant workload when applied to a large area. Moreover, the overfitting and global optimization problems are important limitations when using machine learning to construct flood susceptibility maps. A database management system has recently been integrated with decision support systems to build flood susceptibility maps. Multiple-criteria decision-making (MCDM) makes it possible to determine the optimal variables that influence the probability of flood occurrence (Nguyen et al. 2020). In particular, the analytical hierarchy process (AHP) is widely applied to comprehensively assess susceptibility. It is a simpler technique which can accurately determine flood-susceptible areas by careful evaluation of different influencing factors (Vojtek and Vojteková 2019a). This technique, particularly when combined with GIS, can take into account a large number of parameters in order to obtain precise results that closely represent reality. This method can be applied in large regions. Luu et al. (2020) used MCDM to assess the flood risk in Quang Binh Province in Vietnam. Das (2020) mapped susceptibility in the Western Ghat coastal belt in India utilizing AHP and multi-source geospatial data. Vojtek and Vojteková (2019a) also constructed a map for Western Ghat, using AHP and multi-source geospatial data, and used AHP and GIS to assess flood susceptibility in Slovakia. Dano (2020) used AHP to determine flood-susceptible areas and propose mitigation strategies in Jeddad in Saudi Arabia using expert judgements. However, there is no previous research in the study area that used expert opinion to analyze the factors causing flood. Expert opinion can support deciders to construct necessary strategies to reduce flood damage. Therefore, the application of the AHP method combined with GIS is appropriate for a medium-sized watershed such as Nhat Le‑Kien Giang, which lies in a humid tropical region that is characterized by rugged topography, high urban growth, and the effects of climate change.

The assessment of flood susceptibility in regions in the process of spatial planning has been an important part of previous research. It is particularly important in the context of urban growth. Spatial planning is a tool for managing adaptation to, and the effects of, climate change, as well as reducing the negative effects of flood. For developed areas, flood risk can be reduced by developing effective early warning systems and response and adaptation plans, while for developing areas, building restrictions in areas prone to flood is necessary for sustainable spatial planning.

This study aimed to identify the areas most vulnerable to flood in the Nhat Le‑Kien Giang River basin using AHP and GIS to reduce the consequences of flood. The novelty of this study is that the first time the flood susceptibility map is constructed in the Nhat Le – Kien Giang watershed using the AHP technique. This technique in this study can be applied in other regions in Vietnam. AHP will be established using the conditioning factors and the evaluation of their importance for the probability of flood occurrence. GIS facilitates the analysis and processing of spatial data, as well as facilitating the analysis and evaluation of the AHP results. This is the first time the proposed method has been used. The hypothesis is that the AHP technique would be successful in constructing the FSM. It was also hypothesized that the proportion of urban land in the high and very high flood susceptibility areas would increase. The findings of this study can be used to support planners in developing strategies for managing and mitigating flood risks in Vietnam and around the world.

2. MATERIALS AND METHODS

2.1. Study area and Flood inventory mapping

The Nhat Le-Kien Giang River basin in central Vietnam covers approximately 2,650 km^2 (**Fig. 1**) with the mean altitude about 234 m and the mean slope 20.7 m. The climate divided by the rainy season lasts from August to November and the dry season from December to July. The average annual precipitation is 2000–2500 mm; however, 65–70% is focused in the rainy season.
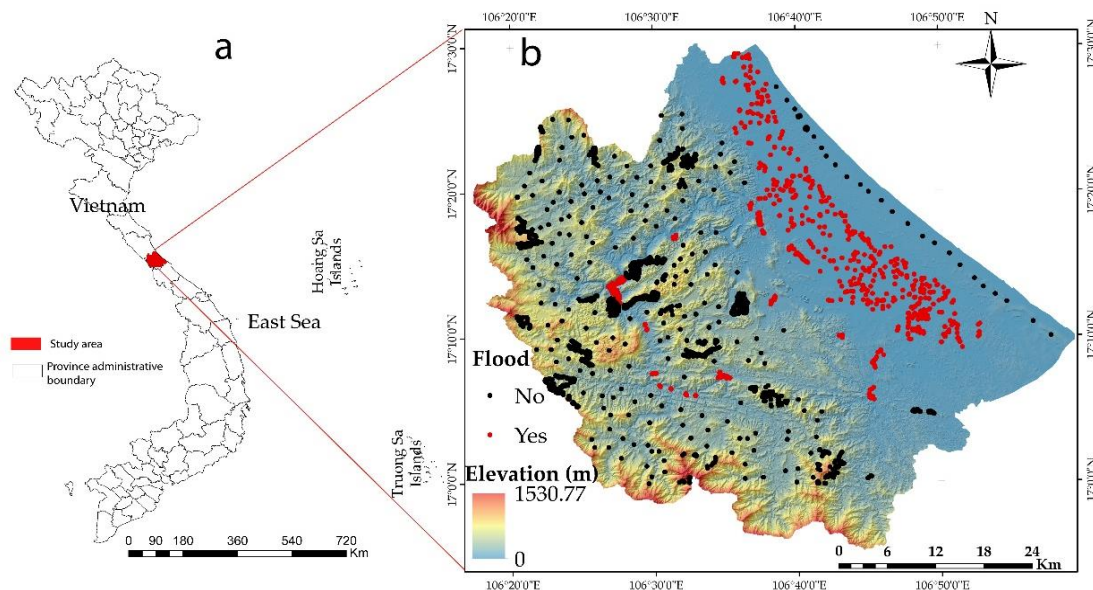
Fig. 1. – a) The location of the Nhat Le – Kien Giang basin in the Vietnam; b) The flood mark and the elevation of the Nhat Le – Kien Giang River basin.

Approximately 75% of the study area is covered by forest, although this coverage has been shrinking rapidly in recent years due to urban expansion, agricultural development, and illegal logging. This increases the likelihood of flood. The region was affected by three major floods in October 2020, which caused significant damage to property and loss of life (**Fig. 2**).

Fig. 2. – The flood marks in the Nhat Le – Kien Giang in 2020 (left) and 2013 (right). These flood marks were recorded by GPS with X and Y coordinates in 2015 and 2022 in Hong Thuy commune – Le Thuy district, Quang Binh, Vietnam. These are two historic floods in Vietnam, have significant effects on human life and the country's economy. *Source: Huu Duy Nguyen 2022 and 2015.*

Floods generally occur from September to November; this is also the period when extreme climatic events such as storms and tropical depressions occur. In many cases, floods occur in combination with these major flood events. Major floods in the study area, such as in 1999, 2010, 2013, 2016, 2020 and 2021 cause major material and human damage. The development of a GIS flood location database is a fundamental step in analyzing flood susceptibility. It shows an overview of the relationship between historical flood locations and input variables. In this study, flood locations were extracted from the Vietnam Disaster Management Authority archives. Flood marks were also collected from the field mission. In addition, to improve the accuracy of the method, flood samples were obtained using a Sentinel 1A image on 18 October 2020. In final, 502 flood samples were obtained in the basin. Also, to improve the quality of the FSM, 1462 non-flood points were selected in high-altitude areas which did not flood. All points were used to build flood inventory mapping using ArcGIS 10.4 software.

2.2. Flood conditioning factors

Conditioning factors play a crucial role in constructing a flood susceptibility model because they represent the connections between past flood events and topographical, climatic, hydrological, and anthropogenic conditions (Gudiyangada Nachappa et al. 2020, Nguyen et al. 2021). In this study, the conditioning factors were collected from sources such as official data from government organizations and remote sensing data. Nine conditioning factors were selected from the literature reviews: elevation, slope, aspect, curvature, rainfall, land use and land cover (LULC), normalized difference vegetation index (NDVI), distance to river, and distance to road (**Fig. 3**).

Elevation, slope, aspect, and curvature – at a resolution of 10m – were obtained from DEM, which was constructed using a 1/50,000 m topographical map. Distance to river and distance to road were extracted from a 1/50,000m topographic map, which is available from Ministry of Natural Resources and Environment of Vietnam. The Euclid Distance method in GIS was used to create the Distance to Road and Distance to River for the flood assessment. The LULC map is also available from this department. Rainfall was built from data collected from ten hydrological stations in the province. The IDW method in GIS was used for the interpolation of rainfall data to build the rainfall map. While NDVI was calculated from a Landsat 8 OLI map (disponible in https://earthexplorer.usgs.gov/). All these factors were analyzed spatial distribution and converted to GRID raster format with 10 m resolution using GIS. Each factor was reclassified and flood susceptibility impact was ranked from *very low* to *very high*, using the natural break method, on the ArcGIS platform (**Tab. 1**).

Table 1.

Flood susceptibility impact of each factor.

	Flood susceptibility level				
	Very low	**low**	**Moderate**	**High**	**Very high**
Elevation	> 728.79	469.5–728.79	270.51–469.5	101.67–270.51	0–101.67
Slope	> 37.5	27.32–37.57	17.36–27.32	6.8–17.36	0–6.8
Aspect	> 286.58	211.76–286.58	141.17–211.76	70.58–141.17	0–70.58
Curvature	> 1.88	0.42–1.88	-0.74–0.42	-2.2–(-0.74)	- 52–(-2.2)
Rainfall	< 2816.7	2816.7–2986.2	2986.2–3159.4	3159.4–3347.4	3347.4–3573
NDVI	> 0.55	0.36–0.55	0.13–0.36	-0.05–0.13	-0.3–(-0.05)
Distance to river	> 2641	1579.9–2641	896–1579.9	377.29 - 896	0–377.29
Distance to road	> 7952.5	4696.7–7952.5	2455.13–4696.7	907.33–2455.13	0–907.33
LULC	Forest	Barren	Agricultural	Urban	Water
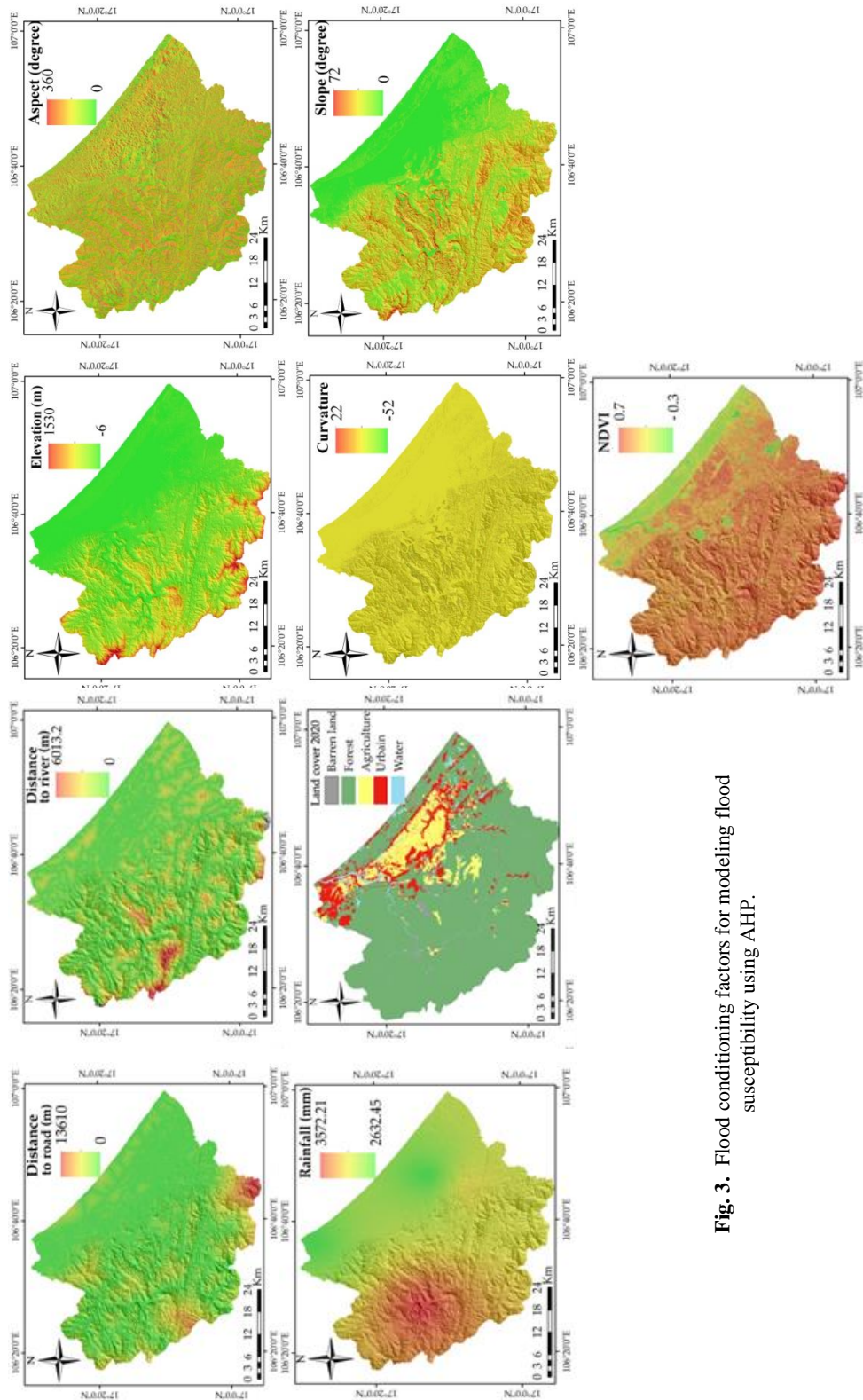
Fig. 3. Flood conditioning factors for modeling flood susceptibility using AHP.

It should be noted that areas with very low flood susceptibility is roped in by 1 and very high is roped in by 5. The natural break method is based on Jenk's optimization formulation to reduce the variability of each class. It generates clusters that have similar values separated by breakpoints and suits values that have been distributed irregularly or have a tendency to cluster at an end point in the distribution. The natural break method has been used in several previous studies on flood susceptibility. LULC was grouped by the spatial distribution and characterization of each type of LULC. It was classified as very low (forest), low (barren land), moderate (agriculture), high (urban) and very high (water). **Table 1** shows the impact of flood susceptibility of each factor.

3. ANALYTIC HIERARCHY PROCESS (AHP)

In this study, AHP was used to compute the weights of the criteria to build the FSM. The advantage of this method is the direct use of expert opinions, the simple integration of the GIS and the consistency of judgment. However, this method has general limitations related to subjective opinions in the evaluation. It is considered the limit in most MCDA methods.

AHP is the theoretical measure of tangible criteria, which was proposed by T. Saaty in 1980. The weights of the criteria were evaluated by pairwise comparison matrices. AHP can support decision-making processes by quantifying alternative priorities by decision makers. This technique has been widely applied in several fields such as economy, environment, transport. In recent years, it is applied in the study of flood susceptibility.

The AHP technique is carried out by four main steps: i) construction of the hierarchical model; ii) the construction of the matrices for the pairs for the criteria based on the subjective judgment of the experts; iii) calculation of criteria weights; iv) model consistency assessment.

i) Construction of the hierarchical model: In the Nhat Le ‑ Kien Giang River basin, several factors influence the flood occurrence. These factors have been divided into four groups: Geo-environment, hydrology, climate and anthropogenic activity. So, nine factors elevation, rainfall, LULC, distance to river, NDVI, slope, aspect, curvature, and distance to road were selected for the hierarchical model.

ii) the construction of the matrices for the pairs based on the subjective judgment of the experts: nine factors conditionings are compared in pairs to calculate the weighting of each factor; the importance of each factor is measured on a scale of 1 to 9 (**Tab. 2**). The weights of the criteria in the AHP method have been evaluated based on the subjective judgment of several experts (Li et al. 2013) or on the experience of the author (Luu et al. 2020). In this study, we used the author's experience.

iii) calculation of the weights of the criteria: after the construction of the matrices for the pairs, several techniques are used to calculate the weights like normalization and vector. In this study, we applied the normalization technique (**Tab. 3**). It should be noted that the higher the weights of the criteria, the more the criteria influence the flood susceptibility.

Table 2.

Pair-wise comparison matrix.

Factors	Elevation	Rainfall	LULC	Distance to river	NDVI	Slope	Aspect	Curvature	Distance to road
Elevation	1	2	3	3	4	4	5	6	8
Rainfall	0.5	1	2	2	3	3	4	5	6
LULC	0.33	0.5	1	2	3	3	3	4	5
Distance to river	0.33	0.5	0.5	1	2	2	2	3	4
NDVI	0.25	0.33	0.33	0.5	1	2	2	2	3
Slope	0.25	0.33	0.33	0.5	0.5	1	2	2	2
Aspect	0.2	0.25	0.33	0.5	0.5	0.5	1	2	2
Curvature	0.16	0.2	0.25	0.33	0.5	0.5	0.5	1	2
Distance to road	0.125	0.16	0.2	0.25	0.33	0.33	0.5	0.5	1

Table 3.

Normalized factor weights

Factors	Elevation	Rainfall	LULC	Distance to river	NDVI	Slope	Aspect	Curvature	Distance to road	Weight (w_i)
Elevation	0.31	0.37	0.37	0.29	0.26	0.24	0.25	0.23	0.24	0.29
Rainfall	0.15	0.18	0.25	0.19	0.20	0.18	0.2	0.19	0.18	0.2
LULC	0.10	0.09	0.12	0.19	0.20	0.18	0.15	0.15	0.15	0.15
Distance to river	0.10	0.09	0.06	0.09	0.13	0.12	0.1	0.11	0.12	0.1
NDVI	0.07	0.06	0.04	0.04	0.06	0.12	0.1	0.07	0.09	0.08
Slope	0.07	0.06	0.04	0.04	0.03	0.06	0.1	0.07	0.06	0.06
Aspect	0.06	0.04	0.04	0.04	0.03	0.03	0.05	0.07	0.06	0.05
Curvature	0.05	0.03	0.03	0.03	0.03	0.03	0.025	0.03	0.06	0.04
Distance to road	0.03	0.03	0.02	0.02	0.02	0.02	0.025	0.01	0.03	0.03

iv) Model consistency assessment: Consistency ratio (CR) is used to examine consistency in expert judgment in the process of comparing factors. CR < 0.10 indicates acceptable consistency, while CR > 0.10 indicates inconsistency. It is measured by the following equation (Hammami et al. 2019):

$$CR = \frac{CI}{RI}$$

The following equation calculates the consistency index (CI) (Hammami et al. 2019):

$$CI = \frac{\lambda_{max} - n}{n - 1}$$

λ_{max} is the maximum eigenvalue of the pairwise comparison matrix (n x n). The maximum eigenvalue λ_{max} is always greater than or equal to the number of rows or columns n. The more consistent the rating, the closer the calculated value of λ_{max} is to n (Ghosh et al. 2018):

$$\lambda_{max} = \sum_{i=1}^{n} wi * \sum_{j=1}^{n} a_{ii}$$

RI is a random index which depends on the number of factors used in the comparison matrix.

In this study, λ max = 9.26, CI = 0.03, RI = 1.54, and CR = 0.023 (satisfying < 0.10), meeting the requirements of the pairwise comparison matrix. The flood susceptibility map was constructed using the following equations

Finally, the flood susceptibility index (FSI) was measured by the following equation:

FSI = elevation*0.29 + rainfall*0.2 + LULC*0.15+ distance to river*0.1 + NDVI*0.08 + slope*0.06 + aspect*0.05 + curvature*0.04 + distance to road*0.03*

The FSI was generated by multiplying the weights for each factor and then totaling the results. It should be noted that each factor has been divided by five classes: 1-very low, 2-low, 3-moderate, 4-high and 5-very high. The process was done using ArcGIS Spatial Analyst's raster calculator. FSI was separated into five levels: very low, low, moderate, high and very high, using the natural break method.

4. RESULTS

4.1. Validation of flood susceptibility mapping

The receiver operating characteristic (ROC) was used to validate the AHP model. It was drawn by 1-specificity on the X-axis and sensitivity on the Y-axis (Janizadeh et al. 2019). The area under the receiver operating curve (AUC) represents the accuracy of the model. If the AUC value is 1, then the model is perfort (Choubin et al. 2019, Dodangeh et al. 2020). **Fig. 4** shows the ROC and AUC of the AHP model. In this study, 1964 flood and non-flood points were overlaid on the FSM in ArcGIS software to assess the accuracy of the FSM. The result indicates the model's acceptable precision (AUC = 0.95) for predicting flood susceptibility (**Fig. 4**).
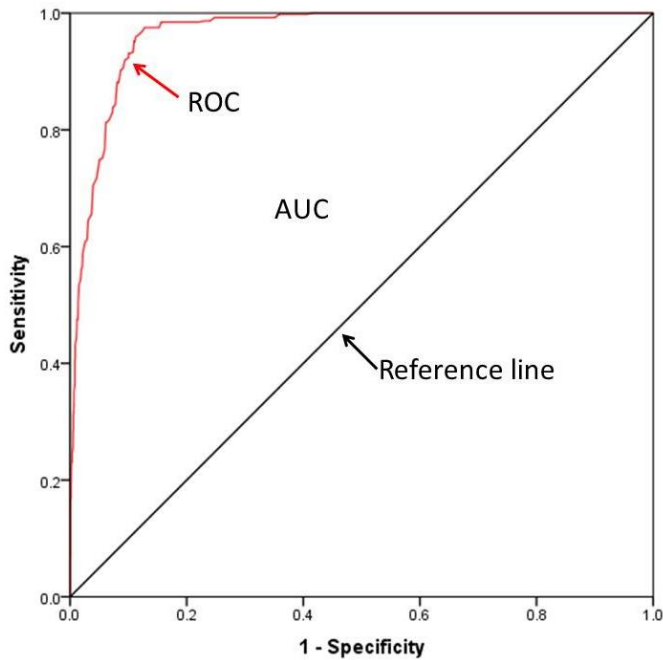
Fig. 4. ROC and AUC of the AHP model.

4.2. Flood susceptibility mapping

After the weighting calculation of all conditioning factors, the final FSM was produced using the AHP technique. The flood level was separated into five classes using the natural break method: very low, low, moderate, high, and very high (**Fig. 5**). The map was crossed with the land-use maps for 2005 and 2020 to understand the rate of each category of LULC corresponding to flood susceptibility levels (**Tab. 4**). Residential area in the high and very high flood susceptibility classes increased from 95 km² in 2005 to 251 km² in 2020. The area in the moderate class increased from 17.4 km² to 30.6 km² over the same period, with the low zone growing from 0.56 km² to 1.07 km². The agricultural area with very high flood susceptibility increased from 81.1 km² to 89.5 km².

On the other hand, in the high areas, these surfaces decreased from 164.2 km² to 150.3 km²; in the moderate areas, the fall was from 25.4 km² to 11.2 km²; and the low areas also saw a decline from 3.7 km² to 0.15 km². Regarding forested areas, the high and very high zone dropped from 638 km² to 610 km², the medium zone increased from 588 km² to 624 km², and surfaces in the low and very low zones increased 10 km².
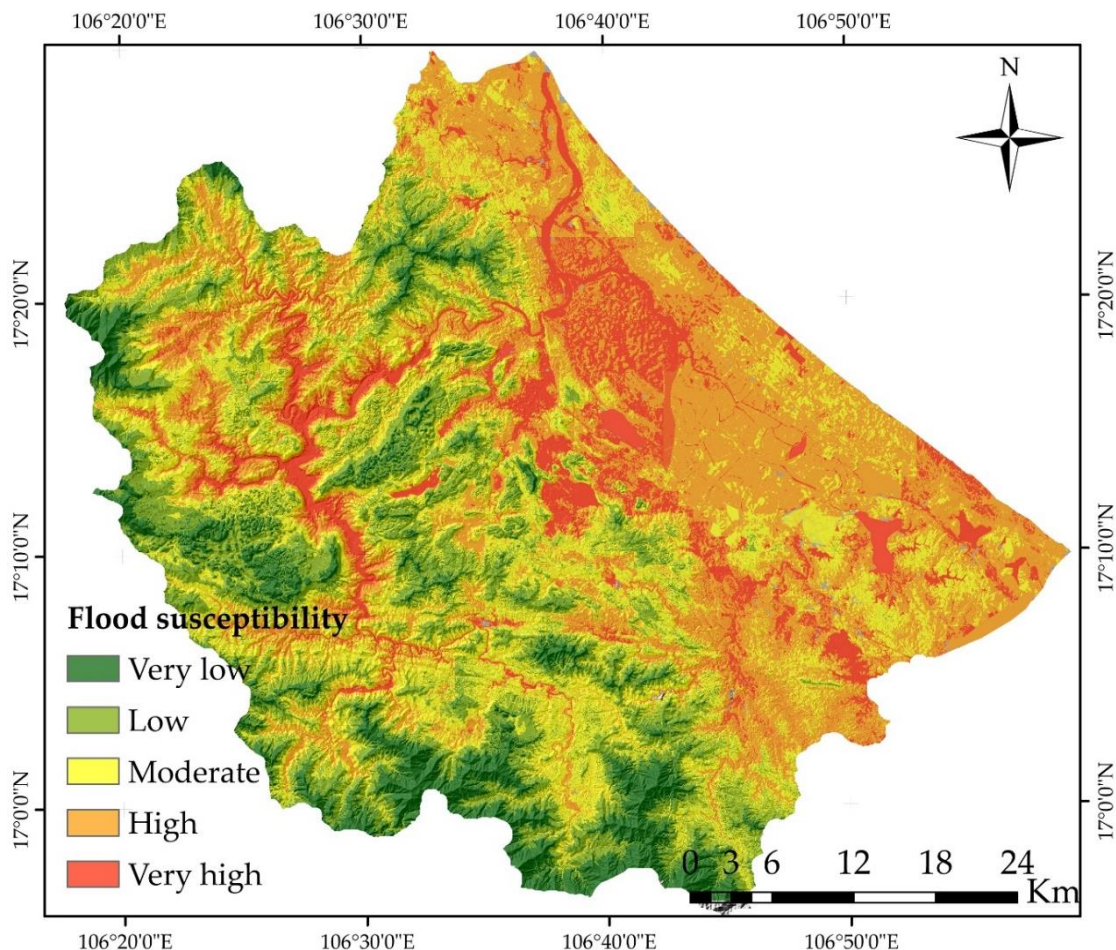
Fig. 5. Flood susceptibility in the Nhat Le–Kien Giang watershed.

Table 4.

LULC and their respective flood susceptibility classes.

	LULC	Flood susceptibility class				
		Very low (km²)	Low (km²)	Moderate (km²)	High (km²)	Very high (km²)
2020	**Agricultural**		0.15	11.2	150.3	89.5
	Residential	0.13	1.07	30.6	156.3	59
	Forest	227.5	496	623.9	532.9	77.9
2005	**Agricultural**	0	3.7	25.4	164.2	81.1
	Residential	0	0.56	17.4	77.1	18.1
	Forest	225.6	487.4	588	519.7	118.3

5. DISCUSSION

Flood is now the world's most dangerous natural hazard, both in terms of damage to humans and economies, and its impact is increasing as our climate changes (Khosravi et al. 2019). Although the number of studies on flood susceptibility has been increasing in Vietnam and around the world, there is still a lack of detailed research on flood susceptibility at the local level (Luu et al. 2020). Therefore, this study has developed a comprehensive approach to the construction of flood susceptibility maps, particularly in study area, using the AHP technique. The findings of this study can support planners in developing flood management and mitigation strategies.

Floods have significantly impacted the socio-economic conditions of the populations in the watershed and the wider Quang Binh province. A comprehensive approach to global and local flood management can reduce the negative effects of flood. Before a flood occurs, we can reduce risk and discuss mitigation activities in areas with high susceptibility to flood with the use of a highly accurate flood susceptibility map.

The residential area in the high and very high flood susceptibility zones increased. These areas are generally located in the littoral zone. This trend is typical of many regions around the world. Güneralp and Seto (2013) pointed out that around 40% of the world's cities will be in the high flood risk zone in 2030. Güneralp et al. (2015) anticipated that approximately 82% of the urban area of Southeast Asia would be located in areas with a high frequency of flood. Nguyen et al. (2018) reported that urbanization increased rapidly between 2003 and 2020 in the flood zone in Vietnam's Gianh River watershed. Abdelkarim et al. (2019) pointed out that urban expansion was increasingly rapid in Tabuk City, Saudi Arabia in the period 1975‑2018. Lin et al. (2020) calculated that the urban area would increase sharply between 2015 and 2050 in China's Pearl River Delta, especially in the flood zone. Our study provides further evidence of urban growth in the flood zone.

AHP is ostensibly a simple technique: it depends on the authors' experience in determining the weightings for each of the different criteria, which leads to some uncertainty. It is appropriate for local and regional flood susceptibility research; however, the problem for scientists is deciding the appropriate number of conditioning factors and the order of priority of these factors to best describe the flood phenomenon in the study area. In this study, ten conditioning factors were selected, which is similar to previous studies. Hammami et al. (2019) used eight conditioning factors ‑ namely land use, elevation, lithology, rainfall, drainage density, slope, soil, and groundwater level ‑ to assess flood susceptibility in Tunis. Vojtek and Vojteková (2019b) built a map in Slovakia using seven factors: river density, distance to river, elevation, slope, flow accumulation, curve number, and permeability level. Swain et al. (2020) used 22 factors, divided by physical, climatic, hydrological condition and human activity to assess the FM in Bihar, India. Souissi et al. (2020) constructed a map in an arid area of southeastern Tunisia using eight factors: drainage density, distance from drainage, elevation, slope, land use, rainfall, lithology, and groundwater. There is no universal guide to selecting conditioning factors for analyzing flood susceptibility. However, studies seem to show that it is best to use more than six factors to avoid single-factor dominance. In addition, the number of factors must also be modified based on the local topographical, climatic, hydrological, and anthropogenic conditions.

Key to correctly mapping the susceptible area is determining the importance of the conditioning variables. In this study, elevation and rainfall had the greatest influence on the probability of flood occurrence, because elevation represents the reaction to runoff and the capacity of water accumulation, and rainfall is the trigger for flood. The importance of the other factors (LULC, distance to river, NDVI, slope, aspect, curvature, and distance to road) were diminished. These results are consistent with previous studies (Santangelo et al. 2011, Souissi et al. 2020).

The flood susceptibility levels of the Nhat Le Kien Giang watershed obtained in this study is consistent with previous studies used machine learning to establish the flood susceptibility map. They showed that the eastern plain is sudden and has a high potential for the occurrence of flooding.

Faced with the effects of flood, people have three main strategies: withdrawal, resistance, or adaptation (Zaninetti et al. 2014). Withdrawal is only carried out when the actual or presumed effect of the flood passes a threshold of danger to life and/or irreversible damage to the territory in question. However, previous researchers have pointed out that planned personal withdrawal strategies are very expensive and rare. Resistance strategies, such as the construction of dykes and dams, are pursued to reduce the effects of flood. However, urban development and population growth in the flood zone increases the flood risk, and in many cases, the failure of dykes causes significant damage to human life, communities, and the local economy. In recent decades, adapting land use to encompass the potential flood hazard has been an appropriate strategy in many countries. In the Nhat Le‑Kien Giang River watershed, although the local government has developed strategies to address the flood risk issue through mitigation, preparedness, response, and recovery, these efforts do not influence land-use planning in the absence of strong planning rules.

Flood susceptibility map in this study provides useful information which can support policymakers and planners in developing strategies for flood management and mitigation at both national and regional levels. The flood susceptibility map can also provide the inhabitants with crucial precise information on the flood situation in the region.

Multi-criteria decision analysis methods in general, and AHP in particular, face limitations due to the subjectivity involved in the selection of the order of importance of the conditioning factors. In addition, this study has limitations related to the data. Several conditioning factors (climatic, hydrological and anthropic factors) were extracted by Landsat 8 with 30m resolution. While previous studies have pointed out that Sentinel 2A with 10m resolution can exhibit more climatic, hydrological and anthropogenic characteristics (Nguyen et al. 2022). In recent years, floods strongly influence urban growth and climate change. Therefore, studies of the effects of these elements on flooding are very necessary in the future, can support land planning decision makers.

6. CONCLUSIONS

FSM is an essential tool for planners working toward more sustainable territory. Nine conditioning factors, namely elevation, slope, aspect, curvature, NDVI, rainfall, distance to river, and distance to road were selected to build FSM.

The AHP technique was utilized to establish the weighting of each of the conditioning factors. They were prioritized in the following order: elevation, LULC, distance to river, NDVI, slope, aspect, curvature, and distance to road. The results show that the AHP technique was able to establish a FSM, with the value of AUC of 0.95. The residential area in the high and very high susceptibility zones increased from 95 km² in 2005 to 251 km² in 2020. In recent years, flood risk management strategies have risen up the list of priorities for decision-makers around the world, especially in the context of global warming. However, a number of the world's watersheds and at-risk regions have yet to be assessed for flood risk mitigation and management plans. The methodology used in this study can be applied to assess flood susceptibility in any region. Therefore, the results of this study provide essential information to support local authorities as they plan to diminish damage to both human life and property.

REFERENCES

Abdelkarim, A., Gaber, A., Youssef, D.A., Pradhan, B. (2019) Flood hazard assessment of the urban area of Tabuk City, Kingdom of Saudi Arabia by integrating spatial-based hydrologic and hydrodynamic modeling. *Sensors,* 19 (1024).

Abu el-magd, S.A., Ali, S.A., Pham, Q.B. (2021) Spatial modeling and susceptibility zonation of landslides using random forest, naïve bayes and K-nearest neighbor in a complicated terrain. *Earth Science Informatics,* 14, 1227-1243.

Apreda, C., D'ambrosio, V., Di Martino, F. (2018) A climate vulnerability and impact assessment model for complex urban systems. *Environmental Science & Policy,* 93, 11-26.

Bui, D.T., Ho, T.-C., Pradhan, B., Pham, B.-T., Nhu, V.-H., Revhaug, I. (2016) GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. *Environmental Earth Sciences,* 75, 1-22.

Bui, Q.-T., Nguyen, Q.-H., Nguyen, X.L., Pham, V.D., Nguyen, H.D., Pham, V.-M. (2020) Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping. *Journal of Hydrology,* 581, 124379.

Chen, W., Hong, H., Li, S., Shahabi, H., Wang, Y., Wang, X., Ahmad, B.B. (2019) Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *Journal of Hydrology,* 575, 864-873.

Chen, W., Li, Y., Xue, W., Shahabi, H., Li, S., Hong, H., Wang, X., Bian, H., Zhang, S., Pradhan, B., Ahmad, B.B. (2020) Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree, and random forest methods. *Science of The Total Environment,* 701, 134979.

Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F., Mosavi, A. (2019) An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Science of the Total Environment,* 651, 2087-2096.

Costache, R. (2019) Flood susceptibility assessment by using bivariate statistics and machine learning models-a useful tool for flood risk management. *Water Resources Management*, 33, 3239-3256.

Dano, U.L. (2020) Flash Flood Impact Assessment in Jeddah City: An Analytic Hierarchy Process Approach. *Hydrology*, 7(1), 10.

Das, S. (2020) Flood susceptibility mapping of the Western Ghat coastal belt using multi-source geospatial data and analytical hierarchy process (AHP). *Remote Sensing Applications: Society and Environment,* 20, 100379.

Dewan, A.M., Islam, M.M., Kumamoto, T., Nishigaki, M. (2007) Evaluating flood hazard for land-use planning in greater Dhaka of Bangladesh using remote sensing and GIS techniques. *Water resources management,* 21, 1601-1612.

Dodangeh, E., Choubin, B., Eigdir, A.N., Nabipour, N., Panahi, M., Shamshirband, S., Mosavi, A. (2020) Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. *Science of the Total Environment,* 705, 135983.

El-Batran, M., Aboulnaga, M. (2015) Climate Change Adaptation: An Overview on Challenges and Risks in Cities, Regions Affected. *Costs and Benefits of Adaptation, and Finance Mechanisms*, 725-764.

Falah, F., Rahmati, O., Rostami, M., Ahmadisharaf, E., Daliakopoulos, I.N., Pourghasemi, H.R. (2019) Artificial neural networks for flood susceptibility mapping in data-scarce urban areas. *Spatial modeling in GIS and R for Earth and Environmental Sciences*, 323-336.

Gudiyangada, N.T., Tavakkoli, P.S., Gholamnia, K., Ghorbanzadeh, O., Rahmati, O., Blaschke, T. (2020) Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. *Journal of Hydrology*, 590, 125275.

Ghosh, A., Kar, S.K. (2018) Application of analytical hierarchy process (AHP) for flood risk assessment: a case study in Malda district of West Bengal, India. *Natural Hazards,* 94: 349-368.

Güneralp, B., Güneralp, İ., Liu, Y. (2015) Changing global patterns of urban exposure to flood and drought hazards. *Global environmental change*, 31, 217-225.

Güneralp, B., Seto, K. (2013) Futures of global urban expansion: uncertainties and implications for biodiversity conservation. *Environmental Research Letters,* 8, 014025.

Hammami, S., Zouhri, L., Souissi, D., Souei, A., Zghibi, A., Marzougui, A., Dlala, M. (2019) Application of the GIS based multi-criteria decision analysis and analytical hierarchy process (AHP) in the flood susceptibility mapping (Tunisia). *Arabian Journal of Geosciences,* 12.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., Kanae, S. (2013) Global flood risk under climate change. *Nature climate change*, 3, 816-821.

Hong, H., Liu, J., Bui, D.T., Pradhan, B., Acharya, T.D., Pham, B.T., Zhu, A.-X., Chen, W., Ahmad, B.B. (2018) Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). *Catena*, 163, 399-413.

Hostache, R., Lai, X., Monnier, J., Puech, C. (2010) Assimilation of spatially distributed water levels into a shallow-599 water flood model. Part II: Use of a remote sensing image of Mosel River. *Journal of hydrology,* 390, 257-268.

Janizadeh, S., Avand, M., Jaafari, A., Phong, T.V., Bayat, M., Ahmadisharaf, E., Prakash, I., Pham, T.B., Lee, S. (2019) Prediction success of machine learning methods for flash flood susceptibility mapping in the Tafresh watershed, Iran. *Sustainability*, 11, 5426.

Kabenge, M., Elaru, J., Wang, H., Li, F. (2017) Characterizing flood hazard risk in data-scarce areas, using a remote sensing and GIS-based flood hazard index. *Natural hazards*, 89, 1369-1387.

Khoirunisa, N., Ku, C.-Y., Liu, C.-Y. (2021) A GIS-Based Artificial Neural Network Model for Flood Susceptibility Assessment. *International Journal of Environmental Research and Public Health*, 18, 1072.

Khosravi, K., Nohani, E., Maroufinia, E., Pourghasemi, H.R. (2016) A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. *Natural hazards,* 83, 947-987.

Khosravi, K., Shahabi, H., Pham, B.T., Adamowski, J., Shirzadi, A., Pradhan, B., Dou, J., Ly, H.-B., Grof, G., Ho, H.L., Hong, H., Chapi, K., Prakash, I. (2019) A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. *Journal of Hydrology*, 573, 311-323.

Lee, S., Kim, J.-C., Jung, H.-S., Lee, M.J., Lee, S. (2017) Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. *Geomatics, Natural Hazards and Risk*, 8, 1185-1203.

Li, G.-F., Xiang, X.-Y., Tong, Y.-Y., Wang, H.-M. (2013) Impact assessment of urbanization on flood risk in the Yangtze River Delta. *Stochastic Environmental Research and Risk Assessment,* 27, 1683‑1693.

Lin, W., Sun, Y., Nijhuis, S., Wang, Z. (2020) Scenario-based flood risk assessment for urbanizing deltas using future land-use simulation (FLUS): Guangzhou Metropolitan Area as a case study. *Science of the Total Environment,* 739, 139899.

Luu, C., Bui, Q.D., Costache, R., Nguyen, L.T., Nguyen, T.T., Tran, V.P., Le, V.H., Pham, T.B. (2021) Flood-prone area mapping using machine learning techniques: a case study of Quang Binh province, Vietnam. *Natural Hazards*, 1-23.

Luu, C., Tran, H.X., Pham, B.T., Al-Ansari, N., Tran, T.Q., Duong, N.Q., Dao, H.N., Nguyen, L.P., Nguyen, H.D., Ta, H.T., Le, V.H., Meding, J.V. (2020) Framework of spatial flood risk assessment for a case study in Quang Binh province, Vietnam. *Sustainability*, 12, 3058.

Mojaddadi, H., Pradhan, B., Nampak, H., Ahmad, N., Ghazali, A.H.B. (2017) Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. *Geomatics, Natural Hazards and Risk*, 8, 1080-1102.

Nachappa, T.G., Piralilou, S.T., Gholamnia, K., Ghorbanzadeh, O., Rahmati, O., Blaschke, T. (2020) Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. *Journal of Hydrology*, 125275.

Narsimlu, B., Gosain, A.K., Chahar, B.R., Singh, S.K., Srivastava, P.K. (2015) SWAT model calibration and uncertainty analysis for streamflow prediction in the Kunwari River Basin, India, using sequential uncertainty fitting. *Environmental Processes,* 2, 79-95.

Nguyen, H.D., Ardillier-Carras, F., Touchart, L. (2018) Les paysages de rizières et leur évolution récente dans le delta du fleuve Gianh. *Cybergeo: European Journal of Geography*, 876.

Nguyen, H.D., Nguyen, Q.-H., Du, Q.V.V., Nguyen, T.H.T., Nguyen, T.G., Bui, Q.-T. (2021) A novel combination of Deep Neural Network and Manta Ray Foraging Optimization for flood susceptibility mapping in Quang Ngai province, Vietnam. *Geocarto International*, 1-25.

Nguyen, H.D., Bui, Q.-H., Nguyen, Q.-H., Nguyen, T.G., Pham, L.T., Nguyen, X.L., Vu, P.L., Nguyen, T.H.T., Nguyen, A.T., Petrisor, A.I (2022) A novel hybrid approach to flood susceptibility assessment based on machine learning and land use change. Case study: a river watershed in Vietnam. *Hydrological Science Journal*, 67.

Nguyen, H.X., Nguyen, A.T., Ngo, A.T., Phan, V.T., Nguyen, T.D., Do, V.T., Dao, D.C., Dang, D.T., Nguyen, A.T., Nguyen, T.K., Luc, H. (2020) A Hybrid Approach Using GIS-Based Fuzzy AHP‑TOPSIS Assessing Flood Hazards along the South-Central Coast of Vietnam. *Applied Sciences,* 10, 7142.

Paule-Mercado, M., Lee, B., Memon, S., Umer, S., Salim, I., Lee, C.-H. (2017) Influence of land development on stormwater runoff from a mixed land use and land cover catchment. *Science of the Total Environment* 599, 2142-2155.

Petrisor, A.-I. (2016) Geographical Information Systems as Environmental, Landscape, and Urban Planning and Research Tools. *Romania as a Case Study*, 233-249.

Pham, B., Tran, P., Nguyen, H., Qi, C., Al-Ansari, N., Amini, A., Ho, S.L., Tran, T.T., Hoang, P.H.Y., Ly, H.-B., Prakash, I., Bui, T.D. (2020) A Comparative Study of Kernel Logistic Regression, Radial Basis Function Classifier, Multinomial Naïve Bayes, and Logistic Model Tree for Flash Flood Susceptibility Mapping. *Water,* 12, 1-21.

Prasad, P., Loveson, V.J., Das, B., Kotha, M. (2021) Novel ensemble machine learning models in flood susceptibility mapping. *Geocarto International*, 1-23.

Qiangsheng, H., Xiaorong, H. (2018) An Integrated Approach to Evaluate Urban Adaptive Capacity to Climate Change. *Sustainability,* 10.

Rajib, A., Liu, Z., Merwade, V., Tavakoly, A.A., Follum, M.L. (2020) Towards a large-scale locally relevant flood inundation modeling framework using SWAT and LISFLOOD-FP. *Journal of Hydrology*, 581, 124406.

Ren, J., Ren, B., Zhang, Q., Zheng, X. (2019) A novel hybrid extreme learning machine approach improved by K nearest neighbor method and fireworks algorithm for flood forecasting in medium and small watershed of loess region. *Water*, 11, 1848.

Santangelo, N., Santo, A., Crescenzo, G.D., Foscari, G., Liuzza, V., Sciarrotta, S., Scorpio, V. (2011) Flood susceptibility assessment in a highly urbanized alluvial fan: the case study of Sala Consilina (southern Italy). *Natural Hazards and Earth System Sciences*, 11, 2765-2780.

Souissi, D., Zouhri, L., Hammami, S., Msaddek, M.H., Zghibi, A., Dlala, M. (2020) GIS-based MCDM‑AHP modeling for flood susceptibility mapping of arid areas, southeastern Tunisia. *Geocarto International,* 35, 991-1017.

Swain, K.C., Singha, C., Nayak, L. (2020) Flood susceptibility mapping through the GIS-AHP technique using the cloud. *ISPRS International Journal of Geo-Information*, 9, 720.

Tehrany, M.S., Pradhan, B., Jebur, M.N. (2014) Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *Journal of hydrology*, 512, 332-343.

Towfiqul, I.A., Talukdar, S., Mahato, S., Kundu, S., Eibek, K.U., Pham, Q.B., Kuriqi, A., Nguyen, T.T.L. (2021) Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers,* 12, 101075.

UNISDR C. (2015) The human cost of natural disasters: A global perspective.

Vojtek, M., Vojteková, J. (2019a) Flood susceptibility mapping on a national scale in Slovakia using the analytical hierarchy process. *Water*, 11, 364.

Vojtek, M., Vojteková, J. (2019b) Flood Susceptibility Mapping on a National Scale in Slovakia Using the Analytical Hierarchy Process. *Water*, 11, 364.

Yariyan, P., Janizadeh, S., Tran, V.P., Nguyen, H.D., Costache, R., Le, V.H., Pham, T.B., Pradhan, B., Tiefenbacher, J.P. (2020) Improvement of best first decision trees using bagging and dagging ensembles for flood probability mapping. *Water Resources Management,* 34, 3037-3053.

Youssef, A.M., Pradhan, B., Hassan, A.M. (2011) Flash flood risk estimation along the St. Katherine road, southern Sinai, Egypt using GIS based morphometry and satellite imagery. *Environmental Earth Sciences,* 62, 611-623.

Zaninetti, J.-M., Ngo, A.-T., Grivel, S. (2014) The social construction of vulnerability to the risk of flooding in Vietnam. *Mappemonde*, 42:17p.

IDENTIFYING CLIMATE CHANGE VULNERABILITY BASED ON LAND COVER INDICATORS: A CASE STUDY IN SURABAYA, INDONESIA

Floriberta BINARTI[1]* (iD) **, Albertus Joko SANTOSO[2]** (iD)

ABSTRACT :
Surabaya is facing the threat of climate change indicated by the increase in air and surface temperature. The city has a risk of sinking by 2050 if the global warming cannot be resolved. Several related studies established that the change in land cover and land use is accompanied by the increase in surface temperature, which will be addressed in this present study. Therefore, this study aimed to examine the impact of land use/cover on the increase in air/surface temperature and investigate the contribution of land cover indicators to climate change in Surabaya as the basis of the identification of spatial climate change vulnerability. Data were collected from satellite images obtained over a long period and processed with GIS-based software to obtain an overview of changes. Mined long-term historical climate data and satellite imagery were processed into a land surface temperature map (LST), describing the tendency of climate change. The satellite imagery data from 2013 to 2021 was used to have an overview of land use and land cover changes based on indicators of built-up area (NDBI), surface imperviousness (NDISI), vegetation (NDVI), and water (NDWI). The contribution of each indicator to the surface temperature was analyzed using the multivariate regression method. The significant contribution of the land cover indicators to the surface temperature as the results means that NDBI, NDISI, NDVI, and NDWI can be used as indicators in climate change vulnerability assessment. The sequential contribution weights to the surface temperature are NDISI, NDWI, NDVI, and NDBI. Furthermore, the climate change vulnerability map of Surabaya City was developed based on the contribution weights, which the pattern of vulnerability levels corresponds to the pattern of water index values.

Key-words: Climate change vulnerability, Land cover indicator, Satellite imagery, Surface temperature.

1. INTRODUCTION

The occurrence of climate change is marked by an increase in global surface temperature of 0.86ºC from 2006 to 2015, accompanied by frequent hot airwaves and an increase in the frequency and intensity of rainfall (IPCC, 2019). The 2015 Paris Climate Agreement requires countries to limit global warming to 1.5ºC by 2050 due to its increased risks to health, livelihoods, food security, and water supplies (IPCC, 2021). The Meteorology, Climatology, and Geophysics Agency results from 1960 to 2021 showed an increase in air temperature in Indonesia from 0.8 to 1.4 º C (BMKG, 2022). Aside from the temperature rise, global warming is also indicated by the increase in sea level, whereby Jakarta, Surabaya, and other capital cities in the northern coast of Java are predicted to sink in 2050, as stated in Climate Central (2022).

According to the International Panel on Climate Change, the adverse impact risk of climate change depends on the hazard, exposure, and vulnerability (Allen et al., 2018). In the context of climate change, vulnerability is the degree to which a system is amenable to the hostile impacts of climate change, including climate variability and extremes. (IPCC, 2007; Gumel, 2022). A system is considered vulnerable when it is exposed and shows sensitivity to climatic changes with low adaptive capacity. Hence, assessment of vulnerability to climate change is very important to identify hot spots

---

[1] *Department of Architecture, Universitas Atma Jaya Yogyakarta, Jl. Babarsari 44 Sleman 55281, Indonesia, floriberta.binarti@uajy.ac.id. Corresponding author**

[2] *Department of Informatics, Universitas Atma Jaya Yogyakarta, Jl. Babarsari 44 Sleman 55281, Indonesia, joko.santoso@uajy.ac.id*

of climate change that require urgent attention to lessen the climate change impacts for sustainable development (Schneiderbauer et al., 2020).

Identifying the right indicators for each component of the vulnerability function is an important step in the development process of climate change vulnerability assessment methodologies. Many studies identified the assessment indicators from the perspective of the disaster impacts of climate change (Ludena et al., 2015; Nguyen et al., 2016; Schneiderbauer et al., 2020; UN-Habitat, 2019), while only a few did from the potential causes (Delaney et al., 2021). United Nations Habitat manual, for example, developed the current disaster risk profile (risk index) to map the most vulnerable areas. The indicators for measuring vulnerability are exposure to hazards, ecosystem, socio-economic, and infrastructure components (UN-Habitat, 2019). The objective of disaster impact-based assessment indicators is to define the degree of adaptive capacity and further propose an effective adaptation method to climate change. To achieve the goal of sustainable development, however, climate-resilient trajectories should combine adaptation and mitigation (Denton et al., 2014). The goal of mitigation is to alleviate the exposure and reduce the vulnerability to climate change (IPCC, 2007). Potential causes-based assessment indicators aim to formulate effective mitigation strategies. For mapping the climate change vulnerability of aquatic-riparian ecosystems, Delaney et al. (2021) chose some exposure indicators classified into hydrology, precipitation, and temperature. Whether disaster impact- or potential cause-based indicators, the assessment method should be guided by five principles – i.e., simplicity, measurability and availability of data, inclusiveness, comprehensiveness, and spatial relevance (Ludena et al., 2015).

To identify climate change vulnerability at a local scale (a city), an understanding of climate change issues at a national scale could help to figure out the major causes. Schneiderbauer et al. (2020) showed the urgency of a vulnerability assessment at a national scale before identifying local-specific drivers of vulnerability and appropriate adaptation measures. According to the 2020 ND-GAIN Country Index, Indonesia is identified as vulnerable to climate change impacts with the rank of 97nd out of 181 countries. Indonesia is exposed highly to flooding (ranked 17th most at risk from this natural hazard) and tropical cyclones (ranked 23rd). High maximum temperatures with an average monthly maximum of around 30.6°C occur regularly (World Bank & ADB, 2021). It was reported that a significant proportion of the greenhouse gas (GHG) emissions in Indonesia emanate from land use change, which represented 52.3% of total GHG emissions in Indonesia between 2000 to 2017 (World Bank & ADB, 2021). At the global scale, a review of 116 studies on the role of land use and land cover change in climate change vulnerability assessments conducted by (Santos et al., 2021) mentioned that 34% of the studies assumed climate change and land use/cover change would act addictively, while 66% allowed for interactive effects. Moreover, land use/cover is an environmental factor, which is one of the four internal vulnerability factors mentioned by the United Nations (2014) (Esperón-Rodríguez et al., 2016), that relevant to reduce the disaster.

Two questions arise regarding indicators of climate change vulnerability for Surabaya – i.e., (1) Does land use/cover determine climate change indicated by the temperature increase? (2) Can land cover indicators be used to assess climate change vulnerability? (3) How much does the contribution of each land cover index to climate change? Therefore, this study was conducted to answer the questions based on the following objectives:

(1) To examine the impact of land use/cover in Surabaya on climate change;
(2) To investigate the contribution of each land cover indicator to climate change in Surabaya City;
(3) To develop a climate change vulnerability map of Surabaya to support decision-makers in prioritizing the implementation of climate change mitigation strategies.

2. STUDY AREA

The study area is situated in Surabaya City, East Java, Indonesia (**Fig. 1**). As the second largest city in Indonesia, the population in Surabaya increased yearly. According to the Central Bureau of Statistics of East Java record, at the beginning of 2019, the registered population was 3,095,026 people, increasing by about 0.52% from 2010 to 2017 (BPS Jawa Timur, 2018). JICA stated that the land in Surabaya is used for agriculture (1.6%), non-irrigation agriculture (0.03%), settlements (39%),

ponds (11.4%), water (2.2%), industry (8.5%), green areas and recreation (8.5%), public facilities (7.1%), forests, mangrove forests and swamps (5.8%), and commercial areas (4.6%); while the remaining (8.3%) are vacant (Savitri et al., 2019). A study using remote sensing data from 1996 to 2015 revealed the land use change from mangrove forest into pond land. The pond area changed from 673 ha in 1996 to 3,139.66 ha in 2015 (Savitri et al., 2019).

There were 23 events consisting of tornadoes, drought, forest and land fires, and floods in 2008-2017 reported by Surabaya City Disaster Risk Assessment Document for 2019-2023. During this period, Surabaya experienced whirlwinds (7 events), drought (2 events), forest and land fires (1 event), and flooding (Kurniawati et al., 2020). Kurniati & Nitivattananon (2016) stated that the growth trend towards East Surabaya raises urban heat island (UHI) – the city experiences much warmer temperature than the surrounding rural areas - in Surabaya with a temperature difference of ±1.4℃, while Jatayu & Susetyo (2017) mentioned that a 6.62℃ rise between 2001 and 2016. According to Syafitri et al. (2020), the UHI deviation of ±1.59℃ in East Surabaya is correlated with changes in land use, building density, and sky view factor. However, Pratiwi & Jaelani (2020) illustrated the fluctuation of the surface temperature from the processing of satellite imagery showing that the average surface temperature in 2002, 2014, and 2019 was 29.09℃, 26.89℃, and 27.13℃, respectively.

Fig. 1. Location of Surabaya City on the map of South East Asia (source: Google Map).

3. DATA AND METHODS

3.1. Data sources

The rise of global air/surface temperature change indicates global warming that causes climate change. Historical air temperature data can be obtained from weather stations. However, the data has low accuracy in reflecting regional temperatures and lack appropriate spatial resolution. Surface temperature data sourced from satellite imagery can provide precise information on spatially ground surface temperatures (Firozjaei et al., 2018). Moreover, remote sensing provides end-users with a consistent, repeatable, and relatively inexpensive methodology for land surface temperature, land cover, and vegetation mapping (Azevedo et al., 2016). However, the weakness of satellite images compared to climate data from weather stations is their availability on specific dates with an accuracy level determined by the cloud cover. In many previous studies, GIS-based software was used to classify satellite images to determine the effect of changes and indicators in land use/cover on the climate surface temperature of cities (Majeed et al., 2021; Maleki et al., 2020).

This study used air and surface temperature data from mining climate data from the nearest weather stations and processing Landsat 8 OLI/TIRS imagery over 10 years obtained from Earth Explorer (https://earthexplorer.usgs.gov/). Satellite imagery is also used to produce maps of land cover indicators. Landsat can detect surface temperatures with a higher resolution than other images with thermal channels because it is equipped with 60 m and 100 m resolution infrared channels (Fawzi, 2017). Furthermore, satellite images of Surabaya City were taken at the latitude of 7.32° S and longitude of 112.71° E with a 20 km radius using images with less than 10% cloud cover.

The land use land cover change (LULCC) analysis provides an overview of the locations and areas experiencing land use/cover changes by comparing two satellite images acquired in different years. To determine the changes in land use/cover, this study downloaded the MODIS Land Cover v.6 satellite image for Surabaya from 2013 to 2021 in Earth Explorer. However, the images can be replaced by Landsat 8 OLI/TIRS B4, B5, and B6 images. By using a qGIS feature, LULC analysis comparing two images in specific years can show changes in surface temperature. **Table 1** presents the data sources used in this study.

Table 1.

Data sources.

The kind of data	Date/year	Source/method	Usage
Historical air temperature data	2013-2021	Data mining	To observe the trend of air temperature increases
Landsat 8 OLI/TIRS	2013-2021	USGS - Earth Explorer https://earthexplorer.usgs.gov/	To observe the land cover changes
MODIS Land Cover v.6	2013-2021	USGS - Earth Explorer https://earthexplorer.usgs.gov/	To analyze the land use land cover change

Only less than ten images of the 33 satellite images of Surabaya City obtained from 2013 to 2021 can produce maps without cloud cover. Four dates of retrieved satellite imagery were presented to illustrate the change in temperature surface and land cover indicators. **Table 2** describes the cloud cover, time, and sun elevation data collected for satellite imagery. It shows that the four satellite images were taken at almost the same hour with little sun elevation.

Table 2.

Date, cloud cover, hour, and sun elevation of the satellite images.

Date	Cloud cover	Hour	Sun Elevation
13/08/2013	6.82	02:37:46	54.13
20/09/2015	3.41	02:35:43	62.35
28/09/2018	0.83	02:35:28	63.85
01/10/2019	9.39	02:36:07	64.43

3.2. Classification of satellite image data

We used land cover indicators to observe the land cover change. Yang and Chen (2016) stated that green plot ratio, built-up ratio, impervious surface fraction, and surface admittance are land cover indicators that can be obtained from satellite image processing. The extraction of Landsat 8 OLI/TIRS imagery consisting of nine spectral bands can be used to classify the land cover indicators. Green plot ratios can be generated from normalized difference vegetation index (NDVI) maps. Built-up ratios can be described by normalized difference built-up index (NDBI) maps. Meanwhile, the impervious surface fraction was generated from normalized difference impervious surface index (NDISI) maps. Normalized difference water index (NDWI) was used to describe the surface admittance. Values of LST and the land cover indicators were calculated using the equation explained in the following sub-sub sections. Meanwhile, the analysis of LULCC utilized semi-automatic classification in qGIS, an open source GIS software (Majeed et al., 2021). Furthermore, maps of the LST and land cover indices were classified using the K-Nearest Neighbors (KNN) method. KNN is an ML algorithm widely used to classify land cover indicators from satellite imagery (Binarti et al., 2021; Ge et al., 2020; Jiang et al., 2020).

3.2.1. Land Surface Temperature (LST)

LST maps (in ^{o}C) were obtained by rasterizing satellite images and clustering using Eq. (1)–(5) (Jeevalakshmi et al., 2017).

$$LST = \frac{BT}{\left\{1+\lambda*\left(\frac{BT}{p}\right)*ln(LSE)\right\}} \quad [^{o}C] \tag{1}$$

$$BT = \frac{K2}{ln\left(\frac{K1}{L}+1\right)} - 272.15 \quad [^{o}C] \tag{2}$$

$$LSE = (0.004 * Pv) + 0.986 \tag{3}$$

$$Pv = \left\{(NDVI - NDVImin)\Big/_{NDVImax - NDVImin}\right\}^{2} \tag{4}$$

$$NDVI = \frac{(NIR-RED)}{(NIR+RED)} \tag{5}$$

where:

BT	-brightness temperature;
λ	-average wavelength of band 10;
p	-the multiplying of Planck's constant by Boltzmann constant and velocity of light (14380);
K1 and K2	-band-specific thermal conversion constant;
L	-spectral radiance;
LSE	-land surface emissivity;
Pv	-the proportion of vegetation;
NDVI	-normalized difference vegetation index;
NIR	-near-infrared band;
RED	-red band

3.2.2. Normalized Difference Built-up Index (NDBI)

NDBI in satellite imagery is the ratio map of the built-up to the total area. Xu et al. (2018) developed Eq. (6) using NIR and short-wavelength infrared band (SWIR) as variables.

$$NDBI = \frac{(SWIR-NIR)}{(SWIR+NIR)} \tag{6}$$

3.2.3. Normalized Difference Imperviousness Surface Index (NDISI)

Xu (2010) developed Eq. (7) to determine the relationship between the thermal band (TIR), the near-infrared band (NIR), the middle infrared band (MIR), and the visible band (VIS) on impervious surfaces based on satellite imagery.

$$NDISI = \frac{\left\{TIR - \frac{1}{3}*(VIS + NIR + MIR)\right\}}{\left\{TIR + \frac{1}{3}*(VIS + NIR + MIR)\right\}} \tag{7}$$

3.2.4. Normalized Difference Vegetation Index (NDVI)

Towers et al. (2019) tested NDVI's ability to estimate spatial variability and found that it is more accurate than other vegetation indices. The NDVI value is extracted using Eq. (5).

3.2.5. Normalized Difference Water Index (NDWI)

NDWI estimates water bodies' area, depth, and turbidity (Mcfeeters, 2007). Meanwhile, Eq. (8) is a modification of the basic formula sensitive to changes in the water content in the leaves (Gao, 1996).

$$NDWI = \frac{(NIR - SWIR)}{(NIR + SWIR)} \tag{8}$$

NDWI=(NIR-SWIR)/(NIR+SWIR)

Where SWIR is the short wavelength infrared band (1.24 μ m) value.

3.3. Accuracy assessment

Since surface temperature is the most important variable in identifying the spatial climate change vulnerability index based on land cover indicators, we compared the surface temperature of three points measured using infra-red camera FLIR i5 to the surface temperature of the same locations at the same time resulted from satellite image processing. The infra-red camera FLIR i5 has a thermal sensitivity of less than 0.1°C and the capability to measure temperatures ranging from -20°C to 250°C with a resolution of 0.1°C. Since there is only one package of satellite image data, which is free of cloud and available during climate monitoring (from August 12 to September 12, 2022), we compared the measured surface temperatures to the ones on the LST map on August 12, 2022, 02:48:18.

3.4. Data analysis

Multivariate regression analysis with a 95% confidence level was used to analyze the effect or contribution of each land cover index - i.e., NDBI, NDISI, NDVI, and NDWI - as independent variables and the surface temperature (LST) as the dependent variable. The reliability and validity of the variables were determined by the coefficient (R-squared) and the p-value. R-squared describes the percentage of the response variable. When the p-value is less than the significance level (0.05), the sample data provide sufficient evidence to reject the null hypothesis for the entire population (Andrade, 2019). Furthermore, Beta Coefficient was used to describe the effect strength of each independent variable on the dependent.

4. RESULTS

4.1. Accuracy of land surface temperature calculations.

Table 3 describes the comparison between measured surface temperatures and surface temperatures in the LST map on August 12, 2022, 02:48:18. The surface temperature differences of points 1, 2, and 3 are only 0.54%, 0.03%, and 0.15% respectively.

Results of the accuracy assessment of LST calculations.

Point	Geographical location	Surface temperature		
		Measured (^{o}C)	Satellite image (^{o}C)	Difference (^{o}C)
1	7°46'01"S & 110°24'24"E	33.6	33.42	0.18
2	7°46'52"S & 110°25'24"E	33.0	32.99	0.01
3	7°46'56"S & 110°23'30"E	33.3	33.25	0.05

4.2. Temperature increases and the land cover profile

The temperature increases are described by the air temperature trend (**Fig. 2**) and LST maps (**Fig. 3**). **Fig. 2** illustrates the trend of increasing the maximum, average, and minimum air temperature obtained from mining nine-year climate data for Surabaya from 2013 to 2021. Despite the insignificant increases, the average air temperature rose from less than 27°C to slightly more than 27°C. The air temperature trends show that the greatest increase sequentially was experienced by the maximum air temperatures. The maximum air temperature in 2013 was less than 31°C and in 2021 reached almost 32°C.
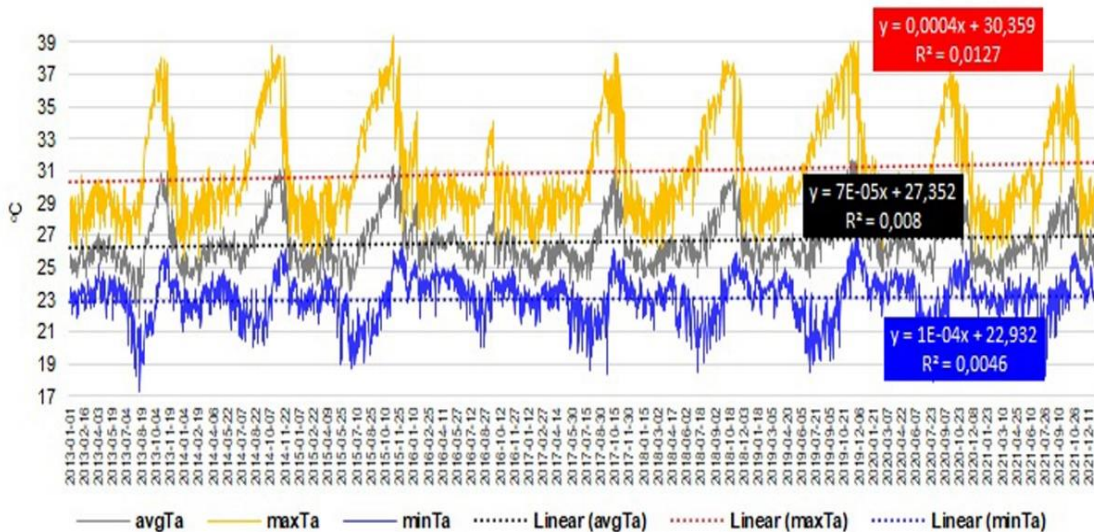
Fig. 2. Trend of temperature increases in Surabaya in 2013-2021.

The LST map was also used to observe climate change on four representative dates from 2013 to 2019, namely 08/13/2013, 20/09/2015, 28/09/2018, and 01/10/2019. **Fig. 3** illustrates the minimum and average increase in surface temperature in 2015 from 29.19^{o}C to 29.41^{o}C and 32.79^{o}C to 32.86^{o}C, respectively. In 2018 the minimum and average surface temperatures decreased to 29.28^{o}C and 32.41^{o}C, respectively. However, in 2019, the minimum and average surface temperatures were 29.71^{o}C and 33.05^{o}C. Although the maximum surface temperature increased in 2018 and 2019, the maximum surface temperature in 2019 at 36.50^{o}C was lower than in 2013 (i.e., 36.75^{o}C). The illustration of the surface temperature fluctuation confirms the results of the study by Pratiwi & Jaelani (2020). Maps of land cover indicators in **Fig. 3** show remarkable changes in the water index (NDWI). From 2013 to 2019, the maximum, average, and minimum water index experienced a gradual decrease from -0.43 to -1.07, 0.06 to -0.11, and 0.77 to 0.51, respectively. A significant increase appears in the maximum value of the built-up index (NDBI) – i.e., from 0.77 in 2013 to 0.84 in 2019. The vegetation index (NDVI) decreased in the minimum and maximum values. However, the average water index increased from 0.15 in 2013 to 0.25 in 2019. A significant increase in the impervious surface index (NDISI) only occurred in the maximum value from 0.71 in 2013 to 0.80 in 2019. The average value increased insignificantly from 0.62 in 2013 to 0.63 in 2019.
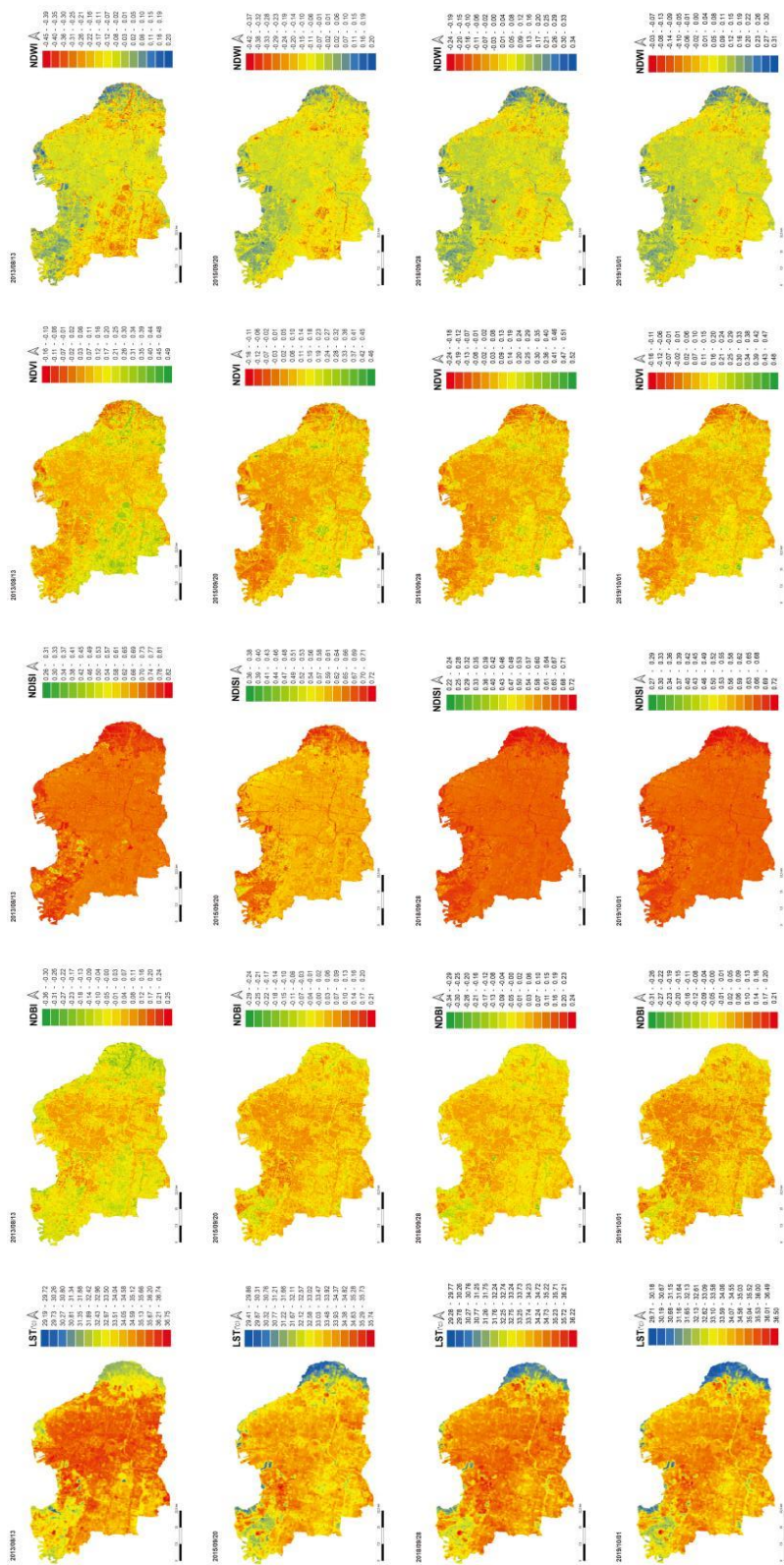
Fig. 3. Map of LST, NDBI, NDISI, NDVI, NDWI Surabaya City on 13/08/2013, 20/09/2015, 28/09/2018, and 01/10/2019

4.3. Land use land cover change (LULCC)

Fig. 4 presents two comparisons between two land cover/use images in 2013 (the duration beginning) and 2015 (the first surface temperature increase) and between two land cover/use images in 2015 and 2019 (the last surface temperature increase). The map of LULCC in 2013 and 2015 shows that the built area dominates Surabaya by 41.3%. Conversely, the map of LULCC in 2015 and 2019 was dominated by changes in agricultural land to the built-up areas by 32.6%. However, the percentage of the total built-up area between 2015 and 2019 was 64%, which is the cumulative of the built-up area without change (26.6%), change of water body (3.3%), green area (1.5%), and agricultural land. The percentage of built-up area in 2019 was much larger than in 2015 (43.3%). The most significant land use change between 2013 and 2015 was from green areas to agricultural land (8.6%), followed by changes from built-up areas to agricultural land (8.2%).
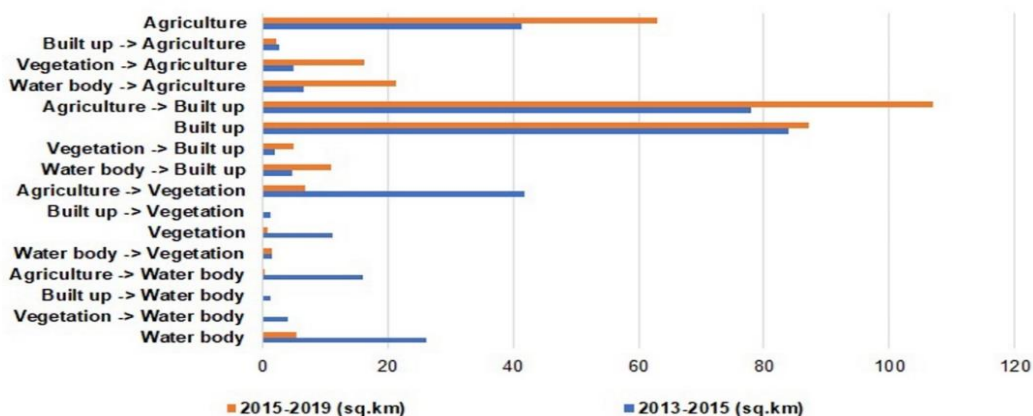
Fig. 4. The land use/land cover change histogram of Surabaya City.

Changes from water body to other functions, which covered 10% of the total area in 2019 and 16% of the total area including the reduction of the water body area from 2015 to 2019, appeared dominantly in **Fig. 5**. The land use changes are detected in the north, northwest, and southeast areas of Surabaya. These maps also depict the change/reduction of vegetation and agriculture to the built-up area found in the southwest and southeast of the study area. This change/reduction consumed 37% of the total area. The increase in a built-up area is only 1% of the total area. Positive land use change (from built-up into vegetation area) only occurred in 2013-2015, which is only 0.44% of the total area. Although some areas in 2015-2019 experienced a change from water bodies into agriculture/vegetation areas prominently as shown in **Fig. 5 (right)**, however, the areas displayed in the 01/10/2019 NDWI map in **Fig. 3** still have high water index values.
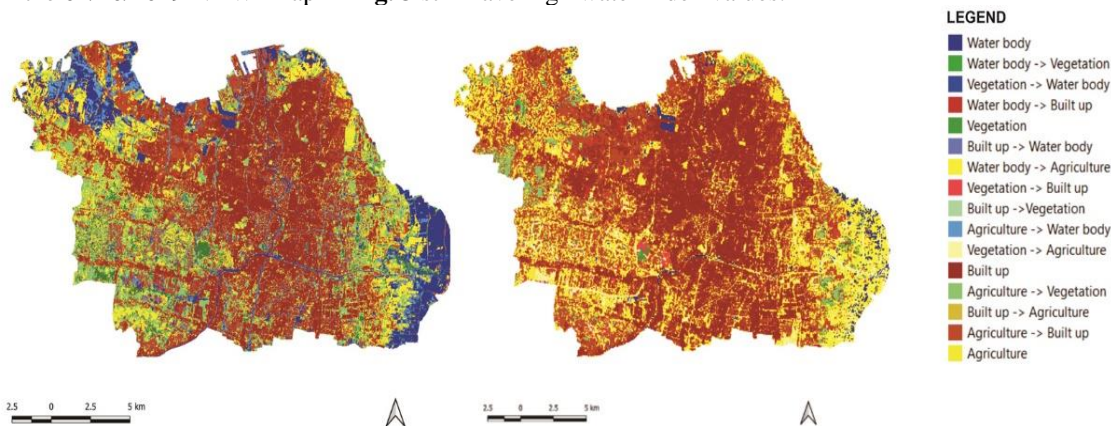
Fig. 5. The land use/land cover change map of Surabaya City: 2013 – 2015 (left) and 2015 – 2019 (right).

4.4. Contribution of each land cover index

The multivariate regression analysis of the contribution of each land cover indicator to land surface temperature was conducted in the latest satellite image data aiming to determine high-level vulnerability areas for the current implementation of climate change mitigations. The results of multivariate regression of five satellite image dates show that satellite imagery of October 1, 2019, has the highest R-squared, as shown in **Table 4**. The R-squared of 0.85 is considered high, which explains that the four land cover indicators on October 1, 2019, affected 85% of surface temperature. Therefore, the contribution weights of four land cover indices were used to develop a climate change vulnerability map. Less than 0.05 of the p-value for all variables indicates that the four land cover indicators determine the surface temperature. Based on the Beta Coefficient of each land cover index, it can be concluded that NDISI has the largest contribution, followed by NDWI and NDVI. This order of contributions also applies to the other four satellite images of the dates.

Table 4.

Coefficient of determination (R-squared) and P-value.

Date	R-squared	β coefficient				P-value			
		NDBI	NDISI	NDVI	NDWI	NDBI	NDISI	NDVI	NDWI
June 26, 2018	0.78	0.19	14.93	-1.40	-1.51	0.00	0.00	0.00	0.00
June 11, 2019	0.81	0.00	17.12	-1.57	-3.26	0.00	0.00	0.00	0.00
July 29, 2019	0.76	0.76	10.48	-3.04	-3.38	0.00	0.00	0.00	0.00
October 01, 2019	0.85	1.13	9.83	-5.33	-5.69	0.00	0.00	0.00	0.00

4.5. Climate change vulnerability map

Based on the contribution weights from the multivariate regression analysis of satellite imagery data obtained on October 1, 2019, we developed the climate change vulnerability (see **Fig. 6**) using an equation derived from the contribution weights – i.e., Vulnerability Level = (9.83*NDISI) - (5.69*NDWI) - (5.33*NDVI) + (0.85*NDBI) – in **Table 4**. The orange and red areas indicate a high-level of vulnerability to climate change. The red areas randomly scattered in the southwest, west, north, east, and southeast must be prioritized firstly for climate change mitigation.
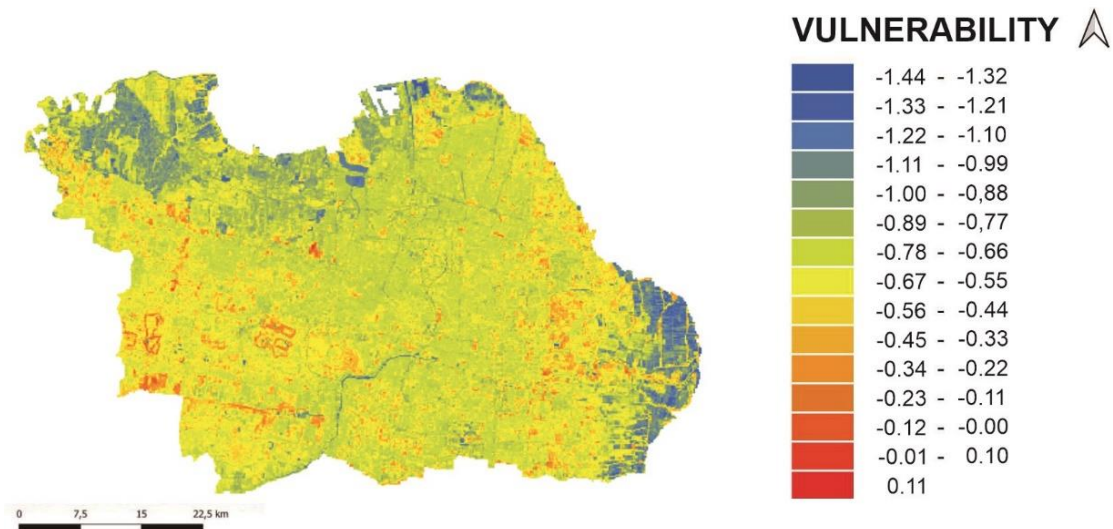
Fig. 6. The climate change vulnerability map of Surabaya City.

5. DISCUSSION

The increase in air and surface temperature during the last nine years indicates climate change in Surabaya. Approximately 1ºC of the increase in average air temperature in Surabaya already exceeds the recently observed global warming trends of 0.2ºC per decade (IPCC, 2019). Moreover, any increase in global temperature (+0.5ºC) was projected to affect human health negatively and heat-related morbidity and mortality (Hoegh-Guldberg et al., 2018).

The trend of air temperature increase in **Fig. 2** corresponds to the gradual reduction in minimum, average, and maximum values of the water index and the increase in the maximum value of the built-up index. The similar pattern of the water index, built-up index, and surface temperature in **Fig. 3** confirms the effect of water index reduction on the (surface) temperature. The highest surface temperature areas located in the center and southern Surabaya are related to the areas with the lowest water index and highest built-up index. An increase in surface temperature can be observed in the east area of Surabaya adjacent to ponds and swamps from 2013 to 2015. Changes from non-built areas (such as water bodies) to built-up areas in the eastern part also appeared on the NDBI and NDWI maps in 2013-2015. The change from a green to a built-up area emerged in the southwest area on the NDBI and NDVI maps. However, the change in surface temperature on the southwest side did not appear prominently on LST maps

Fig. 3 and **Fig. 5** demonstrated that changes in the maximum surface temperature are related to land use changes. The increase in maximum values of the built-up index was in line with the dominance of built-up areas without changes in **Fig. 4**. **Fig. 5 (left)** shows those changing from other land use in 2013-2015 located in the city's center, north and south sides of Surabaya. In 2015-2019, however, the development of built-up areas appeared towards the northwest and east shown in **Fig. 5 (right)**. The percentage of the total area that has changed from a built-up to agricultural land (0.65%), agricultural land to a green area (5.8%), built-up area to a green area (5.2%), and agricultural area to a water body (4.5%). This is in line with the decrease in maximum surface temperature in 2013 and 2015 as shown in **Fig. 4**. Surabaya's urban park development program is one of the causes of this positive change in land use and cover. Setiawati et al. (2021) reported that from 1995 to 2016 the bare land in the southwestern and central part of Surabaya was converted into green parks. Furthermore, the increase in minimum, average, and maximum surface temperatures in 2015 and 2019 is in line with the percentage of land use change into built-up areas (64%) supported by changes in water bodies and green areas to agricultural land at 6.5% and 4.9%, respectively. The change in water bodies into agricultural land obviously can be observed in eastern Surabaya.

Although the impervious surface index contributed most to land surface temperature, the NDISI and LST maps did not reveal similar patterns. Maps of NDWI and NDBI that own similar patterns with LST maps became the second and fourth contributors. However, when the contribution weights were applied to the vulnerability map, the pattern of vulnerability level corresponded to the pattern of the water index values in NDWI maps. Areas with high water index values are the least vulnerable, and vice versa. The study on the relationship between water index and surface temperature on various land surfaces in India (Guha & Govil, 2021) also describes the strong correlation between water index and surface temperature, especially during the post-monsoon season.

Since the correlation between surface temperature and land cover indices could change following the season (Guha & Govil, 2021), future studies on the seasonal impact of land cover indices on the surface temperature would corroborate the contribution weights and develop a more reliable climate change vulnerability map. To use satellite images acquired in every season, more images with higher cloud cover should be included. Consequently, cloud removal in the preprocessing step must be done to lower the risk of loss of information leading to a spatiotemporal discontinuity that degrades the quality and usefulness of satellite images (Hasan et al., 2022).

6. CONCLUSIONS

As one of the cities on the northern coast of Java has a risk of sinking by 2050, Surabaya needs an urgent formulation of climate change adaptation and mitigations. Moreover, an increase in air temperature, as shown by nine-year historical climate data, confirms the occurrence of global warming as an indication of climate change. Since land use/cover change became the significant cause of climate change in Indonesia, land cover indicators can be used as the base for identifying climate change vulnerability. Some findings regarding the trend of air temperature increase, land surface temperature (LST), and four land cover indicators – i.e., built-up index (NDBI), impervious surface index (NDISI), vegetation index (NDVI), and water index (NDWI) – describe the specific characteristic of land use/cover in Surabaya and the impact of land cover on the increase in air and surface temperature. The trend of air temperature increases is in line with the values of the water and built-up index, which are presented in the NDWI and NDBI maps, respectively.

The LULCC maps illustrate the change of water areas mainly to agriculture and built-up areas. However, the multivariate regression analysis results explained the significant contribution of land cover indicators to the surface temperature. Sequentially, land cover indicators that contribute to the surface temperature of Surabaya are the index of surface imperviousness (NDISI), water body/content (NDWI), vegetation (NDVI), and built-up (NDBI). The climate change vulnerability map of Surabaya City developed based on the contribution weight shows the pattern of vulnerability levels corresponding to the water index values in NDWI maps. Future studies on the projected climate change vulnerability map by 2050 will be needed to achieve global warming of less than 1.5°C.

R E F E R E N C E S

Allen, N.M. et al. (2018) Framing and Context. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response, In press. IPCC.

Andrade, C. (2019) The P Value and Statistical Significance: Misunderstandings, Explanations, Challenges, and Alternatives. *Indian J Psychol Med.*, 41(3), 210‑215.

Azevedo J.A., Chapman L. & Muller C. L. (2016) Quantifying the Daytime and Night-Time Urban Heat Island in Birmingham, UK: A Comparison of Satellite-Derived Land Surface Temperature and High Resolution Air Temperature Observations. *Remote Sensing*, 8, 153.A.

Binarti, F., Triyadi, S., Koerniawan, M.D., Pranowo, P. & Matzarakis, A. (2021) Climate characteristics and the adaptation level to formulate mitigation strategies for a climate-resilient archaeological park. Urban Climate, 36, 100811.

BMKG (Badan Meteorologi Klimatologi dan Geofisika) (2022) Ekstrem Perubahan Iklim. BMKG. URL: https://www.bmkg.go.id/iklim/?p=ekstrem-perubahan-iklim.

BPS (Badan Pusat Statistik) Provinsi Jawa Timur (2018) Jumlah Penduduk dan Laju Pertumbuhan Penduduk Menurut Kabupaten/Kota di Provinsi Jawa Timur 2010, 2016 dan 2017. BPS Jawa Timur.

Climate Central (2014) Coastal Risk Screening Tool: Land Projected to be Below Annual Flood Level in 2050. URL: https://coastal.climatecentral.org/map/12/123.9254/10.3001/?theme=sea_level_rise&map_type=coastal_dem_comparison&contiguous=true&elevation_model=coastal_dem&forecast_year=2050&pathway=rcp45&percentile=p50&return_level=return_level_1&slr_model=kopp_2014.

Delaney, J.T, Bouska, K.L., Eash, J.D., Heglund, P.J., Allstadt, A.J. (2021) Mapping climate change vulnerability of aquatic-riparian ecosystems using decision-relevant indicators. *Ecological Indicators*, 125, 107581.

Denton, F., Abeysinghe, A.C., Burton, I., Gao, Q., Lemos, M.C., Masui, T., O'Brien, K.L. & Warner, K. (2014) Climate-resilient pathways: adaptation, mitigation, and sustainable development. In: Field, C.B., Barros, V.R., Dokken, D.J., Mach, K.J., Mastrandrea, M.D., Bilir, T.E., White, L.L (Eds.), *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, 1101‑1131.

Esperón-Rodríguez, M., Bonifacio-Bautista, M., Barradas, V.L. (2016) Socio-economic vulnerability to climate change in the central mountainous region of eastern Mexico. *Ambio*, Mar 45(2), 146-60.

Fawzi, N.I. (2017) Mengukur urban heat island menggunakan penginderaan jauh, kasus di Kota Yogyakarta. *Majalah Ilmiah Globe*, 19, (2), 195.

Firozjaei, M.K., Kiavarz, M., Alavipanah, S.K., Lakes, T. & Qureshi, S. (2018) Monitoring and forecasting heat island intensity through multi-temporal image analysis and cellular automata-Markov chain modelling: a case of Babol City, Iran. *Ecological Indicators*, 91, 155‑170.

Gao, B. (1996) NDWI A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water From Space. *Remote Sensing Environment*, 58, 257‑266.

Ge, G., Shi, Z., Zhu, Y., Yang, X. & Hao, Y. (2020) Land use/cover classification in an arid desert-oasis mosaic landscape of China using remote sensed imagery: Performance assessment of four machine learning algorithms. *Global Ecology and Conservation*, 22, e00971.

Guha, S. & Govil, H. (2021) Relationship between land surface temperature and normalized difference water index on various land surfaces: A seasonal analysis. *International Journal of Engineering and Geosciences*, 6(3), 165-173.

Gumel, D.Y. (2022) Assessing Climate Change Vulnerability: A Conceptual and Theoretical Review. *Journal of Sustainability and Environmental Management (JOSEM)*, 1(1), 22 ‑ 31.

Hasan, C., Horne, R., Mauw, S. & Mizera, A. (2022) Cloud removal from satellite imagery using multispectral edge-filtered conditional generative adversarial networks. *International Journal of Remote Sensing*, 43(5), 1881-1893.

Hoegh-Guldberg, O., Jacob, D., Taylor, M., Bindi, M., Brown, S., Camilloni, I., Diedhiou, A., Djalante, R., Ebi, K.L., Engelbrecht, F., Guiot, J., Hijioka, Y., Mehrotra, S., Payne, A., Seneviratne, S.I., Thomas, A., Warren, R. & Zhou, G. (2018) Impacts of 1.5°C Global Warming on Natural and Human Systems. In: Global Warming of 1.5° C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. Cambridge University Press, Cambridge, UK and New York, NY, USA, 175-312.

IPCC (Intergovernmental Panel on Climate Change) (2007) Climate change 2007: Synthesis report. In: *Contribution of Working Groups I, II and III to the Fourth Assessment Report of Environmental Panel on Climate Change, IPCC*. Switzerland, IPCC, Geneva.

IPCC (2019) Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. IPCC Working Group I Technical Support Unit.

IPCC (2021) Projected Climate Change, Potential Impacts and Associated Risks. URL: https://www.ipcc.ch/sr15/resources/headline-statements/.

Jatayu, A. & Susetyo, C. (2017) , Analisis Perubahan Temperatur Permukaan Wilayah Surabaya Timur Tahun 2001-2016 Menggunakan Citra LANDSAT. *Jurnal Teknik ITS*, 6(2), C429-C433.

Jeevalakshmi, D., Reddy, S.N. & Manikiam, B. (2017) Land Surface Temperature Retrieval from LANDSAT data using Emissivity Estimation. *International Journal of Applied Engineering Research*, 12(20), 9679-9687.

Jiang, F., Smith, A.R., Kutia, M., Wang, G., Liu, H. & Sun, H. (2020) A modified kNN method for mapping the leaf area index in arid and semi-arid areas of China. *Remote Sensing*, 12, 1‑24.

Kurniati, A.C. & Nitivattananon, V. (2016) Factors influencing urban heat island in Surabaya, Indonesia. *Sustainable Cities and Society*, 27, 99-105.

Kurniawati, U.F., Susetyo, C. & Setyasa, P.T. (2020) Institutional assesment through climate and disaster resilience initiative in Surabaya. *IOP Conference Series: Earth Environmental Science*, 562, 012025.

Ludena, C.E., Yoon, S.W., Sánchez-Aragón, L., Miller, S. & Yu, B-K. (2015) *Vulnerability Indicators of Adaptation to Climate Change and Policy Implications for Investment Projects*. Inter-American Development Bank, Technical Note No. 858, Washington DC.

Majeed, M., Tariq, A., Anwar, M.M., Khan, A.M., Arshad, F., Mumtaz, F., Farhan, M., Zhang, L., Zafar, A., Aziz, M., Abbasi, S., Rahman, G., Hussain, S., Waheed, M., Fatima, K. & Shaukat, S. (2021) Monitoring of Land Use‑Land Cover Change and Potential Causal Factors of Climate Change in Jhelum District, Punjab, Pakistan, through GIS and Multi-Temporal Satellite Data. *Land*, 10, 1026.

Maleki, M., Van Genderen, J.L., Tavakkoli-Sabour, S.M., Saleh, S.S. & Babaee, E. (2020) Land use/cover change in Dinevar rural area of West Iran during 2000-2018 and its prediction for 2024 and 2030. *Geographia Technica*, 15 (2), 93-105.

Mcfeeters, S.K., 2007. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17, 1425‑1432.

Nguyen, T.T.X., Bonetti, J., Rogers, K. & Woodroffe, C.D. (2016). Indicator-based assessment of climate-change impacts on coasts: a review of concepts, methodological approaches and vulnerability indices. *Ocean and Coastal Management*, 123, 18-43.

Pratiwi, A.Y. & Jaelani, L.M. (2020) Analisis Perubahan Distribusi Urban Heat Island (UHI) di Kota Surabaya Menggunakan Citra Satelit Landsat Multitemporal. *Jurnal Teknik ITS*, 9(2), C48-C55.

Pratiwi, A.Y. & Jaelani, L.M. (2020) Analisis Perubahan Distribusi Urban Heat Island (UHI) di Kota Surabaya Menggunakan Citra Satelit Landsat Multitemporal. *Jurnal Teknik ITS*, 9(2), C48-C54.

Safitri D.A., Bespalova L.A., Wijayanti F. (2019) Changes in land use in Eastern Surabaya, Indonesia, and their impact on coastal society and aquaculture. *R-economy*, 5(4), 198‑207.

Santos, M.J., Smith, A.B., Dekker, S.C., Eppinga, M.B., Leita˜o, P.J., Moreno-Mateos, D., Morueta-Holme, N. & Ruggeri, M. (2021) The role of land use and land cover change in climate change vulnerability assessments of biodiversity: a systematic review. *Landscape Ecology*, 36, 3367‑3382.

Schneiderbauer, S., Baunach, D., Pedoth, L., Renner, K., Fritzsche, K., Bollin, C., Pregnolato, M., Zebisch, M., Liersch, S., López, M.R.R. & Ruzima, S. (2020) Spatial-Explicit Climate Change Vulnerability Assessments Based on Impact Chains. Findings from a Case Study in Burundi. *Sustainability*, 12, 6354.

Setiawati, M.D., Jarzebski, M.P., Gomez-Garcia, M. & Fukushi, K. (2021) Accelerating Urban Heating Under Land-Cover and Climate Change Scenarios in Indonesia: Application of the Universal Thermal Climate Index. *Frontier in Built Environment*, 7, 622382.

Syafitri, R.A.W.D., Pamungkas, A. & Santoso, E.B. (2020) The Impact of Urban Configuration to the Urban Heat Island in East Surabaya. *IPTEK Journal of Proceeding Series*, 6, 470-477.

Towers, P.C., Strever, A. & Poblete-echeverr, C. (2019) Comparison of Vegetation Indices for Leaf Area Index Estimation in Vertical Shoot Positioned Vine Canopies with and without Grenbiule Hail-Protection Netting. *Remote Sensing*, 11, 1073.

UN Habitat (2019) Climate Change Vulnerability Assessment Manual. United Nations (UN) Habitat, URL: https://unhabitat.org/climate-change-vulnerability-assessment-manual

World Bank and ADB (Asian Development Bank) (2021) Climate Risk Country Profile: Indonesia. World Bank and ADB Report, Washington.

Xu, H. (2010) Analysis of Impervious Surface and its Impact on Urban Heat Environment using the Normalized Difference Impervious Surface Index (NDISI). *Photogrammetric Engineering and Remote Sensing*, 76, 557‑565.

Xu, R., Liu, J. & Xu, J. (2018) Extraction of high-precision urban impervious surfaces from sentinel-2 multispectral imagery via modified linear spectral mixture analysis. *Sensors* (Switzerland) 18, 1‑15.

Yang, F. & Chen, L. (2016) Developing a thermal atlas for climate- responsive urban design based on empirical modeling and urban morphological analysis. *Energy and Building*, 111, 120‑130.

USING DIGITAL TOOLS FOR MONITORING AND ANALYSING SPATIAL VARIATIONS OF POPULATION DISTRIBUTION IN THE CITY OF AL-MADINAH AL-MUNAWARAH, KINGDOM OF SAUDI ARABIA, 2004-2020

Mohamed Ahmed Aly HASSANIEN[1,2] 🆔

ABSTRACT:
This article combines demographic and spatial data to investigate and analyse spatial variations in population distribution in Al-Madinah Al-Munawarah in the Kingdom of Saudi Arabia. Drawing on multiple data sources, it applies a straightforward rationalisation-based methodology based on combining demographic indicators with spatial data using geographic information systems and remote sensing. The findings show that Al-Madinah's population was estimated at 1,500,000 people in 2020; the Lorenz curve and the Gini ratio reveal that the population is unevenly distributed geographically since 95% of the population lives in 18% of the city's total area. The central area is densely populated: about half of the population lives within a 5-kilometre radius. Location quotient values reflect variations between concentrations of Saudis and non-Saudis. Due to the expansion of the city's built-up area, net population density fell from 46 to 32 p/ha between 2004 and 2020; Saudis are more mobile than non-Saudis as their mean centres recently shifted 1,337 and 665 meters south, respectively. The study concluded that Al-Madinah's residents, especially Saudis, are gradually moving from the central business district to the periphery.

Key-words: *GIS, Remote Sensing, City, Net population density, Gini ratio, Mean centre of the population, Standard distance, Location quotient, Al-Madinah Al-Munawarah.*

1. INTRODUCTION

Urban areas are expanding rapidly, with many cities becoming increasingly compact due to population growth (Tungnung & Anand, 2017). Cities are growing in both population and footprint at double previous rates (Angel et al., 2011). These considerable geographical changes, especially the dramatic global increase in urban populations, are affecting natural and man-made systems. Worsening overcrowding, housing shortages, and insufficient infrastructure all require effective city planning and management (Herold et al., 2005): managing the repercussions of urban expansion is a contemporary global challenge. Furthermore, future urban growth and the concomitant appropriation of natural resources has significant consequences for sustainable development plans (United Nations Population Division, 2019).

Geographic inequality is a primary driver of development, suggesting that governments often fail to equitably distribute economic resources (World Bank, 2009). Moreover, many countries face an increasing severity of socio-economic inequalities between or within urban areas due to the interrelationship between spatial population concentration and economic development (Portnov & Pearlmutter, 1999). Population and economic activities, therefore, tend to be concentrated in the central business districts (CBD) of many cities, while the peripheries are characterised by diminishing populations and slower economic development.

The Kingdom of Saudi Arabia (KSA) has undergone rapid urbanisation since the 1950s. The proportion of the population living in urban areas of the KSA increased from 21% in 1950 to 84% by 2018; the urban population is projected to reach 86% by 2030 and 90.4% by 2050 (Alahmadi & Atkinson, 2019; United Nations Population Division, 2019). Therefore, to examine this issue in the context of the KSA, this paper takes Al-Madinah Al-Munwarah, the capital of Al-Madinah Province

---

[1] *Cairo University, Faculty of Arts, Department of Geography, Giza, Egypt: mohamedwf33@cu.edu.eg*

[2] *Taibah University, College of Arts and Humanities, Department of Social Sciences, Al-Madinah Al-Munawarah, Saudi Arabia: mhassanien@taibahu.edu.sa*

and a place of great religious significance to Muslims as it houses the Prophet's Holy Mosque in its centre, as a case study. The Prophet's Mosque is the city's centrepiece and has contributed several times to changes in Al-Madinah's morphology.

The KSA conducted four reliable nationwide censuses in 1974, 1992, 2004, and 2010; a fifth census was intended to take place in 2020, with the planning phase starting on January 10, 2018. However, on March 13, 2020, GAS announced a suspension of the enumeration phase due to the COVID-19 pandemic. Al-Madinah is the fourth most populous city in the KSA. According to the Saudi census conducted by the General Authority for Statistics (GAS), between 1974–2010, Al-Madinah's population increased from 198,000 to 1,100,000. The Ministry of Municipal and Rural Affairs (MMRA) recently estimated the population of Al-Madinah at 1,400,000 and projected an expansion to 2,000,000 by 2030 (MMRA, 2019a). Moreover, the city has experienced high population growth in recent decades (above 3%). Importantly, most of Al-Madinah's inhabitants inhabit the 5,000-hectare area surrounding the Prophet's Mosque at the city's centre. This area is cramped, especially during peak visiting seasons, as around 90% of those performing Hajj and Umrah visit the Prophet's Mosque after their journey to Makkah (Islam's holiest city) (MMRA, 2019a). In 2019, Al-Madinah received almost all of those performing Hajj and Umrah: 2,489,406 visitors during Hajj (GAS, 2019a), and 11,700,368 during Umrah (GAS, 2019b). Furthermore, the morphology of the city, especially its CBD, has changed due to the numerous enlargements to the Prophet's Mosque that could contribute to alleviating the population density in this area.

Accurate mapping of population distribution is essential for policymaking, urban planning (Mossoux et al., 2018), and identifying any dimensions of imbalance between a particular population and the area it occupies. Although spatio-temporal data on population and its determinants are important for understanding and responding to population problems, such long-term data are often difficult to obtain (Wang & Chen, 2016). New tools, however, such as geographic information systems (GIS), remote sensing (RS), and geospatial tools can produce accurate maps that allow data retrieval for customised geographic areas and are not limited to administrative units like censuses (Long et al., 2001; Stewart, 2004; Yang et al., 2009). Combining RS data with GIS technology for population studies is crucial in order to maximise the respective strengths of each approach in analysing certain population aspects (Yagoub, 2006; Benomar et al., 2006; Yang et al., 2009; Weerakoon, 2017).

Al-Madinah's high population density in its central area has led to many problems: overcrowding, traffic congestion, poor infrastructure, and old or dilapidated housing stock. The government has sought to moderate the population density in the CBD and redistribute the population to the periphery. Therefore, to manage the population-related issues effectively, it is necessary to monitor and analyse the changes in the population distribution in Al-Madinah through accurate mapping.

The use of GIS and RS techniques to analyse and model environmental issues has received considerable attention in the critical literature across recent decades. Such techniques have furthered the study of population-related issues such as urban sprawl, urban land use change or land cover, and urban planning (Ibrahim & Sarvestani, 2009; Tungnung & Anand, 2017). Furthermore, these approaches provide effective tools for spatialising population census data; integrating geodemography with GIS capabilities allows researchers to examine and query the spatial components of population data (Yang et al., 2009). The use of GIS in demographic analysis ranges from creating maps to modelling the relationships between population variables (Baudot, 2001). Stewart et al. (2004) used GIS and RS to examine urbanisation in Greater Cairo based on the built-up area identified from census data, finding that population densities tend to decrease in the centre and increase at the periphery. Other scholars have used a dasymetric method to improve the representation of population distribution, obtaining a high-resolution dataset of population distribution, and analysing the accessibility to population centres (Linard et al., 2012; Amaral et al., 2012).

Estimating the number of residents at the neighbourhood level by combining fieldwork data with high-resolution RS offers a solution to systematically mapping population distribution in areas where reliable census data are not regularly available (Mossoux et al., 2018). Benomar et al. (2006) used GIS to estimate the population growth of the Libyan city of Zwarah and found that the population

growth rate was 7.2%, with the city's footprint increasing from 355 to 4,270 hectares between 1980–2000. Ramzi (2012) used RS to predict the future urban growth of the Egyptian city Ras Sudr and found that the city's expansion was due to population growth. Furthermore, Wang and Chen (2016) produced a dataset on China's population distribution and its driving factors using GIS and RS: China's population was densest in the southeastern regions, with availability of water being the primary factor influencing population distribution. Long et al. (2001) proposed a GIS-based measurement of urban population density for China, India, and the USA, providing insight into the ecological impacts of urbanisation. Marti-Henneberg et al. (2016) used digital tools to investigate population density trends in Spain, finding that the population was unevenly distributed across the nation. Yagoub (2006) examined the population distribution in Al Ain city in UAE using GIS and RS: the average difference between the population recorded in 2001 and the estimated population from images was 5% and that the population density decreased further from the CBD.

Al-Madinah's CBD has a relatively long history of being densely populated due to its religious significance and being home to many sectors serving residents and visitors (Ragab, 1979; Makki, 1989; Mohamed, 2011; Al-Mahdy, 2013; Mohamed et al., 2016; Tayan et al., 2017). Ragab (1979) revealed that Al-Madinah's densely populated CBD forms a semi-circle with a radius of 500-metres with the Prophet's Mosque at its centre. In the late twentieth century, Al-Madinah underwent a series of urban expansions due to external and internal migration (Mohamed, 2011). Abdou (2017) reported that between 2000–2016, the overall trend of Al-Madinah's residential mobility was towards its outer zones, motivated by increase in family size, change in marital, rental rates in the centre, or proximity to family, friends, and the workplace.

Previous literature has predominantly focused on the services and urban expansion of Al-Madinah, while overlooking the city's population distribution trends. This study addresses this gap, being (to the best of our knowledge) the first attempt to monitor and analyse the spatial variations in Al-Madinah's population distribution by integrating demographic and spatial data using GIS and RS. The built-up area was calculated for each district in the city, providing a much more comprehensive overview of its population distribution and possibilities for detailed geographical analysis (Marti-Henneberg et al., 2016). The methodology involved combining census data with spatial data to visualise Al-Madinah's population distribution. The 2004 and 2010 censuses by district were used to estimate the population in 2020 in order to provide the demographic data needed for monitoring Al-Madinah's recent population distribution. This study aims to accurately analyse various aspects of the geographical distribution of Al-Madinah's population in 2004 and 2010 in order to predict them in 2020. Additionally, it endeavoured to examine the evolution of Al-Madinah's mean centre of population, standard distance, and location quotient over the past two decades.

2. STUDY AREA

Al-Madinah lies on the western part of the Arabian Shield and extends from 24° 21' 00" to 24° 36' 00"N and 39° 36' 00" to 39° 24' 36"E. It is located on the northwestern side of Saudi Arabia and occupies an almost flat basin surrounded by lava plateaus and hills. The surrounding mountains are 800–1500 meters high, with Wairah being the highest, followed by Uhud (MMRA, 2019a; Matsah & Hossain, 1993). Five valleys pass through Al-Madinah: Al-Aqiq, Al-Aqoul, Bathan, Mahzoor, and Ranounaa. Administratively, Al-Madinah is divided into seven municipalities (see **Fig. 1**).

Al-Madinah has 107 districts and is served by both ring and radial roads; its ring roads include King Faisal Road (the innermost ring road), King Abdullah Road (the second ring road), and King Khalid Road (the third ring road). For the purpose of analysis, the city was divided into four geographic zones:

(i) Zone 1 contains the CBD, is encircled by the innermost ring road and covers 186.5 hectares. Zone 1 contains six districts, with El-Harm at its centre. It is extremely crowded, especially during visiting season, due to its religious significance and the associated concentration of commerce, accommodation, and service industries.

Fig. 1. Location of the study area (*source: Al-Madinah Regional Municipality (ARM), 2020*).

    (ii) Zone 2 forms a ring-shaped area extending from the innermost ring road to the second ring road. This zone covers 5,360 hectares. Zone 2 comprises 25 districts, forms the main residential area, and contains a significant concentration of commercial activities.

    (iii) Zone 3 is a ring-shaped area which stretches from the second ring road to the third ring road, covering 45,947 hectares. Zone 3 comprises 56 districts and is mainly residential. Much of Zone 3 has only recently become residential.

    (iv) Zone 4 comprises the periphery, extending from the third ring road to the city's outer administrative boundary, occupying 177,937 hectares. It includes 20 districts, most of which are uninhabited (see **Fig. 2**).
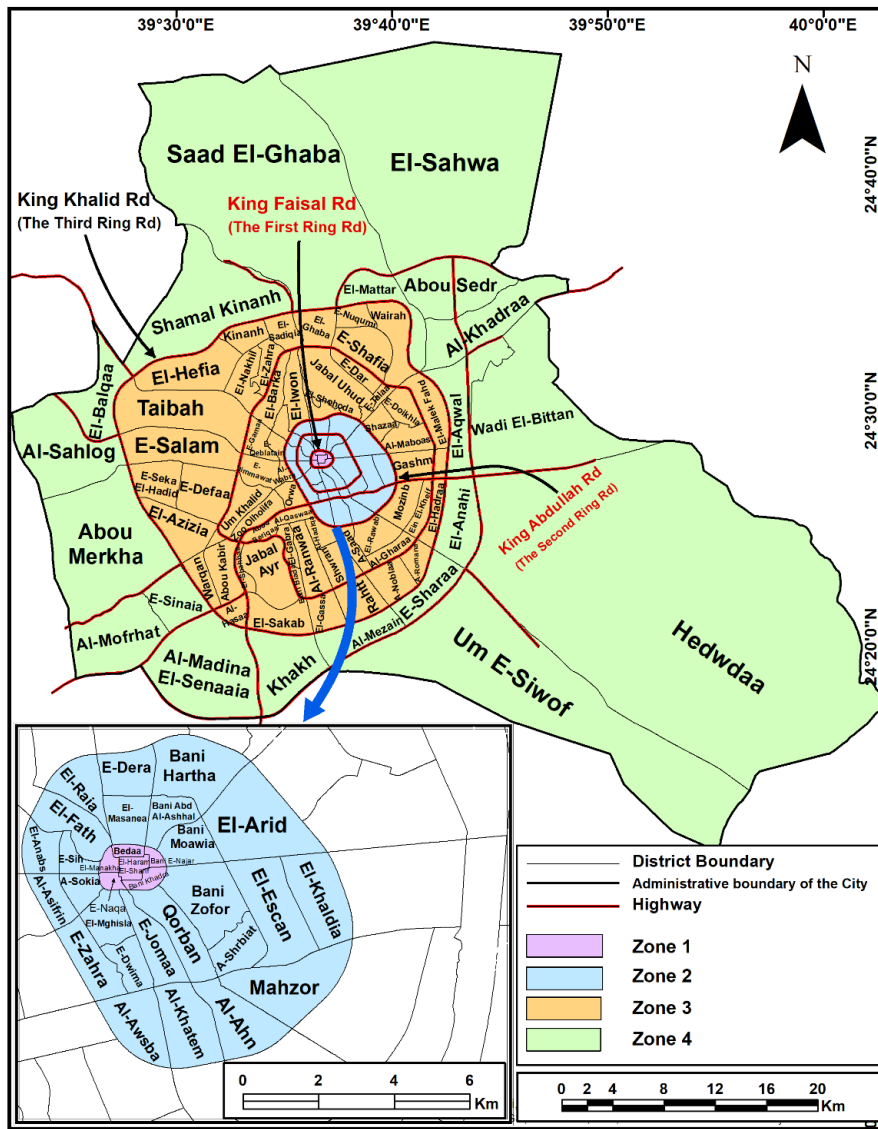
Fig. 2. Administrative map and the four geographic zones of Al-Madinah (*source: As Figure 1*).

3. METHODS AND MATERIALS

The dynamics of spatial population distribution have long attracted academic interest and combined various disciplinary approaches (Voss, 2007). The recent increase in interest in the population sciences has been driven by the increased availability of geospatial data and the emergence of GIScience tools for their analysis, including GIS, spatial analysis, and spatial statistics (Matthews et al., 2019; de Smith et al., 2018). Therefore, in order to analyse and visualise aspects of population distribution in Al-Madinah, the research methodology involved integrating demographic data with spatial data derived from GIS and RS based on district-level information from the 2004 and 2010 censuses. Spatio-temporal demographic data are essential for understanding and responding to

population issues. Unfortunately, it is often difficult to obtain such data in Al-Madinah at the district scale and over an extended period. Beyond the 2004 and 2010 censuses, no data are available on Al-Madinah's population by district. Therefore, the work relied heavily on both censuses, using them to estimate the city's population in 2020. The population estimation approach involved using an equation to determine Al-Madinah's likely population by district on February 2, 2020. Evaluation studies show that mathematical methods may be necessary for the absence of symptomatic data for the postcensal period (Bryan 2004). This approach assumes that population growth occurs in a linear manner and that the rate of change remains consistent: by calculating the growth rate during a baseline period, this same growth rate can be applied to a future, target period (George et al., 2004; Khoiyangbam & Gupta 2015). Initially, the population growth rate of a district was calculated in the baseline period (2004–2010) using the following equation:

$$r = \frac{P_t - P_0}{t * P_0}$$

where r is the population growth rate during the base period (2004–2010), P_t is the population in the launch year, P_0 is the population in the base year, and t is the number of years in the base period. Then, a projection, using this method, was computed using the following:

$$P_t = P_0 * (1 + rt)$$

where P_t is the population in the target year (2/2/2020), P_0 is the population in the launch year, t is the number of years in the projection horizon, and r is the calculated population growth rate for the base period (Khoiyangbam & Gupta, 2015).

Furthermore, the study adopted two assumptions to match Al-Madinah's population structure by nationality and account for recent transformations in socioeconomic and political conditions in the KSA. The first assumption concerned Saudis: it was assumed that the growth rate of each district, shifted from uninhabited by Saudis to inhabited between 2004–2010, was equal to the growth rate of the total number of Saudis in Al-Madinah (0.02091967), and remained stable during the target period (2010–2020). The growth rate of any of the remaining districts in the target period was assumed to be equal to that of the baseline period. The second assumption concerned non-Saudis: it was assumed that the growth rate of each district, changed from uninhabited by non-Saudis to inhabited between 2004–2010, was equal to the growth rate of total non-Saudis in Al-Madinah (0.06674985), and remained constant over the target period. For any of the remaining districts, the target period growth rate was assumed to be lower than the baseline period with a percentage ranging between 0.1–0.7. This expected reduction was grounded in the introduction of new costs attached to foreign employees and dependents: fees for foreigners sponsoring dependents in KSA were introduced in July 2017 and gradually increased over the following three years. These new financial demands forced thousands of migrants (accompanied by family members) to leave the KSA, especially in 2018 and 2019.

RS data were required to accurately analyse the population distribution of Al-Madinah. Due to its ability to provide multi-temporal analysis, RS offers a unique perspective on the evolution of population distribution in cities (Canada Centre for Remote Sensing, 2003; Ramzi, 2012). Combining land cover data with detailed settlement data can redistribute aggregate census counts to enhance the accuracy of gridded population data (Linard et al., 2011). Therefore, the research methodology relied on three RS images that were selected to identify Al-Madinah's built-up areas in 2004, 2010, and 2020, respectively. These three images share the same satellite system, path, row, and spatial and temporal resolution, but have different remote sensor names and identifiers. They were geometrically corrected by the provider on the Universal Transverse Mercator (UTM) projection system, the zone of 37N, and on the World Geodetic Survey of the year 1984 (WGS_84) spheroid and datum (see **Table 1**). These three datasets were carefully selected to match as closely as possible the dates of Saudi censuses and the author's estimate, as only two days separate the image obtained on 17 September 2004 and the third Saudi census on 15 September 2004; seventy-one days separate the

image obtained on 14 February 2010, and the fourth Saudi census on 17 April 2010; the image obtained on 2 February 2020 corresponds to the date the author estimates the postponed 2020 census would have taken place. Thus, these three datasets accurately reflect the built-up area of each district in Al-Madinah in 2004, 2010, and 2020, respectively.

Table 1.

Technical details of the Landsat dataset used in the study.

Sensor Name	Landsat-5-TM	Landsat-7-ETM+	Landsat 8
Acquisition Date	17-SEP-2004	14-FEB-2010	2-FEB-2020
Sensor Identifier	Thematic Mapper (TM)	the Enhanced Thematic Mapper Plus (ETM+)	Landsat Operational Land Imager (OLI) And Thermal Infrared Sensor (TIRS)
Path / Row	170/43	170/43	170/43
Spatial Resolution	30 * 30	30 * 30	30 * 30
Spectral Resolution	7	6	8
Pixel Depth	8 Bit	8 Bit	16 Bit
Cloud Cover (%)	0.00	0.00	0.00
Spatial Reference	WGS_1984_UTM_zone_37N	WGS_1984_UTM_zone_37N	WGS_1984_UTM_zone_37N

Source: The United States Geological Survey USGS's Earth Explorer web browser.

The satellite images were then processed, enhanced, and classified using ArcGIS. They were subjected to supervised classification, since classification of RS images is commonly used as a source of supplementary data. A land cover map is a valuable GIS layer and is essential for disaggregating demographic data (Linard et al., 2011). Supervised classification was conducted for three land-use classes: (i) built-up areas (housing, government offices, educational and health institutions, and recreational and accommodation facilities); (ii) agricultural land (palm groves and green spaces); (iii) uninhabited areas (mountains, hills, and stony plains) (see **Fig. 3**).

The supervised classification started with a training phase by assigning pixel samples belonging to the three aforementioned classes in each image, followed by a decision-making phase where the computer assigned each class label to other, similar pixels in the image. The supervised classification identified the expansion of the built-up area of each district in 2004, 2010, and 2020. Next, the built-up area of each district in the three images was digitized and linked to its population size using GIS tools. This method provides more accurate dot-density maps than methods based on administrative boundaries alone. Dot-density maps were used to represent the actual locations of populations in each district of Al-Madinah: they clearly show where specific data occur, especially in demonstrating how a population is distributed across a given area (Allen, 2022; Gomes, 2017; Newbold; 2010). To summarise, this approach aimed to estimate the population in the built-up area of each district, as GIS and census data were combined to link the population data with particular locations to create population density maps (Ramzi, 2012). Moreover, this method provided accurate maps of population distribution using dot density symbology to represent the population density within a district's built-up area and monitor the spatiotemporal variations of different aspects of population distribution.

While population census data are rarely available at a high spatial resolution (Mossoux et al., 2018), satellite imagery provides a good indication of population density (Rai Technology University, n.d), as a land cover map is perhaps the best single indicator of population density (Dobson et al., 2000). Therefore, the area of built-up area in each district in the three datasets was computed and linked to the population size using GIS tools to obtain the net residential population density.
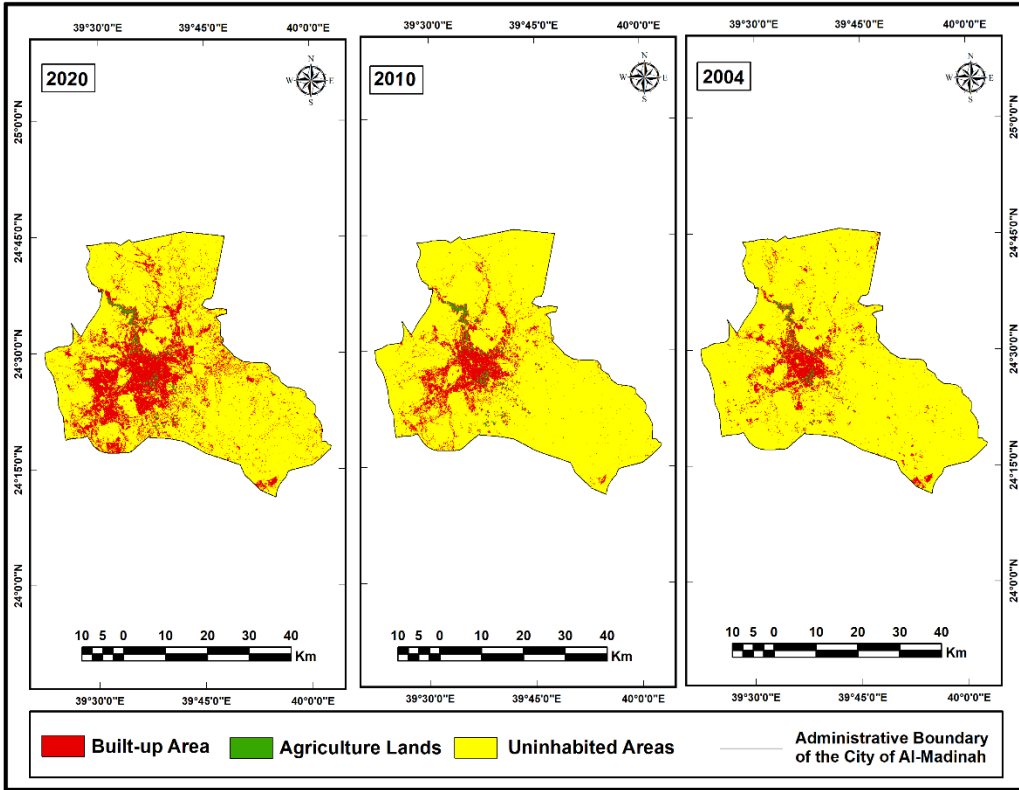
Fig. 3. Supervised classification results for the satellite images of Al-Madinah in 2004, 2010, and 2020, respectively (*source: USGS's Earth Explorer web browser; ARM, 2020*).

Briefly, the population density was calculated for built-up areas covered by Al-Madinah's census units, as calculating net residential densities gives a more accurate representation of population densities and helps planners to analyse urban land use (Long et al., 2001; Weerakoon, 2017).

Overall, the calculation of area of the built-up area of each district in Al-Madinah led to precise calculations for population distribution, net population density, mean centre of population (MCP), and standard distance (SD). Furthermore, the spatial statistics tools provided by the GIS programme were used to determine the MCP and SD for Al-Madinah in 2004, 2010, and 2020, respectively. However, calculations of other indicators of population distribution such as the concentration index (CI), the Lorenz curve, and the Gini Ratio (GR), were based on the administrative boundaries to show clearly the imbalance between the population and available space in Al-Madinah.

The methodology also involved calculating CI and location quotients (LQ). CI is the maximum of the set of values of the difference between the cumulative percentages of population and areas of the geographical units of a region. Algebraically, it is given as:

$$CI = \frac{1}{2} \sum |xi - yi|$$

where *xi* is the percentage of the population and *yi* is the percentage of the area.

LQ compares a population characteristic with the total population in terms of its regional distribution. Mathematically, LQ is written as:

$$LQ = \frac{\frac{xi}{pi}}{\frac{X}{P}}$$

where *xi* is the population of district i with characteristic x, pi is the total population of district i, X is the city's total population with characteristic x, and P is the city's total population.

4. RESULTS AND DISCUSSION

The four reliable Saudi censuses in 1974, 1992, 2004, and 2010 reveal a substantial increase in Al-Madinah's population over the census years: 198,000; 609,000; 920,000; and 1,100,000 people, respectively. The number and proportion of non-Saudis also increased: 43,000 (22%), 176,000 (29%), 280,000 (30%), and 385,000 people (35%), respectively (GAS, 1974-2010). The Madinah Urban Observatory estimated Al-Madinah's population at 1,301,322 people in 2018; the current study estimated the population to be 1,500,308 people in 2020, 20.3% of whom were non-Saudis. Accordingly, the number and percentage of non-Saudis decreased in 2020, reflecting the impact of the previously mentioned new legislation introducing new costs for foreigners.

Regarding the population distribution of Al-Madinah visualised using dot-density maps, **Figure 4** shows the demonstrably uneven distribution of Al-Madinah's populations. The most populous districts are located around the Prophet's Mosque in the CBD, with the population density gradually decreasing towards the periphery. The cities of Al-Madinah and Makkah are similar in urban sprawl as they each expand radially from a central holy site which can be seen to catalyse their population distribution. Looking at the zonal analysis, most residents were concentrated in the two innermost zones (zones 1 and 2), while many parts of zone 3 and most of zone 4 contained large areas which were uninhabited. Zone 1 was the most highly populated, particularly during visiting season; for Al-Mahdy (2013), it represents a 'town inside a city' with 300,000 inhabitants, 80% of whom are temporary residents. The current study reveals that almost 63%, 60%, and 50% of Al-Madinah's population were contained within a five-kilometre circle in 2004, 2010, and 2020, respectively. Al-Madinah's population concentration in the CBD dates back to the early 1970s. A quarter of Al-Madinah's population in 1971 (137,000 people) was estimated to be within a one-kilometre circle of the centre (Ragab 1979). The past two decades have seen spatio-temporal variations in the city's population distribution as inhabitants relocated widely in 2020 due to recent settlement projects. Two new directions of population spread emerged in 2020. The first was to the south, as people became concentrated in districts covered by zone 3, such as north of Mozinb, southwest of El-Gharaa, south of A-Saad, Shwran, Al-Ranwaa, and Al-Hadiqa. This population shift was due to housing plans approved by Al-Madinah for the period 2010–2028 which intended to relocate 80% of the city's population to its southern districts (Mohamed, 2011). The second population spread was to the west, as people started concentrating in the western districts of zone 3 in 2010, a trend that had increased by 2020. Between 1965–1989, Al-Madinah expanded to the west and southeast, beyond the second ring road, avoiding the mountains and expanding into the valleys (MMRA, 2019a). This was to accommodate a wealthy Saudi elite who chose to settle in villas outside the CBD in the 1970s (Ragab 1979) because they could afford private alternatives to public services (Stewart et al., 2004).

Over the past two decades, Al-Madinah has seen increased residential mobility from its two inner zones to the southern, southwestern, and western districts of zone 3. Broadly, Al-Madinah's residential mobility is from the centre to the periphery which is consistent with the most populous city in the KSA in the late 1980s (Riyadh) (Al-Kharif, 1994). Between 2000–2016, some zone 2 districts (such as El-Arid, E-Jomaa, and El-Raia) experienced residential mobility to zone 3 districts located to the south, the southwest, and the west (Abdou, 2017). Saudis and non-Saudis were (and remain) concentrated in the central area within the second ring road in more densely populated unplanned settlements (MMRA, 2019a).
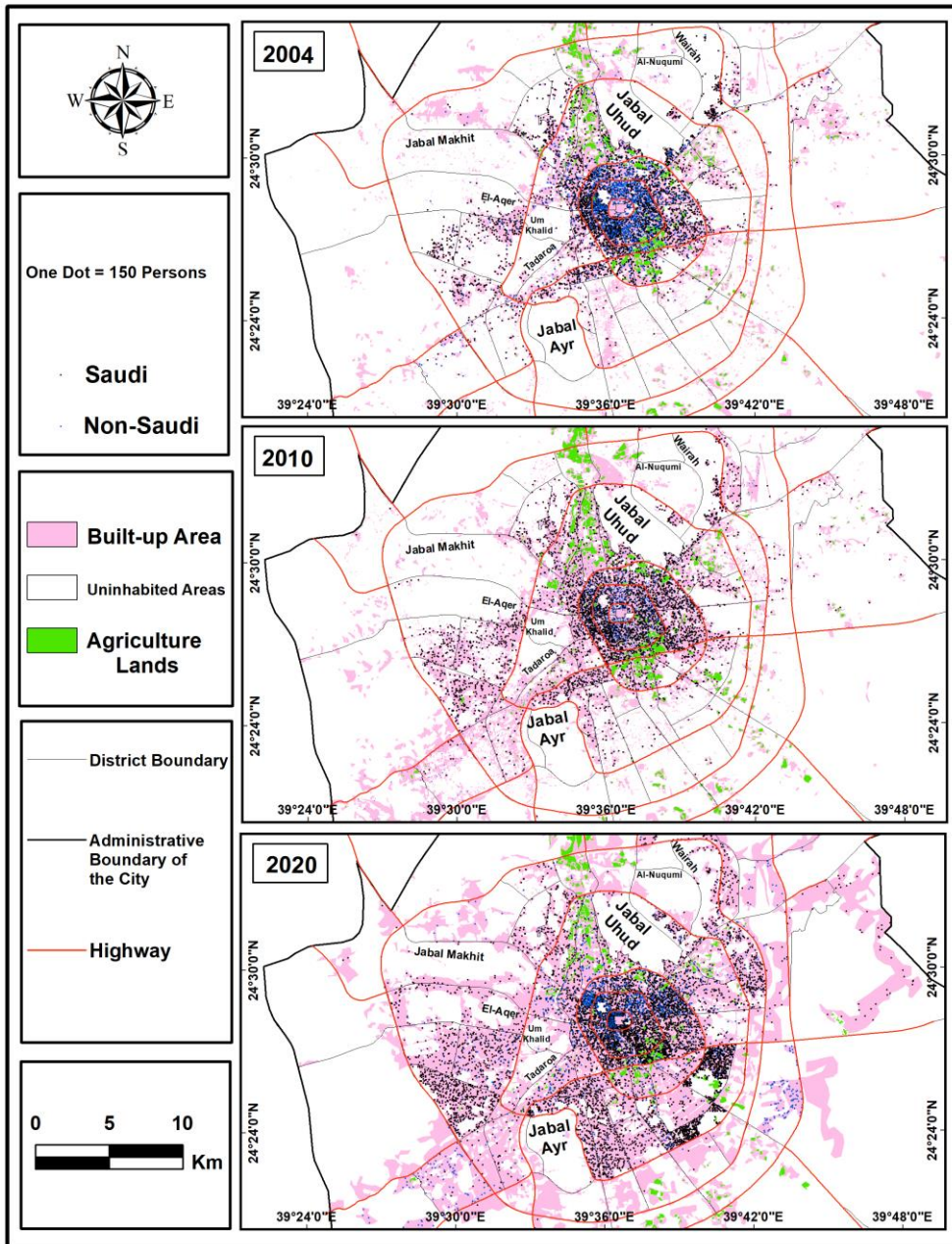
Fig. 4. The evolution of Al-Madinah's population distribution (2004–2020) (*source: GAS 1974-2010; USGS's Earth Explorer web browser; ARM, 2020*).

The 2010 map reflects an increase in the concentration of Saudi inhabitants in the western districts of zone 3. The 2020 map reveals that the concentration of Saudis increased in the southern, southwestern, and western districts of zone 3, especially in Mozinb and Al-Gharaa. However, the vast majority of non-Saudis resided in the two inner zones between 2004-2020, due to proximity to their places of employment in the CBD. In the central area within the second ring road, foreign workers

represent the largest population (MMRA, 2019a). This finding is consistent with the case of Al-Ain city, as the majority of non-citizens were largely clustered near the CBD (Yagoub, 2006). In summary, Saudis have relocated further from the CBD than non-Saudis over the past decade: while many Saudis moved from zone 2 to zone 3, non-Saudis have instead moved between districts within the same zone. Additionally, Al-Madinah's mountains have long served as natural barriers that have curbed urban expansion and subsequently led to the spread of settlements to the north, south, and west. That said, the city has recently expanded north, south, and east along the valleys (MMRA, 2019a), leading to changes in the population distribution map.

Regarding gross population density, continued population growth in a city causes widening administrative boundaries and changes in land cover. Therefore, naturally vegetated land and agricultural land are converted into residential areas characterised by high population density. As Al-Madinah covers a large administrative area (229,430 ha), the city's gross population density is low and slowly increasing, from 4 to 6.5 p/ha between 2004–2020. By comparison, Makkah had a low density of 2.3 p/ha in 2013 (El-Abd, 2018), while the Libyan city of Zwarah had a high gross population density of 39 p/ha in 2000 (Benomar et al., 2006). Al-Madinah's two inner zones recorded the highest densities (see **Table 2**). The significant increase in population density in zone 1 versus a slight decrease in zone 2 between 2004–2020 was due to non-Saudis relocating from zone 2 to zone 1. Conversely, the density of Saudis increased in zones 2 and 3 because many Saudis wished to avoid the overcrowded city centre. These findings are consistent with the case of the city of Riyadh, as it has experienced a decline in population density at an exponential rate from the city centre, with variations between its sectors and zones (Al-Gabbani 1991).

Table 2.
Gross and net residential population densities of Al-Madinah by nationality and zone (2004–2020).

Zones			Zone 1	Zone 2	Zone 3	Zone 4	Al-Madinah
Gross density (people/ha)	2004	SA	6.1	66.6	5.8	0.1	2.8
		Non-SA	30.8	40.8	1.1	0.0	1.2
		Total	36.9	107.3	6.8	0.1	4.0
	2010	SA	24.6	71.2	6.7	0.1	3.1
		Non-SA	140.1	49.1	1.6	0.1	1.7
		Total	164.7	120.2	8.3	0.2	4.8
	2020	SA	348.4	85.4	13.9	0.2	5.2
		Non-SA	280.5	25.3	1.8	0.2	1.3
		Total	628.8	110.7	15.7	0.4	6.5
Net residential density (people/ha)	2004	SA	7.2	90.7	31.1	2.3	31.9
		Non-SA	36.6	55.6	5.8	0.8	13.9
		Total	43.8	146.3	36.9	3.1	45.8
	2010	SA	32.1	90.2	25.3	2.9	29.8
		Non-SA	182.3	62.2	6.1	2.9	16.0
		Total	214.3	152.3	31.4	5.9	45.9
	2020	SA	418.1	95.8	28.4	1.7	25.4
		Non-SA	336.7	28.4	3.6	1.8	6.5
		Total	754.8	124.1	32.0	3.5	31.9

Source: Author's calculation based on: GAS 1974-2010; USGS's Earth Explorer web browser; ARM. (2020).

Considering the entire population of Al-Madinah, **Figure 5** shows that the closer the proximity to the Prophet's Mosque in the CBD, the greater the gross population density; conversely, the further away from the CBD toward the periphery, the lower the gross population density. Therefore, the pattern of population density in Al-Madinah is consistent with location theory, as population density declines in relation to the distance from the CBD. A large area of Al-Madinah's periphery remains underdeveloped and has potential for further residential expansion. Furthermore, the highest

population densities (100 p/ha <) were concentrated in many districts of the two inner zones, particularly those districts proximate to the Prophet's Mosque in 2004 and 2010, while the lowest densities (< 50 p/ha) were found in most districts in zones 3 and 4.
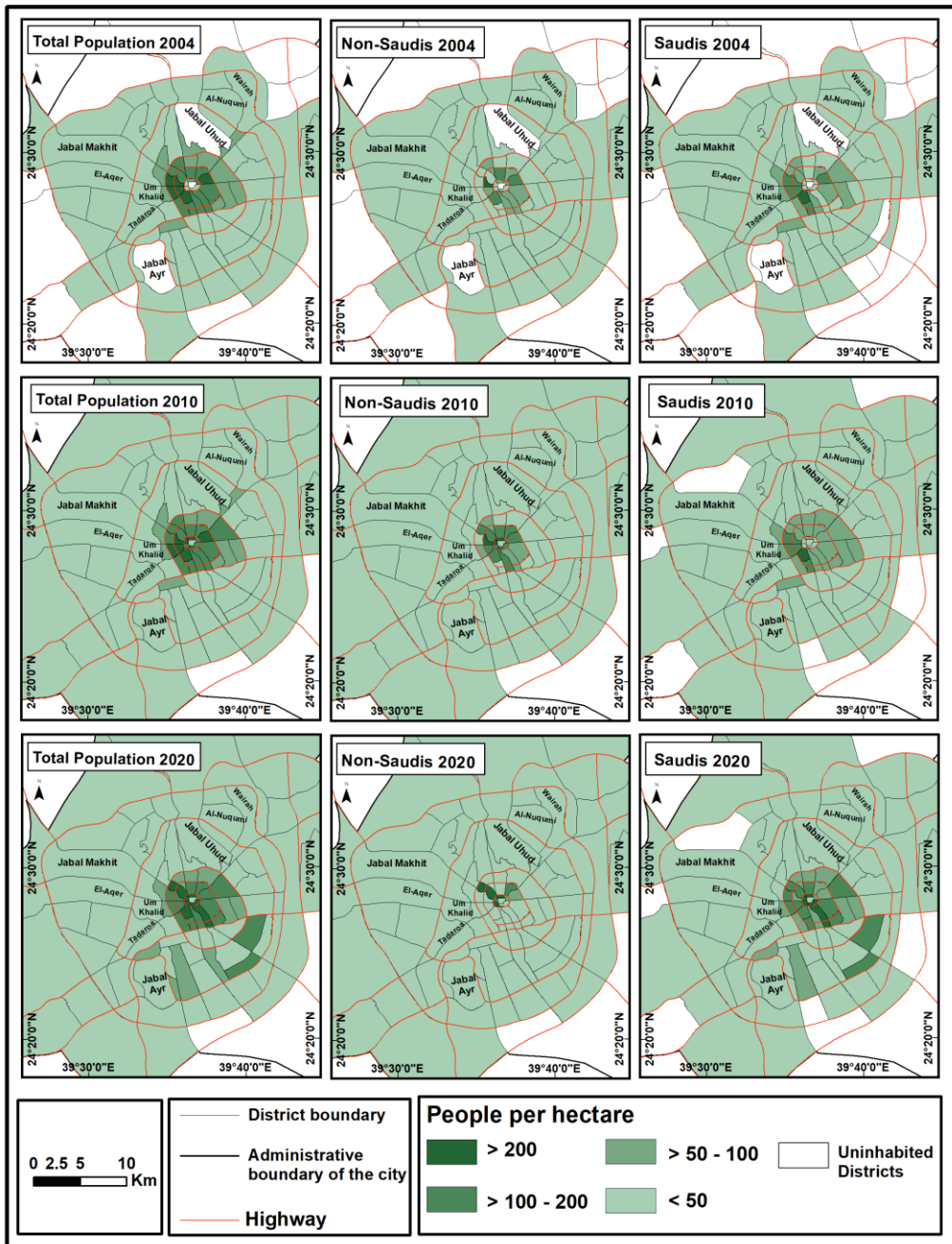

Fig. 5. Evolution of population gross density of Al-Madinah (2004–2020).
(Source: GAS 1974-2010; ARM, 2020).

By 2020, population density had increased in eight districts of zone 3, such as Al-Gharaa, Mozinb, and Al-Wabra due to the relocation of Saudis from the CBD to the southern and western districts of zone 3.

Net residential density in Al-Madinah rose from 13 to 46 p/ha between 1962–1978 (Makki, 1985) before falling from 46 to 32 p/ha between 2004–2020, while Al-Madinah's residential area increased from 20,098 to 47,022 hectares. Therefore, Al-Madinah had a net density lower than the UN-Habitat recommended population density of 150 p/ha (UN-Habitat, 2014), in common with many other Saudi cities in 2016, such as Jeddah (55 p/ha), Dammam (70 p/ha), Riyadh (48 p/ha), and Makkah (46 p/ha) (MMRA, 2019a). The central area of Al-Madinah within the second ring road has remained the most densely populated, with 142, 154, and 144 p/ha in 2004, 2010, and 2020, respectively. These densities almost fit the UN-Habitat recommended density. Zone 1 is more densely populated than zone 2, with a net density of 214 vs 152 p/ha in 2010, and 755 vs 124 p/ha in 2020, respectively.

Variations in population densities provide guidelines for interpreting the geographical distribution of economic activities (Marti-Henneberg et al., 2016). Therefore, the concentration of businesses catering to visitors in the CBD has created an imbalance in the distribution of services and facilities between the CBD and the periphery of Al-Madinah. Most visitors elect to stay in the CBD to be near the Prophet's Mosque and accessibility services and facilities (Makki, 1989). Moreover, both Saudi and non-Saudi residents seek to live in the centre as both a desirable location and for proximity to the mosque where they perform their five daily prayers. In 2020, the net population density in zone 3 was twice the gross population density, reflecting population mobility into zone 3, primarily by Saudis. For the entire population of Al-Madinah, **Figure 6** shows that districts with high net density are spread over a larger area compared to gross density. Aligning with the findings from the gross density maps, most of these districts are in the two inner zones. Conversely, some of these districts are in zone 3. Moreover, the net population density in these districts was extremely high, peaking with E-Sih in 2004 (541 p/ha), Bedaa in 2010 (410 p/ha), and E-Naqa in 2020 (2,243 p/ha). By 2020, six districts had a population density of > 500 p/ha, two of which were located in zone 3. Recently, relatively high population densities have been recorded in zone 3, with the number of districts with > 50 p/ha increasing from 9 to 15 between 2010–2020. **Figure 6** indicates an important distinction in the population density patterns of Saudis and non-Saudis: districts with high non-Saudi populations were confined to the two inner zones in 2004 and 2010, but in 2020 were limited to the eight districts north and west of the Prophet's Mosque (likely due to thousands of accompanying migrants departed the KSA after the declaration of the new fees). Meanwhile, high population densities of Saudis expanded to Zone 3 within the research timeframe. In 2020, relatively high densities of non-Saudis (50 p/ha) were concentrated in a narrow circle less than two kilometres in radius around the CBD, while high densities of Saudis occupied a ring-shaped area over ten kilometres in radius.

Statistically, Al-Madinah's net and gross densities differed during the research timeframe. The proportion of the difference between both densities for each district varied widely between 5–99%, 2–99%, and 1–96%, with means of 59%, 55%, and 39%, and standard deviations of 32%, 33%, and 29% in 2004, 2010, and 2020, respectively. This difference is because some districts have more non-residential areas than others. There are also variations between the four zones, with the differences ranging from 1–33%, 1–30%, 3–96%, and 30–97%, with means of 16%, 10%, 43%, and 75%, and standard deviations of 10%, 8%, 23%, and 17% in zones 1, 2, 3, and 4, respectively in 2020. The high proportion of the differences and means in zone 3 or 4 is due to many districts having non-residential areas such as mountainous, agricultural, and open areas. Therefore, planning for Al-Madinah's population redistribution or resource allocation based solely on gross density would lead to poor decision-making.

Al-Madinah's concentration index (CI) has fallen slightly, being 85.3%, 82.6%, and 81.6% in 2004, 2010, and 2020, respectively. However, all of these figures show a high population concentration in certain districts, as the higher the CI, the more concentrated the population. Approximately 81% of the city's population would have to be redistributed to more peripheral districts to achieve an exact correspondence between population and area. CI in Al-Madinah shows a

greater population concentration than the Gulf city of Al-Ain (55%) (Yagoub, 2006). The CI of Al-Madinah has been gradually declining over the past two decades, reflecting a modicum of success of government efforts to alleviate population concentration in the CBD.
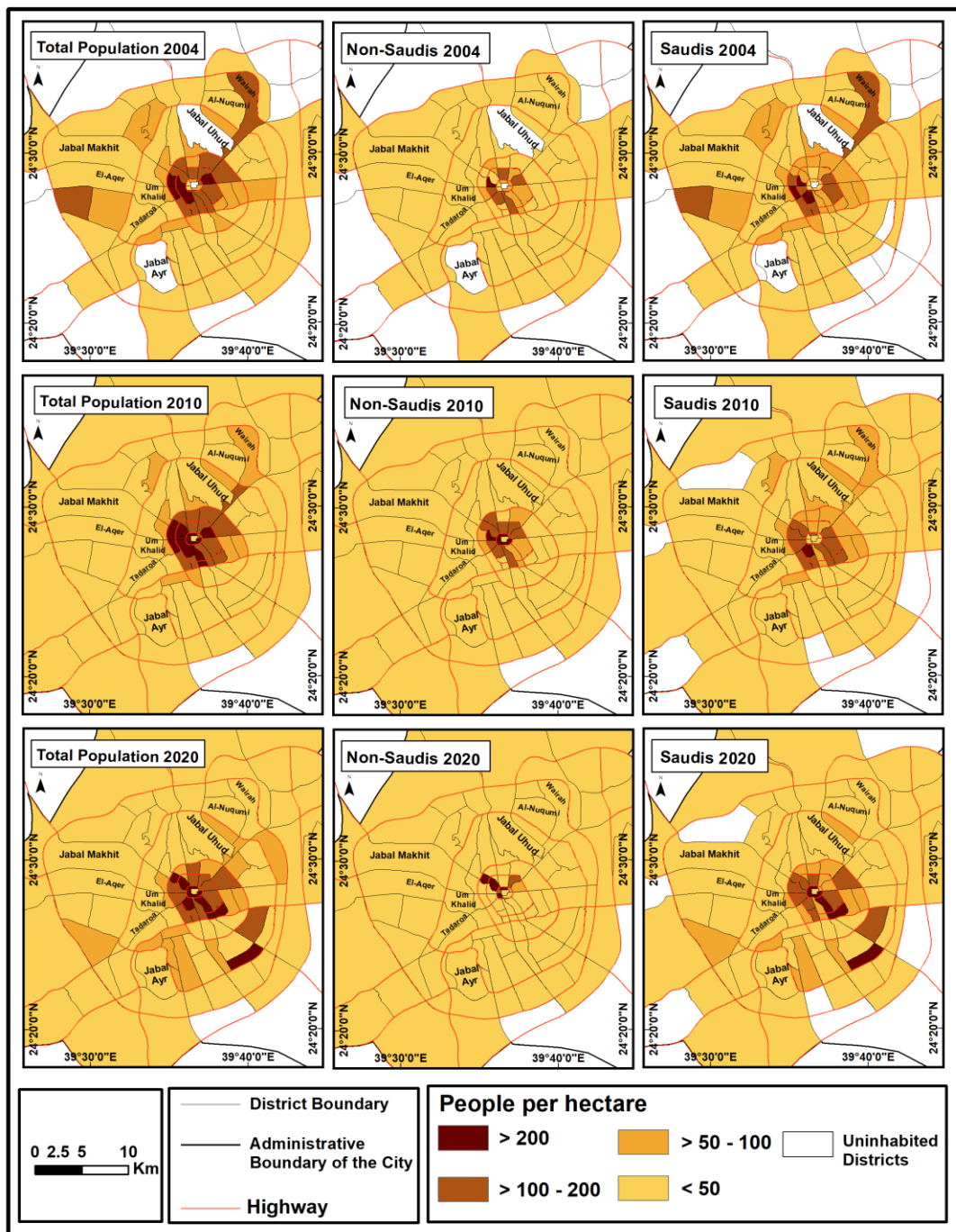
Fig. 6. Evolution of Al-Madinah's net residential population density (2004–2020).
(Source: GAS 1974-2010; USGS's Earth Explorer web browser; ARM, 2020).

Furthermore, Al-Madinah's population was (and remains) unevenly distributed geographically: 95% of the population lived within 11%, 13%, and 18% of the city's total area in 2004, 2010, and 2020, respectively. However, the area occupied by the population is gradually increasing in the twenty-first century (see **Table 3**).

<div align="right">**Table 3.**</div>

**Evolution of the relationship between the proportion of the population
and land area occupied in Al-Madinah (2004-2020).**

% of population	% of land area		
	2004	**2010**	**2020**
95	11	13	18
75	3	4	8
50	1	1	3

Source: Author's calculation based on: GAS. (1974-2010); Al-Madinah Regional Municipality. (2020).

Similarly, the Lorenz curves (**Fig. 7**) show an imbalance between population and area: there is a significant deviation from the perfect equality line of even distribution. Based on the Lorenz curve, the Gini Ratio (GR) can be calculated. Al-Madinah's GR was 0.946, 0.936, and 0.892 in 2004, 2010, and 2020, respectively. This was higher than in Al-Ain (0.732) (Yagoub, 2006). Al-Madinah's GR is close to 1, indicating an almost complete inequality between population and area, with the vast majority of the population located in the central area. Although Al-Madinah's GR is decreasing, there remain large swathes of land which could accommodate a significant proportion of the population.
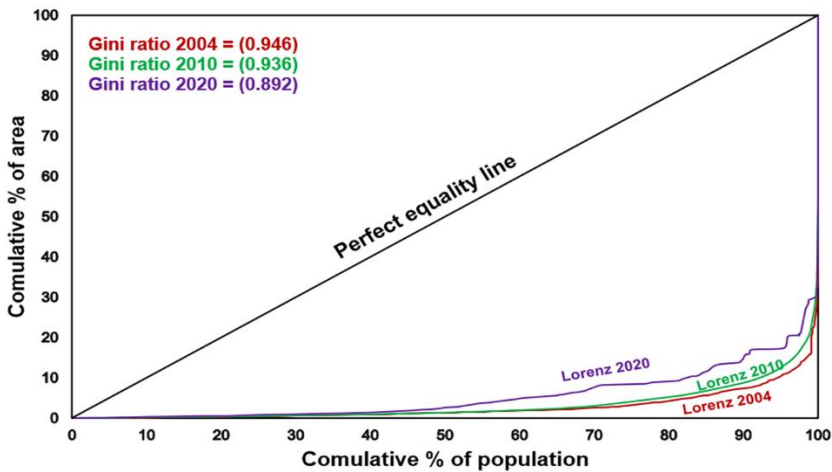
Fig. 7. The Lorenz curves for measuring population concentration in Al-Madinah (2004–2020).
(Source: GAS 1974-2010; ARM, 2020).

The calculation of the MCPs for Al-Madinah, shown in **Figure 8** and **Table 4**, is based on the population distribution over the inhabited area. Although the MCPs were typically close to El-Haram district between 2004–2020, they shifted from west to south, indicating movement from the centre to the south, as Al-Madinah's current urban sprawl follows its major transport routes radiating outwards from the CBD to valleys like Al-Aqiq in the south. Moreover, the south of Al-Madinah is free of topographical constraints and is a preferred residential location for religious considerations: it symbolises the direction from which the Prophet Muhammad (PBUH) first entered Al-Madinah, migrating from Makkah. Regarding the total population, the MCP moved from El-Manakha district in 2004 and 2010 to E-Jomaa in 2020, 746 meters south of El-Haram district. The MCP for Saudis moved from El-Manakha district in 2004 to E-Naqa in 2010, then to E-Jomaa in 2020, 894 meters south of El-Haram. However, the MCP for non-Saudis shifted a short distance from El-Haram district in 2004 to El-Manakha in 2010, then to Bani Khadra in 2020, 173 meters south of El-Haram. Saudis

are more mobile than non-Saudis: the MCP for Saudis shifted 1,337 meters south between 2010–2020 while that of non-Saudis moved just 665 meters south. The likely explanation for this is that housing in the south of Al-Madinah, especially outside the second ring road, is expensive for most non-Saudis but suitable for many Saudis as there are many new residential developments such as Al-Alia, Awali-Quba, and Jawhrat Al-Madinah, which predominantly consist of villas.
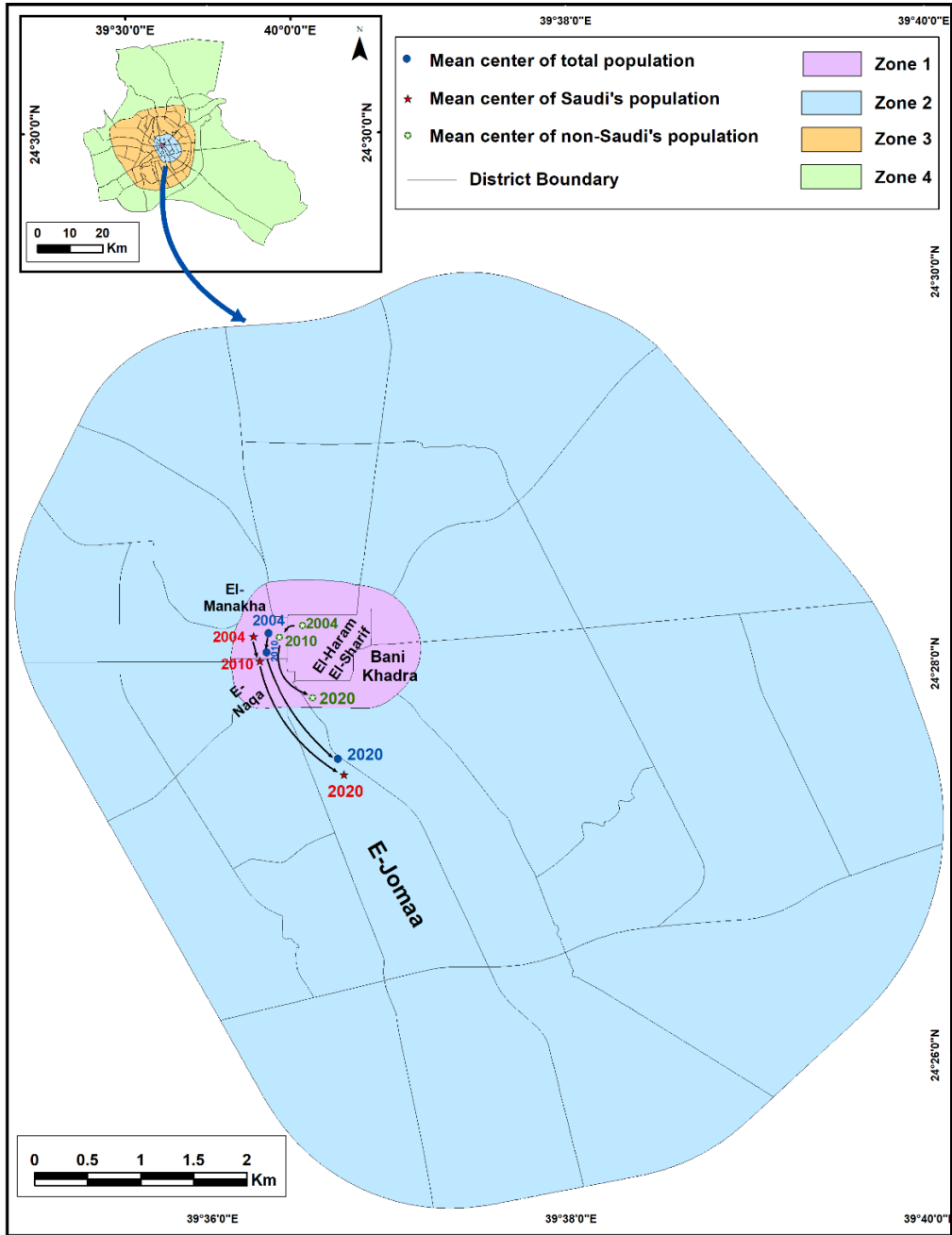
Fig. 8. Mean centre of population for Al-Madinah: 2004–2020.
(Source: GAS 1974-2010; USGS's Earth Explorer web browser; ARM, 2020).

Table 4.

Mean centre of population for Al-Madinah: 2004-2020.

Nationality	Year	North Latitude	East Longitude	Location	
				Zone	District
Saudis	2004	24° 28' 9"	39° 36' 14"	1	El-Monakha, 323 meters west of El-Haram.
	2010	24° 28' 2"	39° 36' 16"	1	E-Naqa, 253 meters southwest of El-Haram.
	2020	24° 27' 27"	39° 36' 44"	2	El-Jomaa, 894 meters south of El-Haram.
non-Saudis	2004	24° 28' 12"	39° 36' 31"	1	Northwest of El-Haram.
	2010	24° 28' 8"	39° 36' 23"	1	El-Monakha, 83 meters west of El-Haram.
	2020	24° 27' 50"	39° 36' 34"	1	Bani Khadra, 173 meters south of El-Haram.
Total population	2004	24° 28' 10"	39° 36' 20"	1	El-Monakha, 184 meters west of El-Haram.
	2010	24° 28' 4"	39° 36' 19"	1	El-Monakha, 199 meters southwest of El-Haram.
	2020	24° 27' 32"	39° 36' 42"	2	El-Jomaa, 746 meters south of El-Haram.

Source: Identified by the author using ArcGis 10.2, based on: GAS. (1974-2010); ARM. (2020).

SD is a measure of inhabitants' dispersion from their mean centre. **Figure 9** confirms Al-Madinah's population concentration as approximately two-thirds of the total population, Saudis, and non-Saudis were concentrated inside small circular areas with radii of 5.2, 5.7, and 3.7 kilometres in 2004 compared to 6.7, 6.8, and 6.5 kilometres in 2020, respectively. This reveals an abnormal population distribution pattern, as the population is confined to a small area surrounding the Prophet's Mosque. This small area, though, is a locus of business, accommodation and services. **Figure 9**, however, shows the wide distribution of Saudis and non-Saudis in 2020 compared to 2004.

This article provides an accurate examination of Al-Madinah's population distribution, net population density, MCP, and SD by identifying and calculating the area of the built-up area of each district using GIS and RS. However, this data has limitations: for example, the built-up area of each district identified comprises buildings not only for residential purposes but also for commercial, industrial, educational, and health activities. Fortunately, these non-residential purposes occupy a small amount of the built-up area. Additionally, the mismatch in dates between the datasets and the Saudi censuses of 2004 and 2010 may have slightly reduced the accuracy of the built-up area calculation.

LQ is widely used as a geographic index to compare the relative population concentrations of a subregion with those of an entire region. In this research, it is used to measure the distribution of Saudis and non-Saudis within a district compared to the total population. There was significant variation in LQs for Saudis (see **Figure 10a**) and non-Saudis (see **Figure 10b**). The local concentrations of Saudis were higher than expected (above 1) in many districts of zones 2 and 3, revealing an over-representation of Saudi citizens, while Saudis were under-represented (below 1) in all districts of zone 1, except Bani Khedra, and many districts of zone 4.

However, the local concentrations of non-Saudis were higher than expected in all districts of zone 1 and many of zone 2. Non-Saudis were most prevalent (above 2) in all districts of zone 1, except Bani Khedra, and some districts of zones 3 and 4. This is likely because zone 1 is characterised by older housing stock (Mohamed et al., 2016; MMRA, 2019a) with more reasonable rents which may be preferable for foreign workers.
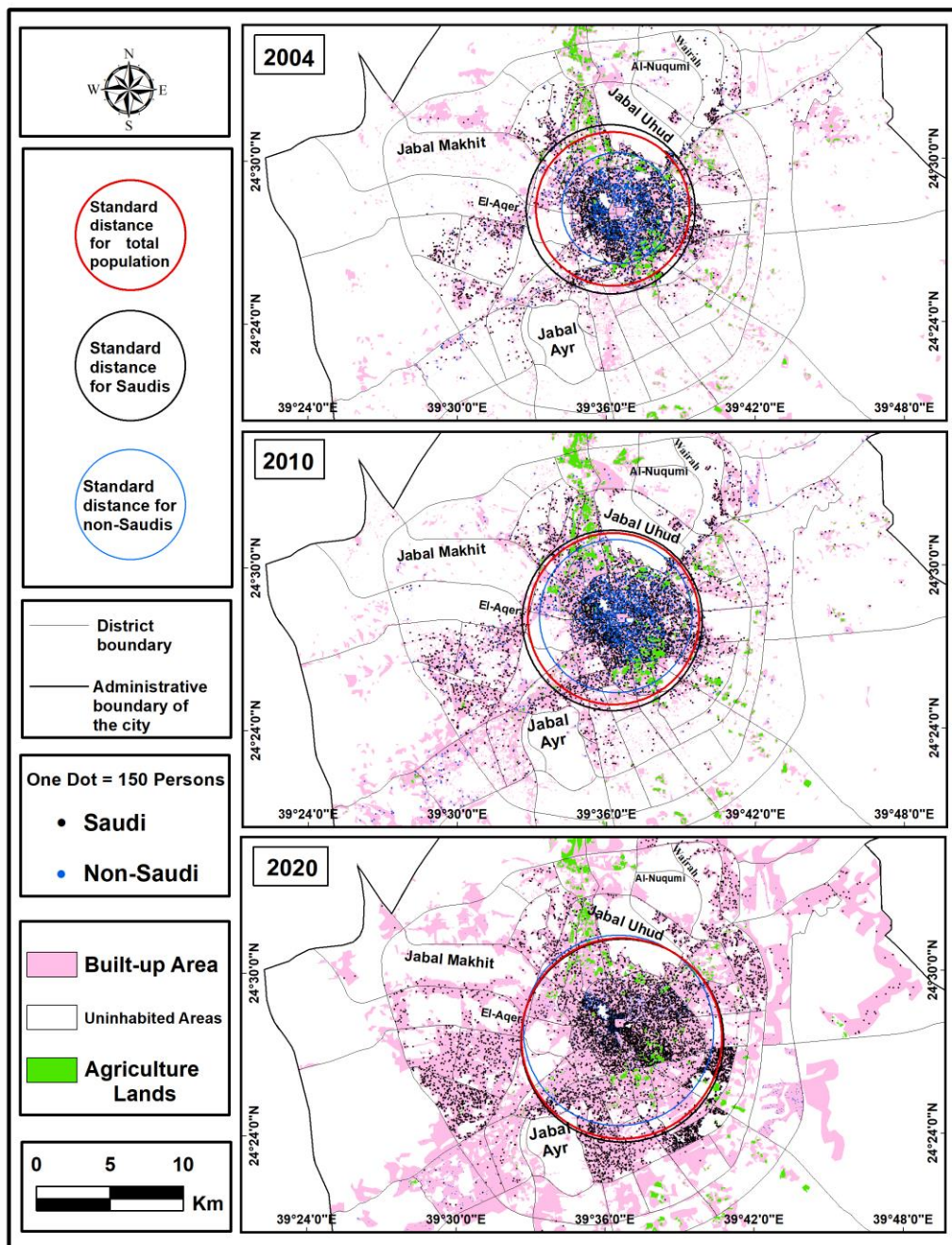

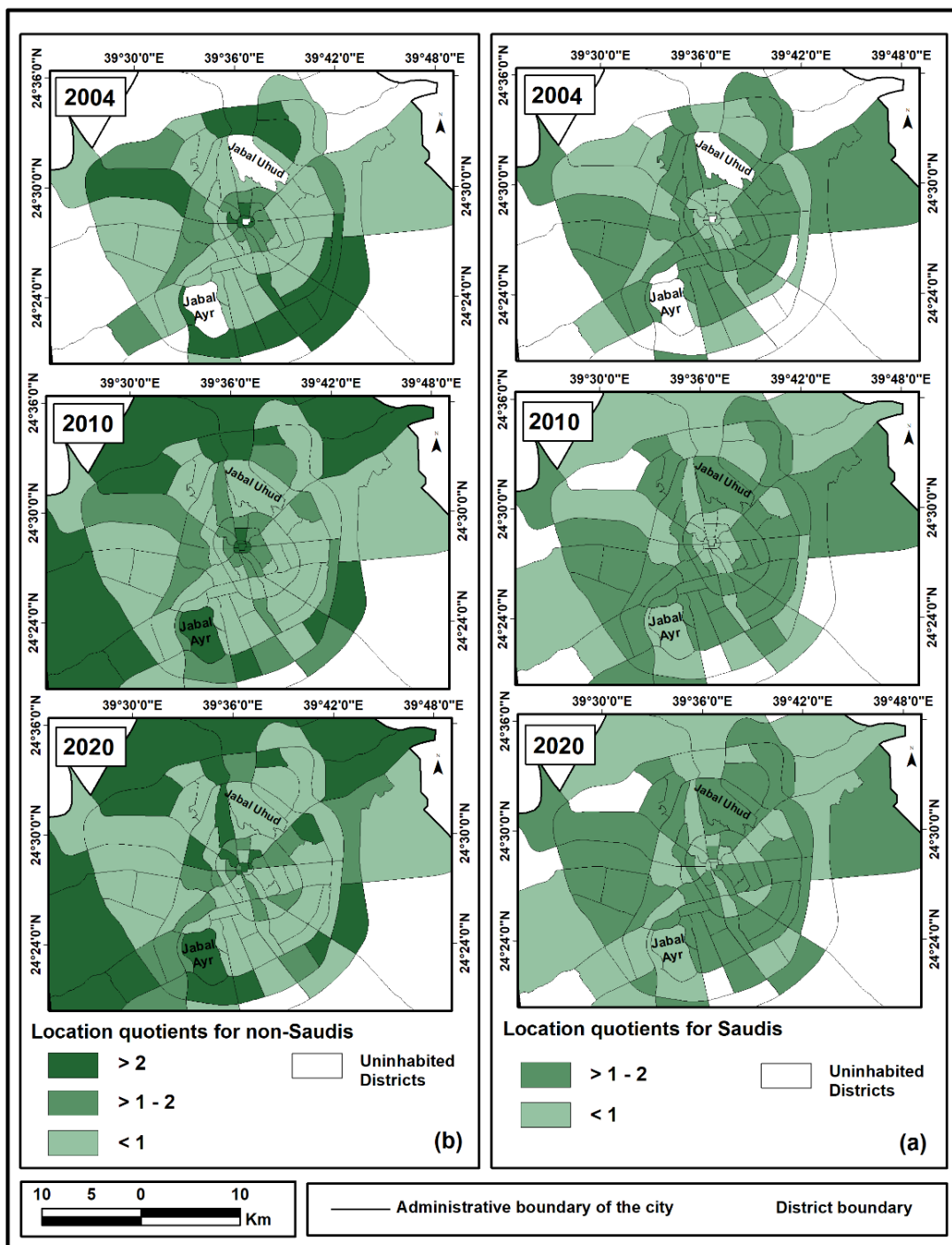**Fig. 9.** Standard distance for the population of Al-Madinah: 2004–2020. *(Source: As Figure 8).*

Fig. 10. Location quotients for Saudis and non-Saudis in Al-Madinah: 2004–2020.
(*Source: GAS 1974-2010; ARM, 2020*).

5. CONCLUSION

Al-Madinah is the fourth most populous city in the Kingdom of Saudi Arabia, with 1,500,000 inhabitants. Most residents have settled in the central area, putting pressure on infrastructure. The current study analysed the population distribution of Al-Madinah by combining demographic data with spatial data using GIS and RS. The findings reveal that the diverse populations living in Al-Madinah are unevenly distributed geographically. While there is a concentration in the centre around the Prophet's Mosque in the CBD, the population density gradually decreases towards the periphery. Like the city of Riyadh in the late 1980s (Al-Kharif, 1994), Al-Madinah has recently experienced residential mobility from the central area to the periphery, especially to the southern, southwestern, and western districts. At least half of the city's population resides in a small circular area with a radius of 5 kilometres surrounding the CBD. Recently, the city has witnessed significant population mobility, especially to the south and west. In line with many cities, Saudi citizens form the bulk of the relocating population, who are financially secure enough to avoid the congested CBD (Yagoub, 2006; Stewart et al., 2004).

The second half of the twentieth century saw an increase in Al-Madinah's net population density, while the last two decades have seen a decline from 46 to 32 p/ha due to the expansion of its built-up areas. The central area within the second ring road is densely populated (144 p/ha in 2020). Most high-density districts are in the central area within the second ring road; however, recently, districts outside the second ring road are also becoming densely populated. Variations in the population densities of Al-Madinah's zones or districts are due to the geographical distribution of different economic sectors (Marti-Henneberg et al., 2016). In 2020, population density increased in some districts between the first and third ring roads due to the relocation of Saudis from the CBD to these districts. High population densities were concentrated within a radius of ten kilometres of the CBD for Saudis, but within two kilometres of the CBD for non-Saudis. Furthermore, this study found that the percentage difference between the net and gross population densities for each district in Al-Madinah varied considerably between 1–96%: the CI (81%) showed a higher population concentration than many other Gulf cities; the Lorenz curves and GR values confirm the imbalance between population density and area in Al-Madinah over the past two decades; almost 95% of the population now lives in 18% of the city. The mean centre of population (MCP) is close to the district of El-Haram and has recently shifted from the west to south of the Prophet's Mosque because the south is free of topographical constraints and is a preferred residential area for religious reasons. Saudis are more mobile than non-Saudis: their MCPs have recently shifted 1,337 and 665 meters southward, respectively. Although the SD confirms the population concentrations in Al-Madinah, it reflects their recent spread over a large area. Moreover, the LQ values show variations between the local concentration of Saudis and non-Saudis. They were higher than expected for Saudis in many districts between the first and third ring roads, while they were most prevalent for non-Saudis in almost all districts within the first ring road. This likely reflects the impromptu residential areas characterised by older housing stock with more reasonable rents which may be preferable for foreign workers.

Overall, Al-Madinah's residents, especially Saudis, are gradually moving from the CBD to the periphery. Based on the results of this study, the city's authorities should implement further residential projects, especially in the periphery of Al-Madinah, allocating some areas specifically to non-Saudi citizens. Finally, city officials should give more consideration to economic growth and service development in the periphery of Al-Madinah and seek to improve quality of life for those residing in the CBD.

R E F E R E N C E S

Abdou, A. A. (2017) *Residential mobility in Al-Madinah Al-Munawarah (2000-2016): A geographical study*. Cairo, The Egyptian Geographical Society.

Alahmadi, M. & Atkinson, P. M. (2019) Three-fold urban expansion in Saudi Arabia from 1992 to 2013 observed using calibrated DMSP-OLS night-time lights imagery. *Remote Sensing*. [Online] 11 (19), 1-19. Available from: https://doi.org/10.3390/rs11192266 [Accessed 10th May 2022].

Al-Gabbani, M. (1991). Population density pattern and change in the City of Riyadh, Saudi Arabia. *GeoJournal*. [Online] 24, 375–385. Available from: https://doi.org/10.1007/BF00578259 [Accessed 11th January 2022].

Al-Kharif, R. M. (1994). *The residential mobility in the city of Riyadh: a study in trends, causes, and characteristics*. Riyadh, King Saud University.

Allen, J. (2022). Temporal transitions of demographic dot maps. *International Journal of Cartography*. [Online] 8 (2), 208-222. Available from: https://doi.org/10.1080/23729333.2021.1910184 [Accessed 28th November 2022].

Al-Madinah Regional Municipality. (2020) *Digital map of Al-Madinah's city*. Al-Madinah Regional Municipality.

Al-Mahdy, O. (2013) *Medina: Reviving Place Identity through Public Space*. M.Sc., Waterloo, Ontario, Canada, the University of Waterloo. [Online] Available from: https://uwspace.uwaterloo.ca/handle/10012/7275 [Accessed 13th May 2022].

Amaral, S., Gavlak, A. A., Escada, M. I. S. & Monteiro, A. M. V. (2012) Using remote sensing and census tract data to improve representation of population spatial distribution: case studies in the Brazilian Amazon. *Population and Environment.* [Online] (34), 142–170. Available from: https://doi.org/10.1007/s11111-012-0168-2 [Accessed 11th April 2022].

Angel, S., Parent, J., Civco, D. L., Blei, A. & Potere, D. (2011) The dimensions of global urban expansion: Estimates and projections for all countries, 2000-2050. *Progress in Planning*. [Online] 75 (2), 53-107. Available from: https://doi.org/10.1016/j.progress.2011.04.001 [Accessed 15th November 2021].

Baudot, Y. (2001) Geographical analysis of the population of fast-growing cities in the Third World. In: Donnay, J. P., Barnsley, M. J. & Longley, P.A. (eds.) *Remote sensing and urban analysis*. [Online] London, Taylor and Francis. Available from: https://www.semanticscholar.org/paper/Geographical-Analysis-of-the-Population-of-Cities-Baudot/cc6140123bc4473ba1a67e7ec2aba3e8e161afd6 [Accessed 18th May 2022].

Benomar, T. B., Biant, F. & Shalgam, A. M. (2006) Application of GIS for population analysis case study of Zwarah, Libya. *Journal of Applied Science*. [Online] 6 (3), 616-621. Available from: https://doi.org/10.3923/jas.2006.616.621 [Accessed 19th May 2022].

Bryan, T. (2004). Population estimates. In: Siegel, J. S. & Swanson, D. A. (eds.) *The methods and materials of demography*. Amsterdam, Elsevier Academic Press, pp. 523-560.

Canada Centre for Remote Sensing. (2003). *Fundamentals of remote sensing tutorial*. [Online] Available from: https://www.pdfdrive.com/fundamentals-of-remote-sensing-ccrs-tutorial-e8562741.html [Accessed 26th November 2022].

de Smith, M. J., Goodchild, M. F. & Longley, P. A. (2018). *Geospatial analysis: comprehensive guide to principles techniques and software tools,* (6th ed.). [Online] Available from: https://www.spatialanalysisonline.com/extractv6.pdf [Accessed 15th March 2021].

Dobson, J. E., Bright, E. A., Coleman, P. R., Duree, R. C., & Worley, B. A. (2000). LandScan: A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing*. [Online] 66 (7), 849–857. Available from: https://www.researchgate.net/publication/267450852_LandScan_A_Global_Population_Database_for_Estimating_Populations_at_Risk [Accessed 9th May 2022].

El-Abd, A. Z. (2018). Population concentration in the holy city of Makkah in light of the environmental challenges. *Journal of the Center for Geographical and Cartographic Studies*. [Online] 26, 5-48. Available from: https://mkgc.journals.ekb.eg/article_156813_135e79a37f176712df32a3c4a5c8a2bd.pdf [Accessed 5th January 2022].

General Authority for Statistics. (1974-2010) *General censuses of population and housing in 1974, 1992, 2004, and 2010*. Riyadh, GAS. [Online] Available from: https://www.stats.gov.sa/en/13 [Accessed 26th March 2021].

General Authority for Statistics. (2019a) *Hajj statistics 2019*. Riyadh, GAS. [Online] Available from: https://www.stats.gov.sa/en/28 [Accessed 29th April 2022].

General Authority for Statistics. (2019b) *Umrah statistics bulletin 2019*. Riyadh, GAS. [Online] Available from: https://www.stats.gov.sa/en/862 [Accessed 29th April 2022].

George, M. V., Smith, S. K., Swanson, D. A. & Tayman, J. (2004). Population projections. In: Siegel, J. S. & Swanson, D. A. (eds.) *The methods and materials of demography*. Amsterdam, Elsevier Academic Press, pp. 561-601.

Gomes, E. (2017) Creating a dot density map: Resident population in Mainland Portugal. *The Cartographic Journal*. [Online] 54 (2), 157-162. Available from: https://doi.org/10.1080/00087041.2016.1148106 [Accessed 25th May 2022].

Herold, M., Couclelis, H. & Clarke, K. C. (2005) The role of spatial metrics in the analysis and modeling of urban land use change. *Computers, Environment and Urban Systems*. [Online] 29 (4), 369-399. Available from https://doi.org/10.1016/j.compenvurbsys.2003.12.001 [Accessed 18th May 2022].

Ibrahim, A. L. & Sarvestani, M. S. (2009) Urban sprawl pattern recognition using remote sensing and GIS - Case study Shiraz city, Iran. *2009 Joint Urban Remote Sensing Event*. [Online] 1-5. Available from: https://doi.org/10.1109/URS.2009.5137528 [Accessed 19th May 2022].

Khoiyangbam, R. S. & Gupta, N. (2015) *Introduction to Environmental Sciences*. New Delhi, TERI.

Linard, C., Gilbert, M., Snow, RW., Noor, AM. & Tatem, AJ. (2012) Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS ONE.* [Online] 7 (2), e31743. Available from: https://doi.org/10.1371/journal.pone.0031743 [Accessed 25th April 2022].

Linard, C., Gilbert, M. & Tatem, A.J. (2011) Assessing the use of global land cover data for guiding large area population distribution modelling. *GeoJournal*. [Online] 76, 525–538. Available from: https://doi.org/10.1007/s10708-010-9364-8 [Accessed 26th April 2022].

Long, J. F., Rain, D. R. & Ratcliffe, M. R. (2001) Population density vs. urban population: comparative GIS studies in China, India, and the United States. *The IUSSP Conference in Salvador*, Session S68 on "Population Applications of Spatial Analysis Systems (SIS)", 18-25 August 2001, Brazil.

Madinah Urban Observatory. (2018) *Background information: population size*. [Online] Available from: https://muo.mda.gov.sa/muosystem/Home# [Accessed 26th March 2021].

Makki, M. S. (1985) *Atlas of Al Madinah Al Munawwarah*. Riyadh, King Saud University.

Makki, M. S. I. (1989) The pattern of distribution of economic activities and services in the central area of Al-Madinah. *Journal of King Saud University (Arts)*, 1 (2), 219-251.

Marti-Henneberg, J., Franch-Auladell, X. & Solanas-Jiménez, J. (2016) The use of digital tools for spatial analysis in population geography. *Frontiers in Digital Humanities*. [Online] 3. Available from: https://doi.org/10.3389/fdigh.2016.00009 [Accessed 24th May 2022].

Matsah, M. I. & Hossain, D. (1993) Ground conditions in Al-Madinah Al-Munawarah, Saudi Arabia. Earth Science, *Journal of King Abdulaziz University*. [Online] 6, 47-77. Available from: https://www.kau.edu.sa/Files/320/Researches/51348_21538.pdf [Accessed 13th March 2021].

Matthews, S; Bacon, R., Lewis-McCoy, R., & Logan. E. (2019) *Spatial analysis*. Oxford Bibliographies Online – Sociology. Available from: https://doi.org/10.1093/OBO/9780199756384-0058 [Accessed 28th April 2022].

Ministry of Municipal and Rural Affairs. (2019a) *Future Saudi cities: Madinah city profile*. [Online] Available from: https://unhabitat.org/sites/default/files/2020/04/madinah.pdf [Accessed 4th January 2022].

Ministry of Municipal and Rural Affairs. (2019b) *Saudi Cities Report 2019*. [Online] Available from: https://unhabitat.org/sites/default/files/2020/05/saudi_city_report.english.pdf [Accessed 4th January 2022].

Mohamed, A. I., Alginahi, Y. M. & Kabir, M. N. (2016). Modeling and simulation of traffic flow: a case study - first ring road in downtown Madinah. *International Journal of Software Engineering & Computer Systems*. [Online] 2, 89-107. Available from: https://dx.doi.org/10.15282/ijsecs.2.2016.8.0019 [Accessed 26th January 2016].

Mohamed, O. M. A. (2011) *Spatial analysis of settlement changes and their current and future trends in Al-Madinah Al-Munawarah (1369 AH/ 1950 AD - 1450 AH / 2028 AD).* Cairo, The Egyptian Geographical Society.

Mossoux, S., Kervyn, M., Soulé, H. & Canters, F. (2018) Mapping population distribution from high resolution remotely sensed imagery in a data poor setting. *Remote Sensing*. [Online] 10 (9), 1-19. Available from: https://doi.org/10.3390/rs10091409 [Accessed 25th January 2022].

Newbold, K. P. (2010). *Population geography: tools and issues*. New York: Rowman & Littlefield Publishers.

Portnov, B. A. & Pearlmutter. D. (1999) Sustainable urban growth in peripheral areas. *Progress in Planning.* [Online] 52 (4), 239-308. Available from: https://doi.org/10.1016/S0305-9006(99)00016-1 [Accessed 25th April 2022].

Ragab, O. S. (1979) *Al-Madinah Al-Munawarah: economics of the place, population, and morphology*. Jeddah, Dar E-Shorwq for Publishing, Distribution, and Printing.

Rai Technology University. (n.d.) *Fundamentals of general cartography*. [Online] Available from: https://www.pdfdrive.com/fundamentals-of-general-cartography-e34313718.html [Accessed 27th March 2021].

Ramzi, A. I. (2012) Studying population of Egypt based on census data and geographic information system. Presented at: *Proceedings of the 2nd International Conference on Egyptian Science Through Ages*, 9-11th October 2012, Cairo, Egypt. [Online] Available from: https://www.academia.edu/29757139/Urban_Expansion_and_Population_Growth_in_Ras_Sudr_City_Using_Remotely_Sensed_Imagery [Accessed 14th May 2022].

Stewart DJ, Yin Z-Y, Bullard SM, MacLachlan JT. (2004) Assessing the spatial structure of urban and population growth in the Greater Cairo Area, Egypt: A GIS and imagery analysis approach. *Urban Studies.* [Online] 41 (1), 95-116. Available from: https://doi.org/10.1080/0042098032000155704 [Accessed 14th May 2022].

Tayan, O., Alginahi, Y. M., Kabir, M. N. & Al Binali, A. M. (2017). Analysis of a transportation system with correlated network intersections: A case study for a central urban city with high seasonal fluctuation trends. *IEEE Access*. [Online] 5, 7619-7635. Available from: https://doi.org/10.1109/ACCESS.2017.2695159 [Accessed 30th May 2022].

The United States Geological Survey USGS's Earth Explorer web browser. [Online] Available from: https://earthexplorer.usgs.gov/ [Accessed 10th May 2020].

Tungnung, J. Z. & Anand, S. (2017) Dynamics of urban sprawl and landuse change in Imphal of Manipur, India. *Space and Culture, India.* [Online] 5 (2), 69-83. Available from: https://doi.org/10.20896/saci.v5i2.271 [Accessed 20th May 2022].

UN-Habitat (2014) *A New strategy of sustainable neighbourhood planning: Five principles, UN-Habitat, Nairobi, Kenya*. [Online] Available from: https://unhabitat.org/sites/default/files/documents/2019-05/five_principles_of_sustainable_neighborhood_planning.pdf [Accessed 4th January 2022].

United Nations Population Division (2019) *World urbanization prospects: The 2018 revision.* New York, United Nations Population Division, p. 4. [Online] Available from: https://population.un.org/wup/publications/Files/WUP2018-Report.pdf [Accessed 10th May 2022].

Voss, P. R. (2007) Demography as a spatial social science. *Population Research and Policy Review.* [Online] 26 (5-6), 457–476. Available from: https://doi.org/10.1007/s11113-007-9047-4 [Accessed 18th May 2022].

Wang, L. & Chen, L. (2016) Spatiotemporal dataset on Chinese population distribution and its driving factors from 1949 to 2013. *Scientific Data.* [Online] 3, 160047. Available from: https://doi.org/10.1038/sdata.2016.47 [Accessed 28th April 2022].

Weerakoon, P. (2017) GIS integrated spatio-temporal urban growth modelling: Colombo Urban Fringe, Sri Lanka. *Journal of Geographic Information System*. [Online] 9 (3), 372-389. Available from: https://doi.org/10.4236/jgis.2017.93023 [Accessed 30th April 2022].

World Bank. (2009) *World development report 2009: Reshaping economic geography*. World Bank. [Online] Available from: https://openknowledge.worldbank.org/handle/10986/5991 [Accessed 24th April 2022].

Yagoub, M. M. (2006) Application of remote sensing and geographic information systems (gis) to population studies in the gulf: A case of al ain city (UAE). *Journal of the Indian Society of Remote Sensing.* [Online] 34 (1), 7–21. Available from: https://doi.org/10.1007/BF02990743 [Accessed 25th May 2022].

Yang, M., Wang, S. Zhou, Y. Wang, R. & Zeng, C. (2009) Population spatialization in Gansu Province based on RS and GIS. *2009 Joint Urban Remote Sensing Event*. [Online] 1-6. Available from: https://doi.org/10.1109/URS.2009.5137641 [Accessed 28th April 2022].

INTEGRATED USE OF OPTICAL AND RADAR DATA FOR CROPLAND MAPPING OVER THE MOUNTAIN SLOPE AREA IN BOYOLALI, INDONESIA

Vidya Nahdhiyatul FIKRIYAH[1] , Nirma Lila ANGGANI[1] , Umar El Izzudin KIAT[1] , Fithrothul KHIKMAH[2] , Wildan Abdul ARROYAN[1] and Muh Faqih RIZKI[1]

ABSTRACT:
Mapping agricultural land cover data is important as an effort to support national food security, especially in Boyolali, Central Java Province, Indonesia which is one of the national rice granaries. However, mapping in the mountain slope area using optical data only is challenging due to cloud cover. The development of remote sensing technology has encouraged the possibility to integrate data with different sensors. This data integration is needed to optimize the ability to detect and map cropland that has a variety of characteristics. Therefore, this study aims to identify the cropland through the integration of time-series optical and Synthetic Aperture Radar (SAR) data. Detection of cropland was carried out using 2021 data. Polarisation of VV, VH, and ratio of VV/VH data was derived from the Sentinel-1, whereas image indices of Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Soil Adjusted Vegetation Index (SAVI) data were obtained from Sentinel-2. Data of Sentinel-1 and Sentinel-2 was combined and several features were selected based on their importance score. Random Forest (RF) classification was then performed. The result show that the mapping using integrated data could improve the accuracy. This indicates the possibility of data to be implemented in further studies such as the cropland type mapping and the estimation of food productivity.

***Keywords:** Optical, Synthetic Aperture Radar, Integration, Cropland, Random Forests.*

1. INTRODUCTION

The need for accurate land cover data, especially cropland, cannot be separated from its important role to support strategic planning in a region. Data on where and when the cropland planted, as well as the accessibility are critical as a basis for effective measure on maintaining national food security (Susilo & Harini, 2018). Since land cover extent is dynamic, therefore, the observation should be done in a timely manner. In particular, the Indonesian government has launched the *Nawacita* program in 2015 where the national food self-sufficiency becomes one of priority targets in national development. For that reason, the accurate data on cropland is highly essential not only for national but also the global purpose as it is also mentioned in the Sustainable Development Goals (SDGs).

Remote sensing technology plays a role in extracting and monitoring land cover which is known to have variations in spectral, temporal, and spatial resolution. Remote sensing with optical sensors has been widely used for land monitoring (Gumma et al., 2019; Piao et al., 2021; Sarono et al., 2015), but the image quality and accuracy of mapping using this sensor depends on atmospheric conditions. Primarily the use of optical data is a challenge due to high cloud interference in tropical countries. To overcome these cloud problems, there is a Synthetic Aperture Radar (SAR) data which has the advantage of recording that is not affected by weather conditions. In other words, remote sensing has been developed in different specifications, so there are more approaches in land detection, such as combining optical and SAR imagery (Joshi et al., 2016).

[1]*Faculty of Geography, Universitas Muhammadiyah Surakarta, Indonesia, *Corresponding author: vidya.n.fikriyah@ums.ac.id; nla624@ums.id; uei665@ums.id; e100170108@student.ums.ac.id; e100190281@student.ums.ac.id.*

[2]*Photogrammetry and Geoinformatics, The Stuttgart Technology University of Applied Sciences, Germany, 12khfi1mpg@hft-stuttgart.de*

The complexity in agricultural area creates a challenge in mapping cropland using optical data (Fritz et al., 2015). Not only that, specifically rice crop is also cultivated in different land-climatic and ecosystems following the sequence in topography, from upland, rainfed lowland, irrigated, to flood prone area (Kuenzer & Knauer, 2012). Despite the relatively high revisit time of Sentinel-2, the observation of cropland in the cloud-prone tropical and sub-tropical area using this data only is still challenging (Singha et al., 2019). Thus, data integration with Sentinel-1 that is independent of sun illumination (Kuenzer & Knauer, 2012) is potential to help mapping cropland in such condition. Recently, the combination of optical and SAR data crop has been employed in the subtropical area using backscatter from Sentinel-1 and NDVI from Sentinel-2 data (Cai et al., 2019). Also, observation of paddy field in the tropical lowland coastal area has also been conducted using Sentinel-1 and 30-m resolution of Landsat data (Arjasakusuma et al., 2020), making a limitation in capturing smallholder fields. In this context, to date the ability of integration Sentinel-2 and Sentinel-1 to map cropland in a tropical mountain slope region has not been given much attention.

Furthermore, in the last decade, the use of machine learning algorithms for remote sensing has received a lot of attention. The application is mainly related to the study of land cover/use classification. Algorithms commonly used to date include support vector machines (SVM), decision trees, random forests (RF), and convolutional neural networks (CNN) (Sheykhmousa et al., 2020). These methods have advantages in terms of complex pattern retrieval capabilities and informative features of remote sensing satellite imagery. In this case, Random Forest (RF) was implemented for classification as it is known to improve the classification accuracy for time-series data without overfitting problem (Piao et al., 2021).

Given the above background, the general objective of this study is to map cropland based on the integration of optical and SAR time-series data. In addition, to capture the timely or seasonal spectral variability of different land cover, the multi-temporal remote sensing data is required. Feature selection will be performed to choose the most informative inputs for classification and RF classifier would be implemented for mapping. Then, a comparison of accuracy would be conducted among classification results using Sentinel-1, Sentinel-2, and the integration of them.

2. STUDY AREA AND DATA

2.1. Study area

The study was carried out in Regency of Boyolali, Province of Central Java, Indonesia. According to the department of regional statistic (BPS Boyolali, 2021), rice crop is the dominant crop in Boyolali (**Fig.1**) and its harvesting area reached 47,760.01 hectares with a total production of 225,425.92 ton in 2020. This makes the regency becomes the ninth largest rice producer in Central Java. Boyolali is situated in the slope area of two volcanos, the Mount Merapi (2,910 m) and Merbabu (3,145 m) in the western side (**Fig.2**). The elevation varies between 75-1,500 msl.
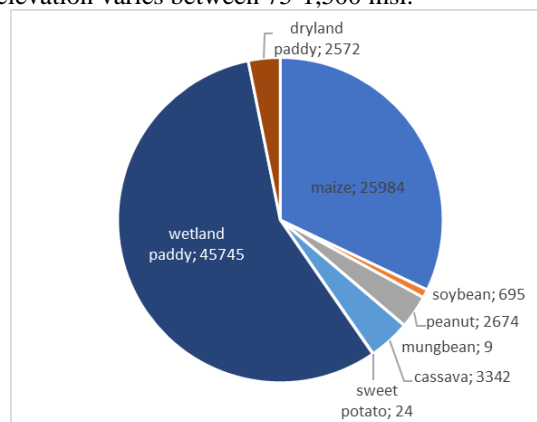
Fig. 1. Harvested area (hectares) of different crop in Boyolali in 2019 (*Source: Regional statistics, 2020*).
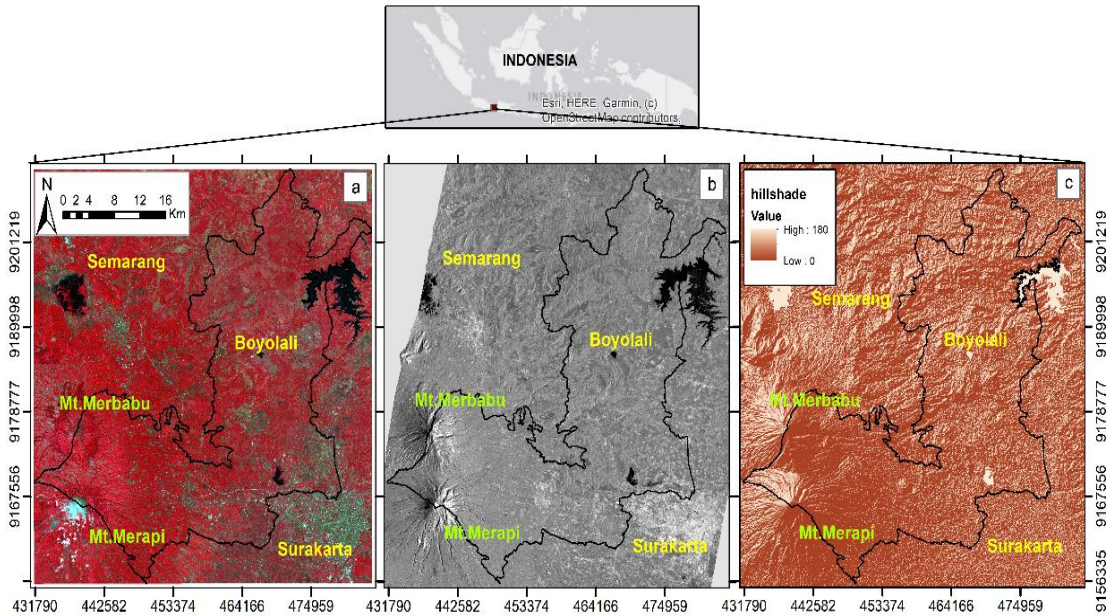
Fig. 2. Location of Boyolali, Indonesia shown in: a) Sentinel-2 data (RGB composite using NIR, red, and green band, respectively); b) Sentinel-1 VV data; and c) Hillshade from ASTER GDEM data.

2.2. Synthetic Aperture Radar (SAR) data collection and processing

One of the products of the Sentinel satellite project is Sentinel-1. It uses a C-band SAR sensor where object identification is based on backscatter. In this study, Sentinel-1 time-series data will be used to extract backscattering of land cover classes. The data selected is Sentinel-1A Level 1 Ground Range Detected (GRD) with dual polarisation (VV and VH) which was downloaded from https://scihub.copernicus.eu/dhus/#/home. The data acquisition period is April, July, and August 2021, taking into account a whole cropping season from rice crop calendar since it is the dominant crop in the study area (**Fig.1**). The characteristics of data used in the study is presented in **Table 1**.

Table 1.

Specifications of Sentinel-1 data used in the study.

Satellite	Sentinel-1A
Height/inclination	693 km/98.18º
Wavelength	C-band (3.75 – 7.5 cm)/ 5.405 GHz
Polarisation	VV+VH
Pixel spacing	10 m
Acquisition date	April, July, and August 2021

GRD product is Sentinel-1A data that has been calibrated radiometrically. The next step is the use of Refined Lee filters to remove speckle effects (Argenti et al., 2013). A geometric correction was done through the Sentinel application platform (SNAP) program, which is a software developed by European Space Agency (ESA). For this purpose, the Shuttle Radar Topography Mission (SRTM) 1 arc-second resolution data was used directly in SNAP. Finally, the image is then cropped using the Boyolali administrative boundary from the Global Administrative Area (GADM). Then, three datasets of polarisations were produced (VV, VH, and ratio of VV/VH) to explore the spectral variations of land covers.

2.3. Optical data collection and processing

Sentinel-2 level 2A was used as the optical data. This data has been corrected radiometrically and geometrically, as its numbers represent the reflectance at the Bottom of Atmosphere (BOA) level so that it can be used immediately. We applied the function to collect monthly composite image in Google Earth Engine (GEE) environment following steps in https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR. Median function was used for the composite image.

To filter cloudy scenes, a maximum cloud percentage of 30% was set as the limit because no good image is available below that. The time-series images were then cropped based on the administrative coverage of Boyolali Regency using administrative data from GADM. Specifications of data used for the study is presented in **Table 2**.

Table 2.

Specifications of Sentinel-2 data used in the study.

Satellite	Sentinel-2A
Wavelength (spatial resolution)	Band 3, 4, and 8 for Green, Red, and NIR respectively (10 m)
Level	Level-2A
Acquisition date	April, July, and August 2021

3. METHODS

3.1. Image Indices

In the next step after the pre-processing of Sentinel-2 data, three vegetation indices were produced. Those vegetation index algorithms were used for the RF classification input. Sentinel-2 images were processed and its pixel values were transformed into Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Soil Adjusted Vegetation Index (SAVI) as stated in formula 1-3. NDVI is one of the most commonly used for vegetation studies, while SAVI has capability to minimize the effect of soil brightness (Sashikkumar et al., 2017). In addition, NDWI is sensitive to the moisture of plants and soil enabling better discrimination between crop and surface water body (Bhattacharya et al., 2021).

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{1}$$

$$NDWI = \frac{Green - NIR}{Green + NIR} \tag{2}$$

$$SAVI = \frac{NIR - Red}{NIR + Red + L} * (1 + L) \tag{3}$$

where, NIR, Red, and Green is the pixel values for each band, respectively, and L is the soil adjustment factor (0.5). SAVI was used to normalize the subtractive soil variations by applying an adjustment factor (L) to reduce the soil background variations. The L values were adjusted lower as vegetation cover increased (Mostafiz et al., 2021). L value 0.5 was used in this study, considering the medium levels of vegetation cover in the whole area and for crop detection.

3.2. Image classification process

For this study, three RF classification schemes of multi-temporal data were conducted using (1) image indices (NDVI, NDWI, and SAVI) from Sentinel-2 dataset only, (2) VV, VH, and ratio of VV/VH dataset only, and (3) the integration of the first and second dataset. There were five classification classes used, namely cropland, built-up, water body, bare land, and forest. Plots of all classes were visually identified using a high-resolution imagery in Google Earth Pro. In total, 100 plots were taken for each built-up, forest, and cropland class, whereas, 50 plots were identified for

each water body and bare land. Plots were then divided into 70% for the training data in RF classification and 30% for the accuracy measurement test. **Fig. 3** shows the distribution of training and validation plots.
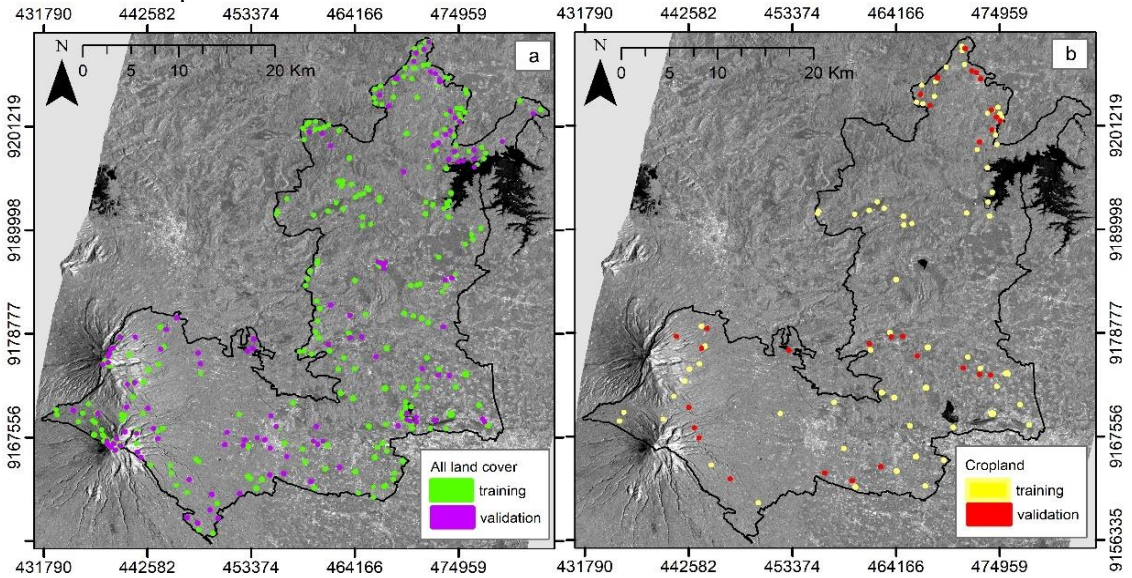
Fig. 3. Plots distribution for a) all land cover types and b) cropland only.

After creating a stack containing Sentinel-1 and 2 images, we calculated the variable importance to assess the relevant inputs for random forest classification. Variable importance scores were calculated based on Gini coefficient in ArcGIS Pro 2.9.0. The number indicates the frequency of a variable is responsible for a split in the model Feature selection entails the most informative features to reduce the effect of high-dimensional data (Akbari et al., 2020). In the RF classification, the maximum *ntree* was set at 50, maximum number tree depth was 30 and maximum number of samples per class was 1000. Converged color and mean digital number were included as the segment attributes. The overview of all steps taken in this study is presented in **Fig. 4**.
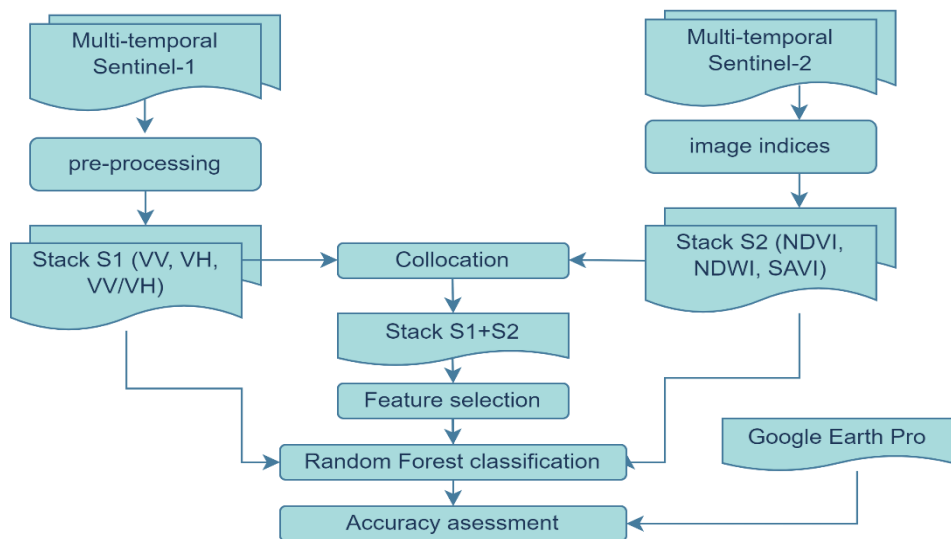
Fig. 4. Steps on the integration of Sentinel-1 and 2 data.

4. RESULTS

4.1. Key Classification in Sentinel-1

The result of pre-processing Sentinel-1 data is as follows. **Fig.5** shows the appearance of built-up, forest, water body, cropland, and bare land observed through Sentinel-1 (VV, VH, and the RGB composite). From the figure, it can be seen that every land cover has distinct characteristics of SAR backscatters. Built-up area gives brighter pixels compared to water body and bare land area. This indicates the higher level of backscatter from built-up area caused by the double bounce and corner reflectance (Deepthi et al., 2018). Meanwhile, the forest and cropland areas were identified having more colour variations in the pixels. In comparison to the cropland, forest areas however, have a slightly brighter tone as their dominant scattering mechanism are from the volume scattering related to the higher biomass in canopies and the double- bounce effect from the vertical trunk structure (Ningthoujam et al., 2016).
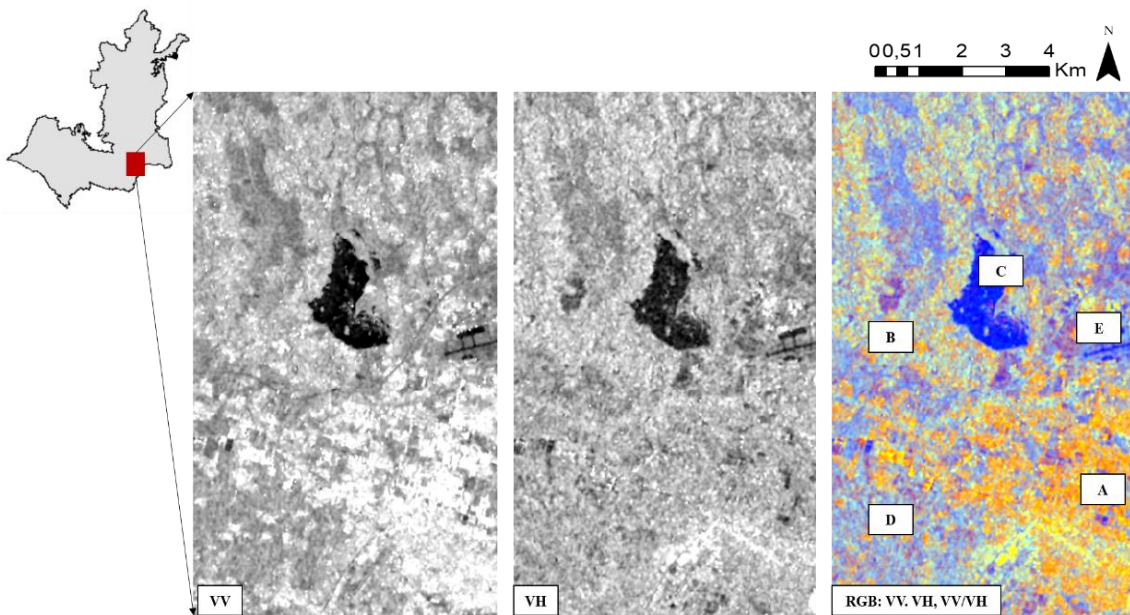
Fig. 5. Visual key interpretation of Sentinel-1 (VV, VH, and RGB composite VV, VH, VV/VH) data, showing A: built-up, B: forest, C: water body, D: cropland, and E: bare land.

Since a time-series data was used, the temporal variation of each land cover class can be observed in **Fig. 6**. The figure shows that cropland has a backscattering pattern that is different from other land covers. This change in backscattering is related to the growth phase of crop (Nguyen & Wagner, 2017). Built-up areas have the highest backscatter and relatively constant throughout the period, while water bodies remain giving the lowest backscatter response among other classes.

The classification results in **Table 3** shows that the classification accuracy of using Sentinel-1 data, giving the overall accuracy of 78%. For the user's accuracy, the value of bare land, cropland, and water body were high (100%, 88.46%, 83.33%, respectively). Meanwhile, high producer's classification accuracies were identified for the class of water body (100%), forest (80%), and built-up area (70%). The result suggests that high accuracy is gained for the stable class such as water body (Piao et al., 2021).
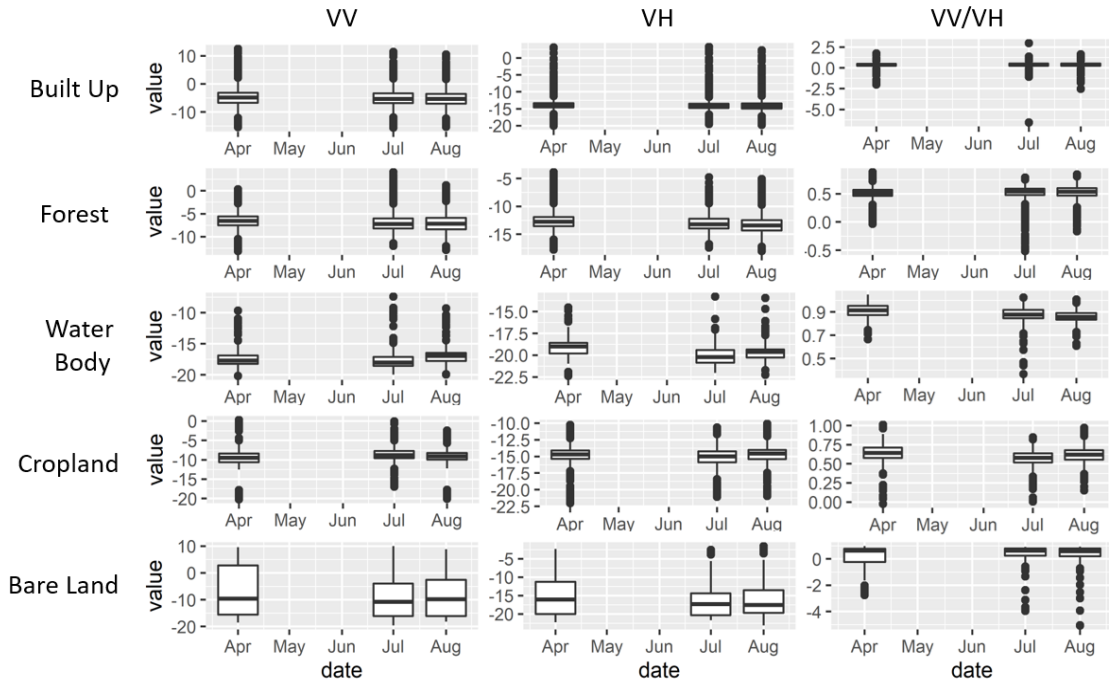
Fig. 6. Value dispersion of training data of every land cover class on VV, VH, and ratio VV/VH data.

Table 3.

Confussion matrix based on Sentinel-1.

Sentinel-1		Reference data					
		Built Up	**Forest**	**Water Body**	**Cropland**	**Bare Land**	**UA**
Classified data	Built Up	21	5	0	4	1	67.74
	Forest	8	24	0	3	0	68.57
	Water Body	0	0	15	0	3	83.33
	Cropland	1	1	0	23	1	88.46
	Bare Land	0	0	0	0	10	100.00
	PA	70.00	80.00	100.00	76.67	66.67	
Overall Accuracy	**78%**						

4.2. Spectral Indices from Sentinel-2

Fig. 7 presents the results of NDVI, NDWI, and SAVI over the study area. As shown in the figure, the distribution value of each index varied during the observation period. Overall, in April, the area was dominated by the high NDVI and SAVI value, especially in the northern part of region. This means that there was a high level of greenness on that period as the rainy season occur. In July, when the dry season occurred, most part of the region were in the middle and low value of all indices. High NDVI and SAVI value, however, were seen in the southwestern part where top of Mount Merapi and Merbabu located (**Fig.2**). For the last observation period, almost all areas were covered by the moderate and low value of NDVI, NDWI, and SAVI. The dryness condition in particular can be detected in the northern part of Boyolali. These value differences could be caused by different planting cycles of crop.
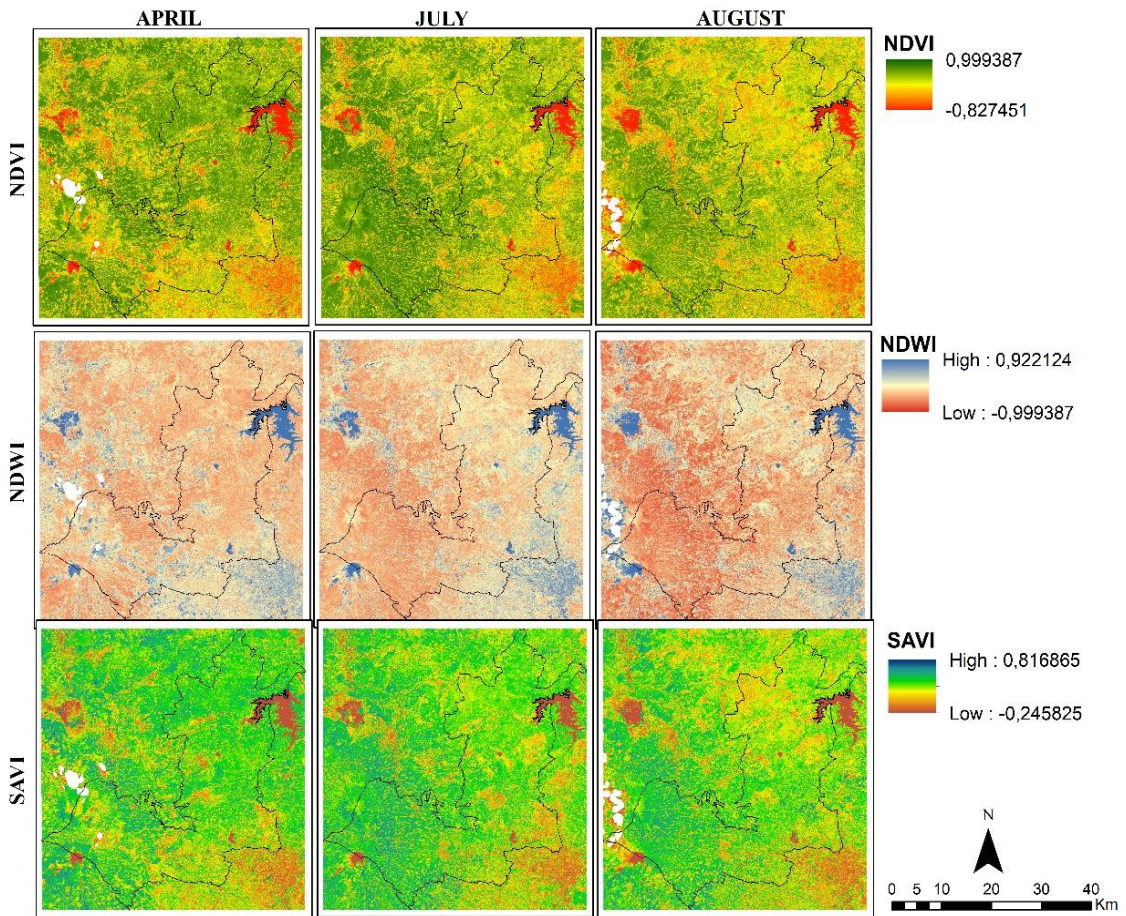
Fig. 7. NDVI, NDWI, and SAVI distribution over the study area in April, July, and August 2021.

Based on the statistical distribution of indices values (**Fig.8**), NDVI and SAVI values of the cropland showed similar ranges of value in the three months of observation. The best result of cropland classification was in August, with the shortest-range value. The median was similar to April's result, while July had the highest range and the lowest median. Built-up and water bodies were well defined in the three-month observations shown by the short range and similar median of NDVI and SAVI value. Forest was better in April and July results because the ranges of values increased in August. Bare land had a similar median value in the three months, but the distribution values varied in several parts of the area.

Furthermore, **Table 4** gives information on the classification accuracy derived from Sentinel-2 only dataset. In general, the classification produced an overall accuracy of 89% which was higher compared to the result of Sentinel-1 dataset (**Table 3**). Bare land and forest gave the highest values with 100% of user's accuracy, whereas built-up and water body area had 100% of producer's accuracy. For the cropland, it produced a user's and producer's accuracy of 89.66% and 83.87, respectively. In our findings, the use of vegetation indices increased the detection accuracy, supporting outcomes from earlier study (Panda et al., 2010). SAVI, in particular, reduces the noise of classification resulting from the soils and the moisture influences of the output (Panda et al., 2010). In this case, soil background influences the spectra of partially vegetated canopies, therefore, vegetation indices are required.
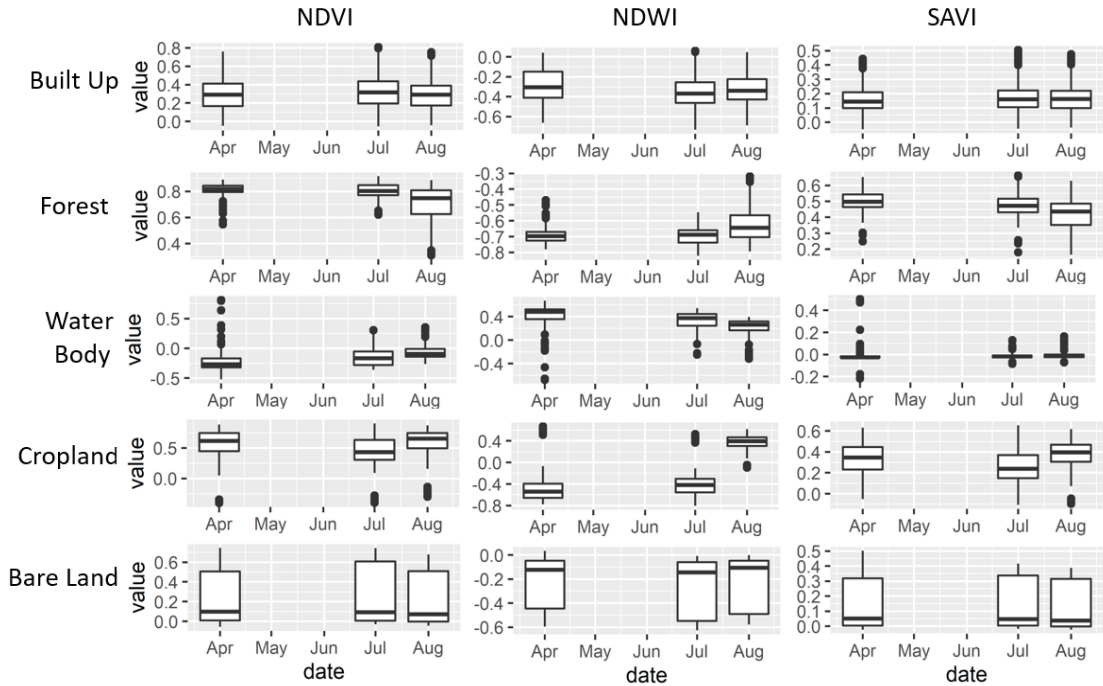
Fig. 8. Value distribution of NDVI, NDWI, and SAVI data for every land cover class.

Table 4.

Confussion matrix based on Sentinel-2.

Sentinel-2		Reference data					
		Built Up	**Forest**	**Water Body**	**Cropland**	**Bare Land**	**UA**
Classified data	Built Up	30	1	0	3	4	78.95
	Forest	0	26	0	2	0	92.86
	Water Body	0	0	15	0	0	100.00
	Cropland	0	2	0	26	1	89.66
	Bare Land	0	0	0	0	10	100.00
	PA	100.00	89.66	100.00	83.87	66.67	
Overall Accuracy		**89%**					

4.3. Cropland Area Distribution

The importance degree of variables used in RF classification is presented in **Fig.9**.
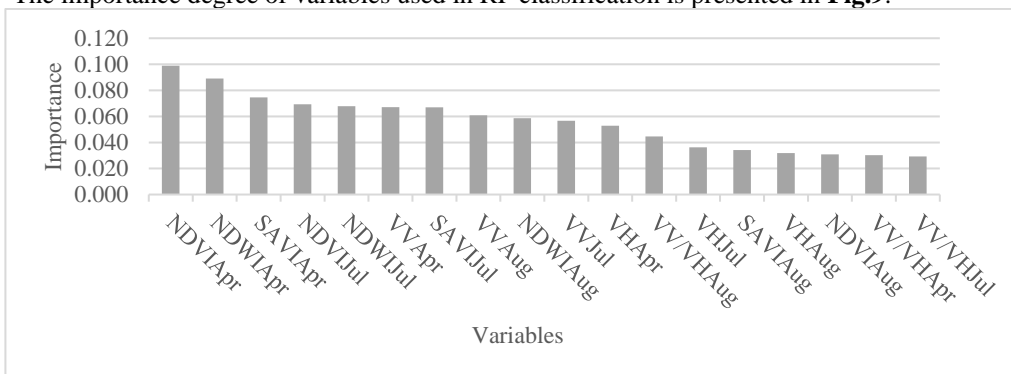
Fig. 9. Value of variable importance in Random Forest classifier.

It is shown that NDVI had the highest score among other 17 variables, followed by NDWI, and SAVI. In the first to ninth position, the high score was dominated by NDVI, NDWI, SAVI, and VV polarisation. From this finding, therefore, we selected those four variables as inputs for RF classification.

Fig.10 and **Fig.11** shows the actual condition of cropland plots in different characteristics of study area. In **Fig.10**, croplands in the upper slope of Mount Merapi and Merbabu are presented, showing the planted crops are dominated with a non-rice crop, such as vegetables and cassava. Meanwhile, the croplands located in the lower slope are shown in **Fig.11**. Croplands near to the water body were characterized by the irrigated rice field, as the high access and availability of water regardless the season (**Fig.11a**), whereas dry croplands were more prominent in the high populated urban area (**Fig. 11b**). Rice fields were also detected in the less populated urban area (**Fig. 11c**). The observed conditions from both figures, we can see that the croplands are attributed to the physical conditions, for example the topography, water, and agroecology characteristic (Widiyanto, 2019).

Fig. 10. Croplands condition in the upper slope of study area taken in August 2022.

Fig. 11. Cropland areas taken in August 2022. Characteristics: (a) near to water body, (b) in high populated urban region, and (c) in less populated urban region.

Implementation of RF classification with Sentinel-1 and Sentinel-2 dataset resulted a land cover map for Boyolali Regency (**Fig.12**) with an overall accuracy of 89% (**Table 5**). Cropland specifically, gained user's and producer's accuracy of 80% and 93.33%, respectively. Looking at the correct classified plots, this accuracy is higher in comparison to the previous scenarios using Sentinel-1 and Sentinel-2 alone for the input. Similarly, with RF method, a high classification accuracy was yielded for cropland mapping conducted by (Phalke et al., 2020).

Confussion matrix based on Sentinel-1 and 2.

Sentinel-1 and 2		Reference data					
		Built Up	**Forest**	**Water Body**	**Cropland**	**Bare Land**	**UA**
Classified data	Built Up	28	0	0	1	4	84.85
	Forest	0	27	0	1	0	96.43
	Water Body	0	0	15	0	0	100.00
	Cropland	2	3	0	28	2	80.00
	Bare Land	0	0	0	0	9	100.00
	PA	93.33	90.00	100.00	93.33	60.00	
Overall Accuracy		**89%**					

From **Fig.12**, it can be seen that dense built-up areas were in southern part of regency where the government officials are located. Water bodies, on the other hand, were only located at certain points and not evenly distributed. This condition may bring influences on the cropland characteristic based on the water use (irrigated and rainfed field) over the area.
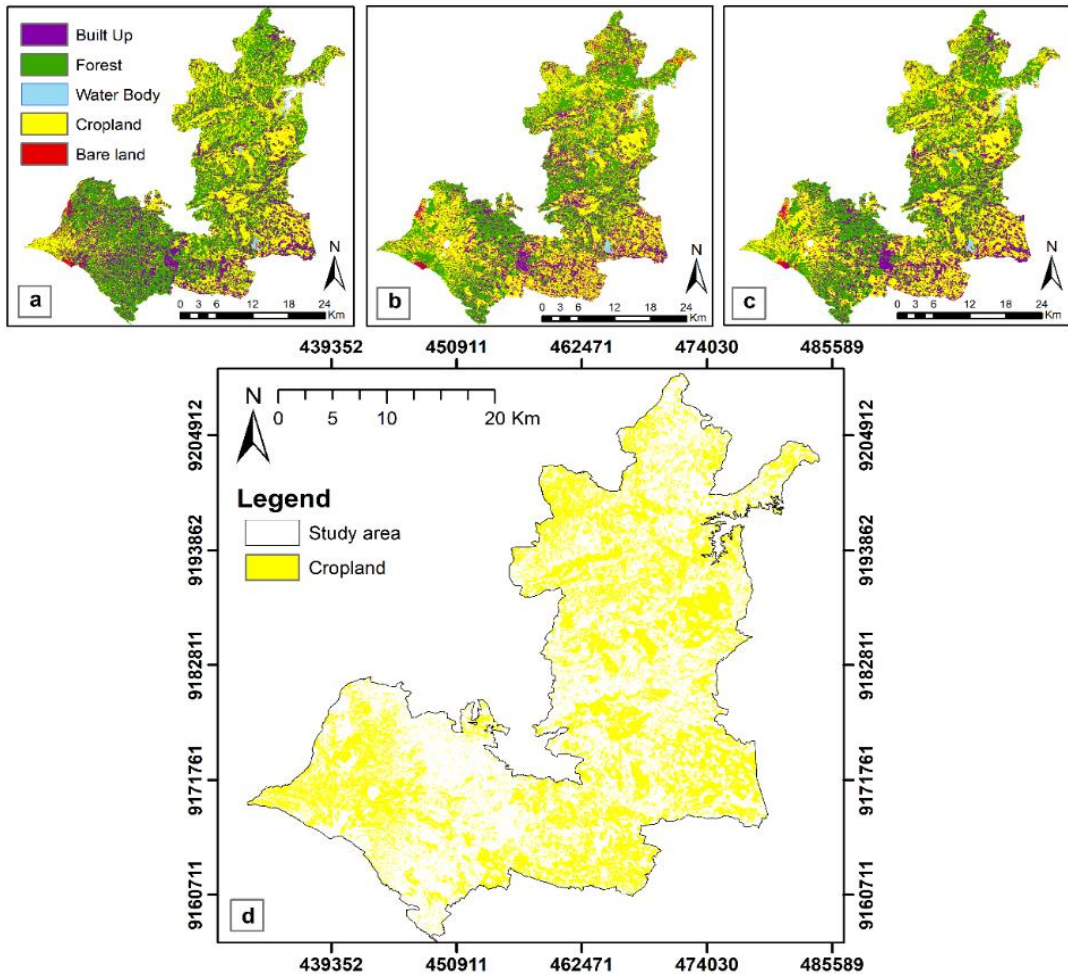
Fig. 12. Land covers from a) Sentinel-1; b) Sentinel-2; c) Sentinel-1 and -2. Figure d showing the cropland distribution based on Sentinel-1 and -2 data.

5. DISCUSSION

In this study, a different approach of mapping cropland area was taken. An integrated of optical data from Sentinel-2 and SAR data from Sentinel-1 was conducted to improve the accuracy of mapping. Based on our observation, we found that the integration of time-series optical and SAR data improved the capability of cropland detection. The proposed approach enables the combination of advantages form each data. Optical sensor of Sentinel-2 uses NIR, and green band that highlight the reflectance of vegetation. Additionally, implementation of vegetation indices also suppresses the influence of soil spectra in red band. On the other hand, Sentinel-1 with SAR sensor is beneficial for the observation in area with high cloud cover as in the mountainous area. The use of different polarisation, its combination, and the textural features could enhance the capability of identification for each land cover class (Haris et al., 2021; Priyono et al., 2022; Qi et al., 2012).

Based on our results, although there was no difference in overall accuracy between data of Sentinel-2 alone and the integration of Sentinel-1 and -2, the correct classified pixels for cropland and its PA were higher when using the integrated data. This implies the effectiveness of selected inputs performance for distinguishing cropland from other classes. Applying feature selection process, therefore, is still worth to do as it provided improvement in the accuracy of cropland detection. According to our findings, the evaluation of relative variable importance showed that all of image indices (NDVI, NDWI, and SAVI) from Sentinel-2 had the high contribution to the classification. The superior importance of NDVI was encouraging the demonstration of this index applied for mapping in a complex cropland region (Estel et al., 2016) and in a smallholder agricultural landscape (Rufin et al., 2022). In addition, only VV backscatter appeared more contributing than other variables from Sentinel-1 data. It is mainly due to VV better separability performance in discriminating water and bare land classes (el Mortaji et al., 2022) as similar finding was also reported by (Abdikan et al., 2016). The present results support the previous observation by Koley & Chockalingam (2022) that using SAR data alone is not sufficient for cropland mapping, despite its wide applications in mapping rice crop specifically because the distinct soil moisture condition in the start of crop growing season (Kuenzer & Knauer, 2012).

For future work, the use of this approach for mapping different cropland type is encouraged to fully understand the temporal, spectral, and spatial dynamic of each crop. Since our study lack information of crop type, this might contribute to the misclassification in the cropland. Apart from that, study on cropland mapping and monitoring is essential in regards to the food security matter. In fact, specifically in the study area from 2017 to 2019, its harvested area has decreased (**Fig. 13**). If the proportion of built-up that surrounds cropland is large and there is no policy preventing it, then this will affect crop productivity and further damage the food security. Findings on this study approach, therefore, could be beneficial for the improvement of cropland data accuracy supporting the regional planning policy.
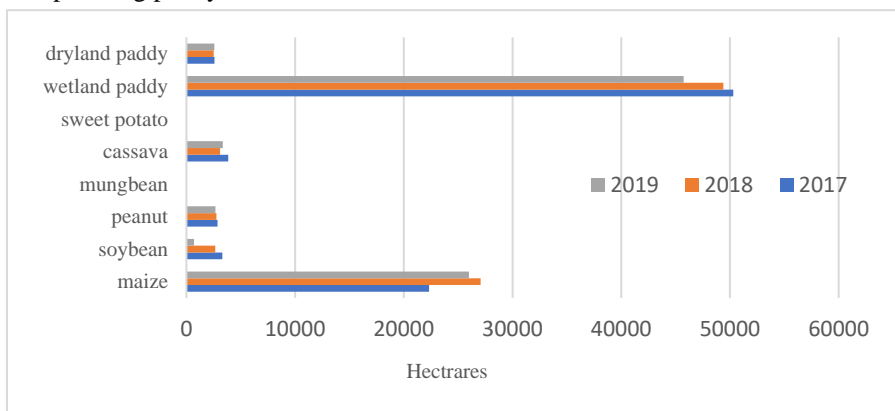
Fig. 13. Crop harvested area (in hectares) in Boyolali Regency from 2017 to 2019 (*Source: Regional statistic, 2018-2020*).

6. CONCLUSIONS

We demonstrated a different approach of mapping cropland area through an integrated of optical data from Sentinel-2 and Synthetic Aperture Radar (SAR) data from Sentinel-1. This particular approach was implemented in the mountainous slope area where generally high cloud cover occurs. Feature selection was also performed resulting NDVI, NDWI, SAVI, and VV polarization as the high importance variables. According to the results, random forest as a machine learning-based classifier could give overall accuracy of 78%, 89%, and 89% from Sentinel-1, Sentinel-2, and the combined data, respectively. Although the same overall accuracy between Sentinel-2 and combined data, the results show that the combination of optical and SAR data could increase the producer accuracy. Therefore, the results of the comparison show that the approach allow an improvement of cropland detection across the different slope area.

ACKNOWLEDGEMENTS

R E F E R E N C E S

Abdikan, S., Sanli, F. B., Ustuner, M., & Calò, F. (2016). LAND COVER MAPPING USING SENTINEL-1 SAR DATA. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLI-B7*, 757–761. https://doi.org/10.5194/ISPRS-ARCHIVES-XLI-B7-757-2016

Akbari, E., Boloorani, A. D., Samany, N. N., Hamzeh, S., Soufizadeh, S., & Pignatti, S. (2020). Crop Mapping Using Random Forest and Particle Swarm Optimization based on Multi-Temporal Sentinel-2. *Remote Sensing 2020, Vol. 12, Page 1449*, *12*(9), 1449. https://doi.org/10.3390/RS12091449

Argenti, F., Lapini, A., Alparone, L., & Bianchi, T. (2013). A tutorial on speckle reduction in synthetic aperture radar images. *IEEE Geoscience and Remote Sensing Magazine*, *1*(3), 6–35. https://doi.org/10.1109/MGRS.2013.2277512

Arjasakusuma, S., Kusuma, S. S., Rafif, R., Saringatin, S., & Wicaksono, P. (2020). Combination of Landsat 8 OLI and Sentinel-1 SAR Time-Series Data for Mapping Paddy Fields in Parts of West and Central Java Provinces, Indonesia. *ISPRS International Journal of Geo-Information 2020, Vol. 9, Page 663*, *9*(11), 663. https://doi.org/10.3390/IJGI9110663

Bhattacharya, S., Halder, S., Nag, S., Roy, P. K., & Roy, M. B. (2021). Assessment of Drought Using Multi-parameter Indices. *Lecture Notes in Civil Engineering*, *131 LNCE*, 243–255. https://doi.org/10.1007/978-981-33-6412-7_18/FIGURES/9

BPS Boyolali. (2021). *Harvesting Area and Rice Production in Boyolali*. Badan Pusat Statistik Boyolali.

Cai, Y., Lin, H., & Zhang, M. (2019). Mapping paddy rice by the object-based random forest method using time series Sentinel-1/Sentinel-2 data. *Advances in Space Research*, *64*(11), 2233–2244. https://doi.org/10.1016/J.ASR.2019.08.042.

Deepthi, R., Ravindranath, S., & Raj, K. G. (2018). Extraction of Urban Footprint of Bengaluru City Using Microwave Remote Sensing. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII–5*(November), 735–740. https://doi.org/10.5194/isprs-archives-xlii-5-735-2018

el Mortaji, N., Wahbi, M., Kazzi, M. A., Alaoui, O. Y., Boulaassal, H., Maatouk, M., Zaghloul, M. N., & el Kharki, O. (2022). High Resolution Land Cover Mapping and Crop Classification in the Loukkos Watershed (Northern Morocco): An Approach Using SAR Sentinel-1 Time Series. *Revista de Teledetección*, *2022*(60), 47–69. https://doi.org/10.4995/RAET.2022.17426

Estel, S., Kuemmerle, T., Levers, C., Baumann, M., & Hostert, P. (2016). Mapping cropland-use intensity across Europe using MODIS NDVI time series. *Environmental Research Letters*, *11*(2), 024015. https://doi.org/10.1088/1748-9326/11/2/024015

Fritz, S., See, L., Mccallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrecht, F., Schill, C., Perger, C., Havlik, P., Mosnier, A., Thornton, P., Wood-Sichra, U., Herrero, M., Becker-Reshef, I., Justice, C., Hansen, M., Gong, P., … Obersteiner, M. (2015). Mapping global cropland and field size. *Global Change Biology*, *21*(5), 1980–1992. https://doi.org/10.1111/GCB.12838

Gumma, M. K., Thenkabail, P. S., Teluguntla, P. G., Oliphant, A., Xiong, J., Giri, C., Pyla, V., Dixit, S., & Whitbread, A. M. (2019). Agricultural cropland extent and areas of South Asia derived using Landsat satellite 30-m time-series big-data using random forest machine learning algorithms on the Google Earth Engine cloud. *Https://Doi.Org/10.1080/15481603.2019.1690780*, *57*(3), 302–322. https://doi.org/10.1080/15481603.2019.1690780

Haris, N. A., Kusuma, S. S., Arjasakusuma, S., & Wicaksono, P. (2021). Comparison of sentinel-2 and multitemporal sentinel-1 sar imagery for mapping aquaculture pond distribution in the coastal region of brebes regency, central java, Indonesia. *Geographia Technica*, *16*(Special Issue), 128–137. https://doi.org/10.21163/GT_2021.163.10

Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., Jepsen, M. R., Kuemmerle, T., Meyfroidt, P., Mitchard, E. T. A., Reiche, J., Ryan, C. M., & Waske, B. (2016). A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring. *Remote Sensing 2016, Vol. 8, Page 70*, *8*(1), 70. https://doi.org/10.3390/RS8010070

Koley, S., & Chockalingam, J. (2022). Sentinel 1 and Sentinel 2 for cropland mapping with special emphasis on the usability of textural and vegetation indices. *Advances in Space Research*, *69*(4), 1768–1785. https://doi.org/10.1016/J.ASR.2021.10.020

Kuenzer, C., & Knauer, K. (2012). Remote sensing of rice crop areas. *Http://Dx.Doi.Org/10.1080/01431161.2012.738946*, *34*(6), 2101–2139. https://doi.org/10.1080/01431161.2012.738946

Mostafiz, R. B., Noguchi, R., & Ahamed, T. (2021). Agricultural Land Suitability Assessment Using Satellite Remote Sensing-Derived Soil-Vegetation Indices. *Land 2021, Vol. 10, Page 223*, *10*(2), 223. https://doi.org/10.3390/LAND10020223

Nguyen, D. B., & Wagner, W. (2017). European Rice Cropland Mapping with Sentinel-1 Data: The Mediterranean Region Case Study. *Water 2017, Vol. 9, Page 392*, *9*(6), 392. https://doi.org/10.3390/W9060392

Ningthoujam, R. K., Balzter, H., Tansey, K., Morrison, K., Johnson, S. C. M., Gerard, F., George, C., Malhi, Y., Burbidge, G., Doody, S., Veck, N., Llewellyn, G. M., Blythe, T., Rodriguez-Veiga, P., van Beijma, S., Spies, B., Barnes, C., Padilla-Parellada, M., Wheeler, J. E. M., … Bermejo, J. P. (2016). Airborne S-Band SAR for Forest Biophysical Retrieval in Temperate

Mixed Forests of the UK. *Remote Sensing 2016, Vol. 8, Page 609*, *8*(7), 609. https://doi.org/10.3390/RS8070609

Panda, S. S., Ames, D. P., & Panigrahi, S. (2010). Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques. *Remote Sensing 2010, Vol. 2, Pages 673-696*, *2*(3), 673–696. https://doi.org/10.3390/RS2030673

Phalke, A. R., & Özdoğan, M. (2018). Large area cropland extent mapping with Landsat data and a generalized classifier. *Remote Sensing of Environment*, *219*, 180–195. https://doi.org/10.1016/J.RSE.2018.09.025

Phalke, A. R., Özdoğan, M., Thenkabail, P. S., Erickson, T., Gorelick, N., Yadav, K., & Congalton, R. G. (2020). Mapping croplands of Europe, Middle East, Russia, and Central Asia using Landsat, Random Forest, and Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, *167*, 104–122. https://doi.org/10.1016/J.ISPRSJPRS.2020.06.022

Piao, Y., Jeong, S., Park, S., & Lee, D. (2021). Analysis of Land Use and Land Cover Change Using Time-Series Data and Random Forest in North Korea. *Remote Sensing 2021, Vol. 13, Page 3501*, *13*(17), 3501. https://doi.org/10.3390/RS13173501

Priyono, K. D., Saifuddin, A., Nugroho, F. S., & Jumadi, J. (2022). Identification of Mangrove Changes in the Mahakam Delta in 2007-2017 Using Alos/Palsar and Landsat. *International Journal of GEOMATE*, *23*(96), 77–84. https://doi.org/10.21660/2022.96.3312

Qi, Z., Yeh, A. G. O., Li, X., & Lin, Z. (2012). A novel algorithm for land use and land cover classification using RADARSAT-2 polarimetric SAR data. *Remote Sensing of Environment*, *118*, 21–39. https://doi.org/10.1016/J.RSE.2011.11.001

Rufin, P., Bey, A., Picoli, M., & Meyfroidt, P. (2022). Large-area mapping of active cropland and short-term fallows in smallholder landscapes using PlanetScope data. *International Journal of Applied Earth Observation and Geoinformation*, *112*, 102937. https://doi.org/10.1016/J.JAG.2022.102937

Sarono, S., Kurniawan, R., Yahya Ahmad Karim, N., Zaky Hadibasyir, H., Nurul Fadhilla, M., Arif Trisnanto, M., & Hamonangan Sinaga, J. (2015). REMOTE SENSING AND GIS INTEGRATION FOR PADDY PRODUCTION ESTIMATION IN BALI PROVINCE, INDONESIA. *Proceedings of the 36th Asian Conference on Remote Sensing 2015*.

Sashikkumar, M. C., Selvam, S., Karthikeyan, N., Ramanamurthy, J., Venkatramanan, S., & Singaraja, C. (2017). Remote Sensing for Recognition and Monitoring of Vegetation Affected by Soil Properties. *JOURNAL GEOLOGICAL SOCIETY OF INDIA*, *90*, 609–615. https://doi.org/10.1007/s12594-017-0759-8

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 6308–6325. https://doi.org/10.1109/JSTARS.2020.3026724

Singha, M., Dong, J., Zhang, G., & Xiao, X. (2019). High resolution paddy rice maps in cloud-prone Bangladesh and Northeast India using Sentinel-1 data. *Scientific Data 2019 6:1*, *6*(1), 1–10. https://doi.org/10.1038/s41597-019-0036-3

Susilo, B., & Harini, R. (2018). Spatial Analysis and Visualization of Geographic Access to Food in the Capital Area of Bulungan Regency, North Kalimantan Province. *Forum Geografi*, *32*(2), 146–155. https://doi.org/10.23917/FORGEO.V32I2.7070

Widiyanto, D. (2019). Local Food Potentials and Agroecology in Yogyakarta Special Province, Indonesia. *Forum Geografi*, *33*(1), 64–81. https://doi.org/10.23917/FORGEO.V33I1.7795

DIRECT GEOREFERENCING IN UNMANNED AERIAL VEHICLE USING QUASI-ZENITH SATELLITE SYSTEM

Agung SYETIAWAN[1] , Yunus SUSILO[2*] , Susilo SUSILO[1] , Surono SURONO[3], Wahono WAHONO[4,5], Yudha Ahmad SIDDIQ[6], Subekti HARTO[7], Yustisi LUMBAN-GAOL[1] , Abdurrahman ABDURRAHMAN[5], Sutrisno SUTRISNO[8] and Abdurrahman ABDURRAHMAN[9]

ABSTRACT:
To produce accurate topographic data, Unmanned Aerial Vehicle (UAV) still rely on Ground Control Points (GCPs) for georeferencing. However, using GCPs has several limitations, among others, related to the cost and time required for field measurements. In addition, not all areas are accessible for GCPs measurements due to poorly accessible terrain or security reasons. Direct georeferencing, a method to determine precise camera position and orientation in UAVs using Global Navigation Satellite System (GNSS) geodetic antenna. Post Processing Kinematic (PPK) or real-time coordinates can be applied to determine the camera position. One satellite that sends corrections to the rover on Earth is the Quasi-Zenith Satellite System (QZSS). This study aims to analyze the orthophoto accuracy of the results of direct georeferencing using precise coordinates from the QZSS satellites. The flight parameter was used at 60% sidelap and 80% overlap on an average flying altitude of 300 m above ground level resulting in 135 photos with a Ground Sampling Distance (GSD) value of 6 cm. The accuracy of direct georeferencing using QZSS horizontally and vertically was 1.134 m and 1.617 m, respectively. Meanwhile, the same metric results using conventional GCPs were 0.417 m horizontally and 0.419 m vertically. With these results, the horizontal accuracy of Direct Georeferencing using corrections from QZSS can be used for large-scale mapping of the 1: 5,000 class 1 scale, while vertical accuracy can be used for large-scale mapping of the 1: 5,000 class 3 scale. Direct georeferencing using QZSS corrections has the potential to support the acceleration of large-scale mapping activities in Indonesia.

Key-words: *Direct Georeferencing, Unmanned Aerial Vehicle, Quasi-Zenith Satellite System, Large-scale mapping*

[1]*Research Center for Geospatial, National Research and Innovation Agency (BRIN), Jakarta-Bogor Street Km. 46, Cibinong 16911, Indonesia, agungsyetiawan@gmail.com, susilosarimun@gmail.com, yust012@brin.go.id*
[2]*Geomatics Engineering, Dr. Soetomo University, Semolowaru 84 Surabaya 60111, Indonesia, yunus.susilo@unitomo.ac.id*
[3]*Magellan Systems Japan, Amagasaki Research Incubator Center, #315 7-1-3, Hyogo, Japan, surono@magellan.jp*
[4]*University of Muhammadiyah Malang, Tlogomas street No. 246, Malang 65145, Indonesia, wahono@umm.ac.id*
[5]*Moto Doro Teknologi, Bukit Cemara Tidar J4 No. 34 Sukun, Malang city, Indonesia, wahono@umm.ac.id, abdurrahman19940607@gmail.com*
[6] *PHW IV, Perhutani Divre Jawa Timur, Terusan Kawi No. 3, Malang, Indonesia, yudhaahmadsiddiq@gmail.com*
[7]*Licensed Cadastral Surveyor Service Office of Muchammad Masykur and partner, Kalila Residence C-10, Malang 65142, Indonesia, subekti_harto@yahoo.co.id*
[8]*Urban Navigasi Indonesia, Alam Bukit Raya resident C6 no 3, Randuagung, Gresik Regency, Indonesia, trisno.urbanav@gmail.com*
[9]*AMZ Geoinfo Solution, Gayungsari XI / 54, Surabaya, Indonesia, Abdurrahman9@gmail.com*
*Corresponding author's email: yunus.susilo@unitomo.ac.id

1. INTRODUCTION

The use of Unmanned Aerial Vehicle (UAV) technology for mapping activities is escalating since it offers aerial photos with very high resolution. Compared to other methods such as satellite sensors, UAV-based photogrammetry provides more benefits regarding data acquisition and the resulting temporal resolution (Liu et al., 2022). With this capability, UAVs are often used for rapid mapping purposes such as disaster management (Restas, 2015), precision farming (Candiago et al., 2015; Syetiawan & Haidar, 2019), landslides monitoring (Godone et al., 2020; D. Turner et al., 2015), and coastal changes (N Long et al., 2016; Nathalie Long et al., 2016).

The accuracy and precision of photogrammetry are highly dependent on various factors, including image quality, camera calibration, flight parameters, image processing algorithms, land cover and surface textures, and the intensity of sunlight (H. Zhang et al., 2019). To produce accurate topographic data, UAVs still rely on Ground Control Points (GCPs) for georeferencing (James et al., 2017; I. L. Turner et al., 2016). Meanwhile, the quality of GCPs depends on their density, which is the number and distribution (Sanz-Ablanedo et al., 2018). Georeferencing is the process of registering bundle adjustment results and photogrammetric processes to a particular coordinate system (Cramer et al., 2000; Kraus, 1993). However, using GCPs has several limitations, among others, related to the cost and time required for field measurements. In addition, not all areas are accessible for GCPs measurements due to poorly accessible terrain or security reasons, for example, swamp, glacier, or military areas.

Along with the development of current navigation satellite technologies, precise geodetic antennas can be integrated into UAV systems. The goal is that each aerial photo can be referenced at precise coordinates, a technique called direct georeferencing. The direct georeferencing process requires the device to have an accurate position up to cm-level to obtain a precise camera position and orientation in UAV measurements (Liu et al., 2022). Positioning computation methods, such as Post Processing Kinematic (PPK) or real-time coordinates where the corrections are obtained from radio signals or directly from satellites, can be applied to determine the camera position. One satellite that sends corrections to the rover on Earth is the Quasi-Zenith Satellite System (QZSS).

Some scholars have studied using QZSS for real-time kinematic positioning combined with multi-GNSS can improve positioning accuracy (Kitamura et al., 2014; Odolinski et al., 2015). QZSS is a regional augmentation satellite owned by Japan that is used as a complement to existing satellite systems like the Global Positioning System (GPS) or Beidou Satellite Navigation System (BDS) (Kubo et al., 2004; Wu et al., 2004). The QZSS satellite was first launched in 2010 and operated along with GPS and allows the combined processing of both systems (Hauschild et al., 2012). Furthermore, the raised number of QZSS satellites, becoming four in 2018, resulted in increased availability, reliability, integrity, and accuracy of positioning covering the Asia-Pacific region (Zaminpardaz et al., 2018; Q. Zhang et al., 2018). The QZSS satellites have three quasi-zenith orbits/QZO (QZS-1, QZS-2, QZS-4), but there is one satellite that has geostationary earth orbits/GEO (QZS-3) (Zhu et al., 2020). By 2023, QZSS is expected to be expanded to a seven-satellite system, which will enable it to provide better positioning, navigation, and timing (PNT) services (Li et al., 2021). It becomes an advantage for the direct georeferencing method in UAVs. This study aims to analyze the orthophoto accuracy of the results of direct georeferencing using precise coordinates from the QZSS satellites. The increased number of QZSS systems is expected to provide positive benefits, especially in rapid mapping applications using UAVs around Indonesia.

2. STUDY AREA

The aerial photo acquisition was carried out in Pandanlandung, Wagir sub-district, Malang Regency, Jawa Timur. The area of research is part of PERHUTANI, mainly dominated by vegetation, having an altitude of approximately 600 m above sea level and a total area of roughly 50 hectares. **Figure 1** illustrates the area of interest (AOI) in this study, where there are up to 100 m height differences between the eastern and western parts. Therefore, the aerial photos were captured at an average flying altitude of 300 m above ground level.
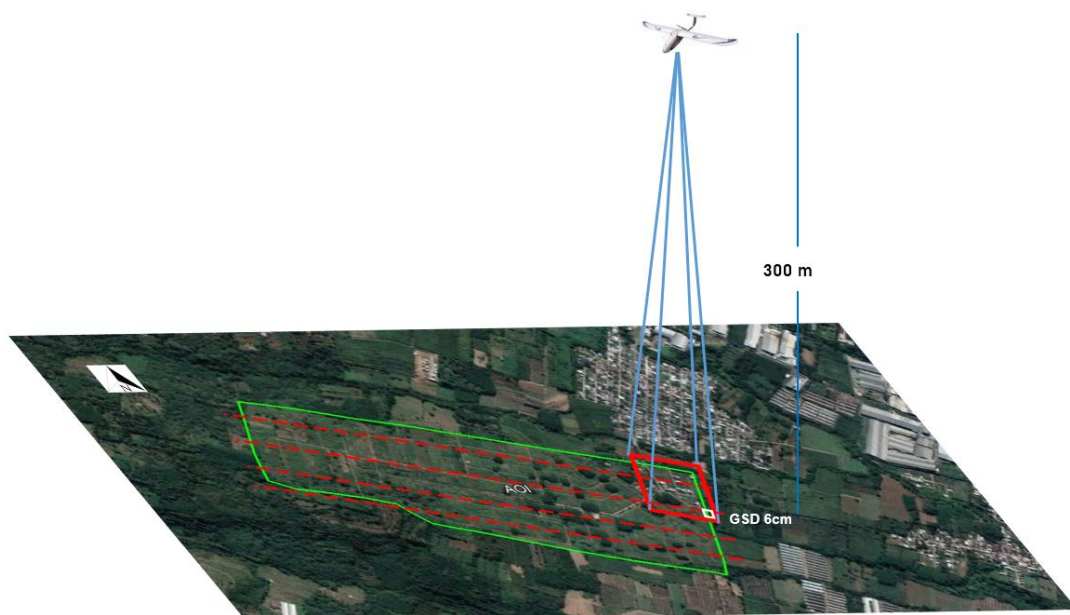
Fig. 1. Data acquisition parameters in the study area.

The flight was planned in the east-west direction with a Ground Sampling Distance (GSD) value of 6 cm. Subsequently, we used 60% sidelap and 80% overlap resulting in 135 photos with a flight duration of approximately 20 minutes.

3. DATA AND METHODS

3.1. Sensor and UAV specifications

For the aerial photo data collection, we used a fixed-wing UAV named FARM Mapper VTOL V2.0 2100 mm wingspan equipped with the Vertical Take-Off and Landing (VTOL) feature. Fixed-wing has the advantage of having more aerodynamic aircraft to fly longer than the multi-rotor type (Boon et al., 2017). **Table 1** (left) provides the complete specifications of the vehicle, and **Table 1** (right) depicts the specifications of the camera used. The UAV used a Sony A6000 type CMOS camera with a resolution of 24.3 Mega Pixels. This study used a mirrorless camera lens. Furthermore, the MSJ antenna was mounted on the top of the fixed-wing to receive the QZSS signal (can be seen in **Figure 2**). The offset between the MSJ antenna and non-metric camera was 13 cm vertically.

Fig. 2. QZSS antenna installation on Fixed-wing UAV.

Table 1.

UAV and Non-Metric Camera Specifications.

FARM Mapper VTOL V2.0		Sony Alpha ILCE A6000	
Airframe:	Epo+Fibercarbon (Wingspan 2100mm)	Lens	E-mount Sony 20mm fixed, F2.8
Flight Controller:	Pixhawk cube orange	Pixel	24.3 MP
GPS + Compass:	GPS here M8N	Sensor type	Sensor CMOS
Control System:	Manual and Auto	Dimension	120x67x45 mm
Radio Telemetry:	RFD 900x	Type	Mirrorless
Servo:	Emax 3504 MD	Sensor Optical	APS-C type (23.5 x 15.6 mm)
Motor:	4120 430 Kv*1 5008 400 Kv*4	Shutter speed	1/4000 to 30sec
Sensor:	Digital Airspeed sensor, Barometer, Magnetometer, 3 Axis Gyroscope	Processor	Bionz X image processor
Battery:	LiPo 16.000 MAh 6S (2unit)		

3.2. Control point parameter

This study compared orthophoto results from the direct georeferencing process using precise coordinates of QZSS and the conventional method using GCPs. Accordingly, we measured control points using geodetic GNSS and divided them into Ground Control Points (GCPs) and Independent Check Points (ICPs), as shown in **Figure 3**. There were 20 control points in total, of which eight points were GCPs, and the rest were ICPs. All control points were distributed evenly over the entire study area. The conventional orthophoto method uses GCPs for the georeferencing process and tests the accuracy of the results using ICPs. In contrast, the direct georeferencing method uses all control points to check the accuracy of the orthophoto result.

Control points were measured using a Leica GS14 and SOUTH Galaxy G1. The observation time of each point was one hour with an interval of five seconds for satellite recording data. The GNSS data processing was performed using the static differential method, which refers to a geodetic control point (ID: TTG.1290). This study used both satellite segments from GPS and GLONASS. **Table 2** presents complete GNSS data processing parameters to obtain the coordinates of each control point.

Table 2.

Control Point Processing Parameter.

Parameter	Description
Positioning mode	Static differential
Base	TTG.1290
Frequencies signal	Dual Frequency (L1 and L2)
Datum	SRGI 2013 (ITRF.08 epoch 2012)
Elevation Mask (°)	10
Satellite data interval	5 seconds
Satellite ephemeris	Broadcast
Satellite segment	GPS and GLONASS
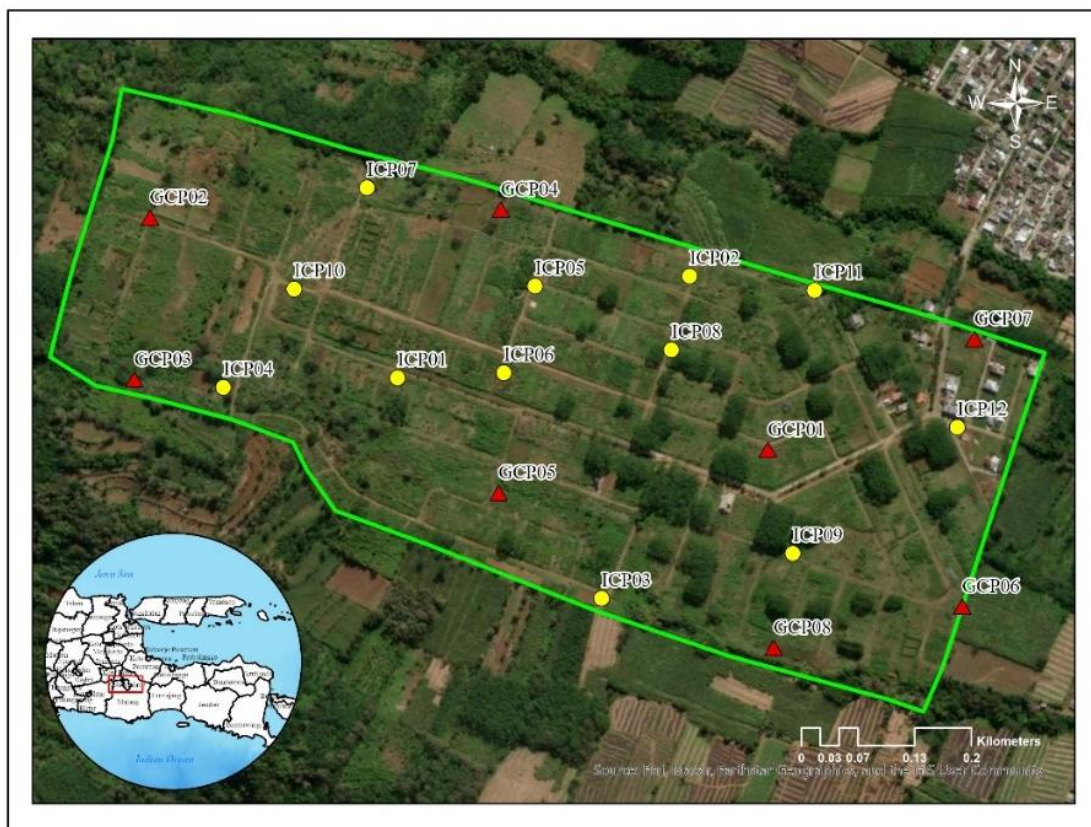
Fig. 3. Control points configuration in the study area. Red triangles represent GCPs, yellow circles are ICPs, and green lines represent areas of interest.

3.3. Accuracy assessment

This study assessed the accuracy of the orthophoto product from the direct georeferencing method and the use of GCPs by testing the geometric accuracy. The results were evaluated by comparing the coordinates of aerial photos with the control coordinates of the field measurements. The provisions for geometrical accuracy refer to the Regulation of the Head of the Geospatial Information Agency (BIG) No. 18 of 2021 regarding procedures for providing geospatial information. The geometric accuracy of the geospatial data is shown in **Table 3**. The geometric accuracy of the map is expressed in Circular Error with a confidence level of 90% (CE90) for the horizontal component and 90% Linear Error (LE90) for the vertical component. Measuring geometric accuracy was applied by calculating the Root Mean Square Error (RMSE) value. The RMSE values for the horizontal and vertical components can be seen in equations 1 and 2. Meanwhile, the calculation of the CE90 and LE90 values can be seen in equations 3 and 4, respectively:

$$\text{RMSE}_{\text{horizontal}} = \sqrt{\frac{\Sigma[(X_{data-}X_{check})^2 + (Y_{data-}Y_{check})^2]}{n}} \tag{1}$$

$$\text{RMSE}_{\text{vertical}} = \sqrt{\frac{\Sigma[(Z_{data-}Z_{check})^2]}{n}} \tag{2}$$

$$CE90 \ = \ 1{,}5175 \ x \ \text{RMSE}_{\text{horizontal}} \tag{3}$$

$$LE90 \ = \ 1{,}6499 \ x \ \text{RMSE}_{\text{vertical}} \tag{4}$$

where n is the total number of checkpoints on the map, X is the coordinates on the axis – X, Y is the coordinates on the axis – Y, Z is the coordinates on the axis – Z, $\text{RMSE}_{\text{horizontal}}$ is the error at horizontal position (XY), and $\text{RMSE}_{\text{vertical}}$ the error at vertical position (Z). The horizontal position in this study refers to the UTM Projected Coordinate System, while the vertical position refers to the height of the WGS84 ellipsoid.

Table 3.
Geometry Accuracy of Geospatial Data Based on the Regulation of the Head of BIG No. 18 of 2021.

No	Aspect	1:5,000 scale			1:1,000 scale		
		Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
1	Spatial resolution (m)	0.25	0.50	0.75	0.05	0.10	0.15
2	Hor. accuracy (CE90) (m)	1	2	3	0.20	0.40	0.60
3	Vert. accuracy (LE90) (m)	0.50	0.75	1	0.10	0.15	0.20

4. RESULTS AND DISCUSSIONS

Figure 4 shows the results of processing control points using commercial GNSS data processing software. The control point processing results show that the horizontal precision ranged from 0.007 m to 0.031 m, while the vertical precision ranged between 0.006 m and 0.098 m. Furthermore, the RMSE varied between 0.013 m and 0.039 m with the fixed solution of phase ambiguity for all control points. The accuracy results were sufficient for georeferencing or as a checkpoint.
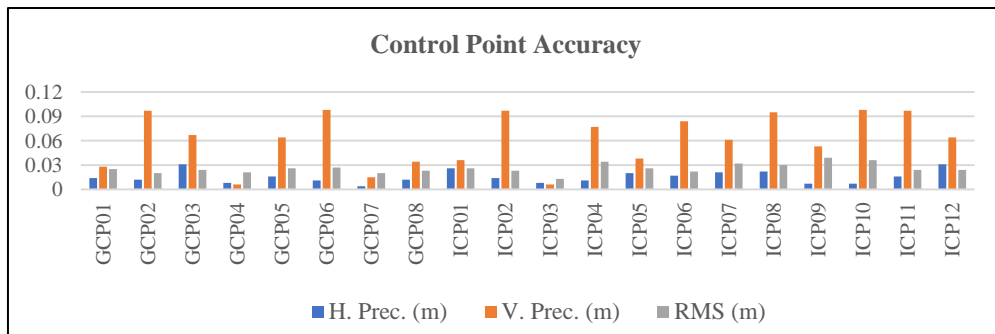
Fig. 4. Control point processing results.

The following process was to process aerial photos, where the geotagging process was first carried out to input QZSS precise coordinates into the photo. The precise coordinates of the QZSS satellite are used to determine the camera principal point position and external orientation parameters. **Figure 5** shows the reconstruction of aligned photos using accurate coordinates from the QZSS satellite. It shows that the is a misaligned photo, indicated by the small overlap area with the other photos. Furthermore, the tie point results of numerous pairs of overlapping photos are used to build dense point clouds (presented in **Figure 6**). The western part of the study area is densely covered with vegetations. In result, some places contain holes in the cloud due to imperfect tie point reconstruction. The total point cloud formed is 7,155,017 for the research area of 50 hectares.
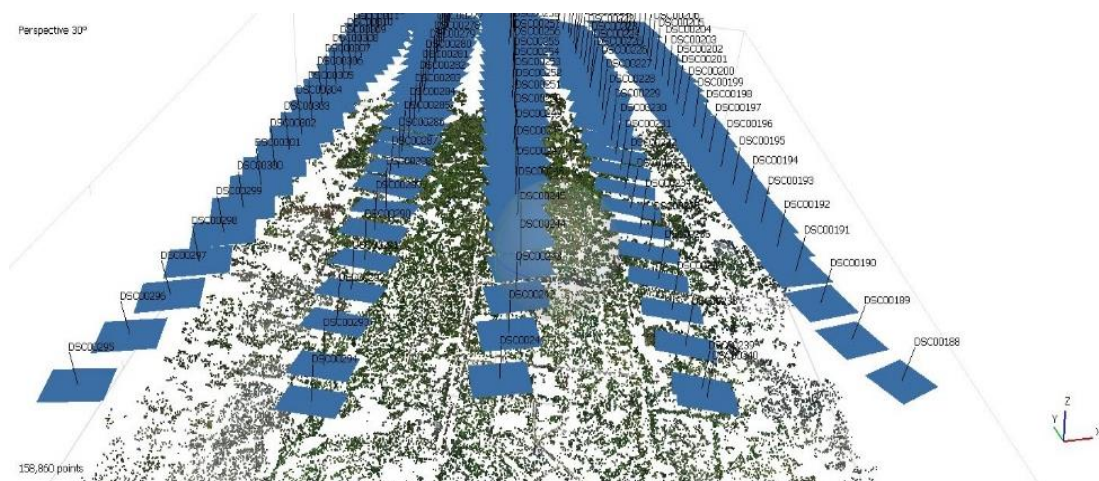
Fig. 5. Reconstruction of aligned photos using precise coordinates from the QZSS satellite.

Fig. 6. Build dense cloud process.

The main difference between direct georeferencing using QZSS and conventional GCP is the process of obtaining accurate coordinates. Before acquiring aerial photo, QZSS receiver needs to initiate satellites for 5-10 minutes to fix coordinates. In contrast, conventional GCP requires a total observation time of around 8 hours (for 8 control points) according to the accuracy of the needed control points. Direct georeferencing based on Inertial Measurement Unit (IMU) and GNSS is preferred due to the time and cost efficiency than GCP field deployment, surveying, and recognition in images.

Figure 7 shows orthophoto mosaic results using direct georeferencing and conventional GCPs. The accuracy of direct georeferencing using QZSS horizontally and vertically was 1.134 m and 1.617 m, respectively (**Table 4**). Meanwhile, the same metric results using conventional GCPs (**Table 5**) were 0.417 m horizontally and 0.419 m vertically. A study by Turner et al. (Turner et al., 2012) showed similar results, where direct georeferencing yielded horizontal accuracy of 1.2 m when conventional GCPs increased the accuracy up to 0.10-0.15 m. The integration of QZSS and GPS can produce accurate Real-Time Kinematic Precise Point Positioning at an accuracy of up to cm level with 4-satellite constellation (Asari et al., 2016). However, in the application of moving vehicles, faster-fixed coordinates are needed. There is still error correction signal from QZSS due to satellite connection to fast-flying UAV and environmental characteristics around the measurement such as

ionospheric effect. Moreover, Syetiawan et al. (Syetiawan et al., 2020) revealed that direct georeferencing could produce a horizontal accuracy of up to 4 cm using post-processing kinematic in a relatively small research area. The direct georeferencing method relies heavily on the accuracy of the GNSS used to record the camera position. Instead of post-processed the camera position, the coordinates were directly corrected from QZSS in this study.
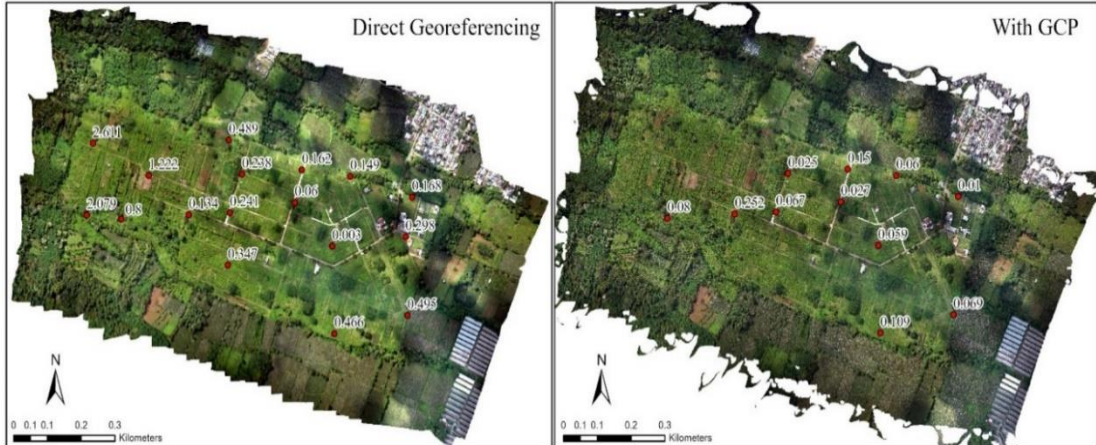
Fig. 7. Residue on horizontal components of direct georeferencing (left) and conventional GCPs (right).
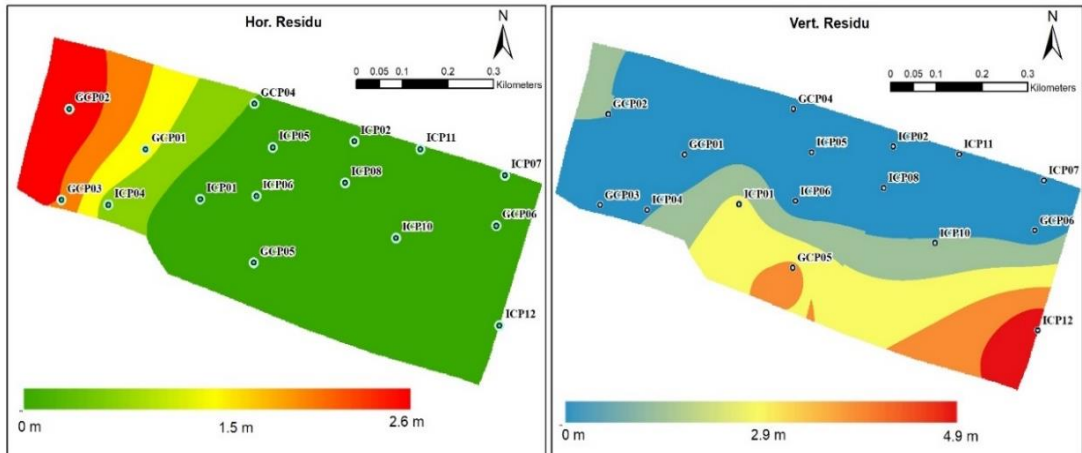
Fig. 8. Interpolation of residues of direct georeferencing method: Horizontal Residu (left); Vertical Residu (right).

The highest residue of the direct georeferencing result in the horizontal component was 2.6 m, while in the vertical component was 4.9 m. **Figure 7** illustrates the distribution of residue on the horizontal component, where GCPs-based mainly differ at the centimeter level, with the most considerable residue of 0.25 m at the ICP01 point. **Figure 8** shows the residual interpolation on the horizontal and vertical components using direct georeferencing. Overall, the most significant residue in horizontal occurred in the western part of the study area, which is steep, with an elevation difference of roughly 100 m compared to the east area.

In addition, direct georeferencing using QZSS corrections can be used for large-scale mapping of 1: 5,000 scale class 2 considering its horizontal accuracy and class 3 based on its vertical accuracy in **Table 3**. On the other hand, the horizontal and vertical accuracy of the conventional GCPs method

was suitable for large-scale mapping of 1: 1,000 scale class 3. Direct georeferencing using QZSS is less accurate than conventional GCP terms of accuracy. Direct georeferencing, however, is more effective at reducing field measurement time while ensuring accuracy. According to Sutanta et al. (2016), only a few areas have completed detailed mapping in Indonesia at scales of 1:5,000 and 1:10,000. Especially considering that Indonesia's land area is almost 2 million square kilometers. The need also increases along with the demand for topographic maps in detailed spatial plans. As a result, direct georeferencing with QZSS satellite correction has the potential to support the acceleration of large-scale mapping activities in Indonesia.

Table 4.

Horizontal and Vertical Accuracy of Direct Georeferencing.

ID point	Δx (m)	Δx^2	Δy (m)	Δy^2	Horizontal in m (Δx^2 + Δy^2)		Δz (m)	Vertical in m (Δz^2)
GCP01	-0.058	0.003	-1.104	1.219	1.223		0.202	0.041
GCP02	-0.081	0.007	-1.614	2.605	2.612		-1.004	1.008
GCP03	0.359	0.129	-1.397	1.950	2.079		-0.701	0.491
GCP04	-0.500	0.250	-0.489	0.239	0.489		-0.717	0.514
GCP05	0.397	0.157	-0.436	0.190	0.347		1.776	3.154
GCP06	-0.157	0.025	0.523	0.274	0.298		0.594	0.353
ICP01	0.023	0.001	-0.365	0.133	0.134		1.524	2.323
ICP02	-0.233	0.054	-0.328	0.108	0.162		-0.543	0.295
ICP04	0.139	0.019	-0.884	0.782	0.801		0.940	0.884
ICP05	-0.105	0.011	-0.477	0.227	0.238		-0.279	0.078
ICP06	-0.090	0.008	-0.483	0.233	0.241		0.440	0.194
ICP07	-0.325	0.105	0.250	0.063	0.168		0.164	0.027
ICP08	0.095	0.009	-0.226	0.051	0.060		0.137	0.019
ICP10	0.038	0.001	0.043	0.002	0.003		1.046	1.094
ICP11	-0.361	0.131	0.135	0.018	0.149		-0.075	0.006
ICP12	0.230	0.053	0.665	0.442	0.495		2.213	4.897
			Total		9.499			15.377
			Variance		0.559			0.961
			RMSE		0.748			0.980
			Accuracy		**1.134**			**1.617**

Table 5.

Horizontal and Vertical Accuracy of Conventional GCPs.

ID point	Δx (m)	Δx^2	Δy (m)	Δy^2	Horizontal in m (Δx^2 + Δy^2)		Δz (m)	Vertical in m (Δz^2)
ICP01	-0.049	0.002	0.500	0.250	0.252		-0.446	0.199
ICP02	-0.315	0.099	-0.224	0.050	0.150		0.150	0.022
ICP04	-0.026	0.001	0.282	0.080	0.080		-0.250	0.063
ICP05	-0.123	0.015	-0.101	0.010	0.025		0.236	0.056
ICP06	-0.256	0.066	0.030	0.001	0.067		0.418	0.175
ICP07	-0.074	0.005	-0.066	0.004	0.010		-0.173	0.030
ICP08	-0.130	0.017	-0.100	0.010	0.027		0.150	0.022
ICP09	-0.190	0.036	-0.270	0.073	0.109		-0.226	0.051
ICP10	-0.207	0.043	-0.128	0.016	0.059		0.084	0.007
ICP11	-0.244	0.060	0.011	0.000	0.060		-0.290	0.084
ICP12	0.042	0.002	-0.259	0.067	0.069		-0.024	0.001
			Total		0.907			0.710
			Variance		0.076			0.065
			RMSE		0.275			0.254
			Accuracy		**0.417**			**0.419**

5. CONCLUSIONS

The Direct Georeferencing method relies heavily on the accuracy of the GNSS used to record the camera position. The horizontal and vertical accuracy of the Direct Georeferencing method using QZSS was 1.134 m and 1.617 m. In accordance with BIG Head Regulation No. 18 of 2021, the horizontal and vertical accuracy is enough to create a large-scale map of 1: 5,000. The challenge ahead is to maintain the stability of the positioning satellite correction to obtain a precise position in real time. As a result, large-scale map creation can be completed quickly while retaining the quality of the final product.

This research still has a limitation because satellite coordinates recording does not coincide with the timing of the camera opening the lens. The UAV vehicle travels at a speed of 8-12 m/s, so it is not easy to synchronize the camera trigger with the sampling frequency of the GNSS receiver perfectly. Therefore, it is expected that there will be a fully integrated system in the future, where the GNSS antenna records the vehicle's position simultaneously with the camera records images.

R E F E R E N C E S

Asari, K., Matsuoka, S., & Amitani, H. (2016). QZSS RTK-PPP Application to Autonomous Cars. *Proceedings of the 29th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2016)*, 2136–2142. https://doi.org/10.33012/2016.14840

Boon, M. A., Drijfhout, A. P., & Tesfamichael, S. G. (2017). Comparison of a Fixed-Wing and Multi-Rotor Uav for Environmental Mapping Applications: a Case Study. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 47–54.

Candiago, S., Remondino, F., De Giglio, M., Dubbini, M., & Gattelli, M. (2015). Evaluating Multispectral Images and Vegetation Indices for Precision Farming Applications from UAV Images. *Remote Sensing*, *7*(4), 4026–4047. https://doi.org/10.3390/rs70404026

Cramer, M., Stallmann, D., & Haala, N. (2000). Direct Georeferencing using GPS/Inertial Exterior Orientations for Photogrammetric Applications. In T. Schenk & G. Vosselman (Eds.), *International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII, Part B3* (pp. 198–205). Amsterdam.

Godone, D., Allasia, P., Borrelli, L., & Gullà, G. (2020). UAV and Structure from Motion Approach to Monitor the Maierato Landslide Evolution. *Remote Sensing* , Vol. 12. https://doi.org/10.3390/rs12061039

Hauschild, A., Steigenberger, P., & Rodriguez-Solano, C. (2012). Signal, orbit and attitude analysis of Japan's first QZSS satellite Michibiki. *GPS Solutions*, *16*(1), 127–133. https://doi.org/10.1007/s10291-011-0245-5

James, M. R., Robson, S., & Smith, M. W. (2017). 3-D uncertainty-based topographic change detection with structure-from-motion photogrammetry: precision maps for ground control and directly georeferenced surveys. *Earth Surface Processes and Landforms*, *42*(12), 1769–1788. https://doi.org/10.1002/esp.4125

Kitamura, M., Ota, T., & Amano, Y. (2014). Improving Availability and Accuracy of Multi-GNSS Positioning Using QZSS. *Proceedings of the 27th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2014)*, 2341–2345. Tampa.

Kraus, K. (1993). *Photogrammetry*. Retrieved from https://books.google.co.id/books?id=dJlQAQAAIAAJ

Kubo, N., Wu, F., & Yasuda, A. (2004). Integral GPS and QZSS Ambiguity Resolution. *TRANSACTIONS OF THE JAPAN SOCIETY FOR AERONAUTICAL AND SPACE SCIENCES*, *47*(155), 38–43. https://doi.org/10.2322/tjsass.47.38

Li, X., Pan, L., & Yu, W. (2021). Assessment and Analysis of the Four-Satellite QZSS Precise Point Positioning and the Integrated Data Processing With GPS. *IEEE Access*, *9*, 116376–116394. https://doi.org/10.1109/ACCESS.2021.3106050

Liu, X., Lian, X., Yang, W., Wang, F., Han, Y., & Zhang, Y. (2022). Accuracy Assessment of a UAV Direct Georeferencing Method and Impact of the Configuration of Ground Control Points. *Drones* , Vol. 6. https://doi.org/10.3390/drones6020030

Long, N, Millescamps, B., Pouget, F., Dumon, A., Lachaussée, N., & Bertin, X. (2016). Accuracy Assessment of Coastal Topography Derived from UAV Images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLI-B1*, 1127–1134. https://doi.org/10.5194/isprs-archives-XLI-B1-1127-2016

Long, Nathalie, Millescamps, B., Guillot, B., Pouget, F., & Bertin, X. (2016). Monitoring the Topography of a Dynamic Tidal Inlet Using UAV Imagery. *Remote Sensing* , Vol. 8. https://doi.org/10.3390/rs8050387

Odolinski, R., Teunissen, P. J. G., & Odijk, D. (2015). Combined BDS, Galileo, QZSS and GPS single-frequency RTK. *GPS Solutions*, *19*(1), 151–163. https://doi.org/10.1007/s10291-014-0376-6

Restas, A. (2015). Drone Applications for Supporting Disaster Management. *World Journal of Engineering and Technology*, *03*, 316–321. https://doi.org/10.4236/wjet.2015.33C047

Sanz-Ablanedo, E., Chandler, J. H., Rodríguez-Pérez, J. R., & Ordóñez, C. (2018). Accuracy of Unmanned Aerial Vehicle (UAV) and SfM Photogrammetry Survey as a Function of the Number and Location of Ground Control Points Used. *Remote Sensing* , Vol. 10. https://doi.org/10.3390/rs10101606

Sutanta, H., Aditya, T., & Astrini, R. (2016). Smart City and Geospatial Information Availability, Current Status in Indonesian Cities. *Procedia - Social and Behavioral Sciences*, *227*, 265–269. https://doi.org/10.1016/j.sbspro.2016.06.070

Syetiawan, A., Gularso, H., Kusnadi, G. I., & Pramudita, G. N. (2020). Precise Topographic Mapping Using Direct Georeferencing in UAV. *IOP Conf. Series: Earth and Environmental Science*, 1–10.

Syetiawan, A., & Haidar, M. (2019). Pemetaan Perkebunan Sawit Rakyat dari Foto Udara Non Metrik Menggunakan Analisis Berbasis Objek. *Majalah Ilmiah Globë*, *21*(1), 53–62. https://doi.org/10.24895/MIG.2019.21-1.990

Turner, D., Lucieer, A., & De Jong, S. M. (2015). Time Series Analysis of Landslide Dynamics Using an Unmanned Aerial Vehicle (UAV). *Remote Sensing* , Vol. 7. https://doi.org/10.3390/rs70201736

Turner, D., Lucieer, A., & Watson, C. (2012). An Automated Technique for Generating Georectified Mosaics from Ultra-High Resolution Unmanned Aerial Vehicle (UAV) Imagery, Based on Structure from Motion (SfM) Point Clouds. *Remote Sensing* , Vol. 4. https://doi.org/10.3390/rs4051392

Turner, I. L., Harley, M. D., & Drummond, C. D. (2016). UAVs for coastal surveying. *Coastal Engineering*, *114*, 19–24. https://doi.org/10.1016/j.coastaleng.2016.03.011

Wu, F., Kubo, N., & Yasuda, A. (2004). Performance analysis of GPS augmentation using Japanese Quasi-Zenith Satellite System. *Earth, Planets and Space*, *56*(1), 25–37. https://doi.org/10.1186/BF03352488

Zaminpardaz, S., Wang, K., & Teunissen, P. J. G. (2018). Australia-first high-precision positioning results with new Japanese QZSS regional satellite system. *GPS Solutions*, *22*(4), 101. https://doi.org/10.1007/s10291-018-0763-5

Zhang, H., Aldana-Jague, E., Clapuyt, F., Wilken, F., Vanacker, V., & Van Oost, K. (2019). Evaluating the potential of post-processing kinematic (PPK) georeferencing for UAV-based structure-from-motion (SfM) photogrammetry and surface change detection. *Earth Surface Dynamics*, *7*(3), 807–827. https://doi.org/10.5194/esurf-7-807-2019

Zhang, Q., Yang, W., Zhang, S., & Yao, L. (2018). Performance Evaluation of QZSS Augmenting GPS and BDS Single-Frequency Single-Epoch Positioning with Actual Data in Asia-Pacific Region. *ISPRS International Journal of Geo-Information*, *7*(5). https://doi.org/10.3390/ijgi7050186

Zhu, S., Yue, D., He, L., Liu, Z., & Chen, J. (2020). Comprehensive analysis of compatibility between QZSS and GPS in Asia-Pacific region: Signal quality, RTK and PPP. *Advances in Space Research*, *66*(2), 395–411. https://doi.org/10.1016/j.asr.2020.04.003

SURFACE HEAT FLUX ASPECT ON THE VARIABILITY OF SEA SURFACE TEMPERATURE AND CHLOROPHYLL-A ALONG THE SOUTHERN COAST OF JAVA

Husein ALFARIZI [1] , *Anindya WIRASATRIYA* [1*] , *Kunarso KUNARSO* [1] , *Muhammad Rais ABDILLAH* [2] *and Dwi HARYANTI* [3]

ABSTRACT:
Indonesia as a region located in the tropics gets a greater heat distribution than the other hemisphere and has an important role in the phenomenon of atmospheric and ocean interactions in the Indo-Pacific region. The heat exchange between the ocean and the atmosphere affects the dynamics of both. The southern coast of Java is known as the upwelling area which is driven by the variability of Ekman transport and Ekman pumping. The present study aims to investigate the effect of heat flux variability on sea surface temperature variability and chlorophyll-a in surface upwelling areas along the Southern coast of Java which was less observed in the previous study. The study was conducted with a quantitative descriptive approach through climatological spatial and temporal data processing for 10 years from 2007 – 2016. The data used are Shortwave Radiation, Longwave Radiation, Latent Heat Flux, Sensible Heat Flux, Sea Surface Temperature, Chlorophyll-a, Surface Wind, and Mixed Layer Depth. The results show that the Southern coast of Java receives an average heat of 547.8 W/m2 per year. Net Heat Flux fluctuations are dominated by heat intake by Shortwave Radiation and heat release by Latent Heat Flux. Net Heat Flux has a very strong relationship with sea surface temperature with the best correlation of 0.84 and 0.83 at lag+2 and lag+3 months indicating that Net Heat Flux plays an important role in modulating changes in sea surface temperature in the next 2-3 months. A significant increase in chlorophyll-a occurred after the Net Heat Flux was positive or there was ocean heating which caused the shoaling of Mixed Layer Depth, resulting in primary productivity in the east monsoon along with nutrient rich entrainment to the surface by EMT and EPV.

Key-words: *Surface Heat Flux, SST, Chlorophyll-a, Upwelling, Southern coast of Java*

1. INTRODUCTION

The oceans play a very important role in balancing the heat energy received from the sun. It is estimated around 93% of the heat received by the earth since almost a century ago is stored in the oceans. Understanding the heat flux exchange is the key to how this balance mechanism occurs (Levitus et al., 2012). This heat exchange between the atmosphere and the ocean is known as the surface heat flux, which indicates the amount of heat absorbed or released in the ocean (Brunke et al., 2011). It is a major factor determining energy content and plays an important role in explaining air-sea interactions on a global and regional scale (Huang, 2016; Cronin et al., 2019). As a region located in the tropics, the Indonesian Seas gets a greater heat distribution than the other area on earth and has an important role in the sea-air interaction in the Indo-Pacific region. Several studies mentioned that the condition of surface heat flux in Indonesia is a key in determining the variability of ocean heat content in the Indian Ocean (Vranes et al., 2002; Lee et al., 2015; Gruenburg and Gordon., 2018). Indonesian waters also act as a region that connects the flow of heat content in the Pacific Ocean and

---

[1] *Department of Oceanography, Faculty of Fisheries and Marine Science, Universitas Diponegoro, Semarang, Indonesia, husein.alfarizi@gmail.com, * Corresponding author anindyawirasatriya@lecturer.undip.ac.id, kunarso@lecturer.undip.ac.id*

[2] *Atmospheric Science Research Group, Institut Teknologi Bandung, Bandung, Indonesia, m.rais@itb.ac.id*

[3] *Department of Marine Science, Faculty of Fisheries and Marine Science, Universitas Diponegoro, Semarang, Indonesia, dwiharyanti@lecturer.undip.ac.id*

the Indian Ocean. The heat received in Indonesian waters or coming from the Pacific Ocean is discharged to the Indian Ocean by the Indonesian throughflow (ITF) current.

The southern coast of Java is located in the eastern part of the Indian Ocean and is part of the outflow route for the ITF current from the Pacific Ocean. The seas along the southern coast of Java are directly opposite to the open sea so it has complex ocean characteristics that are influenced by various factors of atmospheric and oceanic dynamics. The southern coast of Java is also a fertile area with upwelling events that occur during Southeast monsoon season (Wirasatriya et al. 2021). Many studies have been conducted on the upwelling mechanism in Southern Java Sea. Susanto and Marra (2005) demonstrated the effect of 1997/98 El Niño to escalate the Chl-a concentration along the southern coast of Java. Wirasatriya et al. (2020) showed the role of Ekman transport and Ekman pumping to the variability of upwelling along the southern coast of Java. The impact of upwelling along the southern coast of Java for fisheries productivity has been demonstrated by Kunarso et al. (2012); Lahlali et al. (2019); and Sukresno et al. (2018). However, studies on how surface heat flux plays a role in the variability of sea surface temperature and chlorophyll-a concentration have not been widely discussed, particularly along the southern coast of Java. Therefore, this study aims to identify how the distribution of the surface heat flux and its effect on the variability of sea surface temperature and chlorophyll-a as the oceanic parameters for determining the location of upwelling in the southern coast of Java. Understanding the heat flux aspect of the upwelling along the southern coast of Java may contribute to the fisheries management in this area.

2. STUDY AREA

Our study area is located at the southern coast of Java which is part of the Fishing management Zone of Indonesia 573 (WPP RI 573). It is in the southern part of Western Indonesia Region and lies between 7°S - 12°S and between 105°E - 115°E (**Fig. 1**). WPP RI 573 is an important fishing ground for various big pelagic fish such as tuna (Wibowo et al. 2019). The southern coast of Java Sea is at the eastern part of the Indian Ocean and considered as the exit route for ITF from Bali and Lombok Straits.
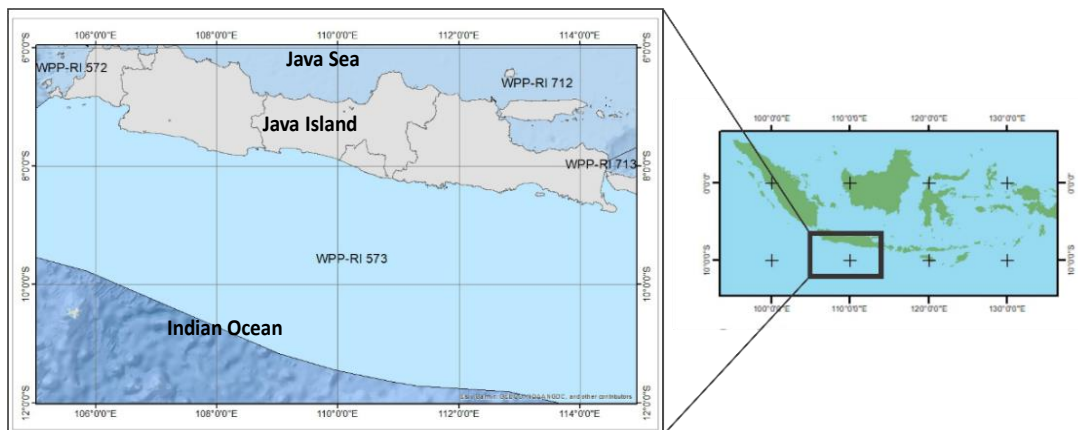
Fig. 1. Study area of the southern coast of Java. WPP-RI 573 is Fishing management Zone of Indonesia 573.

The climatic system is dominated by monsoons because of the difference in seasonal air pressure variability between the Asian and the Australian continents. The characteristic of Southeast (Northwest) monsoon is denoted by southeasterly (northwesterly) wind which blows from Australia (Asia) to Asia (Auistralia), brings dry (humid) air, and causes dry (rainy) season in most areas in Indonesia (Chang et al., 2005; Chang et al., 2006; Alifdini et al., 2021). December, January and February (DJF) represent the Northwest monsoon while June, July and August (JJA) are Southeast monsoon. Transition I and II are represented by March, April, May (MAM) and September, October, November (SON), respectively.

3. DATA AND METHODS

3.1. Reanalysis Data

3.1.1. Surface Heat Flux

This research uses turbulent flux and radiative flux as a component of surface heat flux. Turbulent flux is taken through Objective Analyzed Air Sea Flux (OAFlux) images which consist of sensible heat flux and latent heat flux (Yu and Weller, 2007). For an in-depth look behind the estimation of the turbulent data, humidity and air-ocean temperature data are included. Radiative flux data was obtained through the official website of the International Satellite Cloud Climatology Project (ISCCP) with the product type H Series in the form of shortwave radiation and longwave radiation (Zhang et al., 2004). All data has a spatial resolution of 1° x 1° with the estimated error values available. Data was downloaded from January 1, 2007 to December 31, 2016 with a global coverage area which will then be cut according to the study area coverage.

3.1.2. Sea Surface Temperature (SST)

Surface temperature data used in this study is OISST (Optimally Interpolated Sea Surface Tempetarure) images provided by REMMS (Remote Sensing System). OISST is level 4 data resulting from the combination of microwave (MW) and infrared (IR) sensors so that it has high resolution and complete data (Remote Sensing System, 2017). The downloaded images are daily composite data with a spatial resolution of 0.08° x 0.08° equivalent to 8 km x 8.

3.1.3. Chlorophyll-a (Chl-a)

This study uses chlorophyll-a data obtained from Ocean Color - CCI images (Sathyendranath et al., 2019). This is daily data with a spatial resolution of 0.04° x 0.04° equivalent to 4 km x 4 km.

3.1.4. Sea Surface Wind

This study collects wind data using Cross-Calibrated Multi-Platform Wind Vector Analysis (CCMP) images version 2 provided by REMMS (Remote Sensing System) (Atlas et al., 2011). The downloaded images is daily composite data with a spatial resolution of 0.25° x 0.25° equivalent to 25 km x 25 km.

3.1.5. Mixed Layer Depth (MLD)

The Mixed Layer Depth data used in this study is Level 4 data provided by Marine Copernicus with the product code GLOBAL_MULTIYEAR_PHY_001_030. This data has a resolution of 0.08° x 0.08° and 50 level standard depth (Drévillon et al., 2021).

3.2. Net Heat Flux Balance Calculation

Net heat flux is calculated by summing up the radiative flux and turbulent flux that ocean received and released:

$$Qnet = Qsw + Qlw + Qlh + Qsh + Qadv \tag{1}$$

where Q_{Net}, Q_{SW}, Q_{LW}, Q_{SH}, and Q_{LH} are net heat flux, shortwave radiation, longwave radiation, sensible heat flux and latent heat flux, respectively.

The calculation of net heat flux in this study did not involve the $Qadv$ factor due to data limitations. Positive heat flux means downward flux into the ocean while negative flux means upward flux to the atmosphere (Wirasatriya et al., 2019). The result of this formula will show the heat balance in the ocean that applied to each data grid to produce its spatial and temporal distribution.

3.3. Schematic Flowchart of Method

The method of this study is summarized in the flowchart (**Fig. 2**).
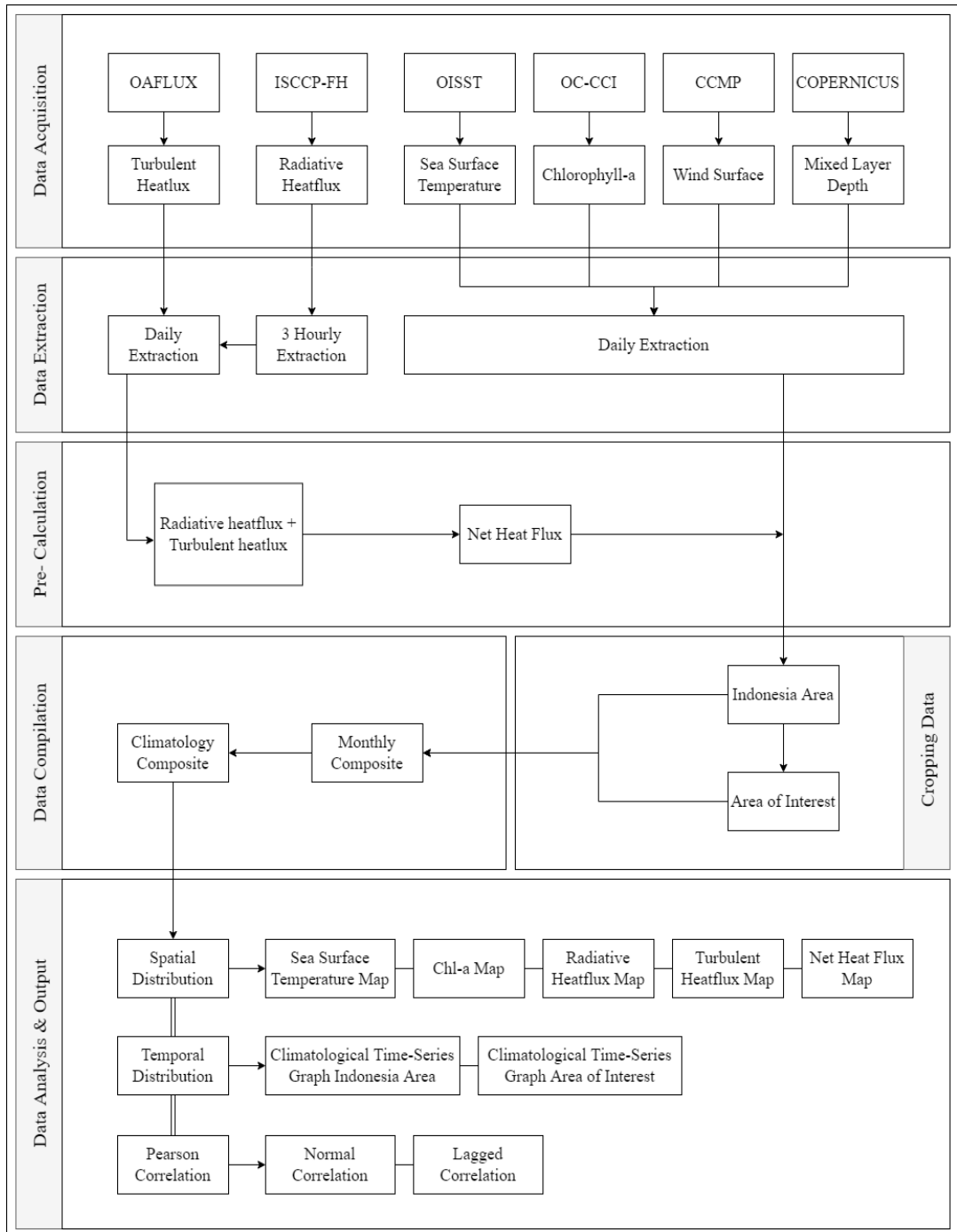
Fig. 2. Schematic flowchart of method.

3.4. Composite Data Calculation

To investigate the seasonal variation of surface heat flux and other parameters, we constructed monthly climatology data. First, all parameters were composited into monthly means and then used to derive monthly climatology by using the following formula (Wirasatriya et al. 2017):

$$\bar{x}b(x,y) = \frac{1}{mh}\sum_{i=1}^{mh} xi(x,y,t) \tag{2}$$

where *xb(x,y)* is monthly mean value or monthly climatology value at position *(x,y)*, *xi* is the value of the data at *(x,y)* position and time *t*.

Moreover, *mh* is number of data in 1 month and number of monthly data in 1 period of climatology (i.e., from 2007 to 2016 = 10 data) for monthly calculation and monthly climatology calculation respectively. Pixel *xi* is excluded in the calculation if it has a gap on data.

3.5. Pearson Correlation Analysis

Pearson correlation is used to determine the relationship between two variables, such as the linear relationship between *NHF* and *SST* or *Chl-a*, as well as other variables. The correlation calculated using the formula:

$$r = \frac{N(\sum XY)-(\sum X \sum XY)}{\sqrt{(N(\sum X^2-(\sum X)^2)-(N(\sum Y^2)-(\sum Y)^2)}} \tag{3}$$

where *r, x, y,* and N are the correlation coefficient values, the value of the first variable, the value of the second variable, and the total count of data respectively.

The output is in the form of values with a range of *0 – (±)1* indicating weak – strong relationship. We also performed lagged correlation to examine the possibility of the delaying time of the correlation among parameters.

4. RESULTS AND DISCUSSIONS

4.1. Results

4.1.1. Seasonal Variation of Sea Surface Temperature

The variation of sea surface temperature for 10 years (2007-2016) along the southern coast of Java shows that the sea surface temperature ranges from 23-30 $^{\circ}$C as shown in **Fig. 3**. The distribution of sea surface temperature along the southern coast of Java has a seasonal pattern that has its own characteristic. In the rainy season (December, January, February), the distribution of sea surface temperature is relatively warm with an average temperature ranging from 28.7 - 28.9 ℃. In the first transitional season (March, April, May), the sea surface temperature along the southern coast of Java reached its warmest temperature in April with an average temperature of 29.2 ℃.

Entering the dry season (June, July, August), the sea surface temperature gradually drops to its lowest temperature in August with an average of 26.2 ºC. In the second transitional season (September, October, November), the sea surface temperature increased again until the average temperature ranged from 26.3 – 28.2 C.
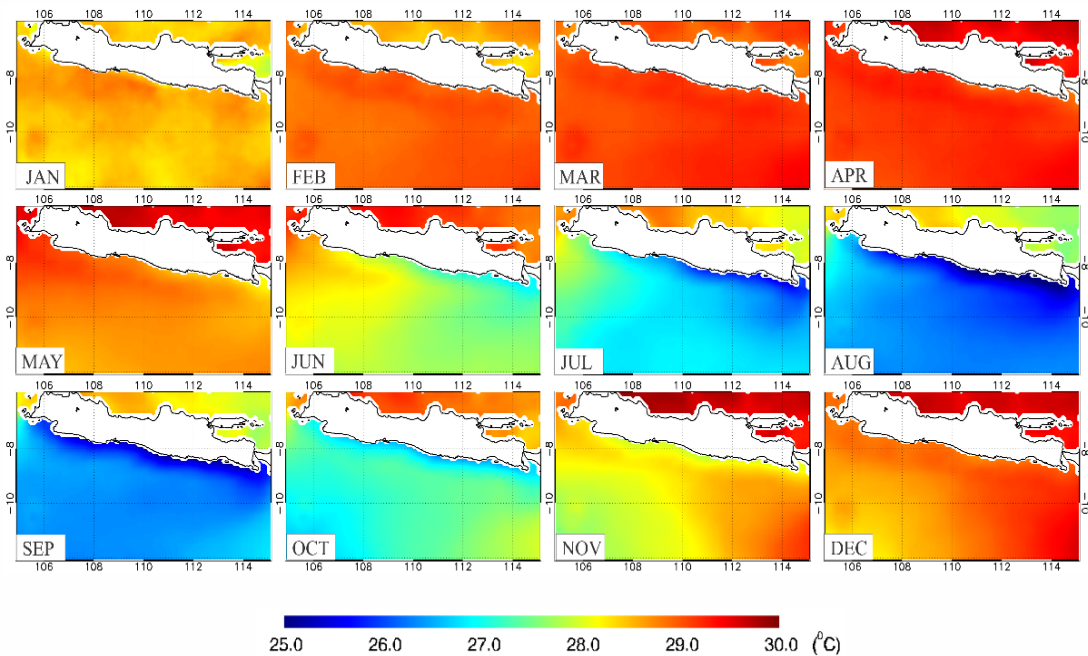
Fig. 3. Monthly climatology of SST along the southern coast of Java.

4.1.2. Seasonal Variation of Chlorophyll-a Concentration

The concentration of chlorophyll-a is the result of primary productivity that occurs in the waters. Primary productivity triggers the growth of phytoplankton as the largest biomass in the marine food chain. Each phytoplankton has a color pigment that is used to carry out photosynthesis, namely chlorophyll-a so that the more phytoplankton in the waters, the greater the concentration of chlorophyll-a. Data retrieval of chlorophyll-a on the satellite using the wavelength reflectance method. The value of chlorophyll-a concentration along the coast in **Fig. 4** could be reflectance data from suspended material in the waters. However, since the study area covers areas near the coast and the high seas, the reflectance data other than chlorophyll-a can be tolerated as an average value.

The seasonal variation of chlorophyll-a for 10 years (2007-2016) along the southern coast of Java shows the distribution of chlorophyll-a ranging from 0.1 – 2 mg/m^3. The distribution of chlorophyll-a only has one peak in one year. Chlorophyll-a concentration reached its highest peak in September with an average value of 0.55 mg/m^3. Meanwhile, the lowest average value was found in February at 0.1 mg/m^3. During the rainy season until the first transition period, the concentration of chlorophyll-a did not show a significant change while an increase began to occur during the dry season between May to September which then fell again in October. The distribution of chlorophyll-a concentrations was higher in areas near the coast, especially in the central and eastern parts of the South Java Sea. In the eastern part, chlorophyll-a concentration consistently spreads towards the offshore up to 20 km from the coastline.

This pattern may be due to the combined influence of the ITF coming out of the Bali strait, The South Equatorial Current (SEC), and The Southern Java Current (SJC) that drift away chlorophyll-a off the coast.
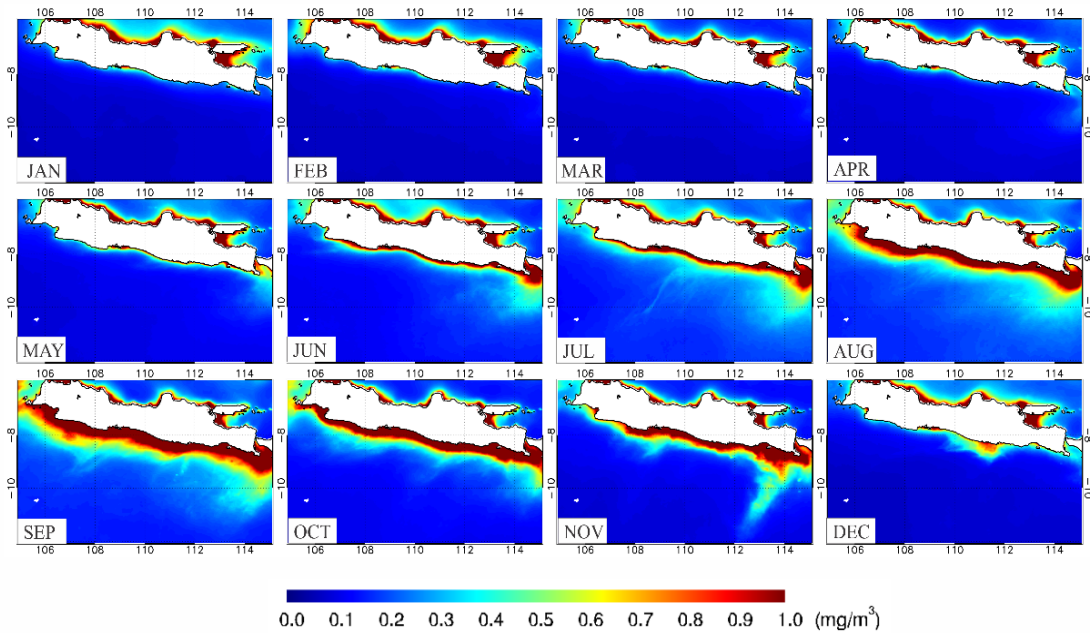
Fig. 4. Monthly climatology of chlorophyll-a along the southern coast of Java.

4.1.3. Relationship among SST, chlorophyll-a and wind speed along the southern coast of Java

Upwelling areas are generally identified using sea surface temperature parameters and chlorophyll-a concentrations in the ocean. Low sea surface temperatures can indicate that there is an water mass lifting of deep sea water to the surface which is rich in nutrients. Nutrients are then used by phytoplankton for primary productivity so that the concentration of chlorophyll-a will increase. Therefore, low sea surface temperatures followed by high chlorophyll-a concentrations indicate the area is an upwelling area. **Fig. 5** shows that along the Southern coast of Java, warm sea surface temperatures are generally found in the rainy season while it decreases during the dry season. Meanwhile, the concentration of chlorophyll-a peaked only in the dry season.

To take a closer look, the wind movement shows a different pattern in the two seasons. The wind moves to the east in the rainy season while in the dry season it moves to the northwest. These two opposing patterns are the result of the monsoon winds movement between the continents of Asia and Australia. This wind movement causes the Ekman mass transport which triggers the displacement of water masses around the coast. With reference to this mechanism, along the southern coast of Java, the downwelling occurs in the rainy season and upwelling in the dry season along the coast. Furthermore, the upwelling intensity is greater in the eastern part of the southern sea of Java than the western part as denoted by the higher chlorophyll-a concentrations and lower SST at the eastern part.

For further analysis, we plot the variation of SST, chlorophyll-a concentration and wind speed in the time series graph which shows the highest chlorophyll-a concentration and lowest SST occur during the highest wind speed during the Southeast monsoon season. High correlation between wind and SST (i.e., -0.88) or wind and Chl-a (i.e., 0.88) indicating that wind plays a major role in the variability of these two parameters.
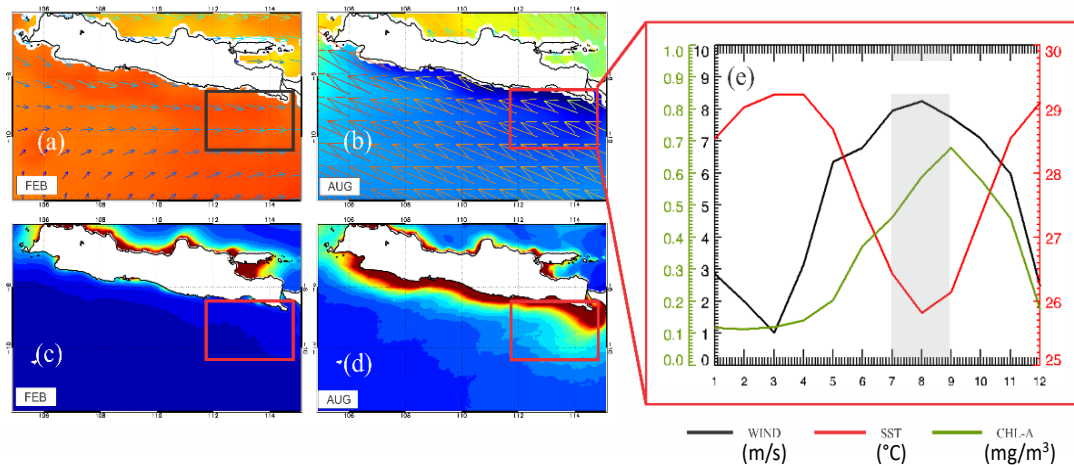
Fig. 5. The location and time of upwelling based on the spatial distribution of (a,b) wind overlaid with SST (c,d), chlorophyll-a concentration and (e) temporal distribution of SST, wind, and chlorophyll-a.

4.1.4. Seasonal Variation of Surface Heat Fluxes

The variation of shortwave radiation in 12 months of climatology shows that the flux distribution into the waters ranges from 140 - 250 W/m^2 as shown in **Fig. 6**. The distribution of shortwave radiation flux increased from December to March with an average flux ranging from 200 – 250 W/m^2 then decreased from April to June with the lowest average flux ranging from 140 – 180 W/m^2. The flux increased again in July to November with the average amount of flux peaking in November to 250 W/m^2. The variation of the heat flux of short-wave radiation is closely related to the annual position of the sun, where the closer the sun is, the greater the incoming heat flux. In the south of Java, the closest position of the sun is in January or October and the farthest one is in June. While the variation of longwave radiation in 12 months of climatology shows the distribution of flux out into the atmosphere ranging from 20 - 80 W/m^2 **Fig. 6**. The spatial distribution shows that the longwave radiation emitted into the atmosphere has increased since January with an average flux of 29.2 W/m^2 until it peaked in October at 53.8 W/m^2 and then decreased again until December to 28 W/m^2. Low radiation fluxes are scattered around the island while high radiation is generally scattered over ocean areas.

Sensible heat flux variation in 12 monthly climatology in shows the flux distribution ranging from -5 to 20 W/m^2. The negative and positive value indicates that the sensible heat flux can occur in two directions, namely receiving heat (negative) or releasing heat (positive). This can be caused by the effect of the resultant air and sea temperatures which are not always positive. The high value of sensible heat flux is generally distributed in the western part of the South Java Sea, while the low value of flux is generally distributed in the eastern part of the South Java Sea.

A significant difference in distribution contours is formed between eastern and western part, especially in June and July. The distribution of sensible heat flux has the highest average value in the rainy season (DJF) in January with an average of 11.2 W/m^2. During the first transitional season (MAM), the sensible heat flux decreased until April reached to 7.3 W/m^2 and rose again in May to 10 W/m^2. In the dry season (JJA), the distribution of sensible heat flux does not change significantly, but the difference in sensible heat flux contours value on the west and east part of the South Java Sea is getting clearer. Sensible heat decreased again in August and reached its lowest point in the second transitional season (SON) in October which reached 5.4 W/m^2 and increased again in November towards the rainy season.
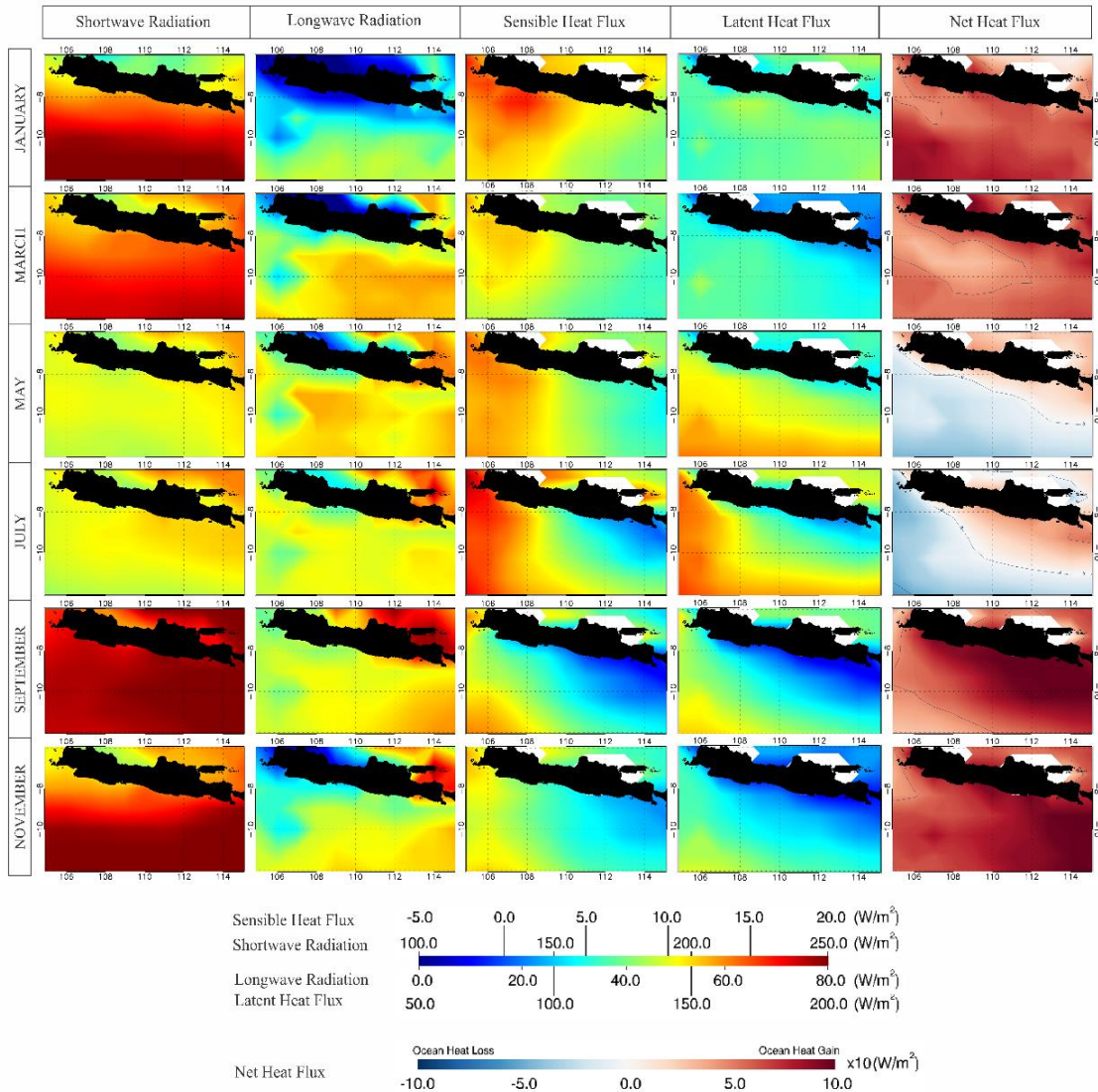
Fig. 6. Monthly climatology of surface heat flux along the southern coast of Java. Positive sign on the shortwave radiation means downward flux, while positive signs on longwave radiation, sensible heat flux, latent heat flux and net heat flux mean upward flux.

Latent heat flux variation in 12 months climatology shows the flux distribution ranging from 50 - 150 W/m². The positive value indicates the release of heat to the atmosphere. The average value per year of latent heat release is 120.74 W/m² which is relatively much larger than the sensible heat flux. The highest latent heat flux release was found in May that reached 141.45 W/m² and has its lowest average value in October, which is 102.78 W/m². In the rainy season (DJF), there is a small fluctuation in the latent heat flux around 20 W/m². In the first transitional season (MAM), there was a significant increase in latent heat release about 40 W/m² towards its highest peak in May. In the dry season (JJA), flux changes are not too significant and only change about 2 - 6 W/m² and there is a decrease in the coast area until the second transitional season (SON). It can also be seen that the heat release in the open seas is generally higher than the coast.

Through the calculation of the total flux coming out and entering the sea surface, the total heat flux value (Net Heat Flux) is obtained. These results indicate whether the sea surface is in a state of balance, positive, or negative. A positive state indicates that the ocean is receiving heat while on the other hand the sea is losing heat to the atmosphere. It can be seen that shortwave radiation and latent heat flux is the main flux that influence the net heat flux variation as shown in **Fig. 7**. Moreover, the South Java Sea received the highest heat in October reaching average 92.8 W/m^2. Total heat then declined since November to its lowest point in June of -14.3 W/m^2. In total, the South Java Sea gain a surplus heat of 547.8 W/m^2 per year.
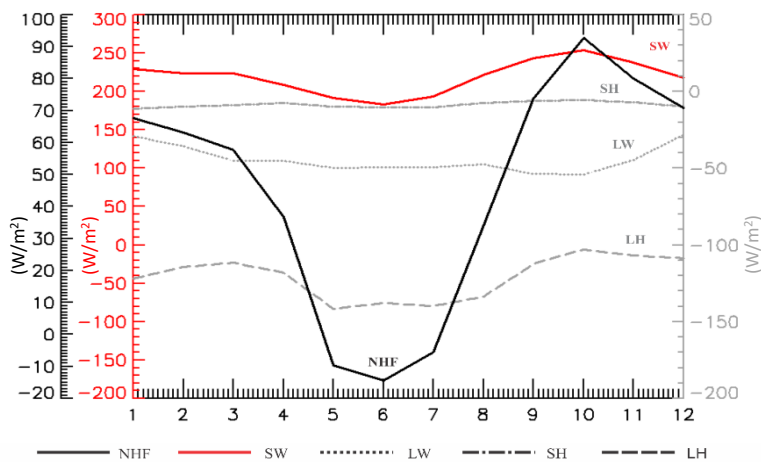
Fig. 7. Time series graph of shortwave radiation, longwave radiation, sensible heatflux, latent heatflux, and net heat flux for the whole study area as shown in Fig. 1. Positive and negative signs mean downward and upward flux, respectively.

4.1.5. Surface Heat Flux Effect on Sea Surface Temperature along the southern coast of Java

In this study, the relationship between the surface heat flux and SST was analyzed using a linear Pearson correlation which was then strengthened by descriptive analysis according to its temporal and spatial distribution. The correlation of each condition is summarized in **Table 1** using 4 months lag conditions between NHF, SST and wind. This condition is determined to see the time lag relationship that occurs between surface heat flux and sea surface temperature. A very strong correlation coefficient between NHF and SST was obtained at lag+2 and lag+3 months with the correlation value of 0.84 and 0.83, respectively. Between NHF and wind, it was obtained at lag+3 months with the correlation value of -0.81. These results indicate that there is a high possibility of delay effects between NHF, SST, and wind. In the lag conditions, the higher the surface heat flux, the higher the sea surface temperature. Meanwhile, the higher the surface wind speed, the lower the surface heat flux.

The heat received by the oceans from shortwave radiation does not necessarily make the oceans always in a hot condition. This is because the sea surface also releases a large amount of heat from the latent heat flux. The release of latent heat is influenced by wind speed and the difference in specific humidity between the sea surface and the atmosphere (ΔQ) according to the Bulk Coare 3.0 formula (Yu et al., 2007; 2008). The heat release in the latent form can be in the form of rains or displacement of water vapors to other areas. In **Fig. 8**, we can see that the variation of latent heat flux is initiated by differences in specific humidity (ΔQ) and enhanced by wind speed given the similarity of patterns between ΔQ and latent heat flux. However, it is difficult to determine which is most dominant factor since both variables are used to calculate the latent heat flux. The greater the wind speed and ΔQ, the greater the heat release that occurs.

Correlation of NHF, SST, wind.

Parameter		Correlation			
		Normal	Lag+1	Lag+2	Lag+3
NHF	**SST**	-0.007	0.5	0.84	0.83
	WIND	-0.007	-0.4	-0.7	-0.81
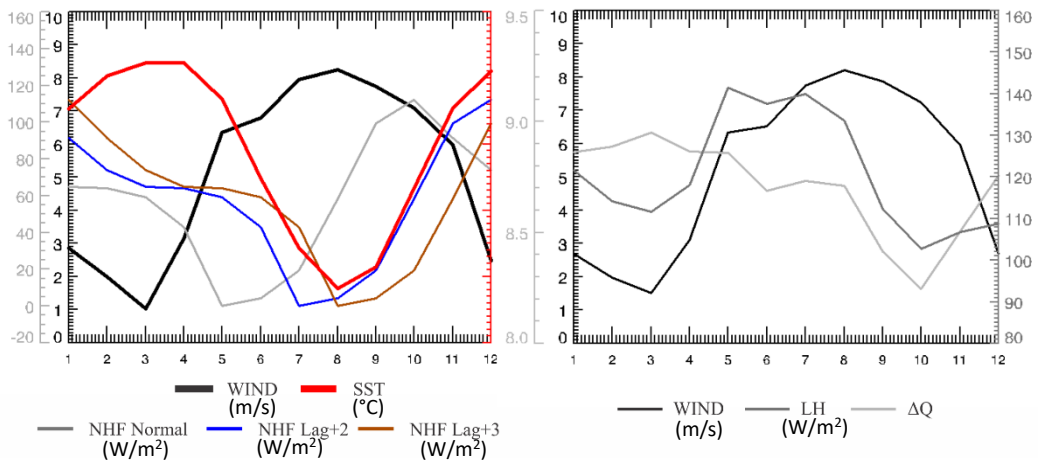
Fig. 8. Timeseries graph of wind, SST and NHF (Left)and wind, latent heat flux, and specific humifdity difference (ΔQ) at the red box shown in Fig. 4. Positive and negative signs of surface heat flux mean downward and upward flux, respectively.

4.1.6. Surface Heat Flux Effect on Chlorophyll-a along the southern coast of Java

The heat flux exchange can affect the thermodynamic processes in the sea surface and column. It can result in changes in temperature, salinity, water layers, or the growth of organisms in the ocean (Ushijima and Yoshikawa, 2019). The results in this study showed that ocean warming contributed to the increase in the chlorophyll-a concentration. The heat received by the oceans is used by autotrophic organisms for photosynthesis. In addition, ocean warming also results in shoaling of mixed layer depth so that uplifted nutrients will be trapped at critical depths (Ghisolfi et al., 2015). This shoaling of MLD causes a decrease in turbulent convection in the water column and stabilizes the mixed layer depth. The shallows MLD can stimulate the growth of phytoplankton by which nutrients is exist (Xu, Y et al., 2020).

A strong positive relationship was found between NHF and Chl-a by 0.71 in dry season as shown in **Table 2**.

Correlation of NHF, Chlorophyill-a, and Mixed Layer Depth.

Parameter		Correlation		
		12 Months	DJF-MAM	JJA-SON
NHF	**CHL-A**	0.35	-0.64	0.71
	MLD	-0.82	-0.51	-0.94

Meanwhile, a moderate negative relationship was found by -0.64 in the rainy season. This difference indicates that there are other factors that influence the chlorophyll-a concentration. This factor may occur in the dry season but it lacks in the rainy season. **Fig. 9.** shows that the increasing Chl-a concentration during dry season coincide with the increase of NHF.
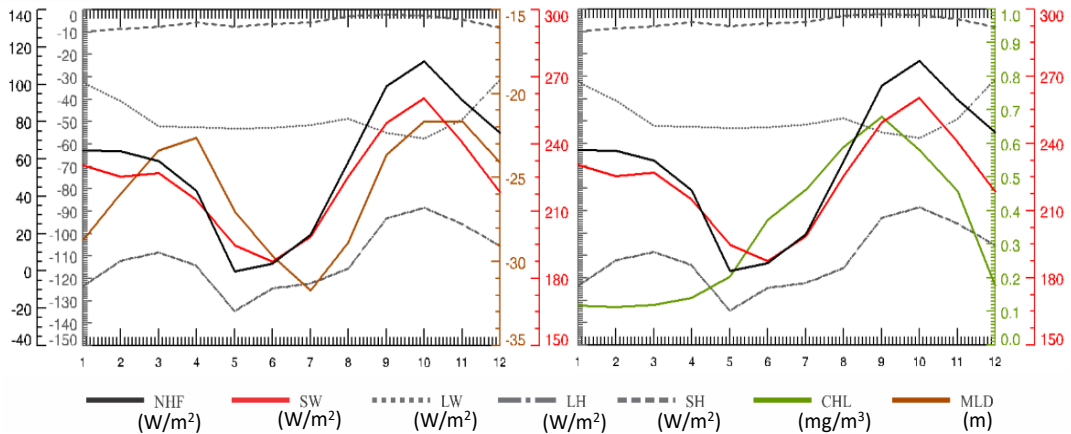
Fig. 9. Time series of surface heat flux, Chlorophyll-a and MLD at the red box shown in Fig. 4. Positive and negative signs of surface heat flux mean downward and upward flux, respectively.

4.2. Discussion

We found that there is a delay two or three month of the correlation between NHF and SST. This means that the decreasing heat gain takes two or three months to affect the decrease of SST. The delay can be caused by heat advection and convection processes that occur between the ocean and the atmosphere. This heat advection process is carried out by ocean currents and wind on the surface which takes time to be moved or released. Meanwhile, the heat convection process takes place vertically through mixing or turbulent convection. The process of heat advection (Thermal Advection) takes about 30 months according to research by Gruenburg and Gordon (2018) which examines the heat content through the Indonesian throughflow current (ITF) in the Makassar Strait to the Eastern Tropical Indian Ocean (ETIO). Given that Southern Java Sea is a small part of the ITF, the possibility of heat advection in this area has a lower time phase given the shorter distance hence the result shows 2 – 3 months delay. This result is consistent with the study conducted by Wirasatriya et al (2019) in the Northern Sea of Java which also found 2 months delay effect between surface heat flux and SST. This finding contradicts to Varela et al. (2016) which states that heat flux is not a driving factor for sea surface temperature variability along the southern coast of Java since they do not consider the delay effect of NHF to SST.

The relation between NHF and chlorophyll-a concentration along the southern coast of Java is related with The Convection Shutdown theory by Smyth et al. (2014) and Ferrari et al. (2016) who said that an increase in chlorophyll-a will occur when the cooling of the ocean by surface heat flux begins to subside and changes to ocean warming or when the NHF changes from negative to positive. This study agrees with the statement which found that significant changes in chlorophyll-a concentration start to occur when NHF switched from negative to positive in May as in **Fig. 9.** However, the concentration of chlorophyll-a is highly dependent on the presence of nutrients in the water column.

The study conducted by Wirasatriya et al. (2020) which examined Ekman transport and Ekman pumping velocity in the same area showed that in the rainy season, nutrient uptake was much lower than during the dry season. This explains why different correlations were found in the rainy and dry seasons between NHF and chlorophyll-a in this study. Although the conditions of surface heat flux and MLD support the growth of phytoplankton, if the water column does not contain nutrients, the growth of phytoplankton will not occur so that the concentration of chlorophyll-a will be low.

The application of satellite data in the present study manages to reveal the spatial and temporal distribution of surface heat flux along the southern coast of Java. However, satellite also has limitation to observe the parameters below sea surface. As explained by Dong et al. (2007), the SST variations are governed through the heat balance not only in the sea surface but also in the mixed layer of the ocean, which is influenced by surface air–sea heat fluxes, horizontal advective and diffusive processes in the mixed layer, and entrainment processes at the base of the mixed layer. Thus, for better understanding of heat budget variation along the southern coast of Java, horizontal advective and diffusive proceses and entranment should be included in the analysis. This task is left for future study.

5. CONCLUSIONS

Remote sensing data have been used to study the spatial and temporal variability of the surface heat flux and its effect on the variability of sea surface temperature and chlorophyll-a in the upwelling potential area along the Southern coast of Java. Climatologically, the seas along the southern coast of Java always gain a surplus heat with an average of 547.8 W/m^2 per year. The largest heat intake comes from shortwave radiation and the largest heat release comes from latent heat flux.

A very strong relationship occurs between surface heat flux and sea surface temperature at lag conditions of 2-3 months. This shows that surface heat flux plays an important role indirectly modulating changes in sea surface temperature which takes 2 - 3 months. The delay may be caused by heat advection and convection processes which are not fully discussed in this study.

A significant increase in chlorophyll-a concentration occurs when the ocean conditions change from cooling to warming by surface heat flux, which is indicated by a change in net heat flux from negative to positive in May. The increase in surface heat flux also affects the shoaling of mixed layer depth, resulting in stratification of the water column which in turn causes the trapping of nutrients at critical depths. If there is a lot of nutrient abundance in the water column, the combination of solar radiation and supportive water column conditions will trigger primary growth in the waters so that the concentration of chlorophyll-a will increases.

R E F E R E N C E S

Alifdini, I., Shimada, T., and Wirasatriya, A. (2021) Seasonal Distribution and Variability of Surface Winds in the Indonesian Seas using Scatterometer and Reanalysis Data. *International Journal of Climatology,* 14 (10), 4825-4843. https://doi.org/10.1002/joc.7101

Atlas, R., Hoffman, R.N., Ardizzone, J., Leidner, M., Jusem, J.C., Smith, D.K., and Gombos, D. (2011) A Cross-Calibrated, Multiplatform Ocean Surface Wind Velocity Product For Meteorological And Oceanographic Applications. *Bullettin American Meteorological Soc*iety February: 157-174 DOI:10.1175/2010BAMS2946.1.

Chang, C-P., Wang, Z., McBride, J., and Liu, C-H. (2005) Annual Cycle of Southeast Asia—Maritime Continent Rainfall and the Asymmetric Monsoon Transition. *Journal of Climate* 18, 287-301. https://doi.org/10.1175/JCLI-3257.1.

Chang, C.P., Wang, Z., and Hendon, H. (2006) The Asian winter monsoon. *The Asian Monsoon*. Springer Praxis Books. Springer, Berlin, Heidelberg, pp 89-127.

Brunke, M. A., Wang, Z., Zeng, X., Bosilovich, M., and Shie, C. L. (2011) An assessment of the uncertainties in ocean surface turbulent fluxes in 11 reanalysis, satellite-derived, and combined global datasets. *Journal of Climate*, 24 (21), 5469–5493. https://doi.org/10.1175/2011JCLI4223.1

Cronin, M. F., Gentemann, C. L., Edson, J. B., Ueki, I., Bourassa, M., Brown, S., Clayson, C. A., Fairall, C., T. Farrar, J., Gille, S. T., Gulev, S., Josey, S., Kato, S., Katsumata, M., Kent, E. C., Krug, M., Minnett, P. J., Parfitt, R., Pinker, R. T., Zhang, D. (2019) Air-sea fluxes with a focus on heat and momentum. *Frontiers in Marine Science.* https://doi.org/10.3389/fmars.2019.00430

Dong, S., Gille, S.T., and Sprintall, J. (2007). An Assessment of the Southern Ocean Mixed Layer Heat Budget. Journal of Climate, 20, 4425-4442. DOI: 10.1175/JCLI4259.1

Drévillon, M., Fernandez, E., Lellouche, J.M. (2021) Product user manual for the Global Ocean Physical Multi Year product GLOBAL_MULTIYEAR_PHY_001_030. Marine.copernicus.eu

Ferrari A.R., Merrifield, S.T. and Taylor, J.R. (2016) Shutdown of convection triggers increase of surface chlorophyll. *Journal of Marine Systems,* 147, 116-122.

Ghisolfi, R.D., da Silva, P., dos Santos, F.T., Servino, R.N., Cirano, M., and Thompson, F.L. (2015) Physical Forcing Mechanisms Controlling the Variability of Chlorophyll-a over the Royal-Charlotte and Abrolhos Banks—Eastern Brazilian Shelf. *PLoS ONE,* 10 (2), e0117082. doi:10.1371/journal. pone.0117082

Gruenburg, L.K and Gordon, A.L. (2018) Variability in Makassar Strait heat flux and its effect on the eastern tropical Indian Ocean. *Oceanography*, 31 (2), 80–87.

Huang, R. X. (2016) Oceanographic Topics: Thermohaline Circulation. In Encyclopedia of Atmospheric Sciences: Second Edition. Elsevier, 4. https://doi.org/10.1016/B978-0-12-382225-3.00281-4

Kunarso, Hadi, S., Ningsih, N. S., and Baskoro, M. S. (2011) Variability of SST and chlorophyll-a at upwelling areas Variabilitas on the variability of ENSO and IOD from the southern Java to Timor. *Indonesian Journal of Marine Sciences*, 16 (3), 171-180–180. https://doi.org/10.14710/ik.ijms.16.3.171-180. In Bahasa

Lahlali, H., Wirasatriya, A., Gensac, E., Helmi, M., Kunarso, and Kismawardhani, R. A. (2019) Environmental aspects of tuna catches in the Indian Ocean, southern coast of Java, based on satellite measurements. *2018 4th International Symposium on Geoinformatics,* 1–6. https://doi.org/10.1109/ISYG.2018.8612020

Lee, S. K., Park, W., Baringer, M. O., Gordon, A. L., Huber, B., and Liu, Y. (2016) Pacific origin of the abrupt increase in Indian Ocean heat content during the warming hiatus. *Nature Geoscience,* 8 (6), 445–449. https://doi.org/10.1038/NGEO2438

Levitus,S., Antonov, J.I.,Boyer, T.P., Baranova, O.K., Garcia, H.E., Locarnini, R.A., Mishonov, Reagan, J.R., Seidov, D., Yarosh, E.S., Zweng, M.M. (2012) World ocean heat content and thermosteric sea level change (0-2000m), 1955–2010. *Geophysical Research Letters*, 39 (2012), 10.1029/2012GL051106

Remote Sensing Systems (2017) MWIR optimum interpolated SST data set. Ver. 5.0. PO.DAAC, CA, USA. Dataset accessed [2020-02-02] at https://doi.org/10.5067/GHMWI-4FR05

Sathyendranath, S., Brewin, R.J.W, Brockmann, C. etal. 2019. An ocean-colour time series for use in climate studies: the experience of the Ocean-Colour Climate Change Initiative (OC-CCI). *Sensors*, 19, 4285. doi:10.3390/s19194285

Smyth, T.J., Allen, I., Atkinson, A., Bruun, J.T., Harmer, R.A. et al. (2014) Ocean Net Heat Flux Influences seasonal to Interannual Patterns of Plankton Abundance. *PLoS ONE*, 9 (6), e98709. doi:10.1371/journal.pone.0098709

Sukresno, B., Jatisworo, D., and Laut, S. P. (2018) Multilayer Analysis of Upwelling Variability. *Jurnal Kelautan Nasional*, 13 (1).

Susanto, R.D., and Marra, J. (2005) Effect of the 1997/98 el niño on chlorophyll a variability along the southern coasts of Java and Sumatra. *Oceanography*, 18 (4), 124–127. https://doi.org/10.5670/oceanog.2005.13

Ushijima Y., and Yoshikawa Y. (2019) Mixed Layer Depth and Sea Surface Warming under Diurnally Cycling Surface Heat Flux in the Heating Season. *American Meteorological Society,* 49, 1769-1787. DOI: 10.1175/JPO-D-18-0230.1

Varela, R., Santos, F., Gómez-Gesteira, M., Álvarez, I., Costoya, X., and Días, J. M. (2016) Influence of coastal upwelling on SST trends along the South Coast of Java. *Plos One*, 11 (9), 1–14. https://doi.org/10.1371/journal.pone.0162122

Vranes, K., Gordon, A.L., and Field, A. (2002) The heat transport of the Indonesian throughflow and implications for the Indian Ocean heat budget. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 49 (7–8), 1391–1410. https://doi.org/10.1016/S0967-0645(01)00150-3

Wibowo, S., Jayawiguna, M.H., and Triyono (2019) Potency of Marine and Fisheries Resources at WPP RI 573. AMAFRAD Press- Badan Riset dan Sumber Daya Manusia Kelautan dan Perikanan. Jakrta, Indonesia. p. 27. e-ISBN : 978-602-5791-78-9. In Bahasa.

Wirasatriya, A., Setiawan, R.Y., and Subardjo, P. (2017) The effect of ENSO on the variability of chlorophyll-a and sea surface temperature in the Maluku Sea. *IEEE Journal of Selected Topics on Applied Earth Observations and Remote Sensing,* 10(12), 5513-5518. **DOI:** 10.1109/JSTARS.2017.2745207.

Wirasatriya, A., Setiawan, J. D., Sugianto, D. N., Rosyadi, I. A., Haryadi, H., Winarso, G., Setiawan, R. Y., and Susanto, R. D. (2020) Ekman dynamics variability along the southern coast of Java revealed by satellite data. *International Journal of Remote Sensing,* 41 (21), 8475–8496. https://doi.org/10.1080/01431161.2020.1797215

Wirasatriya, A., Sugianto, D. N., Helmi, M., Maslukah, L., Widiyandono, R. T., Herawati, V. E., Subardjo, P., Handoyo, G., Marwoto, J., Anugroho, A., Suryoputro, D., Atmodjo, W., and Setiyono, H. (2019) Heat flux aspects on the seasonal variability of sea surface temperature in the Java Sea. *Eco. Env. and Cons*, 25(1), 434–442.

Wirasatriya, A., Sugianto, D. N., Maslukah, L., Ahkam, M. F., Wulandari, S. Y., and Helmi, M. (2020) Carbon dioxide flux in the Java Sea estimated from satellite measurements. Remote Sensing Applications: *Society and Environment,* 20 (7). https://doi.org/10.1016/j.rsase.2020.100376

Wirasatriya, A., R. Dwi Susanto, Kunarso Kunarso, Abd. Rasyid Jalil, Fatwa Ramdani and Ardiansyah Desmont Puryajati. (2021) Northwest monsoon upwelling within the Indonesian seas. *International Journal of Remote Sensing*, 42 (14), 5437-5458. DOI: 10.1080/01431161.2021.1918790

Xu, Y., Travis, M., and Oscar, S. (2020) Physical processes controlling chlorophyll-a variability on the Mid-Atlantic Bight along northeast United States. *Journal of Marine Systems, 212, 103433.*

Yu, L., and Weller, R.A. (2007) Objectively Analyzed air-sea heat Fluxes for the global oce-free oceans (1981–2005). *Bulletin American Meteorological Society*, 88, 527–539

Yu, L., Jin, X., and Weller, R. A. (2008) Multidecade Global Flux Datasets from the Objectively Analyzed Air-sea Fluxes (OAFlux) Project: Latent and sensible heat fluxes, ocean evaporation, and related surface meteorological variables. Woods Hole Oceanographic Institute, Massachusetts.

Yu, Lisan, Jin, X., and Weller, R. A. (2007) Annual, seasonal, and interannual variablity of air-sea heat fluxes in the Indian Ocean. *Journal of Climate*, 20 (13), 3190–3209. https://doi.org/10.1175/JCLI4163.1

Zhang, Y-C., Rossow, W.B., Lacis, A.A., Oinas, V., and Mishchenko, M.I. (2004), Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data. *Journal of Geophysical Research,* 109, 1-27. doi: 10.1029/2003JD004457

A GIS-BASED ANALYSIS FOR ECOTOURISM SUITABILITY IN A GEOLOGICAL COMPLEX AREA OF CARPATHIANS

Judith LAKATOS [1] , Zsolt MAGYARI-SÁSKA [1] , Stefan DOMBAY [1]

ABSTRACT: The purpose of this study is to contribute to the sustainable development of ecotourism through spatial analysis. The aim of the study was to identify suitable locations for various types of ecotourism activities and infrastructure development. This was achieved using multiple criteria analysis and spatial thinking as the foundation for decision-making. The principles of sustainable development were applied and translated into specific criteria and factors for a study location in the Carpathians. The presence or absence of criteria was transformed into GIS data as logical layers, which were later combined to produce the final results. The outcome offers several scenarios for the effective utilization of the ecotourism potential in the protected area of the geological complex. Spatial analysis identified locations suitable for eight types of ecotourism activities, represented in ecotourism suitability maps. The combined suitability map for ecotourism activities and infrastructure covers an area of 113 km^2 in the study area, the geological complex of Racos (Romania) and can be used by authorities for their various spatial development plans. Given that the results are feasible, we believe that this methodology can be adapted to other locations by adjusting the parameter values for various factors.

Key-words: *Spatial planning, Multiple criteria analysis, GIS, Ecotourism, Geosite, Racos, Romania*

1. INTRODUCTION

In the past years the whole tourism market and ecotourism as well has suffered due to the COVID-19 pandemic. Most of recent studies are concentrated of the negative effects of the situation created in different locations around the world (Vidal, M.D. et al., 2021; Chang, C-N. et al., 2022; Persada, S.F. et al., 2023). However, in some point of view new opportunities appeared especially for ecotourism as people started to prefer more remote and preferably natural, less crowded places. This study is a response to the current changing process trying to reveal new opportunities for the study area. Ecotourism is one of the most intensively developing form of tourism. Researches related to this form of tourism appeared first time, in the late 1980s, but after two decades, in 2002 the United Nations declared it the International Year of Ecotourism, gaining ever more importance as one of the crucial objectives being long term sustainable development (Whitelaw et al., 2014).

This kind of long-term thinking in tourism planning is absolutely essential and it can be said that the study is a sort of a response to the current situation in the study area, where the number of tourists has started to increase in the past years based on the evaluation of local authorities and local touristic information office. Hereinafter this study will strengthen the concept that the analysis of ecotourism possibilities can scientifically contribute to the development of ecotourism (Băltărețu, 2011; Dombay et al., 2008), which as a result confirms what can be done and what is worth doing in the region based on spatial modelling. Spatial analysis is crucial in the assessment of tourism potential because it enables a more comprehensive understanding of the relationships between different variables that influence the potential for tourism development. These variables can include physical and natural resources, accessibility, infrastructure, cultural heritage, and social and economic factors. By analyzing these variables in a spatial context, it becomes possible to identify areas of high tourism potential and quantify their importance (Ruda, 2016; Ruda and Pokladníková, 2016).

Additionally, spatial analysis provides a powerful tool for evaluating the potential impact of tourism activities on specific areas, such as protected natural areas or cultural heritage sites

---

[1] *Babeş-Bolyai University, Faculty of Geography, Gheorgheni University Extension, 535000 Gheorgheni, judit.nyulas@ubbcluj.ro; zsolt.magyari@ubbcluj.ro; stefan.dombay@ubbcluj.ro*

(Kusumayudha, 2021). This enables decision-makers to make informed decisions about the allocation of resources and development of infrastructure, ensuring that tourism development is sustainable and responsible. Spatial analysis also enables the integration of data from multiple sources and can be used to create visual representations of data, making it easier to understand complex relationships and patterns. Multi-criteria analysis present in our research is a largely used tool for finding suitable locations. It has been applied in various fields, such as urban green spaces planning (Gelan, 2021), industrial site selection (Rikalovic et al., 2014), landfill site selection (Abujayyab et al., 2017), public school site selection (Prasetyo, 2018). Tourism and, specifically, ecotourism activities and development can be planned based on multi-criteria evaluation in GIS. However, the term "multi-criteria analysis" integrated in GIS and combined with spatial analysis doesn't denote a single specific methodology. It is often coupled with fuzzy logic (Omardezh et al., 2022; Ronizi et al., 2020) or the analytical hierarchy process (Ahmadi et al., 2015; Mansour et al., 2020).

Ecotourism planning in general may consider various aspects of the natural surroundings, infrastructural facilities, accessibility and also vulnerability of the site. Depending on the specific objectives of a study it may emphasize or omit some aspects. Although there are studies which lacks on the spatial aspect of planning (Shang, 2020), most of them use spatial data for which factors and constrains are defined to detect those locations which satisfactorily facilitate the desired conditions (Fung and Wong, 2007; Sánchez-Prieto et al., 2021; Indriyani et al., 2020).

1.1. Study site

The study region is represented by the Racoș Geological Complex as a central point of attraction with various points of interest around it. Is positioned in the Carpathians between longitude 25°37'25.54"E and 25°47'23.41"E having the latitude coordinates between 45°99'86.17"N and 46°06'26.52"N with an altitude range from 450 m to 815 m. The study area contains the geological complex- protected natural area, a geological area of national importance of 95.2 ha (2006) and is integrated into the Natura2000 Homoroadelor Hills since 2007 (**Fig. 1**).

Fig. 1. Localization and arial view of the study region.
Source: Google Earth.

From a scientific point of view the Geological Complex of Racoș a complex area, millions of years old, presenting itself as an open book for Geology and a real "paradise" for geologists. This unique natural formation (Cioacă, 2002) is located in the Perșani Mountains, in the county of Brașov, Romania, in the commune of Racoș and represents a particularly picturesque and natural wild area (Albotă and Fesci, 1980) as presented in **figure 2**.

Fig. 2. Natural potential of the study area
Source: romtur.ro.

1.2. Objectives of the study

The primary objective of this study is to contribute effectively to the foundation of a new concept of ecotourism planning for the study area, a concept based on spatial thinking (Haidu and Haidu, 1998), which will underpin approaches to a more conscious decision-making process. Develop a GIS assessment approach that considers both the current response of the study area and the resulting solutions as part of a long-term development plan of the site which.

To achieve the desired result the following objectives have been set:

   i.   establishing, creating, editing and managing geospatial master data in the GIS with various datasets holding the necessary criteria;

   ii.   identification the possible locations for ecotourism objectives studied as a result of different spatial analyses;

   iii.   integrating the obtained ecotourism possibilities into a general map.

2. DATA AND METHODS

The methodology applied to this study is suitability assessment, based upon employing multiple criteria. This GIS data-based analysis can model the suitable locations for the development of ecotourism activities and infrastructure.

2.1. Data

The GIS data was created for the purpose of elaborating a complex set of layers in a spatial database (Islam et al., 2019). A variety of spatial data from different sources was collected to be applied by the assessment methodology. This includes base data: elevation model, hydrography, various routes and vegetation, and also derived data such as slope, exposition and visibility.

Primary, and the most essential basic GIS data is the digital elevation model, abbreviated as DEM (CDMEA, 2016). It plays an important role for the elaboration of the study. As existing DEM data has a very rough resolution for this local analysis, we decided to create an own model. The primary data sources for DEM modelling were the contour lines, resulted by digitizing them from a basic map of scale 1:5000.

The DEM was obtained by interpolating the vector data using the TIN model (Triangulated Irregulated Network) in QGIS (Cutts and Gasser, 2018). After the first results the need to correct/update appeared because of the changes of relief caused by exploitation and by the fact that the acquired topographic map was made before mining started. For this reason, it was necessary to find appropriate solutions. To solve this problem several onsite elevation measurements were made which were integrated in the new DEM generation process, resulting a final DEM with a resolution of 1 meter.

The elaboration of various other above mentioned primary data occurs through the process of interrogation (using plugin QuickOSM) and vectorization using mainly the basic map. Further data as terrain slope, exposition, visibility resulted by the terrain analysis based on DEM data. Vegetation cover was achieved from GisLounge, having the global land cover in GeoTIFF format. It has been recently released by Impact Observatory (IO) and ESRI.

2.2. Assessment based on multiple criteria

The methodology assessment based on multiple criteria is based on the establishment of criteria in the first phase, which are brought to a comparable form, i.e. a reclassification and then put together by combining them (Mohd and Ujang, 2016).

Once the baseline data is acquired and prepared as appropriate it will be followed by the establishment of the criteria considered being an important step of the analysis (Fung and Wong, 2007). By defining more criteria based on ecotourism experts having GIS and field knowledge the analysis results may have a higher usability being a better model of the reality. Prior to this analysis, a paperwork was done where highlighting tourist prospecting method, the field method, SWOT as actual situation analysis tool, and tourist potential evaluation of the area.

After criteria have been established the research procedure conducted to operations to carried out the assessment methodology, which is GIS related.

The phases of the analysis process consist of four main steps:

1. *establishing of criteria and parameters*, conditions which must be fulfilled for an optimal functionality of ecotouristic objectives. Parameter's setting for each criteria are site-specific.

2. *generation of the criteria maps/layers,* according to the type of criteria. The criteria are set for those spatial data that have been created and prepared previously. All these data are materialized in a different type of layers.

3. *reclassification of layers,* where every criteria layer follows a rule defined by parameters value, based on this each cell will receive a new value. It is a yes / no condition, represented by logical levels, where 0 - represents the fact that condition is not satisfied, and 1 - represents the fact that the condition is satisfied. By this the space is delimited in areas that are not suitable and in areas that are suitable for the analyzed activity, according to criteria.

4. *combination of reclassified layers*, application of a condition system between different logical raster layers. During these operations, mathematical or logical operations are performed between the corresponding cells of different layers that represent the same area at the same resolution (Imbroane, 2018). The result of the operation is reflected in a new raster layer.

On this basis, a land suitability analysis was carried out individually for each ecotouristic activity postulated by the interest groups in the area.

2.3. Defining criteria and parameters value

This step is a major step in the methodology, which requires as much knowledge of geoinformatics as of tourism. Suitability is estimated with the help of several criteria. These factors can be GIS data as well as non-GIS data. Different criteria are identified for specific ecotourism infrastructure / activities. Most of the exact values in criteria are based on field expertise and logical reasoning as follows.

Viewpoints, lookout points are elements for observing the natural environment in natural areas, such as the study area (relief, flora, fauna, etc.), and are basic elements of the ecotourism infrastructure. They are important destination points, generally an integral part of thematic or

educational trails. According to the point of view, a total of seven types of criteria are established for modelling the ecotourism infrastructure (**Table 1 - A**), elevation, vegetation, slope, visibility, distance, and area. Parameters values were defined for each of the factors according to the specificity of infrastructure/ activity and site characteristics. Where the hight, *elevation* has been set to greater than 700 m, to be above the geological complex. It makes absolute sense for the *vegetation* criteria to have the parameter without trees and bushes, first for all having the viewpoint function to see around. Another criterion *visibility,* which means visibility from a point or points to a given area/area of interest. In other words, „what area can see from point x to each point from all directions" (Imbroane, 2018). In conclusion visibility has an area-specific parameter and is set according to the object want to see. *Terrain slope* and *area*, linked to the construction needs, were set to have a flat terrain and at least 64 m² surface. Accessibility, one of the most important tourism characteristics after attractivities, is expressed in *distance* from the roads, where 200 m is a fair distance, not far away, easily can be reached from the roads. Viewpoint distance from center of the area was established to 5km, ecotourism point of view to be accessed by foot. It can be affirmed that these criteria are valid anywhere in the world, and the parameters value depends on the study terrain conditions. Where for example the height is specific to the study area, first depends on the altitude range.

Table 1

Criteria and factors for assessing suitability for various ecotouristic facilities and activities
A. Factors for viewpoint

#	Criteria/factors	Value (parameters)
a	elevation	h > 700 m
b	vegetation	no trees and bushes
c	terrain slope	degree < 3^0
d	distance from castle	d <5 km
e	visibility	Geological Complex; Sükösd-Bethlen Castle
f	distance from roads	d < 200 m
g	area	A = 64 m²

B. Factors for eco-village

#	Criteria/factors	Value (parameters)
a	terrain slope	degree < 17^0(30%)
b	built area	excluding built-up area
c	industrial zone	excluding industrial area + buffer zone with d = 200m
d	Geological Complex area	excluding the surface of the protected area
e	distance from the DJ131C road	d > 3 km
f	distance from paths/forest roads	d < 200 m
g	distance from the Racoș Gorge / Olt River	50 m < d < 750 m
h	southern part of the gorge	excluding the northern part of the gorge
i	area	A=3 ha (30.000 m²)

C. Factors for kayaking on the Olt river

#	Criteria/factors	Value (parameters)
a	Racoș Gorge	gorge surface
b	„islands" on the surface of the river	excluding surface area of the islands
c	distance from the shore	2m

D. Factors for scientific camp

#	Criteria/factors	Value (parameters)
a	terrain slope	degree < 5^0
b	Geological Complex area	excluding the surface
c	built area	excluding built-up area

d	industrial zone	excluding industrial area + buffer zone with d = 200m
e	distance from Geological Complex	d < 500 m
f	Natura2000	Natura2000 protected area
g	distance from roads	d < 200 m
h	area	A =1.5 ha (15000 m^2)

E. Factors for thematic hikes

	#	Criteria/factors	Value
Common criteria	a	terrain slope	degree<35°
	b	distance from the roads	d < 300 m
	c	Olt River and Smarald Lake	excluding the surface of the river and lake area
I. Geodiversity	**G**	distance from geodiversity attractions: G_1; G_2	d < 300 m
		distance from attractions: G_3; G_4; G_5; G_6; G_7	d < 500 m
		distance from attraction G_8	d < 300 m
		distance from attraction G_9	1<d < 200 m
II. Biodiversity	**B**	distance from biodiversity attractions: B_1; B_2: B_3	d < 300 m
		distance from attraction B_4 (polygon)	d < 500 m
III. Cultural -historical	**CI**	distance from cultural-historical attractions: CI_1; CI_2; CI_3; CI_4; C_5; CI_6; CI_7; CI_8; CI_9	d < 500 m
		distance from attractions: CI_{10}	d < 200 m

F. Factors for pontoon

#	Criteria/factors	Value (parameters)
a	above the lake surface	A= lake polygon
b	distance from paths + pontoon length 10m	D < 30m , D= d_1+ d_2 d_1 = 20 m; d_2 = 10 m
c	aspect	S- E
d	area	A=25m^2

However, the distance being the most used factor in the model, distance from roads, distance from tourism objectives. Accessibility is a key parameter in tourism, and it is expressed in terms of distance from roads, tourism objectives / attractions and others.

To implement an ecotourism activity or infrastructure the following factors were determined in the study depending on the activity and requirements which are listed in the tables below (**Table 1**).

3. RESULTS AND DISCUSSION

With the establishment of the spatial database and applying the analysis methodology in QGIS with proper spatial operations (Wegmann et al., 2020), now sites suitable for ecotourism activities and infrastructure can be identified. Six objectives were selected namely, 1. viewpoint; 2. eco-village; 3. pontoon the lake; 4. scientific camp; 5. thematic hikes; 6. kayaking on the river Olt, which were all outlined in the analysis. A similar assessment is created for each activity, which means that the main steps are similar, the basic idea remains the same, but the factors will change, and so will the analysis tools, too. The suitability analysis process for ecotourism objectives consists of following these three steps of the above presented methodology.

3.1. Reclassifying and combining the criteria layers

Every criterion is represented by one raster layer as a result of the reclassification process. All reclassified raster layers are shown in the **figure 3** for an objective viewpoint, where the white patches represent - areas suitable for a viewpoint and the black ones - areas not suitable for a viewpoint due to the defined parameters.
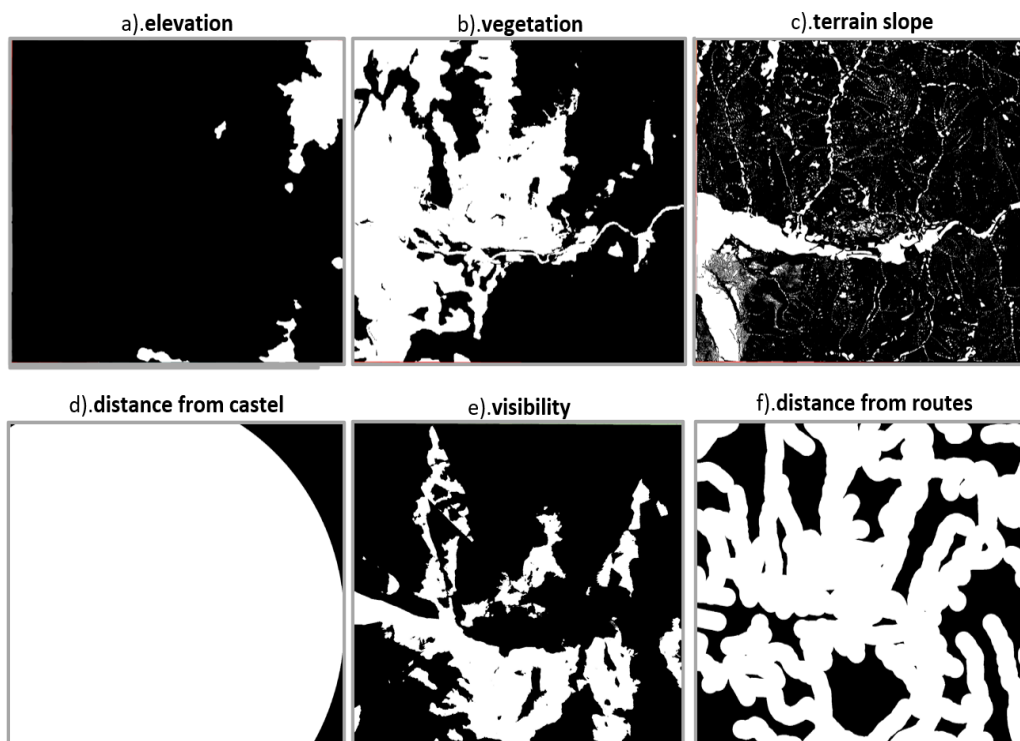

Fig. 3. Reclassified layers of criteria for objective viewpoints.

By combining the resulted logical layers, we are interested in locations that fulfil all the factors for a specific objective. The results for the suitability viewpoint are presented below (**Fig. 4**), where the six reclassified raster layers were combined, and after that filtered by the last condition area.

By applying the multi-criteria evolution methodology, the objectives of the eight types of recreation and ecotourism infrastructure were analyzed and as result possible locations were identified. For each ecotourism objective there is a separate representation illustrated on a single raster layer, this can be transformed into a vector layer for visualization. These images represent suitable locations for viewpoint; eco-village; pontoon the lake; scientific camp; thematic hikes (geological, biological and cultural-historic); kayaking on the river Olt and all these aspects are presented one by one in a map. An important outcome is represented by the results obtained along the way for each type of activity, shown in the **figures 5-11**. The same colors used to mark the selected criteria in these figures are present in the synthetic map representation.

Criterion *area* can be applied only as an additional operation at the end, after spatial analysis of reclassified layers was done. The *surface* is the extra intervening part when suitable areas are identified. For example, to build a viewpoint you need a compact area, small areas with only one or a few pixels should be treated as unsuitable cells. Therefore, the final phase of the evaluation is the identification or with other words the filtering of the results obtained in terms of area, defined by the last criterion g). area.

Fig. 4 – Suitable areas for viewpoints marked with
dark orange

Fig. 5 – Suitable areas for eco-village marked
with mustard yellow

Fig.6 – Suitable areas for poonton marked with pink

Fig. 7 – Suitable areas for scientific camp marked
with orange
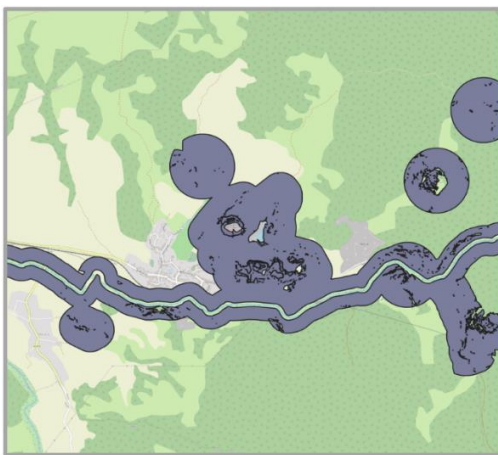
Fig. 8 – Suitable areas for geodiversity hiking marked
with gray

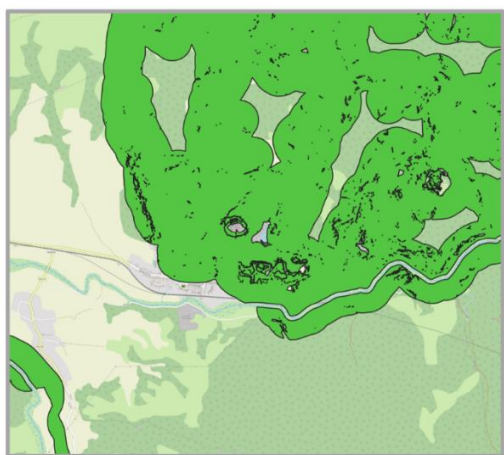**Fig. 9** – Suitable areas for biodiversity hiking
marked with green

Fig. 10 – Suitable areas for cultural and historical hikes marked with purple

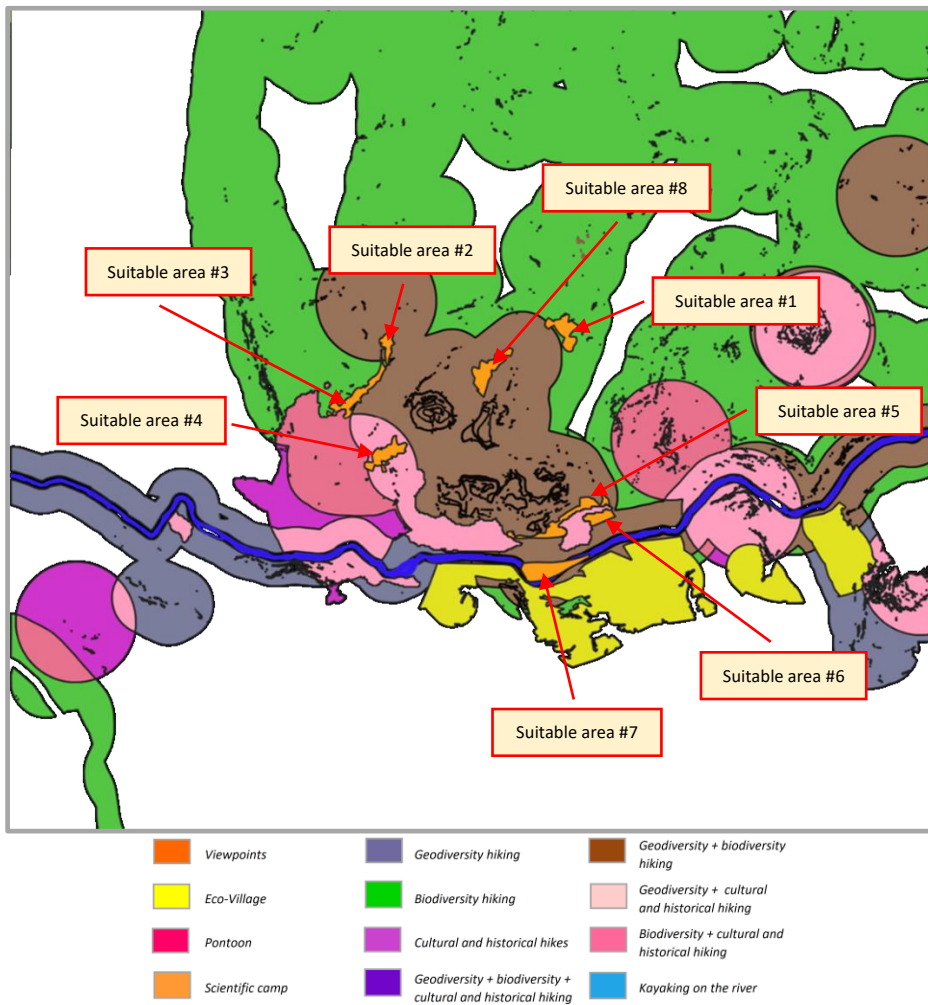**Fig. 11** – Suitable areas for kayaking on the river Olt marked with blue

Fig. 12. Integrated map with the appropriate areas identified for recreational activities and ecotourism infrastructure with suitable areas for Scientific camp.

The **figures 4-11** show the distribution of sites suitable for ecotourism in the study area, which are located on specific areas for each ecotourism infrastructure, where as a result it was found from a few 3 to 10 sites, namely 3 destinations for viewpoint, 3 for pontoon, 5 for eco-village and 8 for scientific camp, and for hiking like 4 geodiversity; 4 biodiversity, 10 cultural and historical hiking, 1 kayak-canoeing on the river Olt with different sizes of area distributed over the study area.

In order to get an overview of the eco-tourism possibilities of the study area, all these locations were combined in a map. The result of the suitability can be seen in just one picture (**Fig. 12**), which shows the eight global ecotourism activities of the studied region. This map is divided into areas having different colors (as shown in the legend), where the colored areas are suitable locations, and the black ones represent unsuitable locations for the chosen activities to analysis. The different colors highlight locations that are suitable for various recreational activities described in the legend.

Analyzing the all surface which are fit for some ecotourism activities/ infrastructure was found out that total is 25,46 % of 113 km^2 suitable for ecotourism recreation. Analyzing further the areas of the study region, the result is that there are locations that are suitable for a single ecotourism objective or even two or more types of objectives. The conclusion is that there are few overlapping areas as possible locations for ecotourism activities. Overlapping areas are locations with considerable values, because more than one eco-tourism activities can be carried out here.

The coverage of each possible location for eco-tourism activities / objectives can be calculated and the result shows an additional piece of information for ecotourism planning. For example, the science camp has got eight possible locations according to the analysis. These are located in different areas that are suitable for other ecotourism activities too, as you can see on the **table 2**. Where area number 4 is strategically located to an eco-tourism point of view, so it can be considered the most valuable area.

Table 2

Results for scientific camp sites - coverage with other areas

suitable scientific camp locations [#]	area [m^2]	biodiversity area [%]	geodiversity + biodiversity area [%]	biodiversity+ cultural-historic area [%]	geodiversity+ biodiversity+ cultural-historic area [%]
Suitable area 1	38629	95.50%	4.50%		–
Suitable area 2	16957	95,4 %	4,6 %		–
Suitable area 3	42019	80,6 %	16,2 %	3,2 %	–
Suitable area 4	39824	–	–	3,3 %	96,7 %
Suitable area 5	52311	–	100%	–	–
Suitable area 6	20151	–	100%	–	–
Suitable area 7	57363	–	100%	–	–
Suitable area 8	44042	–	100%	–	–

Using this methodology there are limitations that the map used is a decisive factor in modelling DEM. Where digitization is a long and time-consuming process. It is very important to establish from the beginning the qualitative level of the data structures (features) for the successful achievement of the objectives in the given project. At the same time, must be completed by field work, on-site measurement, requiring much study and detailed testing. Finally, of course the gains as result far outweigh the limitations. Such a vulnerable point there is the definition of criteria. However, these points did not affect the result of the study.

4. CONCLUSIONS

In this paper the ecotourism potential, suitable ecotourism sites of the Racoş Geological Complex and the surrounding area have been analyzed using a GIS-based methodology, concept based on spatial thinking. The principle of this study was the evaluation based on multiple criteria. Using different criteria levels and parameters value 8 types of ecotourism activity/infrastructure were analyzed to identify the suitable areas for ecotourism. The repartition of the suitability sites on the studied region was represented on a map for each ecotourism activity. Distribution of the well-identified results shown, specific location for each activity from 3 to 10 sites, namely 3 destinations for viewpoint, 3 for pontoon, 5 destinations for eco-village, 8 scientific camp and for hiking - like 4 geodiversity; 4 biodiversity -, 10 cultural and historical hiking, 1 kayak-canoeing on the river Olt with different sizes of area. The result of the integrated map represents the ecotourism suitability of the region, 25.46% of the 113 km^2 being suitable for ecotourism recreation, where the overlying area can be considered the most valuable area.

The applied methodology was able to model different scenarios, solutions for the chosen activities. Criteria plays a very important role, the factors are decisive in determining possible locations. Certainly, defining criteria can sometimes be a subjective approach but based on expert thinking is the appropriate way to achieve results. In conclusion the application of the GIS methodology used, database, adequate resolution, well-defined criteria, the through measurements and the other methods applied led this study to a success.

The objectives of this study are well-identified in the results, support for ecotourism planning, decision-making process, based on scientific criteria and value. The results can be integrated in local development plans an by this they can contribute to the development of ecotourism, showing where which type of activity can be carried out, what can be done and what is worth to be doing in the region.

Furthermore, as new possibilities, two directions can be followed: on the one hand, it can be a horizontal extension, which consists in examining other possible ecotourism activities, and on the other hand, this study can be a model for a new area where ecotourism planning is considered.

In the short term it may have applicability in shaping a better adaptation of planning in the natural environment, and in the long term it may have applicability in all forms of tourism and even in other areas such as the urban environment.

R E F E R E N C E S

Abujayyab, S.K.M., Sanusi, M.S., Yahya, A.S. & Alslaibi, T.M. (2017) GIS Modeling for Landfill Site Selection via Multi-Criteria Decision Analysis: A Systematic Review. In Proceedings of the *International Conference on High Performance Compilation, Computing and Communications* (HP3C-2017). Association for Computing Machinery, New York, NY, USA, 33–38. https://doi.org/10.1145/3069593.3069594

Ahmadi, M., Darabkhani, M.F. & Ghanavati, E. (2015) A GIS-based Multi-Criteria Decision-Making Approach to Identify Site Attraction for Ecotourism Development in Ilam Province, Iran, *Tourism Planning & Development*, 12 (2), 176-189, https://doi.org/10.1080/21568316.2014.913676

Albotă, M., & Fesci, S. (1980) *Munţii Perşani. Ghid turistic*. Bucharest, Sport-Turism Publisher.

Băltăreţu, A. (2011) *Arii protejate. Ecoturism. Dezvoltare Durabilă*. Bucharest, Pro Universitaria Publisher.

Chang, C.-N., Ting-Hsiu, L. & Hao-Chen, H. (2022) Investigating the Acceptance of Marine Ecotourism after the COVID-19 Pandemic in Taiwan, *Sustainability* 14(10), 6116, https://doi.org/10.3390/su14106116

Cioacă, A. (2002) *Munţii Perşani: Studiu geomorfologic*. Bucharest, Fundaţiei România de Mâine Publisher.

Cutts, A., & Grasser, A. (2018) *Learn QGIS*. (Fourth ed.). Birmingham, UK, Packt Publishing Ltd.

Dombay, I., Magyari-Sáska, Z. & László, P.S. (2008) *Ökoturizmus, elmélet és gyakorlat*. Cluj-Napoca, Presa Universitară Clujeană Publisher.

Gelan, E. (2021), GIS-based multi-criteria analysis for sustainable urban green spaces planning in emerging towns of Ethiopia: the case of Sululta town. *Environ Syst Res* 10 (13), https://doi.org/10.1186/s40068-021-00220-w

Fung, T., & Wong, F.-K. (2007) Ecotourism planning using multiple criteria evaluation with GIS, *Geocarto International*, 22(2), 87-105, https://doi.org/1080/10106040701207332

Haidu, I., & Haidu, C. (1998) *S.I.G. Analiză spațială,* Bucharest, *H*G*A* Publisher.

Imbroane, A. M. (2018). *Sisteme informatice geografice. Volumul II - Analiză spațială şi modelare*. Cluj-Napoca, Presa Universitară Clujeană Publisher.

Indriyani, N.M.P & Makalew, A.D.N. (2020), Ecotourism Landscape Planning in Nature Tourism Park of Buyan – Tamblingan Lakes Tabanan and Buleleng Regency Bali Province, *IOP Conf. Ser.: Earth Environ. Sci.* 501, 012037, https://doi.org/10.1088/1755-1315/501/1/012037

Islam, S., Miles, S., Menke, K., Smith, R., Pirelli, L., & Van Hoesen, J. (2019). *Mastering Geospatial Development with QGIS 3.x* (Third ed.). Birmingham, UK, Published by Packt Publishing.Ltd.

Kusumayudha, S.B., Prastistho, B., Zakaria, M.F., Rahatmawati, I. & Setyaningrum, T. (2021), Rock Mass Rating and Feasibility Assessment of Karst Cave Geo-Ecotourism in Tanjungsari District, Gunungkidul Regency, Yogyakarta Special Region, Indonesia, *Geographia Technica* 16(2), 53-68, http://dx.doi.org/10.21163/GT_2021.162.05

Mansour, S., Al-Awhadi, T. & Al-Hatrushi, S. (2020) Geospatial based multi-criteria analysis for ecotourism land suitability using GIS & AHP: a case study of Masirah Island, Oman, *Journal of Ecotourism*, 19(2), 148-167, https://doi.org/10.1080/14724049.2019.1663202

Mohd, Z. H. & Ujang, U. (2016), Integrating Multiple Criteria Evaluation and GIS in Ecotourism: a Review, Int. *Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-4/W1, 351-354, Copernicus Publications, doi: 10.5194/isprs-archives-XLII-4-W1-351-2016

Omarzadeh, D., Pourmoradian, S., Feizizadeh, B., Khallaghi, H., Sharifi, A. & Kamran, K.V. (2022) A GIS-based multiple ecotourism sustainability assessment of West Azerbaijan province, Iran, *Journal of Environmental Planning and Management*, 65(3), 490-513, https://doi.org/10.1080/09640568.2021.1887827

Persada, S.F., Prasetyo, Y.T., Maharani, I.G.A.P., Apriyansyah, B., Ong, A.K.S, Young, M.N., Nadlifatin, R., Setiyati, E.A., Shanti, M., Kumalasari, R.D., Ubud, S., Putra, R.A.K., Putri, H.H.K., Gunawan, G.A., Wijaya, I.B.A., Zunaidah, A., Putri, G.S., Prisca, Y., Redi, A.A.N.P. & Razif, M. (2023) How Tourists Reacted to Ecotourism during COVID-19: Insights on Its Sustainability from a Multivariate Analysis Based on the Case of Banyuwangi, *Sustainability* 15(2), 1440. https://doi.org/10.3390/su15021440

Prasetyo, D.H., Mohamad, J. & Fauzi, R. (2018) A GIS-based multi-criteria decision analysis approach for public school site selection in Surabaya, Indonesia. *Geomatica*. 72(3), 69-84. https://doi.org/10.1139/geomat-2018-0017

Rikalovic, A., Cosic, I. & Lazarevic, D. (2014) GIS Based Multi-criteria Analysis for Industrial Site Selection, *Procedia Engineering*, 69, 1054-1063, https://doi.org/10.1016/j.proeng.2014.03.090

Ronizi, S.R.A., Mokarram, M. & Negahban, S. (2020) Utilizing multi-criteria decision to determine the best location for the ecotourism in the east and central of Fars province, Iran, *Land Use Policy*, 99, https://doi.org/10.1016/j.landusepol.2020.105095.

Ruda, A. & Pokladníková, M. (2016), Map Agebra in Tourism Potential Modelling for Improving Social Issues in Masaryk ́S School Forest Enterprise Křtiny, *Geographia Technica*, 11(1), 67-83, http://dx.doi.org/10.21163/GT_2016.111.08

Ruda, A. (2016), Exploring Tourism Possibilities Using GIS-Based Spatial Association Methods, *Geographia Technica*, 11(2), 87-101, http://dx.doi.org/10.21163/GT_2016.112.09

Sánchez-Prieto, M.C., Luna-González, A., Espinoza-Tenorio, A. & González-Ocampo, H.A., (2021) Planning Ecotourism in Coastal Protected Areas; Projecting Temporal Management Scenarios, *Sustainability*, 13(14), 7528; https://doi.org/10.3390/su13147528

Shang, Y., Sun, Y. & Xu, A. (2020) Rural ecotourism planning and design based on SWOT analysis, *International Journal of Low-Carbon Technologies*, 15(3), 368–372, https://doi.org/10.1093/ijlct/ctaa003

Vidal, M. D., Paim, F.P., Nassar, P.M. & Simonetti, S.R. (2021) Impacts of Covid-19 Pandemic on Ecotourism Segment in Amazonas State, Brazil. *Anais Brasileiros De Estudos Turísticos*, *11*. https://doi.org/10.5281/zenodo.5770250

Wegmann, M., Schwalb-Willmann, J. & Dech, S. (2020) *An Introduction to Spatial Data Analysis. Remote Sensing and GIS with Open Source Software*. Exeter, UK, Pelagic Publishing.

Whitelaw P.A., King B.E.M. & Tolkach D. (2014) Protected areas, conservation and tourism-financing the sustainable dream. *Journal of Sustainable Tourism*. 22(4), 584-603. https://doi.org/10.1080/09669582.2013.873445.

A STUDY ON OIL PALM CLASSIFICATION FOR RANONG PROVINCE USING DATA FUSION AND MACHINE LEARNING ALGORITHMS

Morakot WORACHAIRUNGREUNG [1] 🆔*, Kunyaphat THANAKUNWUTTHIROT [2]* 🆔*,*
Nayot KULPANICH [1] 🆔

ABSTRACT:

Oil palm is a vital force in driving the energy business. In 2020, Thailand had 9,954.27 sq.km. (around 6,220,799 Rai) of oil palm plantations, ranking third in the world after Indonesia and Malaysia. Ranong has the highest oil palm crop yield per Rai in Thailand. Notwithstanding, it is challenging to classify land use accurately and keep it up to date by using only labor, due to the need for a number of laborers and high labor costs. Moreover, land use/land cover cannot use spectral information classification alone. Nevertheless, machine learning is a popular data estimation technique that enables a system to learn from sample data; however, there are few studies on its use for data fusion techniques in order to classify land use/land cover, especially concerning oil palm. Therefore, we aim to apply machine learning and data fusion to classify land use/land cover, especially for oil palm. After a multicollinearity test of spectral information and ancillary variables, Surface Reflectance (SR) of Blue, Near Infrared, SWIR-1, NDWI, NDVI and LST were selected with a threshold of correlation coefficients. A stepwise stack of six inputs was created. The first stack included only Surface Reflectance (SR) of Blue, Near Infrared and SWIR-1. NDWI, NDVI and LST were added later. ID4 (Surface Reflectance (SR) of Blue, Near Infrared, SWIR-1, NDWI, NDVI and LST) in the random forest model resulted in OA being 0.9341 and KC being 0.9239, which was the highest among 12 models. ID4 in the random forest model provided the classification results for oil palm very close to the factual number per the figure of 2.90 sq.km (around 1,814 Rai) from the Department of Land.

Key-words: Oil palm, Ranong, Data fusion, Machine learning, Remote Sensing

1. INTRODUCTION

All industries inevitably need energy to drive their industries. Oil palm is a crucial factor in driving the energy business. From the 1970s to the 2020s, oil palm area has dramatically doubled, and such an increase in oil palm plantations affects the ecosystem. At present, alternative energy has been used to replace fossil fuels, which includes non-renewable petroleum, natural gas and coal as the main sources of electricity and energy used in daily life. Biomass, or biological energy, obtained from palm oil is an alternative as a renewable energy source; for instance, biodiesel renewable energy can replace fuel for future transportation and can also be used as a raw material in soap and foods such as condensed milk, ice cream and butter. This is in line with 17 sustainable development goals as presented by the United Nations (Shaharum N. S., et al., 2020).

In 2020, Thailand had 9,954.27 sq.km. (6,220,799 Rai) of oil palm plantations, which ranked third in the world after Indonesia and Malaysia. The southern region of Thailand has 8,511.36 sq.km. (5,319,602 Rai) of oil palm plantations, and has a total area of 8,077.63 sq.km. (5,048,519 Rai) that can currently produce actual production volume. The province with the largest growing area in the southern region is Krabi, with a total area of 1,873.30 sq.km. (1,170,815 Rai) of oil palm plantation, but the province with the highest crop yield per Rai is Ranong, serving 2,980 kilograms per 1,600 sq.m. (Land Development Department, 2021). Accurate and up-to-date data is important for land management. Notwithstanding, it is quite difficult to classify land use accurately and up-to-date using labor, due to the need of the number of laborers and high labor costs. Using satellite images to classify

---

[1]*Suan Sunandha Rajabhat University, Faculty of Humanities and Social Sciences, Geography and Geo-Informatics Program, 1 U-Thong nok Road, Dusit, Bangkok 10300 Thailand,*
Morakot.wo@ssru.ac.th,nayot.ku@ssru.ac.th

[2] *Suan Sunandha Rajabhat University, Faculty of Fine and Applied Arts, Digital Design and Innovation Program, 1 U-Thong nok Road, Dusit, Bangkok 10300 Thailand, kunyaphat.th@ssru.ac.th*

land use can well solve the problems of the number of laborers and high labor costs. Until the present, satellite images have been used to classify land use in numerous cases (George, Padalia & Kushwaha, 2014).

At present, satellite imagery is applied in the classification of many field crops. The use of satellite images to pinpoint old palm areas is widely popular (Srestasathiern & Rakwatin, 2014; Li, Dong , Fu, & Yu, 2019). To illustrate, Thenkabail used satellite images at 4-meter resolution captured from an IKONOS satellite to study oil palm biomass (Thenkabail, et al., 2004). Gutiérrez used satellite images at 250-meter resolution from MODIS to study oil palm area spanning 939,204 square kilometres (Gutiérrez-Vélez & DeFries, 2013). J. Miettinen studied oil palm plantations across Southeast Asia and Peninsular Malaysia, Sumatra, Java, Borneo, Sulawesi and Mindanao islands (Miettinen, Shi, Tan, & Liew, 2012). This study classifies land use by separating data into 13 layers, including mangrove forests, forests, rain forests and oil palm. By using high-resolution data, it can also help classify oil palm areas. In 2011, Shafri used the maximum likelihood classifier to classify Ganoderma disease infected oil palms with an accuracy of 82 percent (Shafri, Hamdan, & Saripan, 2011). Moreover, Kulpanich et al. used the UAV images to collect relevant data to forecast oil palm yields (Kulpanich, et al., 2022). However, the limitation of UAV images is that the large area will take a lot of time and budget for the operation. From the above statement, it was found that moderate-resolution satellites (MODIS) can classify oil palms, so it is believed that LANDSAT9, with a higher spatial resolution than MODIS, will also be able to classify oil palms.

Regarding the classification algorithm, Morel successfully differentiated forest from oil palm using k-means and an MLC algorithm (Morel, Fisher, & Malhi, 2012). In addition, Cheng successfully classified land cover using LANDSAT and ALOS-PALSAR through SVM and Minimum Distance algorithms (Cheng, Yu, Cracknell, & Gong, 2016). The study found that, for the classification of two areas, SVM was better than Minimum Distance algorithms that give satellite images from LANDSAT and ALOS-PALSAR. Furthermore, Cheng's study covered the areas of Malaysia, Indonesia, Thailand and Nigeria, with an accuracy of over 94 percent for those aforementioned countries (Cheng, et al., 2018). But it found little application using machine learning in classifying oil palms by satellite imagery.

Nowadays, machine learning in classification has been widely adopted (Worachairungreung, et al., 2021; Worachairungreung, Thanakunwutthirot, & Ninsawat, 2019) Machine learning is mostly applied on oil palm classification for interpretation. Nooni used Support Vector Machine learning models to classify oil palm areas (Nooni, et al., 2014). Sitthi used Naive Bayes classifiers to identify what is covered in given areas, (Sitthi, et al., 2016) and Mubin used a convolutional neural network as a deep learning method to identify young and mature oil palm trees) Mubin, et al., 2019). Nevertheless, it is rare to find classification research comparing multiple algorithms given two types of satellite images or more.

The classification of LULC is complex. Currently, data fusion techniques are used to help classify many LULCs. Data Fusion is a method or tool to combine remote sensing data from different sources and multiply them to create new data in order to obtain representative data. Some researchers use a Digital Elevation Model (DEM) and SAR derived features that contribute the most to building damage classification. Classification results showed an overall accuracy of >90% and an average of >67% (Adriano, et al., 2019). Some researchers used a data fusion method and NDVI time-series analysis-based phenology extraction. The Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) technique accurately blended SPOT5 and MODIS NDVI in Shandong Province, China, where counties tested phenology detecting methods with data fusion techniques (Yin, et al. 2019). Some researchers have used mono-temporal and multi-temporal LULC classifications and auxiliary data to determine LULCC in southwest Burkina Faso's varied landscape. Multi-temporal classification outperformed mono-temporal classification in the research area (Zoungrana , et al., 2015). According to the study, data fusion improves classification outcomes. Therefore, this study wanted to use such a technique to classify oil palm. Oil palm is an economically important plant in southern Thailand, but data fusion and machine learning are rarely used to classify oil palms, so in this study, Landsat 9 satellite imagery with 30 meter spatial resolution was used to classify it. In addition, data fusion and

machine learning algorithms such as SVM, Random Forest and CART were used to classify palm, but this study studied only the Ranong dataset. Finally, it is hoped that this study will help classify oil palm in Thailand, as well as other countries in the region.

2. STUDY AREA

Ranong Province is in the southern region of Thailand, with an area of 3,426 sq.km. (2,141,250 Rai). The province is comprised of five districts: Mueang District, La-un-District, Kapoe District, Kra Buri District, and Suk Samran District, with an elevation range between 0-1,388 meters above sea level. Ranong Province is located in the southwest part of Thailand, and entirely influenced by the southwest monsoon. It receives more abundant rainfall than other provinces and it falls most of the year. Most areas of the province are covered by rubber, orchard, forest, mangrove and oil palm. (**Fig. 1**).
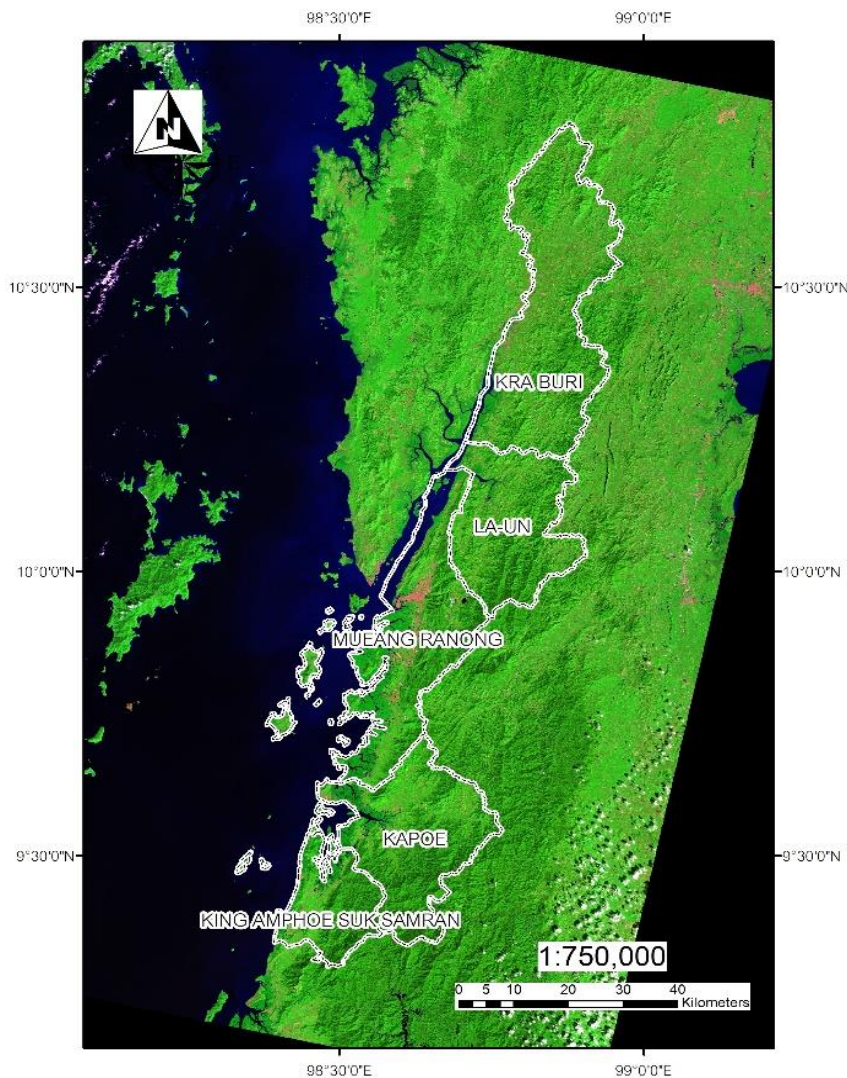

Fig. 1. Ranong Province.

3. DATA AND METHODS

In this study, satellite imagery and data were used. We then take satellite imagery analyzing the land surface, temperature, soil index, water index and vegetation index, and run the analysis results through data fusion. We then use a multicollinearity method to reduce data redundancy. We apply machine learning algorithms by dividing the training and testing datasets into 80:20 proportions. Finally, we compare the results. **Figure 2** shows the overall methodology mentioned above.
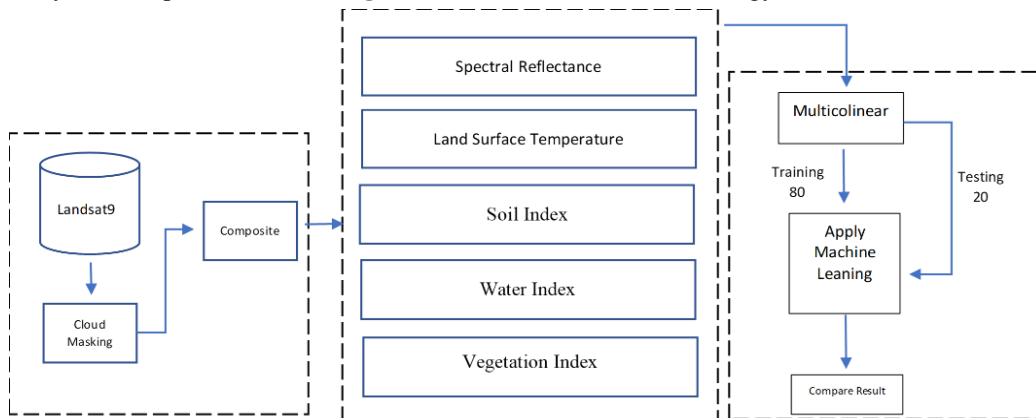

Fig. 2. Overall Methodology

3.1. Landsat 9 Spectral Reflectance Data

In this study, Landsat 9 was selected because the Landsat program continues its mission to capture repetitive observations worldwide for monitoring, comprehending and managing Earth's natural resources, with Landsat 9 under collaboration between the U.S. Geological Survey (USGS) and the National Aeronautics and Space Administration (NASA). Researchers rely on the USGS's Landsat archival data, which has been freely available since 1972, to map changes to the land's surface, but it is required to be pre-processed to make it usable. The researchers chose and altered remotely sensed products using an on-demand interface provided by the USGS Earth Resources Observation and Science (EROS) Center. **Table 1** shows the Landsat 9 details.

Table 1.

Landsat 9 Spectral Reflectance Data

Spectral	Wavelength in micrometers	Resolution in meters
Operational Land Imager		
Band 1—Ultra blue (coastal/aerosol)	0.435–0.451	30
Band 2—Blue	0.452–0.512	30
Band 3—Green	0.533–0.590	30
Band 4—Red	0.636–0.673	30
Band 5—Near infrared (NIR)	0.851–0.879	30
Band 6—Shortwave infrared (SWIR) 1	1.566–1.651	30
Band 7—Shortwave infrared (SWIR) 2	2.107–2.294	30
Band 8—Panchromatic	0.503–0.676	15
Band 9—Cirrus	1.363–1.384	30
Thermal Infrared Sensor		
Band 10—Thermal infrared (TIR) 1	10.60–11.19	100
Band 11—Thermal infrared (TIR) 2	11.50–12.51	100

The Landsat 9 dataset contains surface reflectance from an Operational Land Imager, top of atmospheric (TOA) reflectance, and TOA brightness for temperature in Kelvin, as well as spectral indices, including a Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI), and Normalized Difference Moisture Index (NDMI).

The Landsat 9 scene of path 130 and row 53 was projected using UTM with datum WGS84, and was acquired during a field survey (January 2022). This study adopted SR Bands TOA brightness temperature, Normalized Difference Vegetation, Normalized Difference Vegetation Index, Normalized Difference Water Index, and Soil Index at 30-meter resolution. The researchers only considered the optical bands (2 to 7) for classification among all SR bands. The TOA brightness temperature (only band 10) was utilized to estimate land surface temperature.

3.2. Field Data

Table 2 and **figure 3** show field data. Field data from 1,600 field data points was obtained from the Global Positioning System (GPS). The field data consisted of rubber, oil palm, orchard, forest 1, forest 2, mangrove, built-up area, and water bodies. The data were divided into 2 sets: set 1 was for modeling, and set 2 was for testing accuracy of the model under the proportion of 80:20, respectively.

Table 2.
Field Data

Land Cover Class	NO. Field Data
Bare Soil	36
Rubber	51
Orchard	92
Forest	160
Evergreen Forest	390
Mangrove	245
Oil Palm	420
Built up Area	104
Water Bodies	104

3.3. Auxiliary Variables

In this study, Landsat 9 consisted of SR bands 2-7 and auxiliary data of NDVI, SI and NDWI, including Land Surface Temperature estimation. The purpose of using such auxiliary variables was to improve classification accuracy.

3.3.1. Land Surface Temperature

There are many ways to calculate Land Surface Temperature (LST). Previous research comparing different LST estimation methods favored the Radiative Transfer Equation (RTE)-based technique and band 10 over band 11. (Zhou, et al., 2012; Jiménez-Muñoz, et al., 2014; Santos, et al., 2018; Rehman, et al., 2021). In this study, we used Landsat 9 TIRS Band 10 to estimate LST by using a technique based on the Radiative Transfer Equation (RTE), as described in Equation (1).

$$\text{LST}_{B10} = \tau i(\theta)\text{EiBi}\,(Ts) + \big((1 - Ei)I^{\downarrow}\big) + I^{\uparrow} \qquad (1)$$

i(θ): atmospheric transmission for band 10 when view zenith angle is θ;
Ei: surface emissivity of the band 10;
Bi(Ts) : a ground radiance;
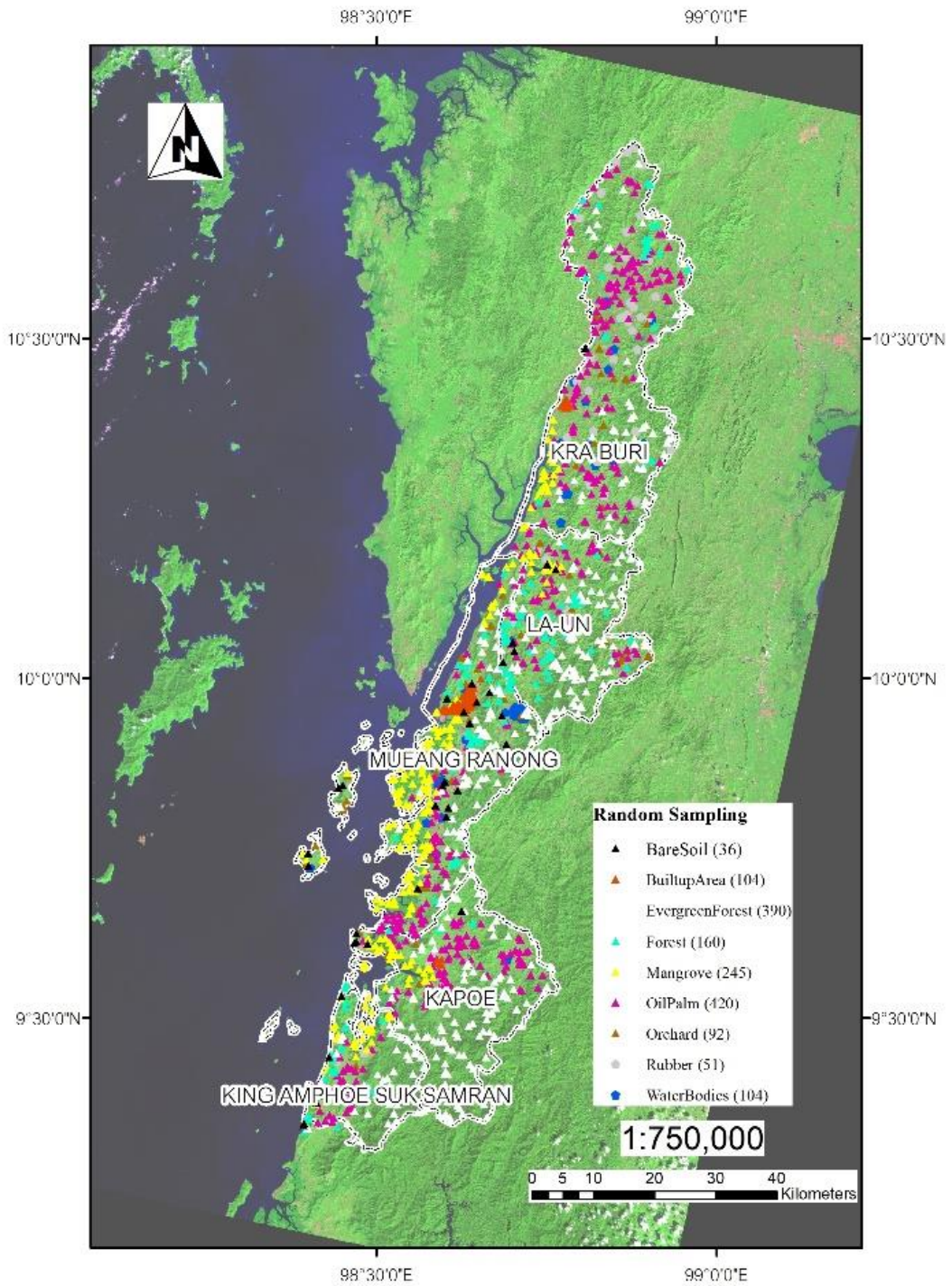I^↓ : Downwelling path radiance;
I^↑ : Upwelling path radiance.

Fig. 3. Field Data.

According to Plank's law, ground radiance Bi(Ts) can be expressed as:

$$Bi(Ts) = \frac{2hc^2}{\lambda_{Bi}^5 \left(\exp \left(\frac{hc}{\lambda_{Bi} k Ts} \right) - 1 \right)} \tag{2}$$

where c is the speed of light (c = 2.9979 x 108 m/s), h is the Planck constant (h = $6.6261 \times 10\text{-}34$ J.s), k is the Boltzmann constant (k = $1.3806 \times 10\text{-}23$ J/K), λ represents the wavelength of TIRS bands (B10 = 10.602), and Ts is TOA brightness temperature.

3.3.2. Soil Index

Soil is a substance with several chemical and physical components (Thenkabail, et al., 2004). In this study, we used the Normalized Difference Soil Index (SI) proposed by Deng (Deng, et al., 2015). They studied the spectrum reflectance of soil samples using Landsat-5 data and discovered that the mean reflectance values of bands NIR, SWIR1 and SWIR2 are greater than those of visible bands. In addition, they examined all potential band normalized difference combinations and concluded that the index obtained from the SWIR2 and green bands are superior for mapping soil information.

$$SI = \frac{Green - SWIR2}{Green + SWIR2} \tag{3}$$

3.3.3. Water Index

McFeeters came up with the NDWI, which is the ratio of the green band to the NIR band (McFEETERS, 1996). Xu revised the NDWI to become the ratio of the green band to the SWIR band (Xu H. , 2006). Using Landsat 8 data, a previous study compared both methods developed by McFeeters and Xu, and found that the best way to map water bodies is to use both Green and SWIR (Du, et al., 2014). Hence, in this study, we applied the equation for the water index made by Xu. Green Bands and SWIR1 were used in this method.

$$NDWI = \frac{Green - SWIR1}{Green + SWIR1} \tag{4}$$

3.3.4. Vegetation Index

The normalized difference vegetation index (NDVI) is one of the most often used vegetation indices, and its value in satellite assessment and monitoring of global plant coverage has been well recognized over the last two decades (Huete, Justice, & Liu, 1994; Leprieur, et al., 2000). The formula is described as follows.

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{5}$$

3.3.5. Multicollinearity Test

Implementing land use classification requires testing for SR and ancillary data redundancy. The solution to such a problem in this study was to use a correlation matrix between the SR bands/ancillary variables. From the study, if the correlation value was greater than 0.7, it was necessary to reduce several redundant variables before importing to a machine learning model, as shown in **figure 4**.
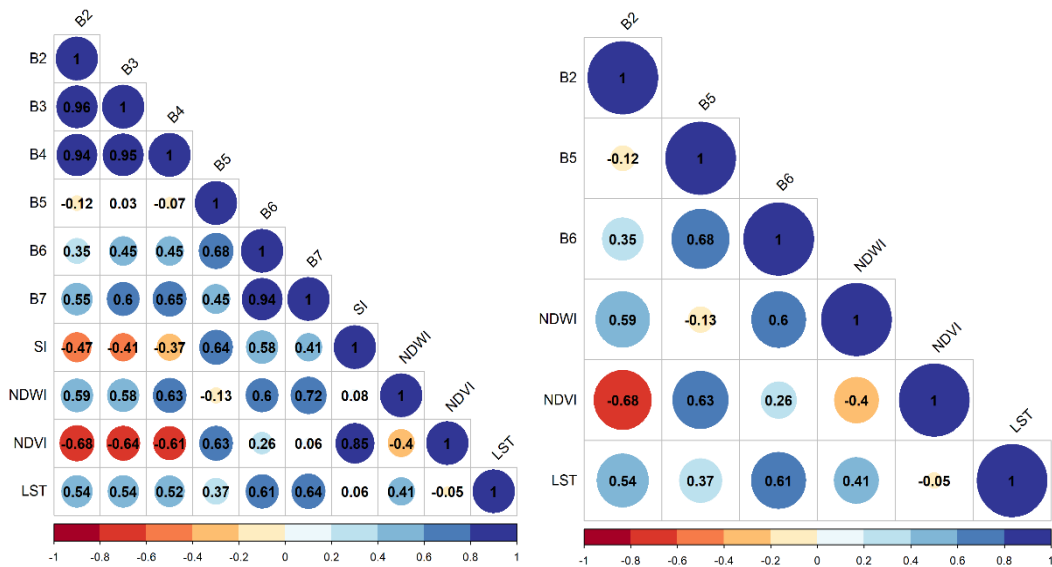
Fig. 4. Results of multicollinearity test to reduce redundancy of SR and ancillary data before importing to a machine learning model.

(B2 = Blue band, B3 = Green band, B4 = Red band, B5 = Near Infrared band, B6 = SWIR-1 band, B7 = SWIR-2 band, NDVI = Normalized Difference Vegetation Index, SI = Normalized Differ-ence Soil Index, NDWI = Normalized Difference Water Index, NDMI = Normalized Difference Moisture Index, LST = Land Surface Temperature)

3.4 Machine Learning Model

3.4.1. CART

The classification and regression tree (CART) is a tree model used in the field of machine learning. CART creates a visualized decision tree to predict a categorical and continuing variable and hence this tree does not create classes of dependent variables. Rather than a classification tree, where an input space of variables is split into subspaces, each of them associates with a class of outputs. Dependent variables are the response values from each observation, given a set of independent variables. As a regression tree does not have preset classes, the output at each stage will be a response value from observations of new dependent variables. A minimization step is applied to create a splitting rule in a tree so that the projected sum variance from two nodes will be deduced.

Proposed by Breiman et al. (Breiman, Classification and Regression Trees, 1984), the classification and regression tree is one of the most adopted methods for handling classification and regression problems. A CART model employs the Gini and least-squared method to deal with categorical and numerical problems, respectively (Breiman, 1996). Given the pth sample is defined as $(I_{(p,1)}, I_{(p,2)}, \ldots\ldots I_{(p,n)}\ldots O_p)$, where $I_{(p,n)}$ is a value of the pth sample with "n" features, and "O_p" is an output value of the sample, minimization of the least-squared deviation under Equation (1) will help create a choice to split up a given tree into branches for a CART regression problem.

$$\frac{1}{N} \sum_{V \in U_r} \left(O_p - \bar{O}_r\right)^2 + \frac{1}{N} \sum_{V \in U_l} \left(O_p - \bar{O}_l\right)^2 \qquad 6)$$

where "U_r" and "U_l" are training data sets of right and left child nodes, and "N" is the total number of training samples, the outcome of the right and left nodes is denoted as $\bar{O}$ and ($\bar{O}_l$).

3.4.2. Random Forest

A random forest contains a set of tree classifiers, each of them is generated by using a random vector sampled discretely from a given input vector with a vote for the most popular class for categorizing an input vector (Breiman, 2001). The random forest then produces a tree by picking or combining features at each node separately. For each feature or combination of features, a technique called bagging to generate a dataset for training by choosing N instances for replacement randomly, given that N is the size of the original training set, is employed. Any instance is characterized by choosing the class having the highest voting score from all tree predictors within the forest. The design process of decision tree requires a selection measure and a pruning method. Several ways to pick characteristics for decision tree induction are available, and most tactics can clearly measure the attributes. The Information gain ratio and Gini index are mostly adopted attributes as selection metrics for induction of a decision tree. The random forest employs the Gini index as an attribute selection matrix which measures the impurity of an attribute about the classes. As depicted in Equation (7), for a given training set T, the Gini index is expressed as follows:

$$\sum \sum_{j \neq i} \left(f(C_i, T)/|T| \right) \left(f(C_j, T)/|T| \right) \tag{7}$$

where $f(C_i, T)/|T|$ is a probability that a selected instance fits into class C_i.

A tree will be formulated each time to its maximum depth by utilizing a mix of features with new training data. The most mature tree remains as-is, this is a benefit of the random forest over other decision tree techniques. Findings reveal that pruning strategies impact the performance of tree-based classifiers rather than the attribute selection criteria (Pal & Mather, 2003).

3.4.3. Support Vector Machine

Introduced by Vapnik et al. (Vapnik, Golowich, & Smola, 1996), support vector machine (SVM) is a supervised classification method to reshape a non-linear environment into a linear one, and make a simple class computable through the generation of a hyper-plane. A kernel function is a mathematical function for transforming data. SVM uses a training dataset to transform an original input into a high-dimensional feature space. A hyper-plane is made from the points of tree classes in the original space of n coordinates. SVM computes the maximum difference across classes to form a classification hyper-plane at the center of the maximum margin. That is, if a point is above the hyper-plane, it is considered as +1; otherwise, it is treated as -1. Support vectors are the training points nearest to the hyper-plane. The new data can then be classified after the decision surface is obtained; such decision surfaces can be utilized to classify auxiliary data. The method is defined over a vector space. The decision surface for a linearly separable space is a hyper-plane, which can be expressed as per Equation (8):

$$wx + b = 0 \tag{8}$$

A vector "w" and constant "b" are derived from a training set of linearly separable items and "x" is an object for characterization. SVM can deal with a problem about linearly restricted quadratic programming such as in Equation (9), and the SVM solution is always globally optimal.

$$\min_{\omega} \frac{1}{2} \parallel \omega \parallel^2 + C \sum_i \xi_i \tag{9}$$

with constraints

$$y_i(x_i w + b) \geqslant 1 - \xi_i \, \xi_i \geqslant 0, \forall i \tag{10}$$

By performing non-linear mapping for linearly inseparable objects, the original input data is converted into a higher dimensional space, and the linearly separating hyper-plane can be found in a new space without any additional computational complexity or quadratic programming problems by

applying a kernel function (Aizerman, 1964). In other words, to compute similarities across vectors in a high-dimensional space for a linearly inseparable problem, the kernel function is applied to reduce those similarities in the original lower dimensional space.

3.5 Preparation of a Training Signature for the Classification of Oil Palm and LULC using a Machine Learning Model

The spectral bands of Landsat 9 and their derived supplementary bands, as well as the ground sample points (representing each forest type and LULC), were opened in an R statistical program by utilizing raster (Hijmans, 2014) and rgdal (Bivand, 2002) packages. Extraction of training points from the stack images made up of all the spectral and derived ancillary bands was used to produce training signatures. Categorization of oil palm and LULC was finally done using a hierarchy. Six input datasets (ID) created for mapping and classifying oil palm and related LULC in Ranong Province were gradually categorized into a classification hierarchy. Each ID was made up of a variety of spectral data (George, Padalia, & Kushwaha, 2014). Blue, NIR and SWIR2 spectral bands were identified using Landsat 9 in the first stage and results were recorded. Next, an auxiliary variable was added to each ID, and its impact on classification accuracy was assessed using the overall accuracy and Kappa coefficient (**Tab. 3**). In this experiment, three machine learning models were used: Random Forest, Support Vector Machine and CART, and an auxiliary variable was examined in terms of its impact.

Table 3.

Surface Reflectance (SR) and Auxiliary Variable

Auxiliary Variable	Description
ID 1	Surface Reflectance (SR) of Blue, Near Infrared, SWIR-1
ID 2	Surface Reflectance (SR) of Blue, Near Infrared, SWIR-1 + NDWI
ID 3	Surface Reflectance (SR) of Blue, Near Infrared, SWIR-1 + NDWI + NDVI
ID 4	Surface Reflectance (SR) of Blue, Near Infrared, SWIR-1 + NDWI + NDVI + LST

4. RESULTS AND DISCUSSION

4.1. Overall Accuracy Results and Kappa Coefficient of the Machine Learning Model

Overall Accuracy simply informs us of the percentage of reference locations that were accurately mapped. The Kappa Coefficient is a statistical test for classifying accurately. Kappa measures how well a categorization fared relative to randomly assigned values. **Table 4** shows overall accuracy results and the Kappa coefficient of Land Use/Land Cover classification in Ranong Province. For ID1 (Surface Reflectance (SR) of Blue, Near Infrared and SWIR-1), it was found that the Random Forest model had OA = 0.9103 and KC = 0.8951. OA and KC would increase with the added variables, as ID4 (Surface Reflectance (SR) of Blue, Near Infrared, SWIR-1 + NDWI + NDVI + LST) resulted in OA = 0.9341 and KC = 0.9239, which were the highest among 12 models. Most interesting was that the CART model under ID2 (Surface Reflectance (SR) of Blue, Near Infrared, SWIR-1 + NDWI) and ID3 (Surface Reflectance (SR) of Blue, Near Infrared, SWIR-1 + NDWI + NDVI) had OA = 0.8646, 0.9239 and KC = 0.8427, 0.9103, respectively, and were higher than the Random Forest model under ID2 and ID3. Details of the OA and KC values of all models can be found in **table 4** and **figure 5**. Only the parts with the highest OA and KC values of map ID4 of CART and RF are shown in **figure 6**. In terms of Bare Soils, some models had low producer accuracy, particularly in the ID1 model (Surface Reflectance (SR) of Blue, Near Infrared and SWIR-1). However, after adding more variables, it was found that producer accuracy of such models was higher. Other models had mixed results of producer accuracy and user accuracy in land use/land cover classification. In terms of oil palm, it was found that every SVM and RF in ID 4 is the model that produces the best results. Forest, mangrove, built-up area and water bodies, as well. The classification details using SVM in Oil Palm SVM give the best results over RF, but RF gives OA and KC of all classifications better than SVM.

Table 4.

**Overall Producer and User Accuracy of Land Use/Land Cover Classification
in Ranong Province.**

Model			Bare Soil	Oil Palm	Orchard	Forest	Evergreen Forest	Mangrove	Rubber	Built Up Area	Water Bodies	OA	KC
RF	ID1	UA	1.00	1.00	0.44	1.00	0.94	0.89	1.00	1.00	1.00	**0.91**	**0.90**
		PA	0.33	0.67	0.80	1.00	0.85	1.00	1.00	1.00	1.00		
	ID2	UA	0.50	1.00	0.57	1.00	0.67	1.00	0.88	1.00	1.00	**0.86**	**0.83**
		PA	0.50	1.00	0.80	0.90	0.92	1.00	0.50	1.00	1.00		
	ID3	UA	1.00	0.67	0.50	1.00	0.91	1.00	1.00	1.00	1.00	**0.91**	**0.89**
		PA	0.50	1.00	0.83	1.00	0.95	1.00	0.56	1.00	1.00		
	ID4	UA	0.50	1.00	1.00	1.00	0.81	1.00	0.83	1.00	1.00	**0.93**	**0.92**
		PA	1.00	1.00	0.80	1.00	0.93	0.86	0.91	1.00	1.00		
SVM	ID1	UA	0.00	1.00	0.38	1.00	0.67	1.00	0.75	1.00	1.00	**0.82**	**0.79**
		PA	0.00	1.00	0.60	1.00	0.92	1.00	0.23	1.00	1.00		
	ID2	UA	1.00	1.00	0.44	1.00	0.71	1.00	1.00	1.00	1.00	**0.86**	**0.84**
		PA	0.67	1.00	1.00	1.00	0.92	0.91	0.50	1.00	1.00		
	ID3	UA	1.00	1.00	0.25	1.00	0.89	1.00	0.91	1.00	1.00	**0.89**	**0.87**
		PA	0.50	1.00	1.00	1.00	0.89	1.00	0.63	1.00	1.00		
	ID4	UA	1.00	1.00	0.60	1.00	0.93	0.93	0.72	1.00	1.00	**0.88**	**0.87**
		PA	0.33	0.40	1.00	1.00	0.93	1.00	0.81	1.00	1.00		
CART	ID1	UA	1.00	0.67	0.40	1.00	0.81	0.94	1.00	1.00	1.00	**0.88**	**0.86**
		PA	0.33	0.67	0.57	1.00	0.87	1.00	0.81	1.00	1.00		
	ID2	UA	1.00	0.50	0.56	1.00	0.75	1.00	0.81	1.00	1.00	**0.86**	**0.84**
		PA	0.67	0.33	0.71	1.00	0.88	1.00	0.68	1.00	1.00		
	ID3	UA	1.00	1.00	0.44	1.00	0.89	1.00	1.00	1.00	1.00	**0.92**	**0.91**
		PA	1.00	1.00	1.00	1.00	0.89	1.00	0.72	1.00	1.00		
	ID4	UA	0.67	1.00	0.60	1.00	1.00	1.00	0.82	1.00	1.00	**0.91**	**0.90**
		PA	1.00	0.60	0.86	0.90	0.90	1.00	0.90	1.00	1.00		

Fig. 5. OA and KC values of all models.

(a) Landsat 9   (b) CART (ID4)   (c) RF (ID4)

Bare Soil Rubber Orchard Forest Evergreen Forest Mangrove Oil Palm
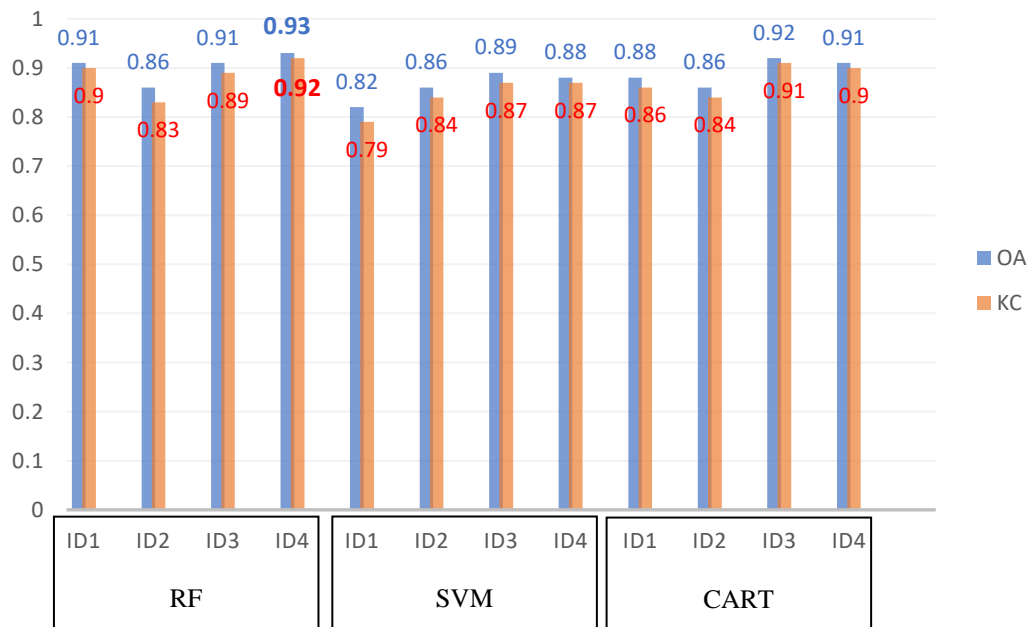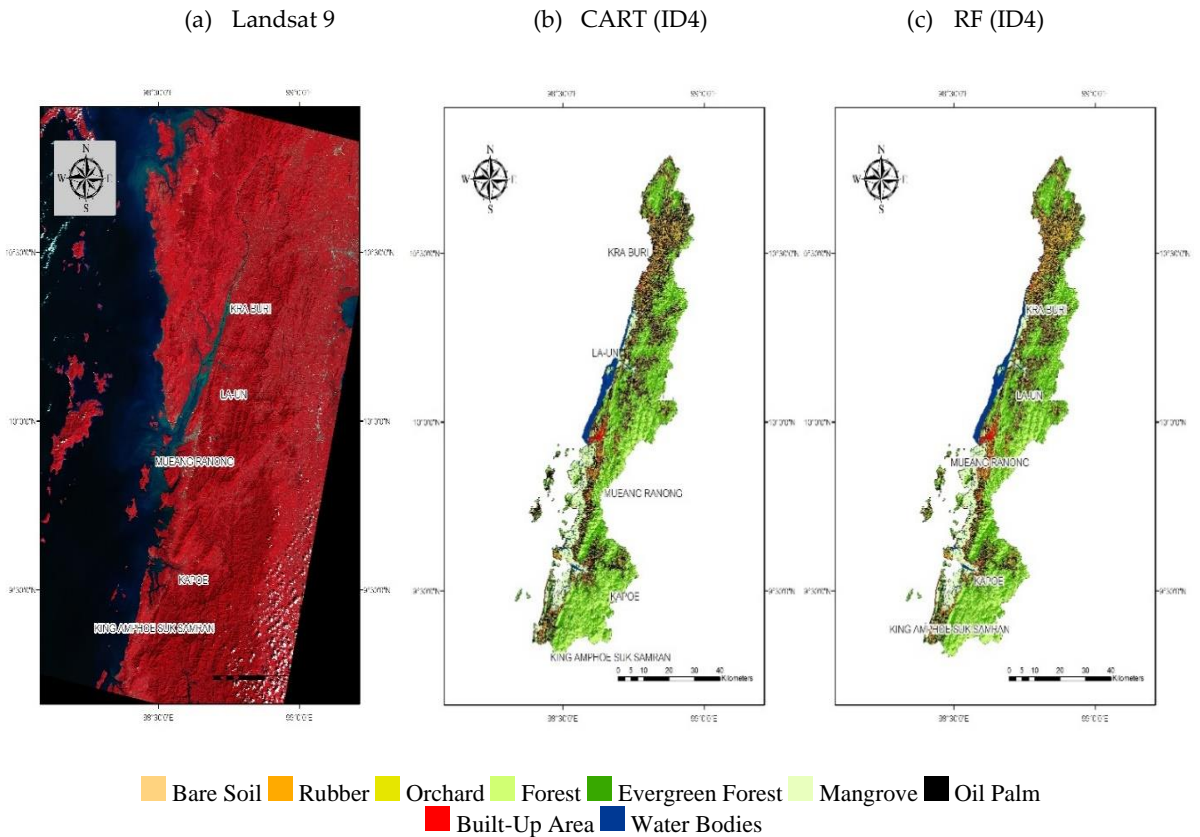Built-Up Area Water Bodies

Fig. 6. Example of the improvements of using Surface Reflectance (SR) and Auxiliary Variable (a) Landsat 9 imagery; (b) Results of using CART and Surface Reflectance (SR) of Blue, Near Infrared, SWIR-1 + NDWI + NDVI + LST; (C) Results of using RF and Surface Reflectance (SR) of Blue, Near Infrared, SWIR-1 + NDWI + NDVI + LST

Table 5 shows the difference in results of LULC classification on oil palms. Compared to the facts from the Department of Land, ID4 (Surface Reflectance (SR) of Blue, Near Infrared, SWIR-1 + NDWI + NDVI + LST) of CART and RF models gave the classification results closest to the facts from the Department of Land. The reason for choosing these two models was that both OA and KC values were the most accurate. When considered at the district level, it was found that the ID4 of RF models provided classification results of oil palms very close to the facts per the figure of 2.90 sq.km. (1,814 Rai) from the Department of Land, especially for Kra Buri District. The discrepancy between classification of oil palms and actual data was only 83,200 sq.m. (52 Rai). Details of such differences can be found in **table 5**.

Table 5.

Difference in the results of LULC classification on oil palms.

District	Land Development Department	CART (ID4)		RF (ID4)	
		Classified	Difference	Classified	Difference
KRA BURI	29,396	29,174	222	29,448	52
MUEANG RANONG	15,227	15,090	137	12,349	2,878
KAPOE	28,578	30,180	1,602	30,398	1,820
LA-UN	14,438	15,090	652	13,299	1,139
SUK SAMRAN	9,170	11,066	1,896	9,500	330
Ranong Province	96,809	100,599	3,790	94,995	1,814

4.2. Mean Decrease Gini

Mean Decrease Gini (IncNodePurity) - is a measure of variable importance based on the Gini impurity index used for calculating the splits in trees. The first feature and B2 and NDVI shows a high Gini impurity index in the classification model; hence, these features are important for detection of land use/land cover classification, as shown in **figure 7**.
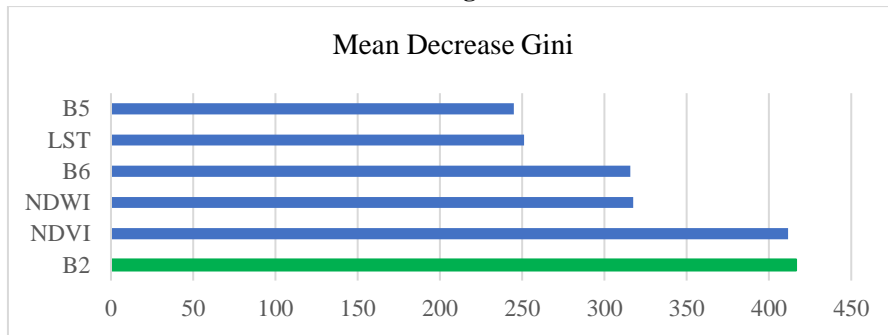

Fig. 7. Mean Decrease Gini of ID4 in a random forest model.

This study demonstrates the capability of using Landsat 9 satellite images for land use/land cover classification. Moreover, using the machine learning model we can precisely classify land use/land cover, especially for oil palms. This study found that multicollinearity is a tool that can significantly reduce variable redundancy. When the remaining variables were used for the Land Use/Land Cover Classification experiment, it was found that variables obtained by fusion data methods such as NDWI, NDVI and LST all resulted in greater accuracy. In addition, machine learning algorithms provide good land use/land cover classification results, especially RF models with the highest OA and KC results. In terms of local economic plants, oil palm, SVM and RF models provided good classification results in both models. This is consistent with Xu's study, which found that Landsat 8 and Sentinel can classify land use/land cover by machine learning (Xu, et al., 2021). Such studies can also classify the life cycle of oil palms using RF as a model, as in this study. In addition, current studies prefer to increase classification accuracy by using the data fusion technique. This technique is a method to bring together different and diversified remote sensing data sources to create new or representative data with the objective of improving data quality, adding more dimensions of data leading to an increase in classification accuracy. This study approach results in more accurate classification than using SR Bands alone. This is consistent with studies by Xu, Shaharum, Poortinga (Xu, et al., 2021; Shaharum N .S ., et al ., 2020; Poortinga, et al ., 2019), especially for Rehman, who found that the multicollinearity test can be used to eliminate factors, removing redundancy of the variables before land use/land cover classification (Rehman, et al., 2021). Besides, Rehman found that adding factors such as indices gradually results in a more accurate model. Like this study, when looking into details, it was found that the SR band 2 NDVI index has more influence on classification, which is consistent with Manandhar's study (Manandhar, Odeh, & Ancev, 2019).

5. CONCLUSIONS

This study highlights the benefits of using more than one data source to create a higher quality dataset and found that machine learning can classify plantations very well. Other researchers can apply such an approach to study other plantations in the future. This study found that, currently, land use/land cover classification cannot use only SR bands alone, so a data fusion technique is necessary to create new or representative data with the aim of improving the quality of information. It found that a machine learning model could also classify land use/land cover precisely. The findings of this study are consistent with several previous studies. Further studies may use a data fusion technique

with SAR data to come up with detailed information reflecting objects on the surface of Earth, or use Google Earth Engine, a massive, systematically compiled data fusion source to further expand the project's success. Such techniques could also be applied to other plants in the future.

ACKNOWLEDGEMENTS

R E F E R E N C E S

Adriano, B., Xia, J., Baier, G., Yokoya, N., & Koshimura, S. (2019). Multi-Source Data Fusion Based on Ensemble Learning for Rapid Building Damage Mapping during the 2018 Sulawesi Earthquake and Tsunami in Palu, Indonesia. Remote Sensing, 11(7), 886. doi:10.3390/rs11070886

Aizerman, M. A. (1964). Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning. Automation and Remote Control. Automation and Remote Control, 25, 821-837.

Bivand , R. (2002). Spatial econometrics functions in R: Classes and methods. Journal of Geographical Systems volume, 4, 405–421. doi:10.1007/s101090300096

Breiman, L. (1996). Bagging predictors. Machine Learning, 24, 123–140. doi:10.1007/BF00058655

Breiman, L. (2001). Random Forests. Machine Learning, 45, 5–32. doi:10.1023/A:1010933404324

Breiman, L. (1984). Classification and Regression Trees. New York: Routledge. doi:doi.org/10.1201/9781315139470

Cheng, Y., Yu, L., Cracknell, A. P., & Gong, P. (2016). Oil palm mapping using Landsat and PALSAR: a case study in Malaysia. International Journal of Remote Sensing, 37(22), 5431-5442. doi:10.1080/01431161.2016.1241448

Cheng, Y., Yu, L., Xu, Y., Lu, H., Cracknell, A. P., Kanniah, K., & Gong, P. (2018). Mapping oil palm extent in Malaysia using ALOS-2 PALSAR-2 data. International Journal of Remote Sensing, 39(2). doi:10.1080/01431161.2017.1387309

Deng, Y., Wu, C., Li, M., & Chen, R. (2015). RNDSI: A ratio normalized difference soil index for remote sensing of urban/suburban environments. International Journal of Applied Earth Observation and Geoinformation, 39, 40-48. doi:10.1016/j.jag.2015.02.010

Du, Z., Li, W., Zhou, D., Tian, L., Ling, F., Wang, H., & Gui, Y. (2014). Analysis of Landsat-8 OLI imagery for land surface water mapping. Remote Sensing Letters, 5(7), 672-681. doi:10.1080/2150704X.2014.960606

George, R., Padalia, H., & Kushwaha, S. P. (2014). Forest tree species discrimination in western Himalaya using EO-1 Hyperion. International Journal of Applied Earth Observation and Geoinformation, 28, 140-149. doi:10.1016/j.jag.2013.11.011

Gutiérrez-Vélez, V. H., & DeFries, R. (2013). Annual multi-resolution detection of land cover conversion to oil palm in the Peruvian Amazon. Remote Sensing of Environment, 129, 154-167. doi:10.1016/j.rse.2012.10.033

Hernández, F. W., Calderón, N. G., & da Silva, P. R. (2022). Oil Palm Yield Estimation Based on Vegetation and Humidity Indices Generated from Satellite Images and Machine Learning Techniques. AgriEngineering(4), 279–291. doi:10.3390/agriengineering4010019

Hijmans, R. J. (2014). Introduction to the 'raster' package (version 2.2-31).

Huete, A., Justice, C., & Liu, H. (1994). Development of vegetation and soil indices for MODIS-EOS. Remote Sensing of Environment, 49(3), 224-234. doi:10.1016/0034-4257(94)90018-3

Jiménez-Muñoz, J. C., Sobrino, J. A., Skoković, D., Mattar, C., & Cristóbal, J. (2014). Land Surface Temperature Retrieval Methods From Landsat-8 Thermal Infrared Sensor Data. IEEE Geoscience and Remote Sensing Letters, 11(10), 1840-1843. doi:10.1109/LGRS.2014.2312032

Kulpanich, N., Worachairungreung, W., Waiyasusri, K., Saengow, P., & Pinasu, D. (2022). Height Measurement and Oil Palm Yield Prediction Using Unmanned Aerial Vehicle (UAV) Data to Create Canopy Height Model (CHM). Geographia Technica, 17(2), 164-178. doi: 10.21163/GT_2022.172.14

Land Development Department. (2021). Appropriate agricultural promotion guidelines according to the proactive agricultural map database (Agri-Map).

Li, W., Dong, R., Fu, H., & Yu, L. (2019). Large-Scale Oil Palm Tree Detection from High-Resolution Satellite Images Using Two-Stage Convolutional Neural Networks. Remote Sensing, 11, 1-11. doi:10.3390/rs11010011

Liu, M., Hu, S., Ge, Y., & Heuvel, G. B. (2021). Using multiple linear regression and random forests to identify spatial poverty determinants in rural China. Spatial Statistics, 42, 100461. doi:10.1016/j.spasta.2020.100461

Manandhar, R., Odeh, I. O., & Ancev, T. (2019). Improving the Accuracy of Land Use and Land Cover Classification of Landsat Data Using Post-Classification Enhancement. Remote Sensing, 1(3), 330-344. doi:10.3390/rs1030330.

McFEETERS, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International Journal of Remote Sensing, 17(7), 1425-1432. doi:10.1080/01431169608948714

Miettinen, J., Shi, C., Tan, W. J., & Liew, S. C. (2012). 2010 land cover map of insular Southeast Asia in 250-m spatial resolution. Remote Sensing Letters, 3(1), 11-20. doi:10.1080/01431161.2010.526971

Mohan, M., Silva, C. A., Klauberg, C., Jat, P., Catts, G., Cardil , A., . . . Dia , M. (2017). Individual Tree Detection from Unmanned Aerial Vehicle (UAV) Derived Canopy Height Model in an Open Canopy Mixed Conifer Forest. Forests, 9(8), 340. doi:10.3390/f8090340

Morel, A. C., Fisher, J. B., & Malhi, Y. (2012). Evaluating the potential to monitor aboveground biomass in forest and oil palm in Sabah, Malaysia, for 2000–2008 with Landsat ETM+ and ALOS-PALSAR. International Journal of Remote Sensing, 33(11), 3614-3639. doi:10.1080/01431161.2011.631949

Mubin, N. A., Nadarajoo, E., Shafri, H. Z., & Hamedianfar, A. (2019). Young and mature oil palm tree detection and counting using convolutional neural network deep learning method. International Journal of Remote Sensing, 40(19), 7500-7515. doi:10.1080/01431161.2019.1569282

Nooni, I. K., Duker, A. A., Van Duren, I., Addae-Wireko, L., & Osei Jnr, E. M. (2014). Support vector machine to map oil palm in a heterogeneous environment. International Journal of Remote Sensing, 35(13), 4778-4794. doi:10.1080/01431161.2014.930201

Pal, M., & Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sensing of Environment, 86(4), 554-565. doi:10.1016/S0034-4257(03)00132-9

Piekutowska, M., Niedbała, G., Piskier, T., Lenartowicz , T., Pilarski, K., Wojciechowski , T., . . . Kosacka , A. C. (2021). The Application of Multiple Linear Regression and Artificial Neural Network Models for Yield Prediction of Very Early Potato Cultivars before Harvest. Agronomy, 5(11), 885. doi:10.3390/agronomy11050885

Poortinga, A., Tenneson, K., Shapiro, A., Nquyen, Q., Aung , K. S., Chishtie, F., & Saah, D. (2019). Mapping Plantations in Myanmar by Fusing Landsat-8, Sentinel-2 and Sentinel-1 Data along with Systematic Error Quantification. Remote Sensing, 11(7), 831. doi:10.3390/rs11070831

Rehman, A. U., Ullah, S., Liu, Q., & Khan, M. S. (2021). Comparing different space-borne sensors and methods for the retrieval of land surface temperature. Earth Science Informatics, 14, 985–995. doi:10.1007/s12145-021-00578-6

Santos, V. G., Cuxart , J., Villagrasa, D. M., Jiménez , M. A., & Simó, G. (2018). Comparison of Three Methods for Estimating Land Surface Temperature from Landsat 8-TIRS Sensor Data. Remote Sensing, 10(9), 1450. doi:10.3390/rs10091450

Shafri, H. Z., Hamdan, N., & Saripan, M. I. (2011). Semi-automatic detection and counting of oil palm trees from high spatial resolution airborne imagery. International Journal of Remote Sensing, 32(8), 2095-2115. doi:10.1080/01431161003662928

Shaharum, N. S., Shafri, H. Z., Ghani, W. W., Samsatli, S., Al-Habshi, M. M., & Yusuf, B. (2020). Oil palm mapping over Peninsular Malaysia using Google Earth Engine and machine learning algorithms. Remote Sensing Applications: Society and Environment, 17, 100287. doi:10.1016/j.rsase.2020.100287

Shen, Y., Mercatoris, B., Cao, Z., Kwan, P., Guo, L., Yao, H., & Cheng, Q. (2022). Improving Wheat Yield Prediction Accuracy Using LSTM-RF Framework Based on UAV Thermal Infrared and Multispectral Imagery. Agriculture, 6(12), 892. doi:10.3390/agriculture12060892

Sitthi, A., Nagai, M., Dailey, M., & Ninsawat, S. (2016). Exploring Land Use and Land Cover of Geotagged Social-Sensing Images Using Naive Bayes Classifier. Sustainability, 8(9), 921. doi:10.3390/su8090921

Srestasathiern , P., & Rakwatin, P. (2014). Oil Palm Tree Detection with High Resolution Multi-Spectral Satellite Imagery. Remote Sensing, 6, 9749-9774. doi:10.3390/rs6109749

Sumathi, S., Chai, S. P., & Mohamed, A. R. (2008). Utilization of oil palm as a source of renewable energy in Malaysia. Renewable and Sustainable Energy Reviews, 9(12), 2404-2421. doi:10.1016/j.rser.2007.06.006.

Thenkabail, P. S., Stucky, N., Griscom, B. W., Ashton, M. S., Diels, J., van der Meer, B., & Enclona, E. (2004). Biomass estimations and carbon stock calculations in the oil palm plantations of African derived savannas using IKONOS data. International Journal of Remote Sensing, 25(23), 5447-5472. doi:10.1080/01431160412331291279

Tsouros, D. C., Bibi, S., & Sarigiannidis, P. G. (2019). A Review on UAV-Based Applications for Precision Agriculture. Information (10), 349. doi:10.3390/info10110349

Vapnik, V., Golowich, S. E., & Smola, A. (1996). Support vector method for function approximation, regression estimation and signal processing. the 9th International Conference on Neural Information Processing Systems (pp. 281–287). Massachusetts: MIT Press. doi:10.5555/2998981.2999021

Wang, C., Morgan, G., & Hodson, M. E. (2021). sUAS for 3D Tree Surveying: Comparative Experiments on a Closed-Canopy Earthen Dam. Forest, 695. doi:10.3390/f12060659

Worachairungreung, M., Ninsawat, S., Witayangkurn, A., & Dailey, M. N. (2021). Identification of Road Traffic Injury Risk Prone Area Using Environmental Factors by Machine Learning Classification in Nonthaburi, Thailand. Sustainability, 13(7), 3907. doi:10.3390/su13073907

Worachairungreung, M., Thanakunwutthirot, K., & Ninsawat, S. (2021). A Study on Estimating Land Value Distribution for the Talingchan District, Bangkok Using Points-of-Interest Data and Machine Learning Classification. Applied Sciences, 11(22), 11029. doi:10.3390/app112211029

Xu, K., Qian, J., Hu, Z., Duan, Z., Chen, C., Liu, J., . . . Xing, X. (2021). A New Machine Learning Approach in Detecting the Oil Palm Plantations Using Remote Sensing Data. Remote Sensing, 13(2), 236. doi:10.3390/rs13020236

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. International Journal of Remote Sensing, 27(14), 3025-3033. doi:10.1080/01431160600589179

Yin, Q., Liu, M., Cheng, J., Ke, Y., & Chen, X. (2019). Mapping Paddy Rice Planting Area in Northeastern China Using Spatiotemporal Data Fusion and Phenology-Based Method. Remote Sensing, 11(14), 1699. doi:10.3390/rs11141699

Zhou, J., Li, J., Zhang, L., Hu, D., & Zhan, W. (2012). Intercomparison of methods for estimating land surface temperature from a Landsat-5 TM image in an arid region with low water vapour in the atmosphere. International Journal of Remote Sensing, 2582-2602. doi:10.1080/01431161.2011.617396

Zoungrana, B. J.-B., Conrad, C., Amekudzi, L. K., Thiel, M., Da, E. D., Forkuor, G., & Löw, F. (2015). Multi-Temporal Landsat Images and Ancillary Data for Land Use/Cover Change (LULCC) Detection in the Southwest of Burkina Faso, West Africa. Remote Sensing, 7(9), 12076-12102. doi:10.3390/rs70912076.

MICROZONATION FOR EARTHQUAKE HAZARDS
WITH HVSR MICROTREMOR METHOD IN THE COASTAL AREAS
OF SEMARANG, INDONESIA

Muhammad Irham NURWIDYANTO [1] , *Muhammad ZAINURI [2]* ,
Anindya WIRASATRIYA [2] , *Gatot YULIANTO [3]*

ABSTRACT:
Semarang coastal area is mainly composed of alluvium sediment, vulnerable to tidal floods caused by land subsidence and seawater inundation. On the other hand, this capital area is also prone to earthquakes evoked by Kaligarang, Semarang, Ungaran 1, and Ungaran 2 Faults. This research aims to examine the earthquake vulnerability index in the coastal area of Semarang using the HVSR (horizontal to vertical spectral ratio) micro tremor method. The study was conducted by measuring the micro tremor signal as many as 110 points spread over the research location with an almost even distribution. Data collection was carried out around existing roads with a distance of about 1000 m west-east and about 500 m north-south. Data was taken with a 3-component TDS seismograph type 303S with a sampling rate of 20 Hz for 10 minutes. The micro tremor data was processed using Software Excel, Pro, and Geopsy data to obtain the dominant frequency values, amplification factor, seismic vulnerability index, Peak Ground Acceleration, and Ground stress-strain. The data obtained is described as a distribution map along with the parameter values using ArcGIS software. The results show that the frequency dominant, amplification, SVI, PGA and GSS vary from 0.13 Hz to 8.44 Hz, from 0.08 to 7.92, from 0.04 gal to 75.74 gal, from 5.75 μs^2/cm to 45.22 μs^2/cm, and from 12.3×10^{-6} – 788×10^{-6}, respectively. The results obtained in the research area that have a vulnerability to earthquakes can be grouped into 3 categories, namely weak, medium and strong categories. Most of the research areas have a vulnerability index to moderate earthquakes, except for the northern and southern parts of West Semarang District, Sayung District, and Genuk District which is in the strong category. Areas with a weak category were found around the eastern part of North Semarang District and Central Semarang District in the northern part, Candisari District, Gayamsari District in the southern part, and Pedurungan District in the southern part. This study was conducted during the rainy season. Therefore, collecting the same data for different season I recommended for further study to investigate the influence of seasonal effects on earthquake-impacted areas.

Key-words: Micro zonation, Earthquake, SVI, Semarang Coastal, HVSR

1. INTRODUCTION

Geographical Location Semarang City is located on the north coast of Central Java, precisely at 6º 50″ – 7º 10″ south latitude and 109º 50″ – 110º 35″ east longitude, with an area of about 373.7 km². Semarang is the capital of the province of Central Java and is a fairly densely populated city, some of which are located on the coast. Semarang is bordered by the Java Sea in the north, Demak Regency in the east, Semarang Regency in the south, and Kendal Regency in the west (BPS Semarang, 2022). The coastal area in the northern part of Semarang City is a coastal alluvial plain that extends from east to west and the topography ranges from 1 to 5 m high above sea level.

_____

[1] *Marine Science Department and Physic Department, Diponegoro University, Semarang Indonesia,* *irhammn@gmail.com*

[2] *Oceanography Department, Faculty of Fisheries and Oceanography, Diponegoro University, Semarang Indonesia,* *Muhammad.zainuri@yahoo.co.id*, *anindyawirasatriya@lecturer.undip.ac.id*

[3] *Physic Department, Faculty of Science and Mathematics, Diponegoro University, Semarang Indonesia,* *gatoty@gmail.com*

The alluvial plain is influenced by coastal and river deposits. In the southern part of Semarang City, there are many hills where volcanic breccia spreads from north to south. These rocks are the result of the eruption of Mount Ungaran, the highest area of Semarang city. The hills of the city of Semarang have a slope of 2% to 40 % and a height of 90 m to 200 m above sea level (Marsudi, 2000). Several active faults have been identified in the Semarang area, such as the Semarang, Kaligarang, Ungaran 1, and Ungaran 2 (Pusgen, 2017). Furthermore, another study defined that the northern and southern of Semarang are separated by the Semarang fault (Hidayat et al., 2014). Several active faults have been identified quite well in the Semarang area, such as the Semarang Fault, Kaligarang Fault Ungaran 1 and Ungaran 2 Faults (Pusgen, 2017). Several studies explain that the fault that divides hight Semarang and low Semarang is known as the Semarang Fault (Hidayat, et al., 2014). The Semarang fault can be recognized from the morphological appearance of the rising fault escarpment that penetrates the Holocene rock (Poedjoprajitno, et al., 2008). The geological structure of Semarang City consists of three parts, namely the joint structure, faults and folds. The fault area is very erosive, has high porosity, and the structure of rock formations is discontinuous and heterogeneous, making it easier for soil migration or landslides.

Geologically, the coastal area of Semarang, Indonesia mainly consist of the alluvium (Qa) deposits which are are quarter old, and only a small part is composed by Damar Formation rocks at the southern part. Alluvium deposits (Qa) are composed of alluvium deposits of beaches, rivers and lakes as shown in **figure 1**. The lithological deposits on the coast consist of clay, silt, sand and a mixture with a thickness of 50 m or more. River and lake deposits consist of gravel, gravel, sand and silt 1 to 3 m thick. The boulders are composed of andesite, claystone and a little sandstone. The Damar Formation of Late Pliocene to Early Pleistocene age is composed of tuffaceous sandstone, conglomerate, volcanic breccia and tuff. Sandstone consists of feldspar minerals and mafic minerals, some tuffaceous and locally there is limestone. As for breccias, the fragments are generally alkaline volcanic rocks and outcrops found in Kedung Mundu, Karanganyar, and Ngadirejo (Poedjoprajitno, et al., 2008). Semarang has active seismic activity, so the Semarang area is likely to be affected by earthquake vibrations, both local and far away.
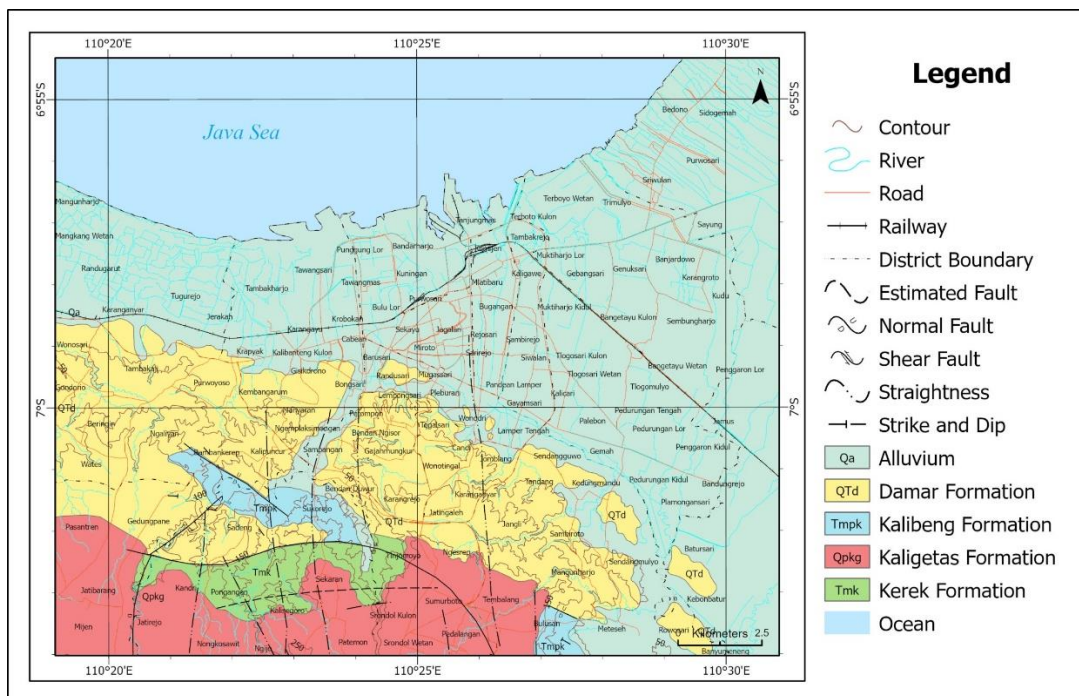

Fig. 1. Geological Map of Semarang City *(Thanden, et al., 1996 in Poejoprayitno, et al., 2008).*

Based on earthquake catalog data from the *Pusat Vulkanologi dan Mitigasi Bencana Geologi* (PVMBG) (Supartoyo, et al., 2014), an earthquake measuring VI to VII MMI occurred in Semarang on January 19, 1856, which caused damage to buildings. The earthquake may be associated with a fault in the Semarang area and is suspected to be active or potentially active in the future.

The HVSR (Horizontal to Vertical Spectral Ratio) is a method commonly used for three-component micro tremor to determine bedrock depth. The parameters used in this method are the amplification factor and natural frequency. Both of these parameters are related to the subsurface physical parameters to identify the geological characteristics of the research area. The HVSR method uses the principle of calculating the spectral ratio resulting from the sum of the horizontal and vertical components (Nakamura, 1989). The value of the amplification factor of a place can be known from the height of the spectral peak of the HVSR curve as a result of micro tremor measurements at that place. The dominant period value or dominant frequency obtained from the HVSR curve has a correlation with the thickness level of the sediment layer (Nakamura, 2008).

Semarang City is an area that is included in the category of a moderate disaster-prone category with a disaster risk index of 183,6 (high level) in 2015 and 120,75 (moderate level) in 2018 (IRBI, 2018). The previous studies shown that the coastal area of Semarang experienced many disasters such as sea water intrusion (Suhartono, et al., 2013, Setyawan, et al., 2016, Supriyadi, et al., 2013, 2016, Wijatno, et al., 2019), tidal flood (Kuriawan, 2003, Wahyudi, 2007, Bakti, 2010, Ramadhani, et al., 2012, Handoyo, et al.,, 2016) and land subsidence (Marsudi, 2000, Suhelmi, 2012, Supriyadi, et al., 2009, Abidin, et al., 2010, 2012, Widada, et al., 2020). However, the vulnerability to the earthquake disaster has not been discussed in the previous study. Thus, it is necessary to study the mitigation of several disasters in the city of Semarang, including earthquakes, and efforts in the context of disaster management. According to Law Number 24 of 2007 concerning Disaster Management, several actions that can be taken in disaster management include prevention, mitigation, preparedness, and response to emergencies. One of the most important steps is disaster mitigation. In the context of disaster risk reduction, on this occasion, we will discuss the micro zonation of earthquake vulnerability in the coastal area of Semarang. It is important to create a zonation for earthquake vulnerability for mitigation because the Semarang Coastal area is the center of activity for government offices, private offices, hotels, industries, services and also housing.

2. RESEARCH METHODS

2.1. Location of data collection

The type of research carried out by acquisition data in the form of measuring micro tremor signals in the form of transient seismic signals with time domain. Data collection was carried out in the Semarang Coastal area, in UTM coordinates, with boundaries Easting 432000-443000 and Northing 9226000-9233000, as shown in **figure 2**.

Data collection was carried out with a semi-grid distribution with a west-east distance of about 500 m and a north-south distance of approx. 1000 m, with the measurement point located around the existing road. Measurement of the micro tremor signal with a 3-component digital seismograph TDS type 303S and recording with a midi data logger type GL 240. Installation of the north direction of the seismograph with a geological compass, data collection for each point for 10 minutes with a sampling rate of 20 Hz. In this study, 110 data points were successfully taken and carried out in February 2020 (Irham, et al., 2021a, b).

The obtained micro tremor data is then converted into .txt format with notepad ++, then the data can be processed with geopsy software to obtain amplification (A_0) and pre dominant frequency (f_0) values. From the values of frequency dominant (f_0), amplification (A_0), Seismic Vulnerability Index (SVI), Peak Ground Acceleration (PGA) and Ground Stress Strain (GSS) can be calculated using the equations described above using Microsoft Excel software. The A_0, f_0, SVI, PGA and GSS data were then plotted into a base map using the ArcGIS software to describe the distribution of these values (Irham, et al., 2021b). The equation for calculating (f_0) and (A_0), the values of SVI, PGA and GSS is described in the next sub-section.
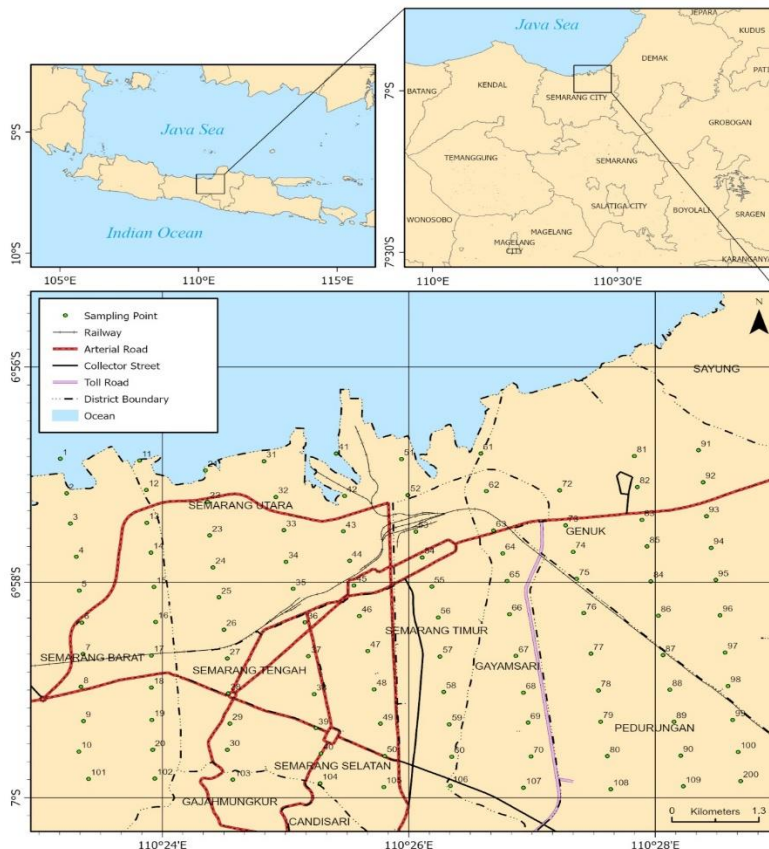
Fig. 2. Location of the research area.

2.2. Dominant frequency and amplification

The dominant frequency is the frequency value that is often recorded when measuring micro tremor signals. The dominant frequency can be considered as the frequency value of the rock layer which can indicate the type and nature of the rock at that point. The dominant frequency value of an area is influenced by the thickness of the sediment layer above the bedrock and the average velocity of the shear wave in that layer. The magnitude of the dominant frequency can be formulated in the following equation (1), Nakamura (1997):

$$f_0 = \frac{v_s}{4h} \tag{1}$$

where v_s is the shear wave velocity of the sediment layer and h is the bedrock depth.

Lachet and Brad (1994) used six simple tectonic models to perform combined simulated tests of contrast changes in shear wave velocity and soil layer thickness. The simulation results show that the peak frequency values change along with changes in geological conditions. From f_o, we can classify the soil types based on the criterion of Kanai (1983) as described in **table 1**. Nakamura (2000) states that the value of the strengthening factor or soil amplification is related to the impedance contrast ratio of the surface layer to the layer below it. If the impedance contrast ratio of the two layers is high, the gain factor value is also high, and vice versa. Arifin (2014) states that the amplification value is directly proportional to the value of the horizontal and vertical spectrum ratio (H/V). The amplification value can increase if the rock has undergone deformation (weathering, folding, or faulting) that changes the physical properties of the rock. In the same rock, the amplification value can vary according to the degree of deformation and weathering of the rock body. Amplification can be formulated with the equation (Arifin, et al., 2014):

$$A_o = \frac{\rho_b v_b}{\rho_S v_s} \tag{2}$$

where b is the density of bedrock (g/cc), v_b is the velocity of wave propagation in the bedrock (m/s), s is the density of soft rock (g/cc) and v_s is the velocity of wave propagation in sedimentary rock (m/s). dt).

Table 1.

Soil Classification based on Dominant Frequency Values (Kanai, 1983).

Soil Clasification Type	Natural Frequency (Hz)	Kanai's Clasification
Type I	6.67 – 20	Tertiary rocks or older. Consist of hard sandy rocks, gravel, etc.
Type II	4 – 6.67	Alluvial Rock, with thickness of 5m, consist of sandy-gravel, sandy hard clay, loam, etc.
Type III	2.5 – 4	Alluvial Rock, with thickness >5m, consist of sandy-gravel, sandy hard clay, loam, etc.
Type IV	< 2.5	Alluvial Rock, originated from top soil delta sedimentation, mud, etc. With Thickness of 30 or more

Amplification is a magnification of seismic waves that occurs due to significant differences in velocity and density between layers. Seismic waves will experience amplification if they propagate from one medium to another medium that is softer than the previous medium. The greater the difference, the greater the magnification experienced by the wave. The amplification values can be grouped as in **table 2** (BMKG in Setiawan, 2009).

Table 2.

Classification of amplification factors (BMKG in Setiawan, 2009).

Zone	Classificatin	Amplification factor value
1	low	A<3
2	medium	3≤A<6
3	high	6≤A<9
4	Very high	A≥9

2.3. Peak Ground Accelleration

Peak Ground Acceleration (PGA) is the greatest value of ground acceleration at a place caused by earthquake vibrations in a certain period of time. The value of ground vibration acceleration that will be taken into account as one part in planning an earthquake-resistant building is the maximum ground acceleration value (Ehsani, 2015). The geological conditions of the soil and the level of sediment compaction in an area greatly determine the size of the PGA value. The more compacted the sediment, the smaller the PGA value in the area.

The PGA value in an area can be measured directly with a PGA measuring device but can also be approximated empirically. There are many empirical approach formulas to calculate the PGA value, but in this study the formula from Kanai (1966) in Douglas (2022) is used, because this formula takes into account the strength of the earthquake in the SR and the distance from the epicenter and the dominant period (T) in the research area being studied.

$$PGA = \frac{5}{\sqrt{To}} 10^{0,6M - p \log R + (0,167 - \frac{1,88}{R})} \tag{3}$$

where PGA is the maximum ground acceleration (cm/s^2), To is the dominant period (s), M is the earthquake magnitude on the Richter Scale (SR), R is the distance from the hypocenter to the measurement point (km), and p is (1.66 + 3.6R). In this study, we used parameters Yogyakarta earthquake data on 27 May 2006.

2.4. Seismic Vulnerability Index

Seismic Vulnerability Index (SVI) can be interpreted as a parameter that can be used to determine the level of vulnerability of an area to the threat of earthquake risk. The seismic susceptibility index with earthquake risk level against earthquake damage shows a linear relationship

(Nakamura, 2008). The value of the seismic vulnerability index at each measuring point is obtained by squaring the amplification value (A_0) and then divided by the dominant frequency value (f_0) obtained in the HVSR spectrum as shown in equation (4) (Nakamura, 1997).

$$SVI = \frac{A_0^2}{f_0} x 10^{-6} \tag{4}$$

where SVI is the Seismic Vulnerability Index (s^2/cm), A_0 is the Amplification and f_0 is the dominant frequency (Hz).

2.5. Ground Stress Strain

GSS (Ground Stress Strain) is the ability of the soil layer material to stretch and shift during the Nakamura earthquake (2000). Areas that have a high GSS value have a high risk of ground motion due to earthquakes, such as land subsidence, ground shaking, and ground stretching. The GSS value and its effect on dynamic soil properties can be summarized in **table 3**. In calculating the GSS of the surface soil layer in an area when an earthquake occurs, it can be done by multiplying the seismic vulnerability index with the maximum soil acceleration which is formulated as in equation (5), (Nakamura, 1997, 2008).

$$GSS = SVIxPGA \tag{5}$$

where SVI is Seismic Vulnerability Index (s^2/cm) and PGA is Peak Ground Acceleration (cm/s^2).

Table 3.

Strain Dependence of Dynamic Properties of Soil (Nakamura, 1996).

Size of strain *(GSS)*	10^{-6} 10^{-5}	10^{-4} 10^{-3}	10^{-2} 10^{-1}
Phenomena	Wave, vibration	Crack, Diff, Settlement	Landslide, Soil compaction, Liquefaction
Dynamic proterties	Elasticity	Elasto-Plasticity	Repeat–Efect, Speed-Effect of loading

3. RESULT AND DISCUSSION

3.1. Dominant frequency and amplification

The dominant frequency of the study area as shown in **figure 3** varies from 0.13 Hz - 7.47 Hz, it can be grouped into 3 categories, namely type IV (0 Hz - 2.5 Hz), type III (2.5 Hz - 4 Hz), and type II (4 Hz-10Hz). In general, almost 90% of the dominant frequencies in the study area are type IV frequencies (low type) which indicates that the sediment in the study area has a thickness of more than 30 m. Some areas have a dominant frequency of type III which are around Candisari and Milo, with a thickness of 10 m - 20 m, while areas that have a dominant frequency of type II are around Candisari and Milo with a thickness of 5 m - 10 m. The amplification map of the research area is shown in **figure 4**. The amplification values ranged from 0.13-5.96. According to Setiawan (2009) (**Table 2**), the amplification values can be grouped into 2, namely low amplification (A < 3) (in **figure 4** shown in blue) which is mostly located in the middle of the study area, and moderate amplification (3<A<6) (in **figure 4** is shown in white) which is located in West Semarang District, North Semarang District, Genuk District, and Sayung District and locally around South Semarang District, Gajahmungkur District, and Mranggen District. The amplification value shows the contrast between the sedimentary layer and the bedrock layer, the higher the contrast, the greater the amplification. The amplification value is influenced by the contrast value of wave velocity and rock density. A low wave speed will have a high amplification which tends to have a low density so that in the event of an earthquake the risk is higher than in an area with small amplification. Small amplification values are usually associated with a small contrast between sediment and bedrock, meaning that sedimentary rocks are denser and seismic waves will propagate faster so they have less risk of earthquakes.
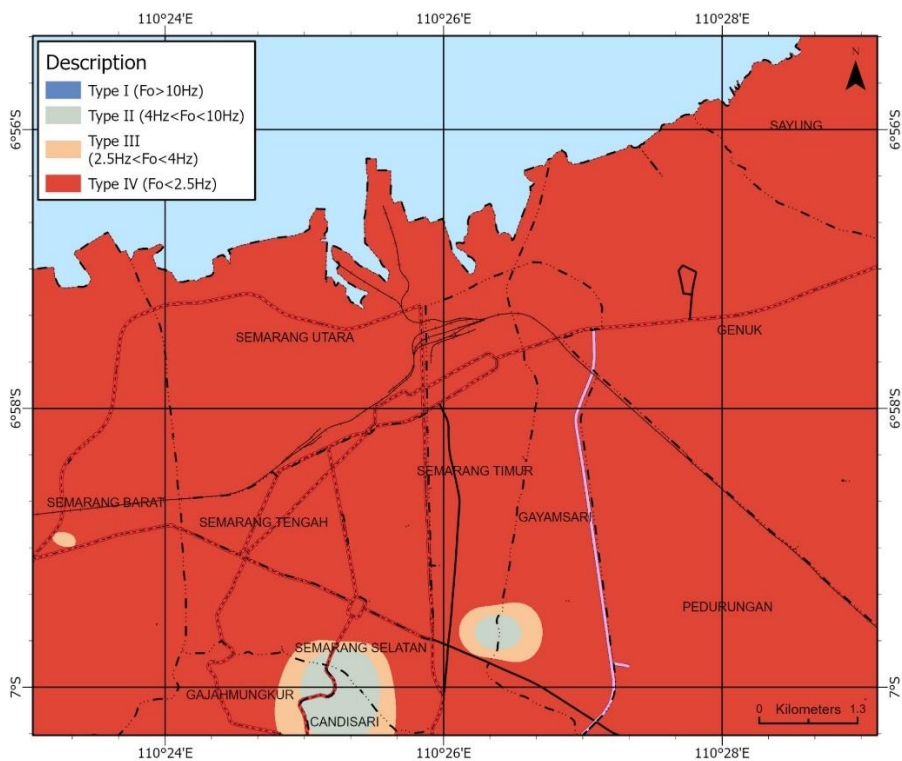
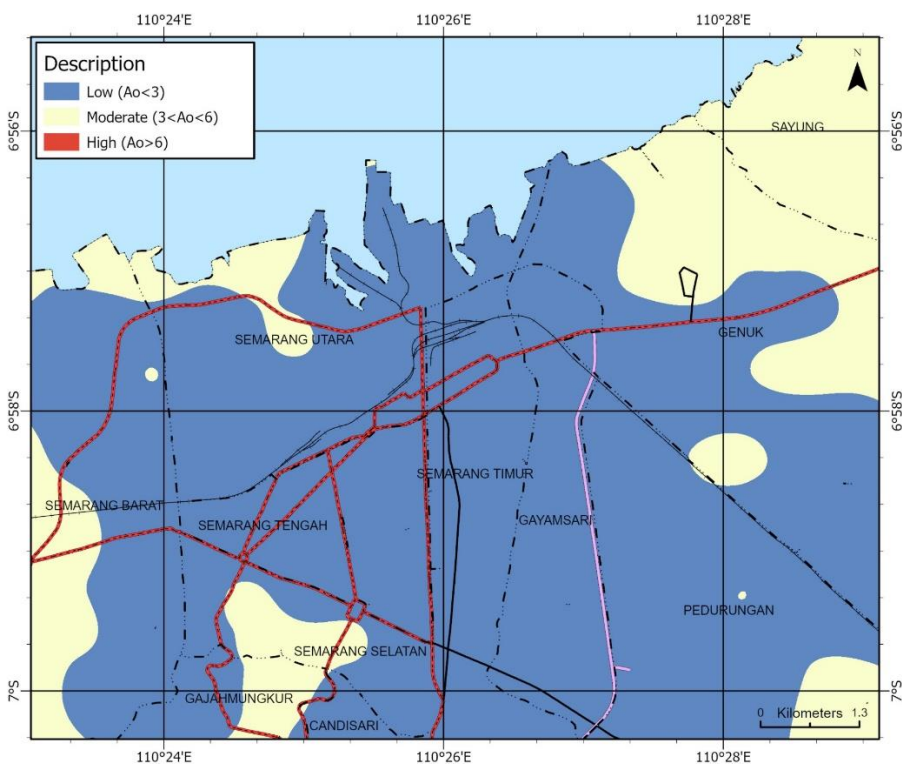**Fig. 3.** Dominant Frequency Map of Semarang Coastal Area.

Fig. 4. Amplification map of Semarang Coastal Area.

3.2. Peak Ground Acceleration (PGA)

The PGA (Peak Ground Acceleration) value is calculated using the Kanai equation (1966 in Douglas 2022) as shown in equation (3) with input parameters: dominant period (T), earthquake strength (M), and epicenter distance (R). The distribution of PGA values in the research area with the source of the Yogyakarta earthquake on 27 May 2006 with a strength (M) of 5.9 SR and an epicenter distance of about 105 km - 120 km is shown in **figure 5**, the PGA value varies between 5.53 gal - 44.15 gal which corresponds to with grades II-VI on the MMI scale.
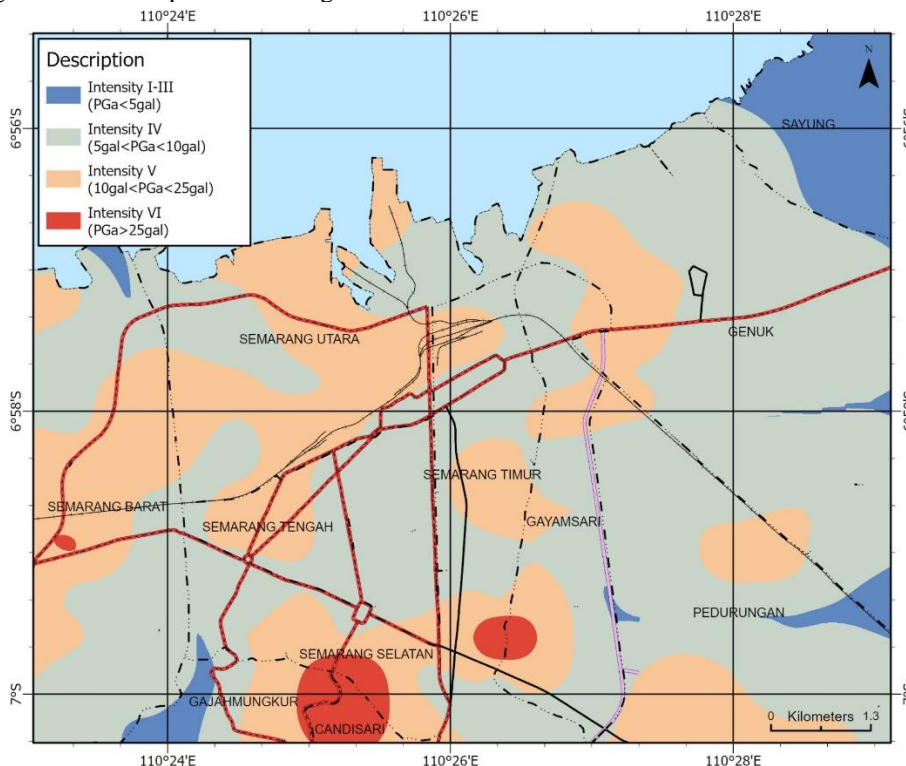
Fig. 5. Peak Ground Acceleration Map of Semarang Coastal Area.

PGA values were grouped into 4 levels with very weak PGA (PGA < 5gal), weak PGA (5<PGA<10 gal), moderate PGA (10<PGA<25 gal), and moderately strong PGA (PGA>25 gal).

Areas with very weak PGA are in the west of Gajahmungkur, to the east of Pedurungan and in Sayung, areas with weak PGA are in the eastern and northern parts of West Semarang, western and eastern North Semarang, Gayamsari District, Pedurungan District, Mranggen District, and Genuk District. During the Yogyakarta earthquake on 27 May 2006, the buildings that are built on the less dense soil structures (i.e.,large PGA values) were areas that suffered severe damage in Bantul (Walter et al., 2008).

3.3. Seismic Vulnerability Index (SVI)

Seismic Vulnerability index (SVI) is calculated based on the Nakamura (2008) approach as shown in equation (4). The SVI value is shown in **figure 6** which varies between 0.41 cm^2/s – 140.27 cm^2/s. Based on Daryono (2014) in this study, SVI values were grouped into 3 parts, namely low SVI (SVI < 5), moderate SVI (5 <SVI < 25) and high SVI (SVI > 25). The higher SVI value indicates the lower level of stability of the sediment structure below the surface, so that if an earthquake occurs, there is a greater risk of damage. Areas with low SVI values are located locally in the central part of the research area, areas with moderate SVI values are almost dominant in the research area covering the central part of West Semarang Sub-District, South Semarang District, North Semarang district,

Candisari district, East Semarang district, Gayamsari District , Pedurungan District and Mranggen District. Meanwhile, areas with high SVI values are located in the northern and southern parts of West Semarang Sub-District, Gajah Mungkur District, South Semarang District, Sayung District and Genuk District. Concerning the case of the Yogyakarta earthquakes on 26 May 2006 in Bantul, the area with more than 20 SVI was the most impacted area with the highest number of casualties, the material collapses, and fatalities (Daryono et al., 2009). Therefore, stakeholders and third parties should monitor the area with 25 SVI in the present study (red color in **figure 6**). Massive socialization regarding earthquake vulnerability must be implemented in local society to mitigate the possible future impacts.
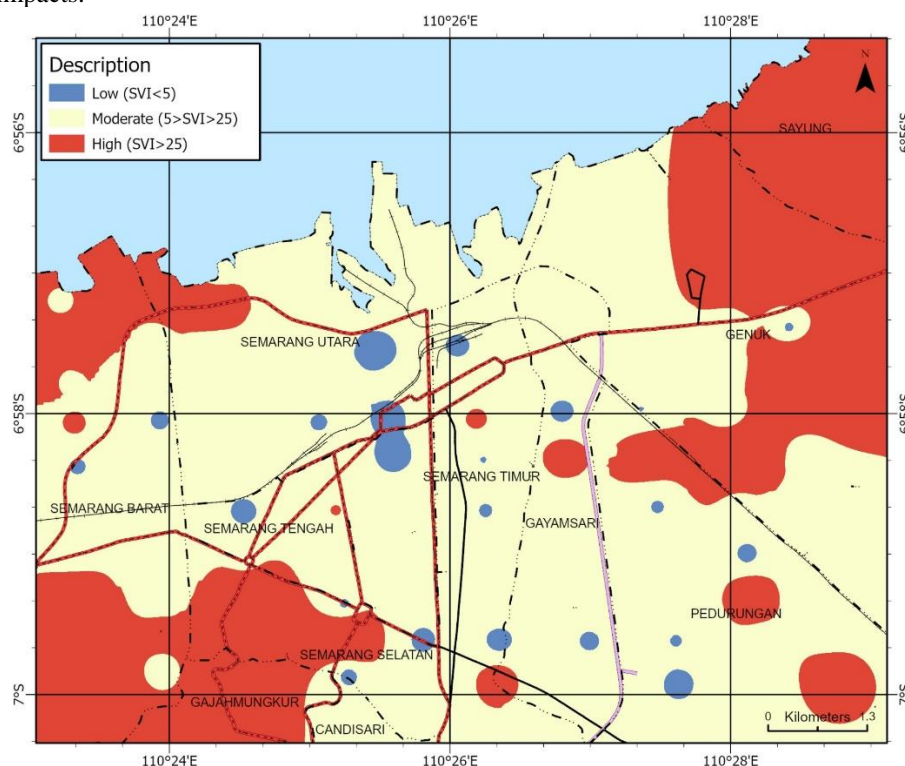

Fig. 6. Seismic Vulnerability index Map of Semarang Coastal Area.

3.4. Ground Stress-Starain (GSS)

Microzonation vulnerabilty to earthquake can be represent by ground shear strain (GSS). GSS is the ability of a material to stretch or shift during an earthquake. Ground shear strain has a relationship with the condition of the soil surface layer. The greater the ground shear strain value, the soil surface layer will be easily deformed, while the small ground shear strain value indicates the soil layer will be more difficult to deform. The distribution of GSS values in the study area is shown in **figure 7** with values varying between $12.3x10^{-6} – 788x10^{-6}$. The GSS value in the study was grouped into three groups, namely low (GSS $< 10^{-4}$), medium GSS value ($10^{-4} <$ GSS $< 10^{-3}$) and high GSS value ($10^{-3} <$ GSS $< 10^{-2}$). Based on the classification of Nakamura (1998), the occurrence of earthquake in Semarang City will cause crack, diff, settlement and soil compaction. Thus, earthquake may worsen the subsidence rate in the coastal area of Semarang.

Zones with low susceptibility to earthquakes are in the central part of the study area shown in blue (**figure 7**), namely in the central part of West Semarang District, the southern part of North Semarang District, the northern part of South Semarang District, the northern part of Gajah Mungkur District, Candisari District, the northern part of East Semarang District, the northern part of Gayamsari District and the western part of Pedurungan District.

Zones with moderate vulnerability are in the western, northern and eastern parts of the study area which are shown in brownish white (**figure 7**). Zones with moderate vulnerability are in West Semarang District, northern North Semarang District, Gajahmungkur District, South Semarang District, northern East Semarang District, northern Gayamsari District, Genuk District, Pedurungan District and Sayung District. There are almost no areas with high vulnerability in the study area.
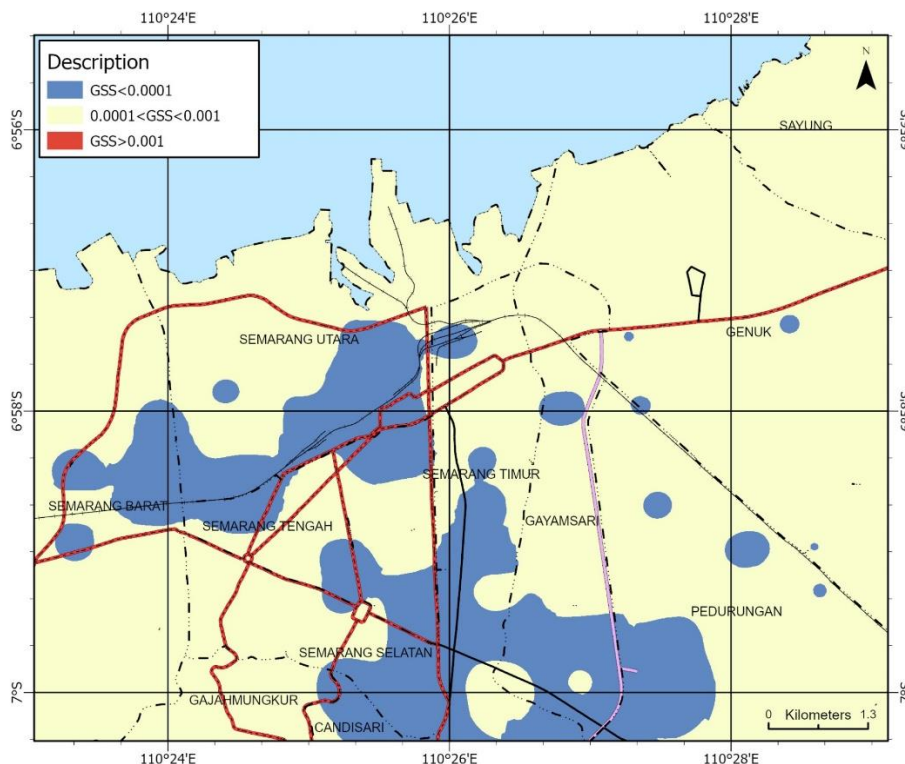

Fig. 7. Ground Stress-Strain Map of Semarang Coastal Area.

4. CONCLUSIONS

The vulnerability level of the Semarang coastal area to earthquakes is classified into weak, moderate, and strong categories. However, based on the ground shear strain (GSS), most of the study area is dominated by the moderate vulnerability. In contrast, several parts of the north to south of Semarang City are weakly vulnerable to earthquakes i.e., central part of West Semarang District, the southern part of North Semarang District, the northern part of South Semarang District, the northern part of Gajah Mungkur District, Candisari District, the northern part of East Semarang District, the northern part of Gayamsari District and the western part of Pedurungan District. Thus, massive socialization is very important to be conducted especially for the people living in the moderate to strong categories of earthquake vulnerability to mitigate the future eartquake impact. It is very important to be noted that, this study was conducted during the dry season. It is very important to collect data during the rainy season to investigate the possibility of its seasonal effect. This task is left for future study.

ACKNOWLEDGEMENT

R E F E R E N C E S

[BPS] Central Bureau of Statistics for Semarang City, 2022. Semarang City in Number, Semarang.

Abidin, H.Z., Andreas, H., Gumilar, I., Sidiq, T.P., Gamal, M., Murdohardono, D., Supriyadi, & Fukuda, Y., 2010. Studying Land Subsidence in Semarang (Indonesia) Using Geodetic Methods. Procceding FIG Congress. Sydney, Australia, 1 – 15.

Abidin, H.Z., Andreas, H., Gumilar, I., Sidiq, T.P., & Fukuda, Y. 2013. Land subsidence in coastal city of Semarang (Indonesia): characteristics, impacts and causes. Geomatics, Natural Hazards and Risk 3, 4(3), 226 – 240.

Arifin, S.S., Mulyatno, B. S., Mujiyono, and Setianegara, R. 2014. Determination of Earthquake-Prone Zones Based on Analysis of the Amplification Value of Hvsrmicrotremor and Analysis of the Dominant Period in Liwa and Surrounding Areas, *Jurnal Geofisika Eksplorasi* Vol. 2 No. 1 .

Bakti, LM, 2010, Study of Tidal Inundation Potential at Semarang City and Its Solutions, Thesis Magister Sipil Engineering Diponegoro University.

IRBI, 2018, Indonesian Disaster Risk Index, Directorate of Disaster Risk Reduction, National Disaster Management Agency, Jakarta.

Douglas, J. 2022. Ground motion estimation equations 1964–2021. Imperial Colleage, London.

Daryono, Sutikno, Sartohadi, J., Dulbahri, Brotopuspito, K.S, 2009, Local assement of effect sites in bantul graben uses a seismic vulnerability index based on microtremor measurements. Journal of Indonesian Disaster Vol. 2 number 1, Mei 2009 ISSN 1978-3450.

Ehsani, N., M.R. Ghaemghamian, M. Faslavi, and E. Haghshenas. 2015. Estimation of Subsurface Structure Using Microtremor in Karaj City Iran. 10th Asian Regional Conference of IAEG.

Handoyo, G, Suryoputro, AAD, Subardjo, P, 2016, Inundation of Rob Floods in North Semarang District, Tropical Marine Journal, *Maret 2016 Vol. 19(1):55–59, ISSN 0853-7291*.

Hidayat, E, 2013, Identification of Active Faults Along the Garang River Path, Semarang, JSD.Geol. Vol. 23 No. 1 Maret 2013

Irham, MN, Zainuri, M, Yuliyanto, and G, Wirasatria, A, 2021a, Measurement of ground response of Semarang coastal region risk of earthquakes by Horizontal To Vertical Spectral Ratio (HVSR) microtremor method, ISNPINSA 2020, *Journal of Physics: Conference Series* 1943 (2021) 012033 IOP Publishing doi:10.1088/1742-6596/1943/1/012033.

Irham MN, T Yulianto, G Yuliyanto, S Widada, 2021b, Estimation of landslide prone zones arround the Semarang Fault based on by combining the slope stability analysis and HVSR microtremor methods. ISNPINSA 2020, *Journal of Physics: Conference Series* 1943 (2021) 012033 IOP Publishing, doi:10.1088/1742-6596/1943/1/012032.

Kanai, K., 1983, *Engineering Seismology*. Tokyo University: Japan.

Kurniawan, L., 2003, Study of Flood Rob in Semarang City (Dadapsari Case), ALAMI Vol 8 no 2.

Lachet, C. and Bard, P.Y., 1994, Numerical and Theoretical Investigations on the Possibilities and Limitations of Nakamura's Technique, *J. Phys. Earth*, 42, 377-397, 1994

Marsudi, 2000, Prediction of Land Subsidence Rates in the Semarang-Central Java Alluvial Plain, Doctoral Dissertation, Bandung Institute of Technology, Bandung Indonesia.

Nakamura, Y. 1997. Seismic Vulnerability Indices for Ground and Structures using Microtremor. World Congress on Railway Research: Florence.

Nakamura, Y. 2008, On The H/V Spectrum. *The 14th World Conference on Earthquake Engineering*. Beijing, China.

Nakamura, Y., 1989, A Method for Dynamic Characteristics Estimation of Subsurface Using Microtremor on The Ground Surface, *Quarterly Reports*, Railway Technical Research Institute, 30(1):25-33.

Nakamura, Y., 2000, Clear Identification of Fundamental Idea of Nakamura's Technique and Its Applications, *The 12th World Conference on Earthquake Engineering*, Auckland, New Zealand.

Poedjoprajitno, S., Wahyudiono, J., and Cita, A., 2008, Kaligarang Fault Reactivity, Semarang, Indonesian Journal of Geology, Vol. 3 No. 3 September 2008: 129-138.

Pusat Studi Gempa Nasional (PUSGEN), 2017, Indonesia Earthquake Source and Hazard Map 2017, Center for Research and Development of Housing and Settlements, Bandung.

Ramadhany, AS, Suryaputra, AAD, Subardjo, P, 2012, Robust Inundation Prone Areas in the Semarang Region, Journal of Marine Research Volume 1, Nomor 2, Tahun 2012, Halaman 174-180 .

Setiawan J.R. 2009. *Seismic Microzonation of Yogyakarta and Surrounding Areas. Thesis,* Bandung Institut Technology.

Setyawan, A., Najib, Aribowo, Y, Trihandini, A., Hastuti, D, Ramadhani, F, Waskito, F, Febrika, G Virgiawan, G, 2016. Sea water intrusion in Kaligawe Semarang Based on Resistivity data, 2nd International Conference on Tropical and Coastal Region Eco development, IOP Conf. Seies: Earth and Environmental Science 55 (2017) doi : 10.1088/1755-1315/55/1/012053.

Suhartono, E, Purwanto, P , S. Suripin, S, 2015, Seawater Intrusion Modeling On Groundwater Confined Aquifer In Semarang, International Conference on Tropical and Coastal Region Eco-Development 2014(ICTCRED 2014), Procedia Environmental Sciences 23 (2015) 110 – 115.

Suhelmi, 2012, Impact Of Land Subsidence On Inundated Area Extensivication At Semarang City, Jurnal Ilmiah Geomatika Vol. 18, No. 1.

Supartoyo, Surono, and Putranto, E.T., 2014, *Catalog of Destructive Earthquakes in Indonesia 1612 – 2014. Center for Volcanology and Geological Hazard Mitigation, Bandung*.

Supriyadi , Khumaedi, dan Yusuf, M, 2013, Penelitian Intrusi Air Laut Di Kawasan Semarang Utara Dengan Metode Gaya Berat Mikro Antar Waktu, Jurnal Fisika Vol. 3 No. 1.

Supriyadi, Khumaedi, Yusuf , M and Agung, W., 2016, Using a Time Lapse Microgravity Model for Mapping Seawater Intrusion around Semarang, The 4th International Conference on Theoretical and Applied Physics (ICTAP) 2014 , AIP Conf. Proc. 1719, 030041-1–030041-8; doi: 10.1063/1.4943736 © AIP Publishing LLC 978-0-7354-1366-5/$30.00

Walter TR et al. (2008). "The 26 May 2006 magnitude 6.4 Yogyakarta earthquake south of Mt. Merapi volcano: Did lahar deposits amplify ground shaking and thus lead to the disaster?". *Geochemistry, Geophysics and Geosystems*, 9(5).

Widada, S., Sugianto, D. N., Zainuri, M., Yulianto, G. dan Yulianto T., 2019, Identification Ground Layer Structure Of Land Subsidence Sensitive Area In Semarang City With Horizontal To Vertical Spectral Ratio Method, IOP Conf. Series: Earth and Environmental Science 246, 012023.

Widada, S., Zainuri, M., Yulianto, G., Satriadi, A, Wijaya, Y.J., 2020 Estimation of Land Subsidence Using Sentinel Image Analysis and Its Relation to Subsurface Lithology Based on Resistivity Data in the Coastal Area of Semarang City, Indonesia, Journal of Ecological Engineering, Volume 21, Issue 8, pages 47–56.

Wijatno, A.B., Kayyis, M, Satriyo, Pujindayati, E.P, 2019, Study of Seawater Intrusion in Deep Aquifers of Semarang Coast Using Natural Isotopes and Hydrochemicals, Indonesian Journal on Geoscience Vol. 6 No 1 page 17-28.

DELINEATION OF COAL FIRE RISK AREAS FROM LANDSAT-8 TIRS DATA: A CASE STUDY OF NA DUONG COALFIELD (NORTH-EAST OF VIETNAM)

Danh-tuyen VU[1] , Tien-thanh NGUYEN[1] , Anh-huy HOANG[2], Thi-thu-huong PHAM[1]

ABSTRACT:
Coal fires are a persistent threat to major coal-producing countries in the world. Thus, it is very important to delineate the potential risks of coal fires. This study presents a method for the delineation of coal fire risk areas from Landsat-8 TIRS data. Land surface temperatures (LSTs) were first retrieved from the Landsat-8 TIRS images. The degree of spatial autocorrelation among these LSTs was then identified using local Moran's *I* statistic. Thermal-related anomalies for the delineation of coal fire risk areas were identified by setting the MEAN+2*SDEV (SDEV is the standard deviation), MEAN+3*SDEV, and MEAN+4*SDEV formulas as thresholds on the local Moran's *I* statistic. These coal fire risk areas were finally validated using known coal fire sites and cross-validated by comparing them with those obtained from hot spot analysis. A case study of the Na Duong coal field (northern Vietnam) has shown that coal fire risks at moderate and high levels were mainly detected in the center of the coal field. The higher values of local Moran's *I* statistic, the higher levels of coal fire risks. These coal fire risks were mainly concentrated around known coal fire sites. These results reveal that Landsat-8 TIRS data can effectively delineate coal fire risk areas.

Key-words: *Land surface temperatures; coal fire risks; local Moran's I statistic; Na Duong coal field (Vietnam), Landsat-8 TIRS data.*

1. INTRODUCTION

Coal fires, which occur on the surface (primarily in coal waste piles) and in underground coal seams and are caused by spontaneous combustion, natural events (lightning, forest fires, and peat fires), and human activities (mining and domestic fires) (Du et al. 2015; Vu and Nguyen 2021a). These fires lead to the emission of greenhouse relevant and toxic gasses, the deterioration of vegetation, land subsidence due to volume loss underground, and to the loss of the valuable resource coal (Kuenzer et al. 2007). Coal fires pose potential risks to the environment, infrastructure, and human health. It is therefore, the delineation of coal fire risk areas plays an important role in controlling and preventing their effects and environmental impacts.

Coal fire risks have been conventionally conducted by temperature measurements in boreholes and using thermal cameras in the 1960s (Greene et al. 1969; Knuth et al. 1968) and 1970s (Ellyett and Fleming 1974; Rabchevsky 1972). Although, the main advantages of these methods are the ability of precise coal fire risk detection, they are nearly impossible to gather enough data over large area, especially for inaccessible areas. The occurrence of high spatial resolution airborne thermal remote sensing has overcome this limitation (Vu and Nguyen 2018). However, the high costs for data acquisition by these airborne scanners have limited their practical applications. Since freely available thermal infrared images acquired by satellite sensors started to become available, automated coal fire risk area delineation from space became easier (Künzer 2014). Recent evidence suggests that thermal infrared data from earth observation by satellites has proven to be a powerful tool to support the delineation of coal fire risks(Vu and Nguyen 2021a; Nguyen and Vu 2019b; Vu and Nguyen 2018). Thus, thermal infrared remotely sensed images were chosen to delineate coal fire risks in this study.

[1]*Faculty of Surveying, Mapping and Geographic Information, Hanoi University of Natural Resources and Environment, Vietnam, vdtuyen@hunre.edu.vn,* C*orresponding author: ntthanh@hunre.edu.vn, ptthuong.tdbd@hunre.edu.vn*

[2]*Faculty of Environment, Hanoi University of Natural Resources and Environment, Vietnam, hahuy@hunre.edu.vn*

The methods involved in the automated delineation of coal fire risk areas from thermal infrared sensed data typically incorporate the identification of changes in the surface feature and by-products of coal fires based on land surface temperatures. Using visible/near-infrared data, Song and Kuenzer (2017) successfully demarcated coal fire risks through the study of spectral reflectance (400-2500 nm) properties of coals, adjacent sediments, metamorphic and pyrometamorphic rocks in Wuda coal fire areas, northern China. Apart from thermal infrared data, short-wave infrared (SWIR) data covered by Landsat TM band 4, 5, and 7, ETM+ SWIR band 7, OLI band 6 and 7, and ASTER band 9 has been also successfully employed to detect coal fires in many coal fields around the world such as China's Rujigou coal field (Huo et al. 2014a; Huo et al. 2014b; Huo et al. 2015) and India's Jharia coalfield (Singh et al. 2017). However, up to now, far too little attention has been paid to the spatial autocorrelation among thermal infrared images-obtained land surface temperatures in the coal fire risk delineation. The spatial autocorrelation can be measured using spatial statistics such as Getis's G, Geary's C, spatial scan, and the Moran's I statistic (Nguyen et al. 2016; Nguyen and Vu 2019a; Nguyen 2017). Among these spatial statistics, the local Moran's I statistic has been the most commonly used for spatial autocorrelation analysis (Tiefelsdorf 2002). To overcome the limitation of commonly used methods for the detection of coal fire risks when the spatial pattern of LSTs is not taken into account, one of the first studies on this issue by (Nguyen and Vu 2019b) has attempted to take hot spot analysis-based spatial autocorrelation into the study of coal fires in the Khanh Hoa coal field, north-east of Vietnam. Later, with the help of the local Moran's I statistic, Vu and Nguyen (2021a) has also successfully identified spatial patterns of LSTs retrieved from remotely sensed data and their relation to coal fires in the Khanh Hoa coal field. The results of Vu and Nguyen (2021a) have shown that the closer the coal fire sites, the higher the spatial autocorrelation level, and there exists a strong degree of positive correlation between the distribution of LST spatial pattern with active coal fire sites. However, the main drawback of this study was the lack of detecting local Moran's I anomalies obtained from LSTs. It is therefore, to overcome this limitation, based on the idea of spatial autocorrelation among LSTs proposed by Vu and Nguyen (2021a), this study aims to delineate coal fire risk areas from Landsat-8 TIRS data using the local Moran's I statistic.

In this study, the first step of this method will retrieve LSTs from the Landsat-8 TIRS data using the radiative transfer equation. It will go on to identify the degree of spatial autocorrelation among these LSTs using the local Moran's I statistic. Thermal-related anomalies for the delineation of coal fire risk areas were then identified by setting the MEAN+2*SDEV (SDEV is the standard deviation), MEAN+3*SDEV and MEAN+4*SDEV formulas as thresholds on the local Moran's I statistic. These coal fire risks were finally validated using known coal fire sites.

2. STUDY AREA

The Na Duong basin is situated near the boundary between the Indochina and Southern China microplates. It is part of the Cao Bang Tien Yen fault system (Pubellier et al. 2003) which parallels the Ailao Shan-Red River Fault Zone at 160 km distance to the South (Madelaine et al. 2018). The Na Duong coal field, an open pit coal mine in the Na Duong basin, is situated in the Na Duong basin, Loc Binh district, approximately 20 km SE of Lang Son province, northern Vietnam (Ducrocq et al. 2015). Geographically, the study area of the Na Duong coal field is situated between 21°45′45″N to 21°43′50″N latitudes and 106°57′50″E to 106°59′40″E longitudes, covering an area of approximately 9.8 square kilometers (**Fig. 1**). Its coal reserve was estimated to be 17.5 million metric tons. The coal seams from the Na Duong coal pit represent mainly the telocollinite subfacies and originated mostly under conditions of a wet forest swamp with a high water level (Wysocka 2009). Na Duong coal is a rare special coal in Vietnam with the characteristics of being able to burn naturally, with a large sulfur content. When it meets water, this coal can be converted into sulfuric acid, if released into the environment, it will affect the environment, health, and the surrounding environment. A study of Hoang (2005) concluded that the working environment in the coal field was contaminated by toxins from the coal fires. Coal fires in the Na Duong coal field were reported in recent years. Particularly, the fire intensity was strong in 2017. Thus, remotely sensed images in this year were employed to delineate coal fire risk areas in the coal field.

3. DATA AND METHODS

3.1. Data

In this study, Landsat-8 TIRS data was used to delineate coal fire risk areas in the Na Duong coal field. The Landsat-8 images used in this study were the Level-1 precision and terrain-corrected product that was corrected geometrically and topographically through co-registration (Chen et al. 2017). The Landsat-8 image (path 127, row 045) distributed by the U.S. Geological Survey was acquired on December 6, 2017, and projected in the UTM Zone N48 and WGS 1984 ellipsoid datum. The image was almost free of clouds. The Na Duong coal fire boundary was provided by Na Duong VVMI coal company limited. In addition, a total of four active coal fire sites collected from the field survey by the retrospective study will be also used for the validation of coal fire risk area delineation (**Fig. 1**).
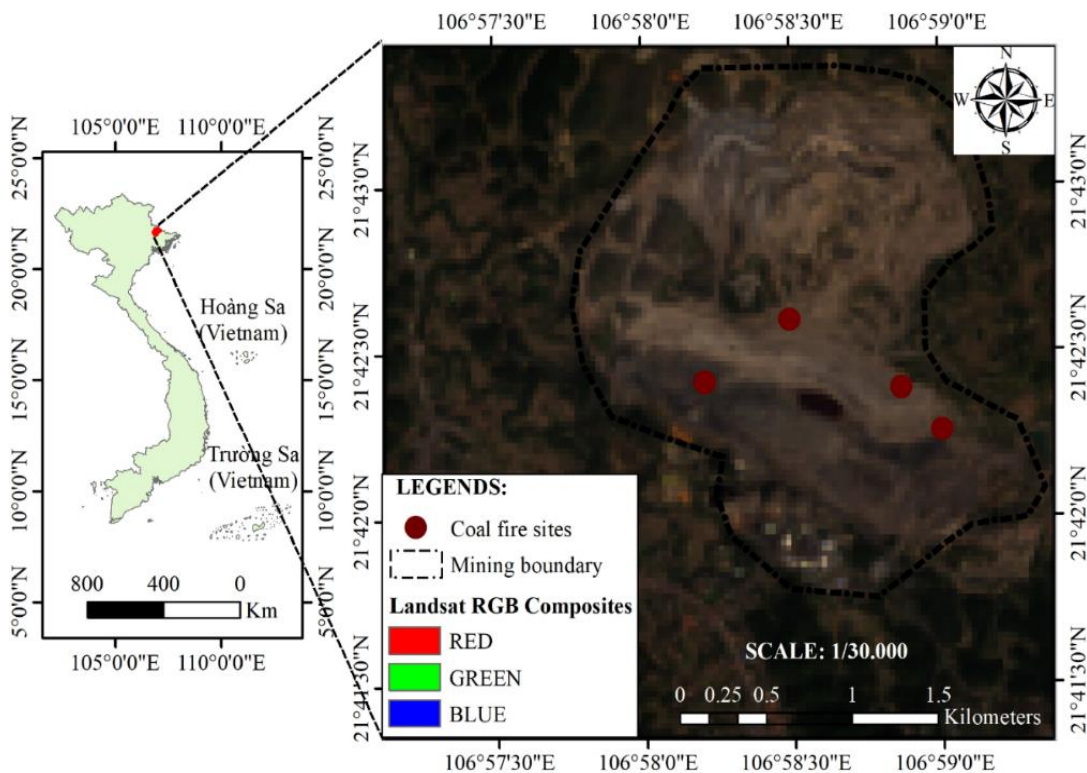

Fig. 1. Study area of Na Duong coal field, Lang Son province, north-east of Vietnam (natural color composite image of Landsat-8 OLI bands 4, 3, 2 in RGB).

3.2. Methods

In this study, the first step of this method will retrieve the land surface temperatures from the Landsat-8 TIRS data using the radiative transfer equation. The degree of spatial autocorrelation among these LSTs was then measured with the help of local Moran's *I* statistic. It will then go on to the identification of thermal-related anomalies for the delineation of coal fire risk areas by setting the MEAN+2*SDEV, MEAN+3*SDEV, and MEAN+4*SDEV formulas as thresholds on the local Moran's *I* statistic. These coal fire risks were validated using known coal fire sites. Finally, Landsat-8 TIRS images acquired in the Na Duong coal field (northern Vietnam) were used to identify low, moderate, and high levels of coal fire risks in the coal field.

a. Land surface temperature retrieval from Landsat-8 TIRS data

The top of atmosphere (TOA) spectral radiance using a conversion equation given as follows (Mishra et al. 2014):

$$L_{TOA,\lambda} = M_L \times Q_{cal} + A_L \tag{1}$$

where: $L_{TOA,\lambda}$ is the TOA radiance $[W/(m^2{\cdot}sr{\cdot}\mu m)]$ at the wavelength λ (μm) measured by the Landsat-8 OLI and TIRS1; M_L is the band-specific multiplicative rescaling; A_L is the band-specific additive rescaling factor; and Q_{cal} is the quantized and calibrated standard product pixel values [DN].

A radiative transfer equation can express the TOA radiance received by a sensor using the following equation (Barsi et al. 2003):

$$L_{TOA,\lambda} = \tau_\lambda \big[\varepsilon_\lambda B_{b,\lambda}(T_s) + (1 - \varepsilon_\lambda)L_{atm,\lambda}^{\downarrow}\big] + L_{atm,\lambda}^{\uparrow} \tag{2}$$

where: $L_{TOA,\lambda}$ is the TOA radiance $[W/(m^2{\cdot}sr{\cdot}\mu m)]$ at the wavelength $\lambda = 10.60\text{-}11.19\mu m$ measured by the Lansat-8 TIRS1; ε_λ is the land surface emissivity; $B_{b,\lambda}(T_s)$ is the blackbody radiance $[W/(m^2{\cdot}sr{\cdot}\mu m)]$ given by the Planck's law and T_s is the LST [Kelvin]; $L_{atm,\lambda}^{\uparrow}$ is the upwelling atmospheric radiance $[W/(m^2{\cdot}sr{\cdot}\mu m)]$; $L_{atm,\lambda}^{\downarrow}$ is the downwelling atmospheric radiance $[W/(m^2{\cdot}sr{\cdot}\mu m)]$ and τ_λ is the total atmospheric transmissivity [dimensionless] between the surface and the sensor.

In this study, the land surface emissivity (LSE) was derived based on the work of (Sobrino et al. 2008) by using the Normal Differential Vegetation Index (NDVI)-based threshold approach. LSE of each land cover type is derived using vegetation fraction (f_v) determined from NDVI. The following equation is used to derive vegetation fraction from NDVI:

$$f_v = \left(\frac{NDVI - NDVI_s}{NDVI_v - NDVI_s}\right)^2 \tag{3}$$

where: $NDVI_s$ and $NDVI_v$ are the pixel values corresponding to non-vegetation cover (bare soil) and full vegetation cover, respectively.

For this present research, the values of $NDVI_s = 0.18$ and $NDVI_v = 0.76$ were extracted from the NDVI histogram using the 5% and 95% confidence levels of NDVI values as proposed by Sobrino et al. (2008). LSE values were calculated in three cases: (i) if $NDVI < NDVI_s$ then the pixel is considered mainly covered by bare soil ($f_v = 0$), and a mean value of 0.97 is assumed for the soil emissivity (ε_s) (Sobrino et al. 2008); (ii) if $NDVI > NDVI_v$ then the pixel corresponds to dense vegetation areas (fully vegetated) ($f_v = 1$), and the vegetation emissivity (ε_v) is given a value of 0.99 (Sobrino et al. 2008); (iii) if $NDVI_s < NDVI < NDVI_v$ then each pixel is considered mixing of bare soil and vegetation, and the emissivity can be derived from the following equation:

$$\varepsilon_i = f_v \cdot \varepsilon_v + (1 - f_v) \cdot \varepsilon_s \tag{4}$$

where: ε_i is the emissivity of pixel i; ε_v and ε_s are the vegetation and soil emissivity; and f_v is the vegetation fraction obtained from equation (3). The LST, T_s [Kevin], is estimated by inversion of Planck's law using equation (5):

$$T_s = \frac{K_2}{ln\left[\frac{K_1}{B_{b,\lambda}(T_s)} + 1\right]} \tag{5}$$

where: K_2 = calibration constant 2 [Kevin]; K_1 = calibration constant 1 $[W/(m^2{\cdot}sr{\cdot}\mu m)]$; $B_\lambda(T_s)$ = blackbody radiance $[W/(m^2{\cdot}sr{\cdot}\mu m)]$; and ln = natural logarithm.

b. Local Moran's I statistic

The spatial pattern of land surface temperatures at each specific location can be identified using the local Moran's *I* statistic. The local Moran's *I* statistic for the LST at location *i* is defined as the following equation (Anselin 1995):

$$I_i = \frac{(x_i - \bar{x})}{\sigma^2} \sum_{j \# i, j \in J_i}^{N} W_{ij}(x_j - \bar{x}) \qquad (6)$$

where: x_i and x_j are LSTs at locations i and j; J_i denotes the neighbourhood set of LSTs at location i; $\bar{x}$ is the mean of land surface temperatures; W_{ij} is a spatial weight matrix used for computing a local Moran's I coefficient at the location i within a given distance d. W_{ij} is a spatial weight matrix for a given distance d and d_{ij} is the distance between the RMDs at locations i and j. If $d_{ij} < d$ then $W_{ij} = 1$ and $W_{ij} = 0$ otherwise.

In this way, only the pairs of sites (i, j) within the stated distance class (*d*) are taken into account in the calculation of any given coefficient. For a given distance class, the weights W_{ij} are written in a ($n \times n$) spatial weighting matrix W_{ij}. Before computing the local Moran's, *I* coefficient, a matrix of geographic distances among land surface temperatures must be first calculated.

The value $I_i = 0$ indicates no spatial autocorrelation; $I_i > 0$ represents positive spatial autocorrelation; and if $I_i < 0$ then the spatial autocorrelation is negative. When there is a positive Moran's *I* coefficient, two types of LST spatial clusters were identified including high-high (H-H) clusters (a high LST surrounded by high LSTs) and low-low (L-L) clusters (a low LST surrounded by low LSTs). If $p(I_i)<\alpha$, $I_i>0$, and $x_i - \bar{x}>0$, then x_i and $x_{j \in Ji}$ belong to a spatial cluster between high LSTs (H-H clusters). If $p(I_i)<\alpha$, $I_i>0$ and $x_i - \bar{x}<0$, then x_i and $x_{j \in Ji}$ belong to a spatial cluster between low LSTs (L-L clusters). If $p(I_i)<\alpha$, $I_i<0$, and $x_i - \bar{x}>0$ then a high LST, x_i, is surrounded by low LSTs, $x_{j \in Ji}$, (H-L outliers). If $p(I_i)<\alpha$, $I_i<0$, and $x_i - \bar{x}>0$ then a low LST, x_i, is surrounded by high LSTs, $x_{j \in Ji}$, (L-H outliers) (Vu and Nguyen 2021a).

c. Identification of coal fire risk areas

The distribution of local Moran's *I* statistic is strongly right-skewed (Nguyen et al. 2016; Nguyen and Vu 2019a). Thus, thermal-related anomalies to delineate coal fire risk areas were identified by setting the MEAN+2*SDEV (SDEV is the standard deviation), MEAN+3*SDEV, and MEAN+4*SDEV formulas as thresholds on local Moran's *I* coefficients. Four classes of coal fire risks are obtained and ranked, accordingly, as follows: (i) no risk, if I_i < MEAN+2*SDEV; (ii) low risk, if MEAN+2*SDEV < I_i < MEAN+3*SDEV; (iii) moderate risk, if MEAN+3*SDEV < I_i < MEAN+4*SDEV; (iv) high risk, I_i > MEAN+4*SDEV.

d. Accuracy assessment

The areas of coal fire risks were first validated based on four coal fire sites collected from the field survey and then cross-validated by comparing them with those obtained from hot spot analysis. The validation was performed using ArcGIS software. In this case, the layer of coal fire sites is laid over that of coal fire risk areas which were obtained from the above-discussed method. The degree of spatial correlation or conformity between the areas of medium-to-high risk levels and the active coal fire sites is the basis for evaluating the accuracy of the proposed method. The higher the degree of spatial correlation or conformity, the higher the accuracy of the proposed method. The cross-validation was then carried out by comparing areas of coal fire risks at different levels delineated from the proposed method with those obtained from hot spot analysis. The Getis-Ord's G_i^* statistic-based hot spot analysis allows for identifying spatial clusters of high (hot spots) and low (cold spots) values at local locations (Kowe et al. 2019). Recently, a recent study of Nguyen and Vu (2019b) has successfully employed this method for detecting coal fires in the Khanh Hoa coal field in north-east of VietNam. Therefore, different coal fire risk levels delineated from hot spot analysis were also used to compare with those obtained from the proposed method.

4. RESULTS AND DISCUSSIONS

4.1. Distribution of LSTs in the Na Duong coal field

Land surface temperatures are indicative of underground coal fires in the coal field (Vu and Nguyen 2021b). Therefore, the study of the distribution of LST plays important role in the detection of coal fire risk from remotely sensed data. It has been observed by Mukherjee et al. (2018) that, during the summer season, water bodies have high temperatures, thus affecting the performance of detection of fire risks. This might lead to the false detection of coal fire risks. Therefore, in this study, Landsat-8 TIRS satellite images were collected in the winter season of 2017. Data from **Fig. 2** demonstrates the spatial distribution of land surface temperatures obtained from Landsat-8 TIRS satellite images in the Na Duong coal field, Lang Son province. As can be seen from **Fig. 1**, land surface temperatures ranged from 16.2°C to 35.3°C. The lowest, highest, and mean temperatures were 16.2°C, 35.3°C, and 19.3°C, respectively. What stands out from **Fig. 1** is that low LSTs were mainly concentrated in areas with hills, dense vegetation covers, and on the outside of the coal field. Whereas, high temperatures above 25°C were mainly concentrated in densely populated areas and especially inside the Na Duong coal mining site. Particularly, there is a clear trend that very high land surface temperatures were detected inside areas of the Na Duong coal field. These high LSTs nearby coal fire sites were much higher than those of surrounding environments. This is consistent with the findings in a recent study by Nguyen and Vu (2019b). Another aspect of interestingness about the data in Fig. 1 is that the radiation heat flux was locally concentrated in the Na Duong coal field, especially nearby active coal fire sites. This is also in line with those reported in previous studies (Nguyen and Vu 2019b; Vu and Nguyen 2021b; Vu and Nguyen 2018). This can be judged that underground coal fires were likely to occur in this area.

4.2. Distribution of local Moran's *I* statistic

With the input data of land surface temperatures obtained from Landsat-8 TIRS data in the Na Duong coal field as discussed in the previous section. The local Moran's *I* statistic was employed to discover the spatial patterns of these LSTs. Data from **Fig. 3** demonstrates the spatial distribution of the local Moran's *I* statistic. It can be seen that the study resulted in the values of the local Moran's *I* varying from -0.09 to 62.12. It's minimum, mean, maximum values, and standard deviation of local Moran's *I* index were -0.09, 1.92, 62.12, and 3.14, respectively. From the data in **Fig. 3**, it is apparent that, similar to those obtained from low land surface temperatures, as discussed in section 4.1, low values of local Moran's *I* statistic were mainly detected in the areas outside the coal mining area.
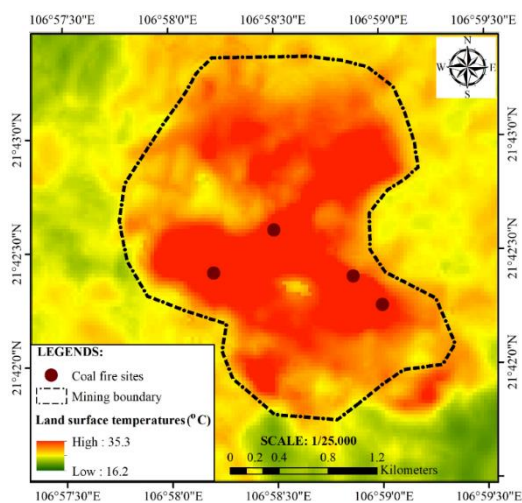

Fig. 2. Spatial distribution of land surface temperatures in the Na Duong coal field.
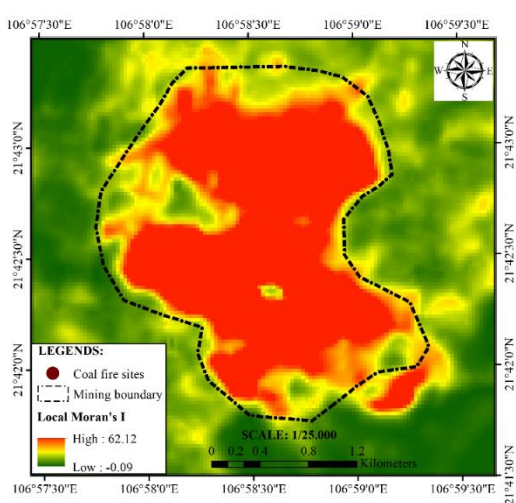
Fig. 3. Spatial distribution of the local Moran's I statistic in the Na Duong coal field.

These areas correspond to different land cover types such as vegetation, bare soil, residential areas, or surface water outside the coal field mining area. The low values of local Moran's *I* statistic represent the spatial clustering of low land surface temperatures (low-low clusters). Whereas, high values of the local Moran's *I* statistic were mainly concentrated in areas with high land surface temperatures above 30°C. These areas were mainly detected in densely populated areas and especially inside the Na Duong coal mining boundary. The most interesting aspect of this figure is that there existed high and very high values of local Moran's *I* inside the Na Duong coal field. These high values of the local Moran's *I* statistic show significant evidence that spatial clustering of high land surface temperatures (high-high clusters) was detected around active coal fires. This proves that areas of high land surface temperatures (thermal anomalies) were concentrated at the time of image acquisition. These high values of Moran's *I* coefficients might be caused due to the existence of the underground coal fires in the coal field.

It can be seen that data from **Fig. 3** can be compared with the data in **Fig. 2** which shows an important difference. The local Moran's *I* statistic highlights the spatial dependence and auto-correlation of land surface temperatures retrieved from Landsat-8 TIRS data. More specifically, the local Moran's statistic highlights areas where high land surface temperatures were highly concentrated (high-high cluster). Typically, although the land surface temperatures were high in the north and northeast of the study area, the local Moran's *I* statistic found there was spatial clustering of high temperatures presenting no concentration of thermal radiation in the north and northwest areas outside the coal field. This is consistent with the fact that no coal fire sites were reported in these areas. Whereas, the most surprising aspect of the data in Fig. 3 is that the local Moran's statistic well detected a high degree of spatial dependence or spatial auto-correlation in areas of high land surface temperatures around coal mine fire points within the mining area. Additionally, the histogram for the local Moran's statistic shows that the data of land surface temperatures does not follow the normal distribution. The distribution of the local Moran's *I* statistic is rather right-skewed due to the existence of high and very high values of the local Moran's *I* statistic. A large number of the local Moran's *I* values were found in the range of 0 to 4. The values of local Moran's *I* which are greater than 10 presenting high land surface temperatures were highly concentrated. This proves that there existed high risk of the underground coal fires in the coal mining area in the Na Duong coal field.

4.3. Analysis of coal fire risk areas

As discussed in the previous section, the distribution of the local Moran's *I* statistic is strongly right-skewed due to the existence of high and very high land surface temperatures caused by active coal fires in the coal field. The anomalies of the local Moran's *I* statistic reflect the high degree of concentration of high land surface temperatures or spatial clustering of high temperatures (high-high clusters). It is, therefore, these anomalies were identified by setting the MEAN+2*SDEV, MEAN+3*SDEV, and MEAN+4*SDEV formulas as thresholds on the local Moran's *I* coefficient. With the values of the mean and standard deviation of 1.92 and 3.14, three different thresholds of 8.74, 12.15, and 15.56 were successfully determined for the local Moran's *I* statistic, respectively. In this case, based on these three thresholds, the areas of coal fire risk areas were delineated and ranked, accordingly, as follows: (i) no risk, if $I_i < 8.74$; (ii) low risk, if $8.74 \leq I_i < 12.15$; (iii) moderate risk, if $12.15 \leq I_i < 15.56$; and (iv) high risk, $I_i \geq 15.56$. These areas of coal fire risk levels were then validated using four active coal fire sites collected from the field survey at the time of image acquisition. From the data in **Fig. 4**, it can be seen that there was a high degree of spatial correlation or good conformity between the areas of coal fire risks and active coal fire sites (high-high clusters). Particularly, coal fire areas at high-risk level were detected around the known coal fire sites. As data in **Fig. 4** shows the further away from the active coal fires, the lower the risks of fires. This is consistent with those reported in previous studies (Vu and Nguyen 2021b; Vu and Nguyen 2018). Additionally, it can be seen that coal fire areas from **Fig. 4** can be compared with those in **Fig. 5** which shows good conformity, especially for coal fire risks at the moderate level. Thus, it can be concluded that these results prove the effectiveness and high accuracy of the proposed method.
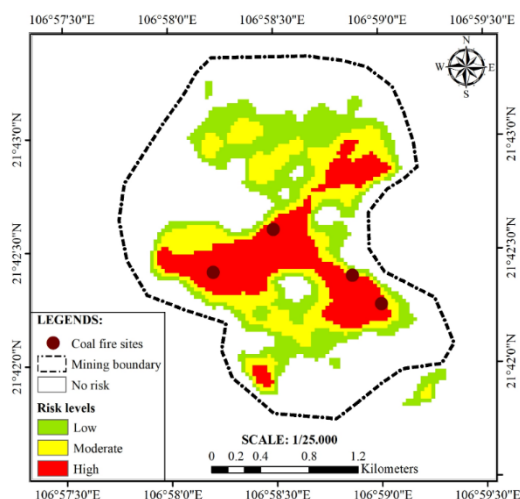
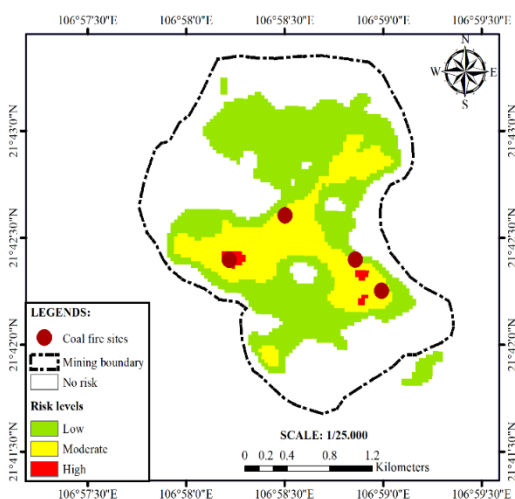**Fig. 4.** Coal fire risk areas delineated using the local Moran's *I* in the Na Duong coal field.

Fig. 5. Coal fire risk areas delineated using the hot spot analysis in the Na Duong coal field.

Data from **Fig. 4** illustrates the spatial distribution of coal fire risk areas in the Na Duong coal field in the winter season of 2017. It can be seen that coal fire risks at moderate and high levels were mainly detected in the center of the coal field. Whereas, coal fire risks at the low level were found around areas of risk at moderate and high levels in the northern part of the coal field. There was almost no risk of underground coal fires in such land cover types as forests, water surfaces, and densely populated areas where the low degree of spatial auto-correlation was identified in areas of low land surface temperatures. Additionally, underground coal fire risks at low, moderate, and high levels were mainly detected inside the coal mining area of Na Duong. Most importantly, coal fire risks at the high level were mainly concentrated near known coal fire sites (as shown in **Fig. 4**). Besides that, the local Moran's *I* anomalies caused by high land surface temperatures were also found in the southeast region of the study area. However, these were false anomalies caused by the high density of the population.

5. CONCLUSIONS

In this study, a method for the delineation of coal fire risk areas from Landsat-8 TIRS data was proposed. Land surface temperatures were first retrieved from the Landsat-8 TIRS data using the radiative transfer equation. The degree of spatial autocorrelation among land surface temperatures was then identified with the help of the local Moran's *I* statistic. Thermal-related anomalies for the delineation of coal fire risk areas were finally identified by setting the MEAN+2*SDEV, MEAN+3*SDEV, and MEAN+4*SDEV formulas as thresholds on the local Moran's *I* statistic. The results of this study show that coal fire risks at moderate and high levels were mainly detected in the center of the coal field. The higher values of local Moran's *I* statistic, the higher levels of coal fire risk. These findings suggest that Landsat-8 TIRS data and the proposed method can effectively delineate coal fire risk areas.

AUTHOR CONTRIBUTIONS

Tien-thanh NGUYEN conceived and designed the study. Anh-huy HOANG and Thi-thu-huong PHAM collected the data. Tien-thanh Nguyen and Danh-tuyen VU performed image processing and wrote the manuscript. All authors discussed the results, and implications and commented on the manuscript at all stages.

R E F E R E N C E S

Anselin L (1995) Local indicators of spatial association—LISA. Geographical analysis 27 (2):93-115

Barsi JA, Barker JL, Schott JR An atmospheric correction parameter calculator for a single thermal band earth-sensing instrument. In: Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International, 2003. IEEE, pp 3014-3016. doi:http://dx.doi.org/10.1109/IGARSS.2003.1294665

Chen G, He Y, De Santis A, Li G, Cobb R, Meentemeyer RK (2017) Assessing the impact of emerging forest disease on wildfire using Landsat and KOMPSAT-2 data. Remote Sensing of Environment 195:218-229

Du X, Cao D, Mishra D, Bernardes S, Jordan TR, Madden M (2015) Self-adaptive gradient-Based thresholding method for coal fire detection using ASTER thermal infrared data, part I: methodology and decadal change detection. Remote sensing 7 (6):6576-6610

Ducrocq S, Benammi M, Chavasseau O, Chaimanee Y, Suraprasit K, Pha PD, Phuong VL, Phach PV, Jaeger J-J (2015) New anthracotheres (Cetartiodactyla, Mammalia) from the Paleogene of northeastern Vietnam: biochronological implications. Journal of Vertebrate Paleontology 35 (3):e929139

Ellyett C, Fleming AW (1974) Thermal infrared imagery of the Burning Mountain coal fire. Remote Sensing of Environment 3 (1):79-86

Greene GW, Moxham RM, Harvey AH (1969) Aerial infrared surveys and borehole temperature measurements of coal mine fires in Pennsylvania. Remote Sensing of Environment, VI:517

Hoang VT (2005) Working environment and health of workers in Na Duong coal mine, Lang Son province. Journal of Preventive Medicine:65-69

Huo H, Jiang X, Song X, Li Z-L, Ni Z, Gao C (2014a) Detection of coal fire dynamics and propagation direction from multi-temporal nighttime Landsat SWIR and TIR data: A case study on the Rujigou coalfield, Northwest (NW) China. Remote sensing 6 (2):1234-1259

Huo H, Jiang X, Song X, Ni Z, Liu L (2014) Coal fires dynamics detection over Rujigou coalfield, Ningxia, NW China. In: Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International, 2014b. IEEE, pp 4512-4515. doi:https://dx.doi.org/10.1109/IGARSS.2014.6947495

Huo H, Ni Z, Gao C, Zhao E, Zhang Y, Lian Y, Zhang H, Zhang S, Jiang X, Song X (2015) A study of coal fire propagation with remotely sensed thermal infrared data. Remote Sensing 7 (3):3088-3113. doi:https://dx.doi.org/10.3390/rs70303088

Knuth W, Fisher Jr W, Stingelin R (1968) Detection, delineation and monitoring of subsurface coal fires by aerial infrared scanning. Geographer HRB-Singer, Inc, a subsidiary of the Singer Company, state College, Pennsylvania, USA:877-881

Kowe P, Mutanga O, Odindi J, Dube T (2019) Exploring the spatial patterns of vegetation fragmentation using local spatial autocorrelation indices. Journal of Applied Remote Sensing 13 (2):024523-024523

Kuenzer C, Zhang J, Li J, Voigt S, Mehl H, Wagner W (2007) Detecting unknown coal fires: synergy of automated coal fire risk area delineation and improved thermal anomaly extraction. International Journal of Remote Sensing 28 (20):4561-4585

Künzer C (2014) Remote and in situ mapping of coal fires: case studies from China and India. Coal and Peat Fires: A Global Perspective Volume 3: Case Studies–Coal Fires 3:57-93

Mishra N, Haque M, Leigh L, Aaron D, Helder D, Markham B (2014) Radiometric cross calibration of Landsat 8 operational land imager (OLI) and Landsat 7 enhanced thematic mapper plus (ETM+). Remote Sensing 6 (12):12619-12638

Mukherjee J, Mukherjee J, Chakravarty D Detection of coal seam fires in summer seasons from Landsat 8 OLI/TIRS in Dhanbad. In: Computer Vision, Pattern Recognition, Image Processing, and Graphics: 6th National Conference, NCVPRIPG 2017, Mandi, India, December 16-19, 2017, Revised Selected Papers 6, 2018. Springer, pp 529-539

Nguyen TT (2017) Use of Moran's I and robust statistics to separate geochemical anomalies in Jiurui area (Southeast China). Bulletin Of The Mineral Research and Exploration 156 (156):179-192

Nguyen TT, Vu DT, Nguyen TLH (2016) Spatial cluster and outlier identification of geochemical association of elements: A case study in Juirui copper mining area. Bulletin of the Mineral Research and Exploration 153 (153):159-167

Nguyen TT, Vu TD (2019a) Identification of multivariate geochemical anomalies using spatial autocorrelation analysis and robust statistics. Ore Geology Reviews 111:102985

Nguyen TT, Vu TD (2019b) Use of hot spot analysis to detect underground coal fires from Landsat-8 TIRS data: A case study in the Khanh Hoa coal field, north-east of Vietnam. Environment and Natural Resources Journal 17 (3):1-10

Pubellier M, Rangin C, Phach PV, Que BC, Hung D, Sang C (2003) The Cao Bang-Tien Yen Fault: Implications on the relationships between the Red River Fault and the South China Coastal Belt. Advances in Natural Sciences 4 (4):347-361

Rabchevsky G (1972) Determination from Available Satellite and Aircraft Imagery of the Applicatility of Remote Sensing Techniques to the Detection of Fires Burning in Abandoned Coal Mines and Unmined Coal Deposits Located in North Central Wyoming and Southern Montana. Contract No SO 211087

Singh A, Raju A, Pati P, Kumar N (2017) Mapping of Coal Fire in Jharia Coalfield, India: a Remote Sensing Based Approach. Journal of the Indian Society of Remote Sensing 45 (2):369-376. doi:https://doi.org/10.1007/s12524-016-0590-5

Sobrino JA, Jiménez-Muñoz JC, Sòria G, Romaguera M, Guanter L, Moreno J, Plaza A, Martínez P (2008) Land surface emissivity retrieval from different VNIR and TIR sensors. IEEE Transactions on Geoscience and Remote Sensing 46 (2):316-327. doi:https://dx.doi.org/10.1109/TGRS.2007.904834

Song Z, Kuenzer C (2017) Spectral reflectance (400–2500 nm) properties of coals, adjacent sediments, metamorphic and pyrometamorphic rocks in coal-fire areas: A case study of Wuda coalfield and its surrounding areas, northern China. International Journal of Coal Geology 171:142-152

Tiefelsdorf M (2002) The saddlepoint approximation of Moran's I's and local Moran's Ii's reference distributions and their numerical evaluation. Geographical analysis 34 (3):187-206

Vu D-t, Nguyen T-t (2021a) Spatial pattern of land surface temperatures and its relation to underground coal fires in the Khanh Hoa Coal Field, North-East of Vietnam. Arabian Journal of Geosciences 14 (3):1-15

Vu D-t, Nguyen T-t (2021b) Spatial pattern of land surface temperatures and its relation to underground coal fires in the Khanh Hoa Coal Field, North-East of Vietnam. Arabian Journal of Geosciences 14 (3):145

Vu TD, Nguyen TT (2018) Spatio-temporal changes of underground coal fires during 2008–2016 in Khanh Hoa coal field (North-east of Viet Nam) using Landsat time-series data. Journal of Mountain Science 15 (12):2703-2720

Wysocka A (2009) Sedimentary environments of the Neogene basins associated with the Cao Bang–Tien Yen Fault, NE Vietnam. Acta Geologica Polonica 59 (1):45-69

Aims and Scope

Geographia Technica is a journal devoted to the publication of all papers on all aspects of the use of technical and quantitative methods in geographical research. It aims at presenting its readers with the latest developments in G.I.S technology, mathematical methods applicable to any field of geography, territorial micro-scalar and laboratory experiments, and the latest developments induced by the measurement techniques to the geographical research.

Geographia Technica is dedicated to all those who understand that nowadays every field of geography can only be described by specific numerical values, variables both of time and space which require the sort of numerical analysis only possible with the aid of technical and quantitative methods offered by powerful computers and dedicated software.

Our understanding of *Geographia Technica* expands the concept of technical methods applied to geography to its broadest sense and for that, papers of different interests such as: G.I.S, Spatial Analysis, Remote Sensing, Cartography or Geostatistics as well as papers which, by promoting the above mentioned directions bring a technical approach in the fields of hydrology, climatology, geomorphology, human geography territorial planning are more than welcomed provided they are of sufficient wide interest and relevance.

Targeted readers:

The publication intends to serve workers in academia, industry and government.
Students, teachers, researchers and practitioners should benefit from the ideas in the journal.

Guide for Authors

Submission

Articles and proposals for articles are accepted for consideration on the understanding that they are not being submitted elsewhere.

The publication proposals that satisfy the conditions for originality, relevance for the new technical geography domain and editorial requirements, will be sent by email to the address editorial-secretary@technicalgeography.org.

This page can be accessed to see the requirements for editing an article, and also the articles from the journal archive found on www.technicalgeography.org can be used as a guide.

Content

In addition to full-length research contributions, the journal also publishes Short Notes, Book reviews, Software Reviews, Letters of the Editor. However the editors wish to point out that the views expressed in the book reviews are the personal opinion of the reviewer and do not necessarily reflect the views of the publishers.

Each year two volumes are scheduled for publication. Papers in English or French are accepted. The articles are printed in full color. A part of the articles are available as full text on the www.technicalgeography.org website. The link between the author and reviewers is mediated by the Editor.

Peer Review Process

The papers submitted for publication to the Editor undergo an anonymous peer review process, necessary for assessing the quality of scientific information, the relevance to the technical geography field and the publishing requirements of our journal.

The contents are reviewed by two members of the Editorial Board or other reviewers on a simple blind review system. The reviewer's comments for the improvement of the paper will be sent to the corresponding author by the editor. After the author changes the paper according to the comments, the article is published in the next number of the journal.

Eventual paper rejections will have solid arguments, but sending the paper only to receive the comments of the reviewers is discouraged. Authors are notified by e-mail about the status of the submitted articles and the whole process takes about 3-4 months from the date of the article submission.