Land Use and Land Cover Change Detection for Water Yield Estimation using Remote Sensing Data in Batu Pahat, Johor

Hong Ziexin[a], Amal Najihah[a], Sunisa Suchat[b], Siwipa Pruitikanee[b], Jinda Kongcharoen[b], Supattra Puttinaovarat [b,*]

[a] Faculty of Earth Science, University Malaysia Kelantan, 16100, Malaysia
[b] Faculty of Science and Industrial Technology, Prince of Songkla University, Surat Thani Campus, 84000, Thailand
Corresponding author: *supattra.p@psu.ac.th

Abstract— **Globally, the accelerating urbanization led by industrialization and population growth causes severe environmental degradation. The urban expansion particularly to the conversion of land activities affects the ecosystem services critically. This study helps to fill in the gap of determining water yield in the urban area to reduce water stress due to the spatial land use change. The research objectives are to quantify the spatial land use change in Batu Pahat, Johor in the year 1999, 2010, and 2018. Second, to identify the water yield of Batu Pahat in the years 1999, 2010, and 2018. Third, to determine the relationship between water yield, vegetation, and urban expansion. The methods used are landscape change, water yield simulation, and statistical analysis by using the software included ENVI, ArcGIS, FRAGSTAT, Annual Water Yield InVEST Model, and Microsoft Excel. Raw satellite images were extracted for the year 1999, 2010, and 2018. The supervised classification of LULC (Land Use and Land Cover) was done based on the region created which are built-up area, cleared land, vegetation, and water bodies. This study generates results for the changes in percentage area for each LULC class. The highest percentage of area in Batu Pahat is vegetation while the cleared land ranked lowest. In conclusion, this study will aid in understanding and provided empirical data result for the urban expansion and water yield in Batu Pahat, Johor by using GIS and remote sensing applications to produce land use and water yield map as final output.**

Keywords—**Water yield; LULC classification and change detection; InVest model; landsat.**

I. INTRODUCTION

Currently, more and more countries worldwide have great impacts on environmental problems and natural disasters which the frequency is increasing, and the situation is more difficult [1], [2]. The crucial cause of such circumstance is the expansion of urban areas resulted from the change of land use and land cover (LULC) [3], [4]. The former vast agricultural area of forest has been changed into the urban area, industrial area, and bare land, which affects the ecological balance. Consequently, many problems arise, for instance, the off-season natural disaster that the preparation is not possible as in the past, the increasing pollution due to the destruction of forest and the construction and tools and equipment [5]. Moreover, one of the crucial problems found in the agricultural country was the draught that directly impacts the agriculturist who does not have sufficient water to utilize in the farm and household.

Additionally, the available water cannot be fully utilized due to its quality resulting from land-use change [6]. The flood that damages partial agricultural areas, so the agriculturist is unable to harvest the product, and the products are affected [7] is also the problem that arises from the lack of land use and land cover change monitoring. Previously, the fieldwork to survey LULC was applied. However, the limitation was that the number of laborers' amount of time was required [8].

However, with the advancement of Geographic Information System (GIS) and Remote Sensing (RS), the data obtained from the Earth Observation Satellite were used for digital image processing using machine learning regardless of human labor. As a result, the LULC data are acquired quickly [8], [9] and used as the input to analyze water yield estimation in couple with other spatial data to determine the relationship between the water yield and the urban growth, as well as the change of LULC for monitoring and preparation. Further, such data might be useful for urban planning. From the

literature review and relevant research, most of them studied the LULC classification using satellite images and the LULC change detection using satellite images by utilizing the machine learning algorithms widely [8], [9]. Nevertheless, the study on the application of LULC and the spatial data to analyze the relationship between the water yield and urban growth and the use of each type of land is very rare. Most studies only emphasized the specific aspect, such as examined only the water yield at each area at the same period. Therefore, the application is limited; for instance, the automatic monitoring of land use is not possible, the data cannot be used for prediction [10], [11].

For this reason, this research proposes the application of machine learning techniques and supervised learning classification. It aims to classify and monitor the LULC change detection using satellite images Landsat 5 and Landsat 7 at each period. Hence, this approach is subject to acquire LULC data to analyze the relationship between the water yield and the urban growth and the use of each type of land in the Batu Pahat district in Malaysia, the studied area.

II. Materials and Method

The main contribution of this research is the classification and detection LULC using Landsat satellite images based on a supervised learning classification algorithm. Moreover, this research consisted of three steps, i.e., data collection and preparation, LULC classification and change detection, and water yield estimation. Each step is described in more detail in the following sub-sections. Fig. 1 depicts the conceptual diagram describing the proposed scheme.
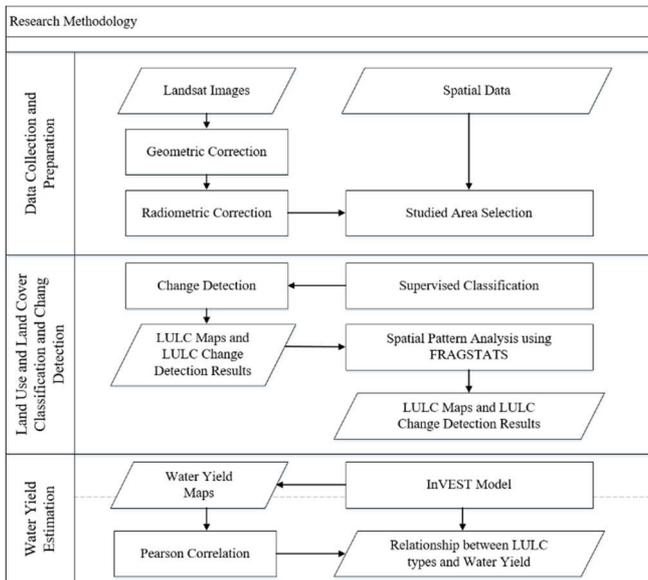
Fig. 1 Conceptual framework diagram of the proposed scheme.

A. Study Area

Batu Pahat is a district in Malaysia (Fig. 2). The main river in Batu Pahat is the Batu Pahat River which also known as Sungai Batu Pahat. According to the official portal of Batu Pahat Municipal Council, the total number of Malaysian citizens in Batu Pahat is around 383,391, while non-Malaysian citizens are around 18,511. The main factor Batu Pahat has been chosen as a study area is the demographic of the area. Batu Pahat is a developing district that mainly focuses on manufacturing, food packaging, and agritourism. The rapid growth of the district had increased the economic growth and the total population. In the meantime, the LULC change dramatically led to rapid urban expansion.

Fig. 2 The map of Batu Pahat district.

B. Data Collection and Preparation

The satellite images of the study area were downloaded from the USGS website, which the satellite imagery had high acceptance and was widely used across the global [12]. The images were obtained from Landsat 7 for the year 1999, Landsat 5 for the year 2010, and Landsat 7 for the year 2018 with only less than 1% cloud coverage for better visualization. Level 1 of the images being chosen as the images are atmospherically corrected which can reduce the influences and the uncertainty in processing with data analysis. The data collection needed for the water yield simulation included precipitation. Precipitation data for the three years was also downloaded from the freely available global dataset through the website climatologylab.org which the Climatology Lab developed as the field data was difficult to obtain for the three years. The datasets included primary climate variables and the derived variables. The temporal information precipitation is inherited from the Climatic Research Unit (CRU). Besides, the average annual reference evapotranspiration data expressed in depth of water in millimeters (mm) also downloaded from the same website as the TerraClimate were well linked to station-based reference evapotranspiration from FLUXNET stations. The file types of both the data used to process are in the raster dataset. The PAWC (Plant Available Water Content) was downloaded from HWSD (Harmonized World Soil Database) developed by the ISRIC-World Soil Information. The database composed of a map in GIS polygon format consists of different soil mapping units linked to harmonized attribute data. Data were extracted and converted into the unit millimeter (mm) for the soil depth and fraction between 0 to 1 for the plant's available water content. Next, the biophysical table will be stored in a .csv file type showing each of the LULC classes in the map produced.

C. LULC Classification and Change Detection

LULC classification based on supervised learning classification using Landsat satellite images. On detecting LULC changes, the classified images were validating on Landsat imaging data acquired in 1998, 2010, and 2018. The LULC changes between subsequent annual periods were determined by comparing extracted LULC types at the respective year. In this paper, the accuracy of the LULC

classification by Maximum Likelihood Classification (MLC) algorithms was assessed by comparing the LULC result extracted against the LULC data from the Malaysia government.

D. Water Yield Estimation

InVEST Model invented by the Natural Capital Project is commonly used in the study because of the ease of use and the data requirement needed to run the modeling is relatively easy [11]. The model was quantified mainly from the LULC maps and the main input data. The water yield model works using the Budyko curve theory [13]. It simulates the water yield at grid level change associating with the biophysical components of the LULC maps. The InVEST model run in gridded format is required the input data, which consists of LULC, precipitation, soil-depth potential evapotranspiration, and PAWC to be in raster form to generate the pixel output [14]. Water yield equation calculated as follow:

$$Y(x) = (1 - \frac{AET(x)}{P(x)}) \cdot P(x) \qquad (1)$$

Where $AET(x)$ is annual real evapotranspiration for the pixel x and $P(x)$ is annual precipitation on pixel x.

For LULC evapotranspiration fragment using an equation as follow:

$$\frac{AET(x)}{P(x)} = 1 + \frac{PET(x)}{P(x)} - [1 + (\frac{PET(X)}{P(X)})^w]^{\frac{1}{w}} \qquad (2)$$

PET(x) is defined as:

$$PET(x) = K_c(l_x) \cdot ET_o(x) \qquad (3)$$

Where $ET_o(x)$ is evapotranspiration from pixel x and $K_c(l_x)$ is an evapotranspiration coefficient associated with the LULC (l_x) on pixel x. $\omega(x)$ is a linear function of $AWC \cdot N/P(x)$, where N is the number of the events per year, and AWC is the volumetric PAWC, and $\omega(x)$ is defined as:

$$\omega(x) = Z \cdot \frac{AWC(x)}{P(x)} + 1.25 \qquad (4)$$

Where AWC(x) is the volumetric (mm) of PAWC. The soil texture and effective rooting depth define AWC(x). Z is a constant of seasonal factors, which record the precipitation and hydrogeological properties [4].

III. RESULT AND DISCUSSION

A. LULC Change Detection Results

The analysis was conducted in the ArcGIS Desktop Application to determine the spatial area change of the study area in the three years. The LULC changes have then been identified by applying the intersect function in the software to determine the transition of the class types from 1999 to 2018. The outcomes of the result were tabulated in Table I and Fig. 3. From the data below, the most abundant class with the highest spatial area is the class vegetation in the three years, which has 67.04% in 1999, 69.78% in 2010, and 75.68% in 2018, showing the area of the class increase gradually year by year. In contrast, the class water bodies decrease incrementally from the year 1999 (16.56%) to 2010 (12.6%) to 2018 (7.42%). According to the vegetation and forest area

were converted to the palm oil plantation in the year 1999 due to the higher economic values. Consequently, the increased agriculture also raised the water demand for the plantation.
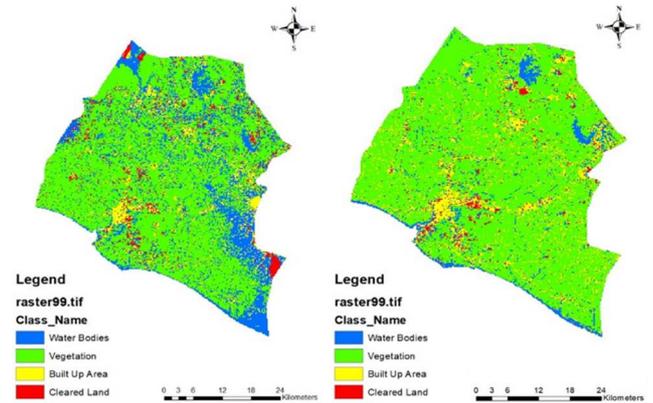
Fig. 3 Land use map in the year 1999 and 2018

However, the water bodies dropped by half from 2010 to 2018 due to the increasing water stress issues from the unsustainable water management and drought season in 2015 to 2016. The built-up area decreases from 12.74% (1999) to 12% (2010) and increases dramatically to 14.53% (2018), while the cleared land increase in the year 1999 from 3.66% to 5.62% in the year 2010 and drop to 2.37% in 2018. Both classes have fluctuated from 1999 to 2018, probably due to the urbanization mostly occurred the internal mitigation within the urban and suburban area. Besides, the strategies of the ninth Malaysia plan included thrust three: improving the standard and sustainability of quality of life.

TABLE I
RESULT OF AREA CHANGE FOR YEAR 1999, 2010 AND 2018

	1999		2010		2018	
	Area (Ha)	%	Area (Ha)	%	Area (Ha)	%
Cleared Land	7256.22	3.66	11142.81	5.62	4701.6	2.37
Built Up Area	25241.3	12.74	23765.31	12	28795.41	14.53
Vegetation	132865.52	67.04	138305.48	69.78	149997.19	75.68
Water Bodies	32829.97	16.56	24979.4	12.6	14698.8	7.42

The conversion of the land class water bodies in 1999 to 2010 to vegetation met 55.40%, which is the highest among the classes in the transition years, while the least area of percentage convert from the water bodies to the cleared land in the cleared land 4.36%. This is clearly caused by the plantation's higher water consumption, which drives to the decrease water class in the area. During the time period, conversion classes to vegetation had the highest area percentage compared to the other classes (cleared land 50.89%, built-up area 45.30%, vegetation 79.04%, and water bodies 55.40%) year 1999 to 2010. The data continue to grow to the year 2018 (cleared land 52.51%, built-up area 44.55%, vegetation 87.90%, and water bodies 47.98%), indicating that the vegetation was the most abundance class in the year LULC in the study area from the year 1999 to 2018. Furthermore, the rapid urbanization and development led to a percentage of the cleared land in the 1999 increase to 2010 then decrease in the year 2018. This can be clearly shown on

the value of transition of cleared land to cleared land in the year 1999 have 17.15% while the transition value of cleared land from 2010 to 2018 had 10.53%.

B. Landscape Metric Results

The FRAGSTATS [15] results (Table III) at the class level show the four different landscape structures and patterns with the four different classes. The class-level metrics were used to quantify the landscape pattern change as the class level was more accurate. The data reflected that the patch density (PD) of water bodies increased by 0.08 from 1999 to 2010 and decreased by 0.43 to 0.65 in 2018. This indicated that the water bodies class was the least dispersed area in 1999 and 2010 but became more dispersed from 1999 to 2018 while built-up area increased gradually and reached highest in the year 2018, showing that built-up area becomes more dispersed in the years. The trend corresponds to the study that mentioned the highest PD had the more dispersed and fragmented patches [16]. In the mean patch area (AREA_MN), vegetation is the highest area among the three years, which increase gradually from the year 1999 (654.88 ha) to 2018 (1616.21 ha). This showed that area of vegetation class expanded gradually and ranked the highest distributed among the three years in the landscape. From 1999 to 2010, the water bodies and cleared land to undergo a fragmentation process as the PD increased while AREA_MN decreased.

However, the vegetation aggregated as the PD decreased, reduced the density stand on the landscape; AREA_MN increased grew the patch of the class on the area, they become clustered together in 2018. While the water bodies from 2010 experienced patches loss in the land as the PD decrease from 1.08 to 0.65, AREA_MN dropped from 11.65 ha to 11.39 ha. Besides, vegetation class also ranked highest in the largest patch index (LPI), which show 62.60% in 1999, reduced to 68.73% in 2010 then increased back to 75.04% in 2018 among the three years. The other class type only had around 1% (0.71% water bodies, 1.33% built-up area, 0.12% cleared land in 2018; 0.82% water bodies, 1.40% built-up area, 0.53% cleared up in year 2010). For 1999, the water bodies had 6.26% in the landscape but dropped in the following years. This scenario showed that vegetation had the largest single patch, which dominates the landscape [17]. This indicated that the spatial distribution of the water bodies dropped from 2010 to 2018, probably due to the drought season in the year, which led to the shrinkage of the water bodies' size. This finding corresponded to the existing studies.

TABLE II
THE RESULT DATA FOR LANDSCAPE METRIC PARAMETER

Year	PD	LPI (%)	LSI (m/Ha)	AREA_MN (Ha)	ENN_MN (m)
2018	2.02	75.03	31.89	49.50	389.86
2010	2.72	68.72	39.98	36.72	343.2
1999	2.30	62.59	39.45	43.40	353.98

For the landscape level, the PD increased from the year 1999 to 2010 then dropped in 2018. The LPI at the landscape level grew gradually from the year 1999 to 2018. LSI value was almost the same for 1999 and 2010 but decreased to 31.8947 m/Ha in 2018. The AREA_MN and ENN_MN

accounted for dropping values from 1999 to 2010 then increased significantly in the year 2018.

C. Water Yield Results

The total amounts of water yield produced varied among the three years, which was 3709.97 mm in 1999, decrease to 2335.58 mm in the year 2010, and rose to 10988.57 mm in the year 2018. This may be because the precipitation of a certain year decreased greatly during the three years. The highest amount of water yielded was in 2018, which was near twice the year 2010. The water depth yielded showed a different range for temporal and spatial distribution with various areas of hectares. The highest and lowest water depth different in range among the three years. In 1999, the highest range between 180 mm to 222 mm covered the region in the northeast region. In 2010, the area with the highest range was between 120 mm to 145 mm, located in the northeast area. In 2018, the highest water depth ranges from 484 mm to 598 mm as the amount of precipitation in the year was the highest concentration in the north and east. In both years 2010 and 2018, a small portion of the area had a higher water depth. The water depth decreased from north-east to west and south area correspond to the monsoon season, bringing the precipitation from west and south to the northeast.

TABLE III
THE RESULT DATA FOR WATER YIELD AND LULC TYPES.

| | Land Class 2018 | | | | | | |
| | 1999 | | 2010 | | 2018 | | AET |
	mm	%	mm	%	mm	%	mm
Clear Land	126.6	3.4	131.9	5.7	260.4	2.4	128.5
Built Up	480.3	12.9	284.9	12.2	1609.8	14.7	138.9
Vegetation	2496.8	67.3	1641.4	70.2	8355.7	76	751.3
Water Bodies	606.9	16.4	277.2	11.9	762.6	6.9	127.8
Total	3710.6	100	2335.4	100	10988.5	100	1146.5

The simulated water yield and the actual evapotranspiration (AET) differ significantly from the LULC types (Table II). The highest AET was the vegetation class with 751.3 mm, while the lowest AET had both cleared land and water bodies classes. The higher the AET, the lower the water yield will be Spracklen et al. [18]. This trend showed that the built-up area had a higher water yield than vegetation but a lower yield than the other classes. However, our results indicate that the water yield from the vegetation area was far higher than the built-up area. This had met the inconsistent with the other existing research that the finding of the built-up area produced higher water yield while lower in vegetation land. As Li et al. [4] stated, the results might explain by some are a portion of the study area are sparse vegetation that enlarges evapotranspiration allow the water infiltration through the vegetation canopy, coverage and roots without capturing the moisture. Similar research Yang et al. [19] and Zhao et al. [20] showed that water yield during flooding due to urbanization increased as the vegetated area could transfer water to the soil.

Furthermore, the highest water yield is on the vegetation class type among the three years. Accordingly, the percentage of water yield for the vegetation increases gradually from 67.3% to 76.04% as the total area of the vegetation grew 67.04% from 1999 to 75.08% in 2018. However, the water yielded by

the built-up decreased 0.75% from the year 1999 and increase to 1609.82 mm in the year 2018, corresponding to the area taken by the built-up area dropped from 25241.3 ha to 23765.31 ha then increase to 28795.41ha in the three years. This trend indicates that the increase of total water yields mainly resulted from the expansion of the vegetation, while the fluctuation of total water yield could be ascribed to the variation of the build-up area. This result corresponds to the spatial area changes that show that the vegetation class increases while the water bodies decrease. This may be due to the water coefficient of the plant in a certain area and have a lower AET value.

D. Statistical Analysis Results

Statistical analysis is used for the data comparison between the LULC types and the water yield. The steps started with checking the datasets were normally distributed through a normality test. The Z-value of the skewness of area is 0.0019, and 0.0017 for the water yield data indicated that the dataset was normally distributed. One justification is gained, the one-way ANOVA test in the process to determine the significance of the value. The popular alpha used in most of the tests is 0.05 due to the difference between means. Conversely, if the p-value is smaller than 0.05, the null hypothesis cannot be concluded as the mean does not exist. The results of the three years were greater than 0.05 showed that all the values are insignificant; no posthoc test was needed.

Pearson correlation was then analyzed between the area of LULC built-up and vegetation type with the hydrological component of water yield (Fig. 4(a) and (b)). Results revealed that the built-up or urban area was stably correlated to the water yield change positively as the correlation coefficient, r^2 value is 0.9773. This shows that the higher the build-up area, the higher the water yield in the class type. In addition, the vegetation type was also positively associated with the water yield change with the r^2 value equals 0.8241. The correlation results indicate that the built-up area had a higher impact on the change of the water yield than the vegetation.

(a)

(b)
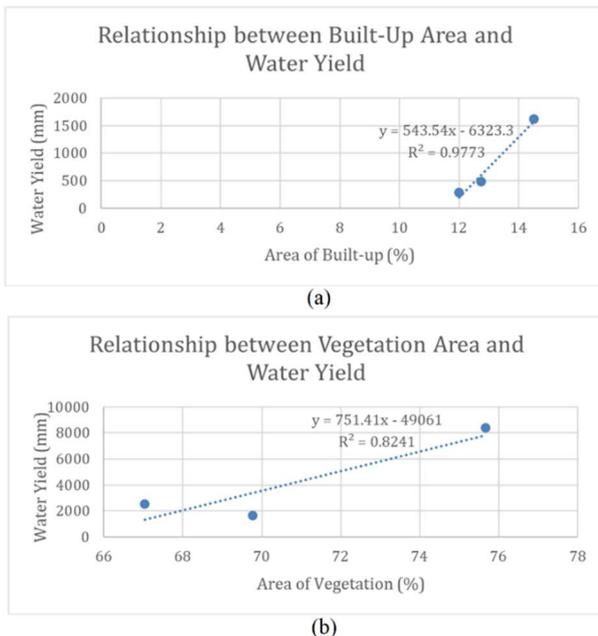
Fig. 4 The graph shows relationship between (a) built-up area and water yield (b) vegetation and water yield.

E. Discussion

Research results illustrated the LULC classification and monitoring method at each period using the satellite image Landsat with the supervised learning classification method. To acquire the data to analyze the water yield estimation to find the relationship between the water yield and use of each type of land proposed. This research provided the data used to prepare and solve the problems and prevent the possible problems about LULC monitoring and water yield, which affected the living and career of people in each area. It was contrary to the studies that proposed only the LULC change monitoring and studies that only presented the method of water yield analysis without monitoring the automatic change of land use [10] [11]. As a result, it delayed the data utilization in water yield analysis or other applications.

IV. CONCLUSION

Land use and land cover classification and monitoring using the satellite image Landsat 5 and Landsat with supervised learning classification technique fastened the monitoring regardless of the fieldwork. Furthermore, the use of data acquired from the data processing and other spatial data, such as precipitation soil-depth potential evapotranspiration and PAWC, to analyze the relationship between the water yield and the use of each type of land using InVEST Model, could predict the water yield in the future for planning the solutions and prevention of problems about water and urban planning control. Further research should apply the high-definition satellite image, such as WorldView-3 and SPOT, to classify LULC due to the high resolution of the results. Moreover, the analysis results of water yield would have a higher resolution. However, other spatial data to be used should also have a high resolution as well.

REFERENCES

[1] C. Grislain-Letrémy, and B. Villeneuve, "Natural disasters, land-use, and insurance." *Geneva Risk Ins Rev*, vol. 44, no. 1, pp. 54-86, 2019.
[2] S. Puttinaovarat, and P. Horkaew, "Flood Forecasting System Based on Integrated Big and Crowdsource Data by Using Machine Learning Techniques." *IEEE Access*, vol. 8, pp. 5885-5905, 2020.
[3] D. Dodman, "Environment and Urbanization. International Encyclopedia of Geography: People," *the Earth, Environment and Technology*, pp. 1–9, 2017.
[4] S. Li, H. Yang, M. Lacayo, J. Liu, and G. Lei, "Impacts of land-use and land-cover changes on water yield: A case study in Jing-Jin-Ji," *China. Sustainability*, vol.10, no. 4, pp. 960, 2018.
[5] Y. Zhou et al., "Simulation of the Impact of Urban Forest Scale on PM2. 5 and PM10 based on System Dynamics." *Sustainability*, vol. 11, pp. 5998, 2019.
[6] M. Camara, N. R. Jamil, and A. F. B, Abdullah, "Impact of land uses on water quality in Malaysia: a review." *Ecological Processes*, vol. 8, no. 1, pp. 1-10, 2019.
[7] S. Puttinaovarat, and P. Horkaew, "Internetworking flood disaster mitigation system based on remote sensing and mobile GIS." *Geomat Nat Haz Risk*, vol. 11, no. 1, pp. 1886-1911, 2020.
[8] A. M. Hersperger et al., "Urban land-use change: The role of strategic spatial planning." *Global Environ Chang*, vol. 51, pp. 32-42, 2018.
[9] S. Puttinaovarat, and P. Horkaew, "Multi-spectral and Topographic Fusion for Automated Road Extraction." Open Geosci, vol. 10, no. 1, pp. 461-473, 2018.

[10] G. Yin et al., "InVEST model-based estimation of water yield in North China and its sensitivities to climate variables." *Water*, vol. 12, no. 6, pp. 1692, 2020.

[11] F. Scordo et al., "Modeling water yield: Assessing the role of site and region-specific attributes in determining model performance of the InVEST seasonal water yield model," *water*, vol.10 no. 11, pp. 1496, 2018.

[12] J. R. Irons, J. L. Dwyer, and J. A. Barsi, "The next Landsat satellite: The Landsat data continuity mission," *Remote Sens Environ*, vol.122, pp. 11-21, 2012.

[13] J. Sinha et al., "Assessment of the impacts of climatic variability and anthropogenic stress on hydrologic resilience to warming shifts in Peninsular India." *Sci Rep-Uk*, vol. 8, no. 1, pp. 1-14, 2018.

[14] H. Bandi Hermawan, I. Agustian, and B. G. Murcitro, "A Model to Predict Plant-available Water Content of Soils at Different Land Units in Bengkulu, Indonesia." *Terra*, vol. 3, no. 1, pp. 10-14, 2020.

[15] S. Lamine et al., "Quantifying land use/land cover spatio-temporal landscape pattern dynamics from Hyperion using SVMs classifier and FRAGSTATS®." *Geocarto Int*, vol. 33, no. 8, pp. 862-878, 2018.

[16] L. Hou, F. Wu, and X. Xie, "The spatial characteristics and relationships between landscape pattern and ecosystem service value along an urban-rural gradient in Xi'an city, China," *Ecol Indic*, vol. 108, pp. 105720, 2020.

[17] J. Yang, L. Shimei, and L. Huicui, "Quantitative influence of land-use changes and urban expansion intensity on landscape pattern in Qingdao, China: Implications for urban sustainability." *Sustainability*, vol. 11, pp. 6174, 2019.

[18] D. V. Spracklen et al., "The effects of tropical vegetation on rainfall," Annu Rev Env Resour, vol.43, pp. 193-218, 2018.

[19] C. Yang et al., "Effects of Vegetation Cover on Hydrological Processes in a Large Region: Huaihe River Basin, China," *J Hydrol Eng*, vol. 18 no. 11, pp. 1477-1483, 2011.

[20] G. Zhao et al., "Evidence and causes of spatiotemporal changes in runoff and sediment yield on the Chinese Loess Plateau," *Land Degrad Dev*, vol. 28 no. 2, pp. 579-590, 2017.