TABLE OF CONTENT

Research Article

Assessment of the impact of climate change on the water resources in Phu Yen province, Vietnam

Nguyen Van Hong[1]*

[1] Sub Institute of Hydro–Meteorology and Climate Change;
nguyenvanhong79@gmail.com

*Corresponding author: nguyenvanhong79@gmail.com; Tel.: +84–913613206

Abstract: This study investigates the impact of climate change (CC) on water resources in Phu Yen province, Vietnam by assessing historical trends of temperature, rainfall and flood simulation scenarios from a report of Vietnam of the Ministry of Natural Resources and Environment in 2020. The method for assessing the impacts on water resources was provided in the report of the Intergovernmental Panel on Climate Change (IPCC) in 2007. Several areas in Phu Yen province could be high risk of flooding in the middle of the century, such as Son Hoa, Song Cau, Tuy An, Dong Hoa and Tuy Hoa districts, where were chosen as study areas. The analytical results indicate that the impacts of climate change significantly influence to inundation on water resources in these localities. These results could help to partially predict impacts of climate change in the future and provide useful information to plan adaptation strategies in water utilization in Phu Yen province for the policy making on sustainable management with future climatic condition.

Keywords: Impact of climate change; Water resources.

1. Introduction

Climate change has been considered as the greatest challenge of the world. It describes a change in the average climate conditions and extreme events are likely to have major impact on human and ecological systems. The water resources are amongst those severely expected affected by the changing climate [1]. In particular, the hydrological variables in terms of precipitation and temperature affect the intensity, frequency and timing of specific rainfall and potential evaporation, river runoff, surface water, groundwater, flooding, drought, saline intrusion, high tide and water demand [2–9]. The number of reports addressing both climate change and water resources [10–11]. In 2006, [12] reported that approximately 94 million people in the world will be strongly affected by, and sensitive to, variability in climate conditions with an increase around 40 cm of sea level. Among them, about 20% population is in the Southeast Asia area, particularly in the Mekong River Delta (Mekong Delta) and the Red River Delta regions in Vietnam. [13] reported that the sea level in Vietnam is rising up in the range of 1.75–2.56 mm every year. Vietnam is in the group of 5 countries most affected by the climate change, of which the Mekong Delta and the Red River Delta are the two most affected regions in Vietnam – once again the level of danger of climate change to Vietnam is affirmed. In addition, the lower Mekong Basin (Vietnam) is considered as one of three deltas in the extremely vulnerable to climate change beside the Ganges–Brahmaputra (Bangladesh) and Nile (Egypt) [15]. Several topics and projects on the assessment of climate variability relating to social economy in Vietnam were studied by many research groups in Vietnam.

For example, [16] investigated the change in nature, people and socio–economy in Ho Chi Minh City linked the variability of climate condition. In addition, [17] developed a guidance: "Assessment of the impact of climate change and identification of adaptation solutions" funded by the United Nations Development Program (UNDP).

In 2015, [18] developed the Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX). [19] investigated the impact of saltwater intrusion and socio–economic factors on agricultural sector in Vung Liem district, Vinh Long province. In particular, in assessing the impact of climate change on water resources (WR), there is a project implemented by the Institute of Hydrometeorology and Environment (2010) in the Mekong Delta sponsored by the Danish Embassy in Vietnam. The vulnerability of water resources also depends on the change in intensity and frequency of extreme hazards, which have been calculated by numerical models such as: climate simulation from the AGCM/MRI (models of the Institute for Meteorology Japan), the PRECIS model of the UK Hadley Center and the CCAM model of the Commonwealth Scientific and Industry Research Organization, Australia (CSIRO), the hydraulic model simulation and the GIS modeling methods [20–21].

Phu Yen is located in the Central of Vietnam, as one of the most beautiful coastal provinces in Vietnam, with coastal islands and a complicated coastline. With the advantage of being a gateway to the sea of the Central Highlands economic region, Phu Yen province has had a strong growth rate, and marine tourism has contributed as a main sector to attract many domestic and foreign investors. So far, the increase in human activities during socio–economic development, and activities of over exploiting GHG sinks in forests and ecosystems enhance greenhouse gases emission, which could enhance the worst influence of climate change in the local area. According to the scenario RCP4.5 by the mid–century, the increase in mean surface temperature every year in Phu Yen province is around 1.5°C [1]. As for rainfall, by mid–century, the sea level is commonly varying from 8.5–17.4% under the RCP4.5 scenarios. It has been reported that the estimated sea level increasing in Phu Yen province in the mid–century (2050) is around 22.5 cm. Consequently, it is forecasted that the risk of flooding also increases in coastal areas, and along the downstream rivers at low–lying terrain exhibit an increasing the flooding rate varies from 5.55% (in 2021) to 6.29% (RCP4.5 scenarios by mid–century), which is much higher than other areas in the whole province [22]. It is notable that the impact of climate change on the rainfall, flooding, sea level rising, and water resources in Phu Yen province could be extremely unpredictable in the future. It could be a scarcity of water resources in the dry season and flood in the rainy season, leading to the difficulties in water supply and demand in Phu Yen. Therefore, it is needed to assess the impact of climate change on the water resources in the province.

Thus, to assess the impact of climate change on water resources in Phu Yen province, The purpose of this study is used the results of data on climate change scenarios (temperature, rainfall, drought, sea level rise [2]) by using modeling method was mainly used in computing and assessing flood inundation impacts on water resources. The research result provides scientific foundation for strategic solutions and improving action plans to respond to climate change, contributing to ensure sustainable development goals in Phu Yen province.

2. Materials and Methods

2.1. Study area

Phu Yen province is located in the South–Central Coast of Vietnam, with a drainage area of 506,057 hectares which is 1.53% of the total area of Vietnam, as shown in Figure 1. The topography of the province is quite diverse, consisting of 70% high region in the west and 30% of the fertile plain in the east such as coastal plains, low hills, coastal gentle hills, highlands and valleys [22]. Currently, water resources in Phu Yen area are come from (i)

rainfall in last 3 or 4 months of the year, with the annual precipitation of 1,700–2,000 mm, and (ii) surface water sources from 17 rivers with catchment areas in range of 100−500 km^2. Among the 3 major rivers including Ky Lo, Ba and Ban Thach River, Ba River is the largest river in the Central area, with a basin area of 13,500 km^2.
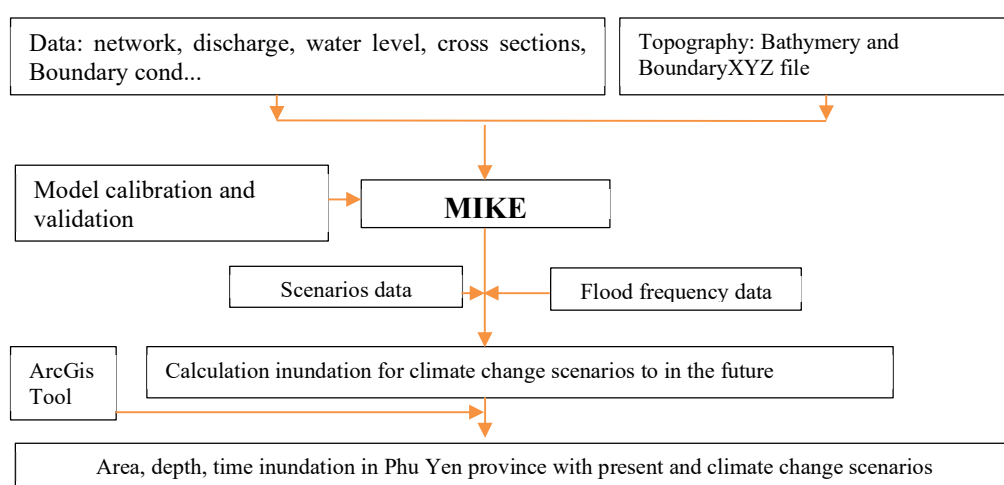
PHU YEN PROVINCE MAP

Figure 1. Map of Phu Yen province.

Although the rainfall distribution over the year is significantly heterogeneous, the rivers are still the main water resource for agriculture and water supply. In particular, 65–75% of total annual rainfall is distributed from October to December in the wet season (or flood season), while January to September of the next year is the dry season, with around 25–35% of annual rainfall, as measured by volume. Additionally, there are two dry periods in April and August with only 2% of annual rainfall. Moreover, it is influenced by diurnal tide regime from 17 to 23 days every month and also by irregular semi–diurnal tide on the remaining days. Since the heavy floods at the downstream of Ba River and the tides of the East Sea are often occurred at the same time of the year, it may make an extreme danger in the downstream area during flood season, especially in the Tuy Hoa rice cultivation area of Dong Cam irrigation system. Therefore, it is necessary to have solutions for water drainage in the downstream areas and especially in the rice cultivation areas and Tuy Hoa city.

2.2. Modeling setup

Figure 2. Oriented framework for flood calculation in Phu Yen province.

The modeling method will be mainly used in calculating and assessing flood impacts on water resources. MIKE of DHI software for the hydraulic calculation of the study area are applied in the study and the used modules consist of MIKE 11HD, MIKE 21FM and MIKE FLOOD. Models are calculated in accordance with the orientation frame, as shown in Figure 2.

2.3. Methods of assessing climate change impacts on water resources

Water resources is a sensitive sector for climate factors such as temperature, rainfall, sea level rise and the increase of the intensity and frequency of extreme climate events. Therefore, climate change has a great impact on water resources (Surface runoff water, reserves of water sources, water quality and quantity). After calculating and analyzing flood inundation levels and affected areas, the analysis of the challenges of water resources in the context of climate change is used in the methods at in the Table 1.

Table 1. Methods of assessing climate change impacts on water resources [17].

Elements of climate	Influenced objects	Risk effects	Evaluation methods and tools
Increased temperature	Surface runoff	Change the operational intensity of atmospheric circulation, water cycle, hydrological regime and other physical cycles	Mathematical model for the correlation between precipitation and temperature and runoff for the basin
Increased Precipitation	Volume of water sources	Increase water volume	Hydrological model

Elements of climate	Influenced objects	Risk effects	Evaluation methods and tools
	Water quality	Water pollution can be spread due to excessive rain causing inundation	Development of flooding maps
Sea level rise	Water source	Increase risk of inundation and soil erosion; change the flow regime in rivers and groundwater; change geomorphology in estuaries.	Models for predicting changes in surface runoff and groundwater Flood mapping method
	Water quality	The level of water pollution increases due to widespread and prolonged inundation	Flood mapping method Vulnerability and impact mapping method
Increase of the intensity and frequency of extreme climate events	Water sources	Inundation increases in some areas	Map overlapping method Vulnerability and impact mapping method

2.4. Spatial interpolation methods

Inverse Distance to a Power and Surfer 10.0 software are utilized to calculated the exposure possibilities of different areas within study scope. The distance inverse method determines values for grid cells by averaging the values of sample points in the vicinity of each grid cell. The closer the point is to the center point (being determined), and the more influence it has.

Interpolation:

$$\hat{Z}(S_0) = \sum_{i=1}^{n} \lambda_i Z(S_i); \ \lambda_i = \frac{d_{ij}^{-p}}{\sum_{i=1}^{n} d_{ij}^{-p}} \tag{1}$$

where d_{ij} is the spatial distance between the two points i and j. The higher the exponent p is, the lower the influence of the distant points is and some consider insignificant, normally $p = 2$. However, in some cases the input data density is high, p must be correspondingly high. In this study, it is expected to apply $p = 5$.

Search Radius: The characteristic of the interpolated surface is also influenced by the search radius that limits the number of sample points used to calculate the interpolated grid cells.

Barrier: A barrier is a set of polylines as an interruption limiting the area to seek sample points. A polyline can be a cliff, a mountain, or some other barriers in the landscape. When this element appears, only the sample points on the same side and the grid under survey are considered. In the case of interpolation for inundation and saline intrusion, the barrier is the boundary of the rivers.

2.5. Mapping method

Techniques of information integration were used for mapping, overlapping layers of weighted information, synthesizing, calculating for main and sub–indexes through ArcGIS 10.1 software on the map, as coordinate system VN_2000.

2.6. Data

In this study, data used in the method for collecting, synthesizing, inheriting are documents of natural, socio–economic and environmental conditions; data on climate change scenarios (temperature, rainfall, drought) (as shown in figures 3–5), topographical data including 119 of cross – sections, the computation mesh of the main river bed, more than 20 river banks of Ba and Ky Lo river systems and sea level rise in Phu Yen province to develop the inundation maps from which to assess the impacts of climate change on the water

resources in Phu Yen province. 270 survey questionnaires of people in 8 cities, towns and districts of Phu Yen province are used to assess which floods have serious impacts in the locality. Inundation and data of climate change and sea level rise scenario are built according to the average emission scenario RCP4.5 in 2030 and 2050 [22].
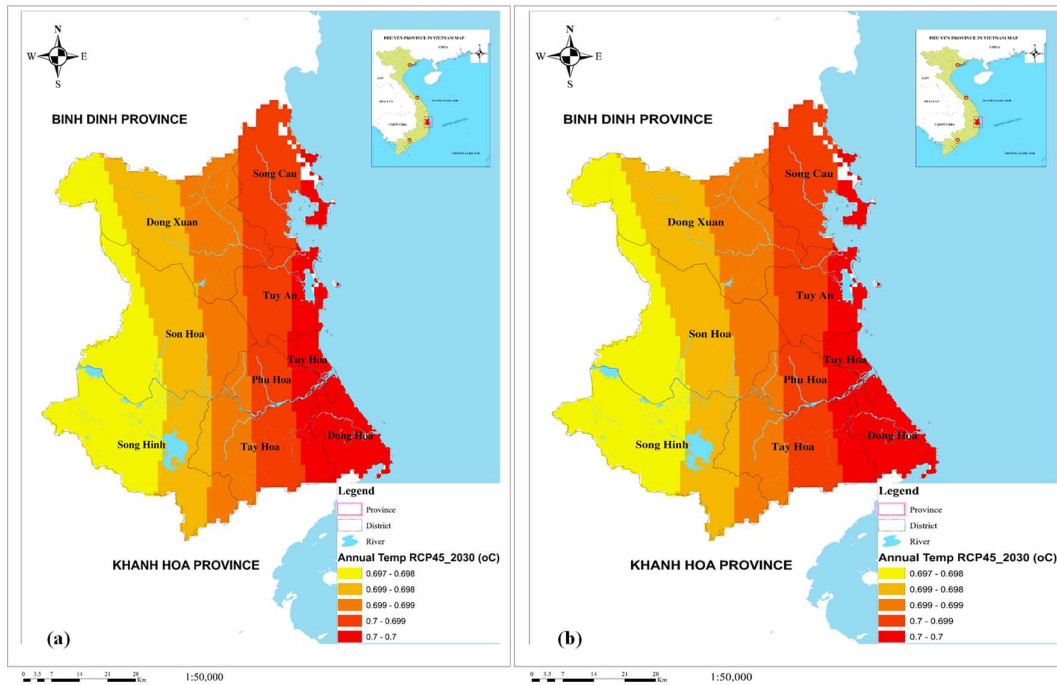
Figure 3. Scenario of annual average temperature change in Phu Yen province: (a) RCP4.5 năm 2030; (b) RCP4.5 năm 2050.
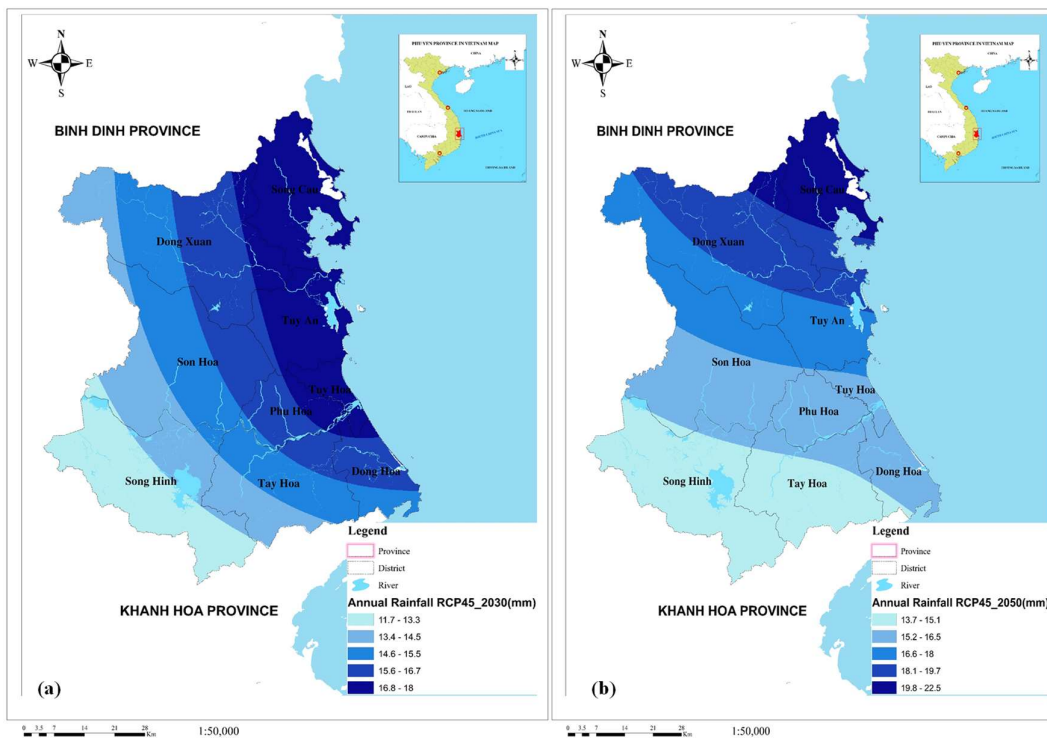
Figure 4. Scenario of annual rainfall change in Phu Yen province: (a) RCP4.5 in 2030; (b) RCP4.5 in 2050.
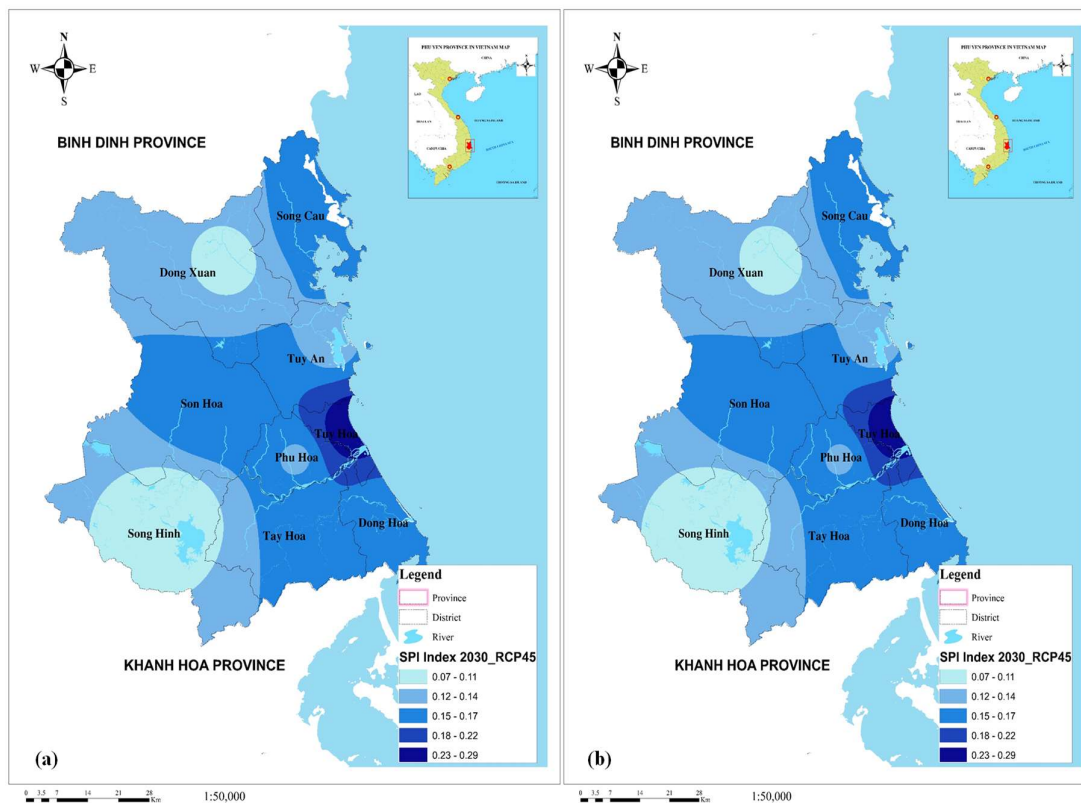
Figure 5. Scenario of changes in the yearly standardized precipitation index (SPI) in Phu Yen province: (a) RCP4.5 in 2030; (b) RCP4.5 in 2050. (Note: SPI Index range: ≥ 2: extremely wet; 1.5 ÷ 1.99: very wet; 1.0 ÷1.49: moderately wet; -0.99 ÷ 0.99: near normal; -1.0 ÷ -1.49: moderately dry; -1.50 ÷ -1.99: severely dry; ≤ -2.0 extremely dry).

3. Results and Discussion

3.1. Model calibration and validationof MIKE 11HD model

Calibration and validation of MIKE 11HD model used the past floods to find the suitable parameters for the simulation scenarios. The calibration and validation process used an observed water level at Phu Lam station during from Feb to July, 2016 and from Feb to July, 2013, respectively. In this study, Nash–Sutcliffe efficient (NSE) and Coefficient of determination (R^2) were used to evaluate the observed and simulated water levels. The results of the calculated and observed water levels are in good agreement in terms of the vibration amplitude, absolute value, and phases during both the calibration and validation processes. The values R^2 ranged from 0.89 to 0.92 for calibration and validation, respectively. It indicates that model results produced for the flow are very good for both periods. The NSE value for calibration and validation of water level at Phu Lam ranged from 0.79–0.80. The simulation results for the hydrodynamic regime using MIKE11HD model were very good in term of performance ratings as revealed by NSE and R^2. The calibration parameters used in the model validation process were bed resistance Manning coefficients which varied from 0.03 to 0.033 ($m^{1/3}$/s). The calibration characteristic parameters such as Manning coefficients (M), time, and time–step is performed by means of gradually.
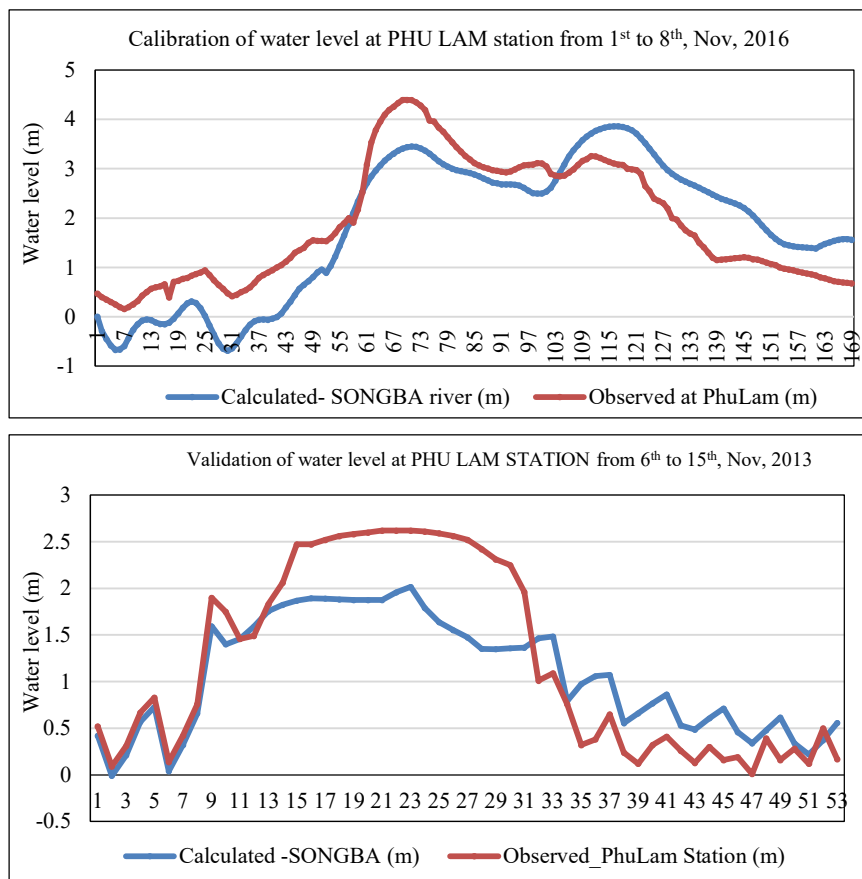
Figure 6. The calibration and validation process used an observed water level at Phu Lam station.

3.2. Model calibration and validation of MIKE FLOOD model

Calibration of MIKE FLOOD model used an observed floods in downstream at Cung Son site, Ba River and Ha Bang site, Ky Lo river. The values R^2 ranged from 0.92 to 0.93 for calibration, respectively (Cung Son site ($R^2 = 0.92$) and Ha Bang site ($R^2 = 0.93$)). The NSE value for calibration of water level at these sites ranged from 0.63–0.71 (Cung Son (NSE = 0.71) and Ha Bang (NSE = 0.32)), from 01st to 06th, December, 2016.
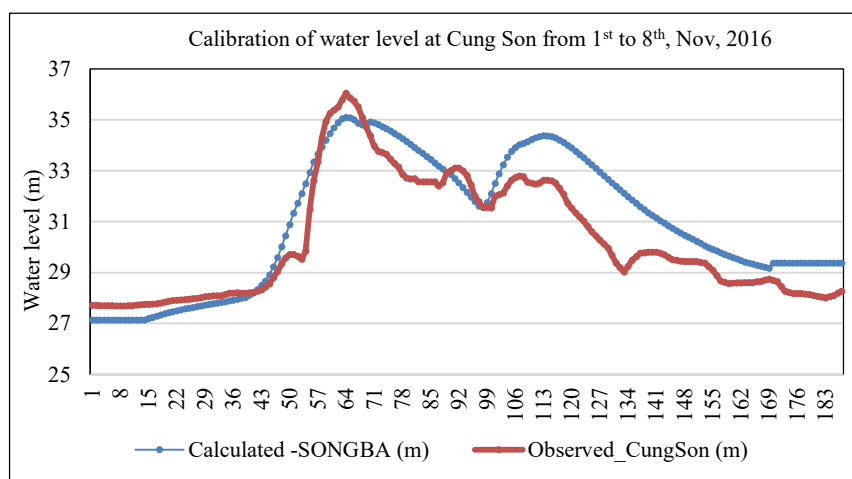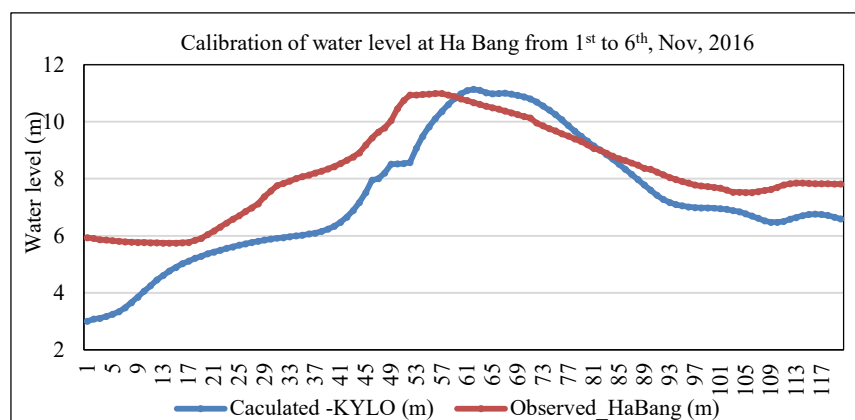
Figure 7. The calibration process used an observed water level at Cung Son and Ha Bang site from 01st to 06th, December, 2016.

Validation of MIKE FLOOD model used an observed floods in downstream at Cung Son site, Ba River and Ha Bang site, Ky Lo river. The values R^2 ranged from 0.92 to 0.96 for validation, respectively (Cung Son site (R^2 = 0.92) and Ha Bang site (R^2 = 0.96)). The NSE value for validation of water level at these sites ranged from 0.72–0.75 (Cung Son (NSE = 0.75) and Ha Bang (NSE = 0.72)), from 04th to 08th, December, 2013.

The validation model results in the flood from 04th to 08th, December, 2013 showed that the MIKE FLOOD model was well–simulated between observed flood and calculated water level at sites. It could be concluded that the MIKE FLOOD hydraulic model was reliable enough to simulate and calculate the flood scenarios and establishing inundation maps.
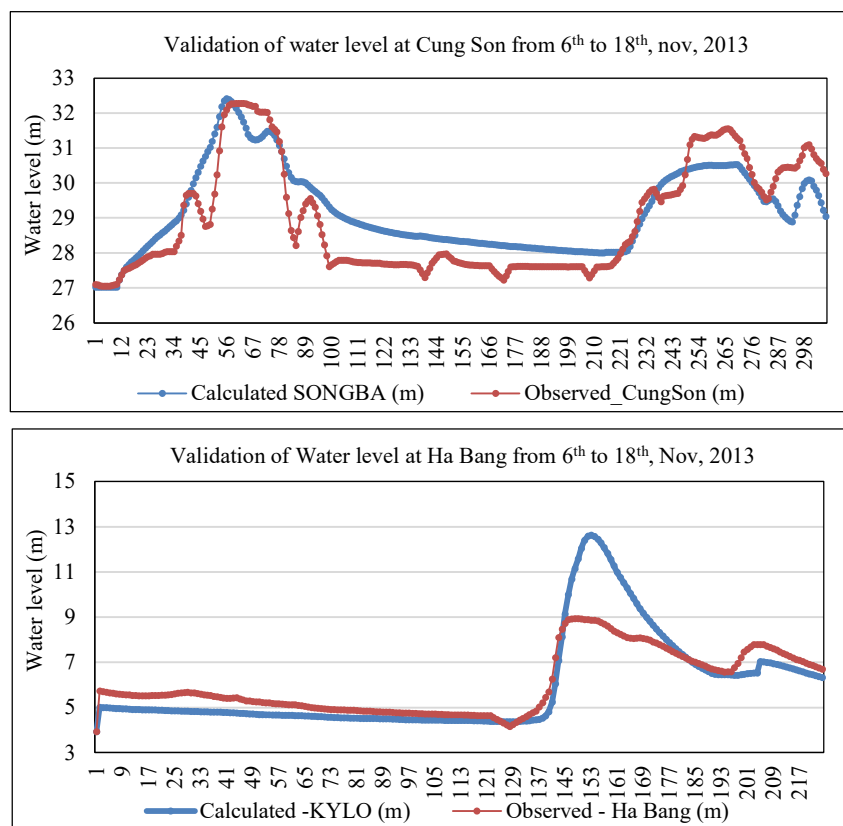
Figure 8. The validation process used an observed water level at Cung Son and Ha Bang site from 04th to 08th, December, 2013.

3.3. Computating flood inundation results

The results of flood inundation in the area are evidenced by the simulation time shown in figure and application of ArcGIS 10.1 software and data layers including flooded layer depths simulated by MIKE FLOOD model, terrain data, administrative boundary data, flood survey data, situational data and so on to establish the inundation map according to the scenarios. The inundation map that illustrates flooding inundation according to the scenario given is shown in Figure 9. Because the territory stretches along the coastline and the river system flows in the West–East direction. The riverbed is steep and narrow at the upstream and wider at the downstream toward the sea. Therefore, the lower coastal area in Phu Yen province is heavily affected by floods. The scenario of climate change (rainfall changes with the seasons), sea level rise is used calculated as the RCP4.5 with scenarios corresponding to the periods 2030, 2050. The results of the calculation flood according to the scenarios represented in Figure 9. The calculated flood inundation results of the scenarios have a relative increase compared to the current flood status, by the rainfall volume in the upstream, sea level rise in downstream of the rivers. The higher the sea level rise scenarios, the greater the flood inundation level and the area of flooding. The areas affected by flood inundation are mainly in Tuy Hoa City, Tuy An District, Cau River Town, Hinh River District and Dong Xuan District.

- Regarding the flood level: the lowest is 0.1 m, the highest is 4.0 m and the maximum flood level is in range of 1.0–3.0 m. Specifically, the flooded area of the whole Phu Yen province with the flooding level of 1.0–2.0 m could be 6,560 and 6,624 ha as predicted in 2030 and 2050, respectively. If the flooding level is from 2.0–3.0 m, the flooded area of the whole province might become to be 7,692 ha (in 2030) and 7,700 ha (in 2050), as shown in Table 2.

- Regarding the flooding scope and flooding rate: the two districts Dong Hoa and Phu Hoa could be largest flooded areas, and highest the flood rates with 11.05% and 13.18% compared to the total natural area of the two districts, respectively (Table 3). On the other hand, the districts Son Hoa, Song Hinh, and Song Cau town have the smaller flooded areas.

Table 2. The flood rea corresponds to flooding levels in Phu Yen province under the RCP4.5 scenario (Unit: hectare).

Time	2030					2050				
Flood (m)	0.1 – 0.5	0.5 – 1	1 – 2	2 – 3	3–4	0.1 – 0.5	0.5 – 1	1 – 2	2 – 3	3–4
City/ District										
Tuy Hoa City	28	84	604	1084	2468	28	80	676	1024	2528
Song Cau Town	76	212	284	48	88	76	212	284	48	88
Dong Xuan District	12	64	260	372	884	12	64	248	376	892
Dong Hoa District	76	464	1580	2428	6428	76	468	1588	2472	6476
Phu Hoa District	48	600	2288	2352	1600	44	560	2224	2368	1696
Son Hoa District	36	12	112	116	388	36	0	112	112	404
Song Hinh District	92	68	152	120	512	92	44	164	108	536
Tay Hoa District	420	564	708	428	424	720	576	784	476	464
Tuy An District	480	460	572	744	1864	480	460	544	716	1920
Province	1,268	2,528	6,560	7,692	14,656	1,564	2,464	6,624	7,700	15,004

Table 3. Flooded area and flooding rate in Phu Yen province under the RCP4.5 scenario (Unit: hectare).

City, district, town	Natural area	Flood area (ha)		The rate of flooded area over the natural area (%)	
		2030	2050	2030	2050
Tuy Hoa City	110,336.04	4,268	4,336	3.87	3.93
Song Cau Town	28,302.83	708	708	2.50	2.50

City, district, town	Natural area	Flood area (ha)		The rate of flooded area over the natural area (%)	
		2030	**2050**	**2030**	**2050**
Dong Xuan District	26,981.64	1,592	1,592	5.90	5.90
Dong Hoa District	99,289.37	10,976	11,080	11.05	11.16
Phu Hoa District	52,292.90	6,888	6,892	13.17	13.18
Son Hoa District	94,737.44	664	664	0.70	0.70
Song Hinh District	66,611.95	944	944	1.42	1.42
Tay Hoa District	43,283.54	2,544	3,020	5.88	6.98
Tuy An District	11,843.62	4,120	4,120	34.79	34.79
Province	**533,679.32**[a]	32,704[a]	33,356[a]	6.13[b]	6.25[b]

Note: (a) = Total area,   (b)= Rate of flooded area over the province
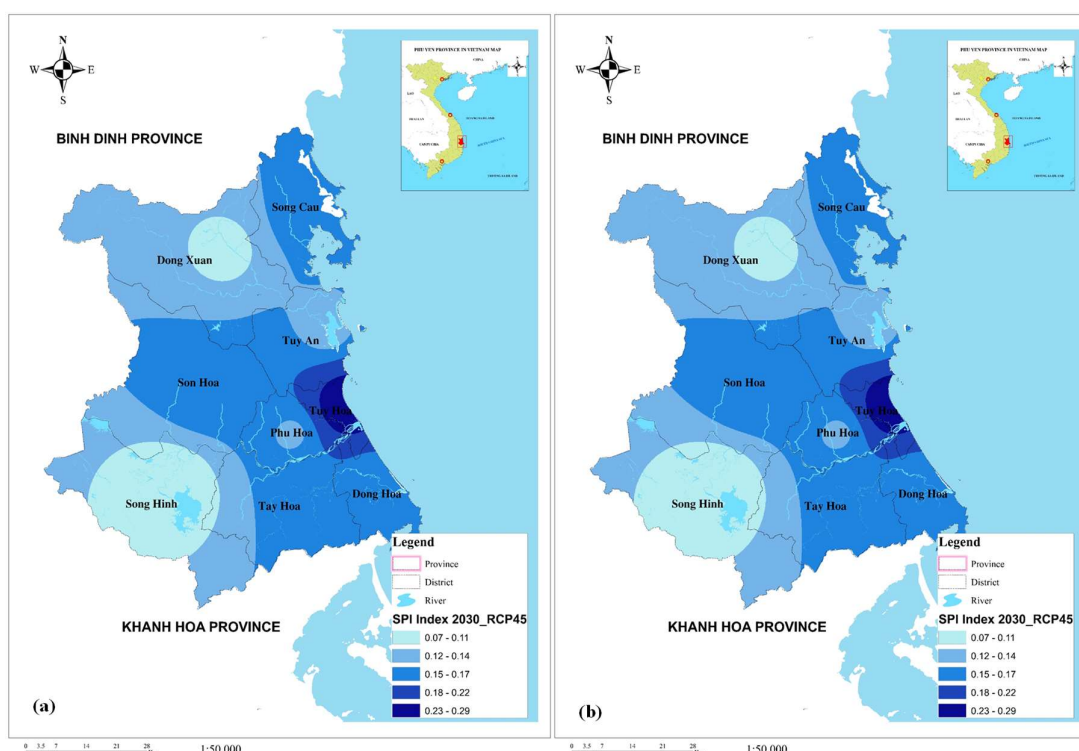
Figure 9. Map of climate change impact indicators in Phu Yen province according to the RCP4.5 scenario: (a) RCP4.5 in 2030; (b) RCP4.5 in 2050.

Regarding the popular flooding depth (0.1–0.5 m) and serious flooding depth (> 2.0 m): It can be observed that the inundation rate in 2030 and 2050 are 3.88 and 4.69 %, respectively, in total flooded area of the whole province, as presented in Table 5.

In the range of popular flooding depth: Tuy An and Tay Hoa district reveal highly proportion of flooded areas with 1.47, 1.28 % (2030) and 1.44, 2.16% (2050), as calculated in the total flooded area of the province, respectively.

In the range of serious flooding depth: Tuy Hoa city, Dong Xuan district and Dong Hoa are localities with higher flooding rates compared to other urban districts in the same depth range. In the years of 2030 and 2050, the above localities have the rate of flooded area over 78% in comparison with the total flooded area of the whole province. In addition, Son Hoa district has flooded areas in 2030 and 2050 accounting for 75.90% and 77.71%, respectively.

Table 4. Flooding rates in Phu Yen province classified by popular and severe flooding depth, RCP4.5 scenario (Unit: %).

City/ District	Popular Flooding Depth (0.1–0.5 m)		Serious Flooding Depth (>2.0 m)	
	2030	**2050**	**2030**	**2050**
Tuy Hoa City	0.09	0.08	81.92	83.22
Song Cau Town	0.23	0.23	19.21	19.21
Dong Xuan District	0.04	0.04	78.89	79.65
Dong Hoa District	0.23	0.23	80.69	80.76
Phu Hoa District	0.15	0.13	57.38	58.97
Son Hoa District	0.11	0.11	75.90	77.71
Song Hinh District	0.28	0.28	66.95	68.22
Tay Hoa District	1.28	2.16	33.49	31.13
Tuy An District	1.47	1.44	63.30	63.98
Province	**3.88**	**4.69**	**68.33**	**68.07**

Water resources are an important factor in the socio–economic development of each locality, but urbanization and population growth have led to a series of consequences such as water pollution, decrease in water quantity. In addition, climate change factors also contribute to changes in the quality and quantity of water resources. Climate change impacts the water environment through the following aspects:

Rainfall changes, increases in the rainy season and decreases in the dry season, along with an increase in temperature leads to a change in the evaporation factor, which is potential to change the flow of rivers as well as underground water flows. Flood and drought have serious impacts on water environment through both water reserve and quality.

Due to the high risk of flooding in 2030 and 2050, the water resources management in Phu Yen province should be careful considered. The high level of inundation will lead to riverside areas in the Ba River basin from Ha Bang station (Dong Xuan district) to Phu Lam station (Tuy Hoa city) being heavily flooded, with the popular flooding levels in range of 0.6–1.15 m.

- Dams on the river such as Lo Gom, Thach Khe, Dong Kho (Song Cau town) is in the flooded area from 0.65 to 1.25 m. In addition, dams namely Dong Lau, Bau Da, Tan Giang Huong, Dinh Ba are also affected in the popular flooding level of 0.15 to 0.25 m.

- The dyke sections of Xuan Hai, Xuan Loc, Xuan Canh (in Song Cau town), Binh Thanh dyke, An Hiep dyke – Phu My and Ngo Ham Thuy dyke (in Tuy An district) are also in the flooded area due to sea level rise, increasing the risk of bank erosion by the dyke in 2030 and 2050 if there is no response approach for climate change.

In addition to analyzing flooding levels and affected areas, the analysis of the strengths, weaknesses, opportunities and challenges of water resources in the context of climate change is the basis for planning adaptation solutions, integrating into water exploitation and use planning in Phu Yen (Table 5).

Table 5. Analysis on the pros and cons, opportunities and challenges of water resources in the context of climate change in Phu Yen province.

Pros	Cons
Hydropower works and irrigation lakes are capable of storing a huge amount of water, supplementing surface water for many areas Ba River – Ky Lo River flow from the Central Highlands, bringing a relatively large amount of fresh water to the locality.	Irrigation infrastructure is incomplete, many areas are in lack of water in the dry season, especially the mountainous areas affected by salt water intrusion and drought.

Pros	Cons
Opportunities	**Challenges**
Heavy rain is an important source of replenishment of surface and groundwater, especially in the rainy season. Irrigation systems are being built to meet the demand of water for production in many different areas.	The dry season is getting more and more severe, leading to a local drought. The water level of the lakes is low in several times, reducing the water supply capacity for production and salinity downstream areas. Flooding caused by tides affects surface water quality, especially in coastal areas such as Tuy Hoa, Song Cau, ... Heavy rains change water quality, water sources containing a lot of pollutants and suspended solid.

4. Conclusion

MIKE software (MIKE11, MIKE 21 and MIKE FLOOD) have been used in this study with rainfall and air temperature data under climate change scenarios. The result shows that: (i) water resources distribution in Phu Yen provice will be potentially affected by the flood due to the sea level rise in some concerned areas including districts with major rivers flowing through and high risk of inundation in the middle of the century such as Son Hoa, Song Cau, Tuy An, Dong Hoa and Tuy Hoa districts; (ii) inundation due to floods will be a source of significant amounts of surface water and groundwater to reduce water pressure during dry times; (iii) inundation caused by the sea level rise will increase the spread of salinity in the surface water, potentially leading to the deterioration of groundwater quality in the coastal areas of the province such as Song Cau, Tuy An, Tuy Hoa and Dong Hoa, etc.

MIKE software is a very useful tool for flood inundation simulation in Phu Yen provice even lack of investigated data on flood traces and MIKE 21 model has not been verified and calibrated. It can be applied effectively to assess impact of climate change on the water resources in other provinces / cities / river basins in Vietnam if measured data on topography, water level and flood traces with high reliability are available.

Author Commitment Statement: The paper submitted with the full knowledge and consent of the author (if any), without any prior publishment or copy from other previous studies; There is no dispute of interest in the authors group.

References
1. IPCC. The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 2007.
2. Ministry of Natural Resources and Environment, Climate change and sea level rise scenarios for Vietnam. Vietnam Publishing House of Natural Resources, Environment and Cartography, 2020.
3. Tan, P.V.; Thanh, N.D. Climate change in Vietnam: Some research results, challenges and opportunities in international integration. *VNU J. Sci.: Earth Environ. Sci.* **2013**, *29(2)*, 42–55.
4. Hong, N.V., Dong, N.P. Simulation of saline intrusion in main rivers of Ba Ria – Vung Tau province under the context of climate change. *VN J. Hydrometeorol.* **2021**, *728*, 67–79.
5. Hong, N.V.; Dong, N.P. Studying on building the flood scenarios in ho chi minh city by the impacts of climate change. *VN J. Hydrometeorol.* **2021**, *729*, 1–13.
6. Hong, N.V.; Nguyen, V.T. The impact of Climate Change on the transportation in Binh Thuan Province. *VN J. Hydrometeorol.* **2021**, *8*, 9–15.

7. Hong, N.V.; Hien, N.T.; Minh, N.T.T.; Toan, H.C. Forecasting saline intrusion under the influence of the northeast monsoon in the Mekong Delta. *VN J. Hydrometeorol.* **2021**, *9*, 23–36.

8. Hong, N.V.; Dong, N.P. Research on assessing the impact of saline intrusion on the water resources in Ho Chi Minh City in the context of climate change. *VN J. Hydrometeorol.* **2021**, *10*, 11–23.

9. Tri, D.Q. Application MIKE 11 model on simulation and calculation for saltwater intrusion in Southern region. *VN J. Hydrometeorol.* **2016**, *671*, 39–46.

10. World Bank. The Impact of Sea Level Rise on Developing Countries: A Comparative Analysis. World Bank Policy Research Working Paper 4136. 2007. Available online: http://go.worldbank.org/775APZH5K0 (accessed on 15 March 2017).

11. MONRE (Ministry of Natural Resources and Environment of Vietnam). Vietnam special report on managing the risks of extreme events and disasters to advance climate change adapatation. Vietnam publishing house of Natural Resources, Environment and cartography, 2015.

12. Nicholls, R.J.; Lowe J.A. Climate Stabilization and Impacts of Sea–Level Rise. Avoiding Dangerous Climate Change. Cambridge University Press, ISBN: 13 978–0–521–86471–8. 2006.

13. Hanh, P.T.T. and Furukawa, M. Impact of sea level rise on coastal zone of Viet Nam. Bulletin of the College of Science, University of the Ryukyus, ISSN: 0286–9640. 2007.

14. Dasgupta, S.; Laplante, B.; Meisner, C.; Wheeler, D.; Yan, J. The Impact of Sea Level Rise on Developing Countries. *Clim. Change* **2009**, *93*, 379–388.

15. IPCC. Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, Eds., Cambridge University Press, Cambridge, UK, 2007, pp. 976.

16. Phung, N.K. Assessment of the climate change impact on nature, people, and socio–economy in Ho Chi Minh City. 2011.

17. Vietnam Institute of Meteorology, Hydrology and Climate Change. Guidance for Assessment of Climate Change Impact and Identification of Adaptation Options. Vietnam Map and Natural Resources and Environment Publishing House. 2011.

18. Thuc, T.; Huong, T.T.T.; Thang, N.V.; Nhuan, M.T.; Tri, L.Q.; Thanh, L.D.; Huong, H.T.L.; Son, V.T.; Thuan, N.T.H.; Tuong, L.N. Special Report of Vietnam on Managing the Risks of Natural Disaster and Extreme Phenomena to Promote Climate Change Adaptation. Vietnam. Vietnam Publishing House of Natural Resources, Environment and Cartography. Hanoi, Vietnam, 2015.

19. Nguyen, Q.H.; Cao, T.Q., Vo, T.P.; Le, V.K.; Vo, Q.M. Evaluation on the effects of saline intrusion and socio–economic factors on agricultural production in Vung Liem district, Vinh Long province. *Can Tho Univ. J. Sci.: Environ. Clim. Change* **2017**, *1*, 64–70.

20. Ministry of Natural Resources and Environment, Updated Nationally determined Contribution (NDC) for Vietnam. Vietnam Publishing House of Natural Resources, Environment and Cartography, 2020.

21. Institute of Meteorology, Hydrology and Climate Change. The Impacts of Climate Change on Water Resources and the Mekong Delta Adaptation Measures. 2010.

22. The People's Committee of Phu Yen province. Project: Developing and updating the Action Plan in response to Climate Change in the period 2021–2030, vision to 2050 of Phu Yen province, 2019.

Check for updates

Research Article

Application of self-organizing maps and K–Means methods to classify summer heat wave weather patterns in Viet Nam

Mai Tran Thi Tuyet[1], Hoa Van Vo[1*], Tuan Le Danh[2]

[1] Regional Hydro–Meteorological Center of Red delta river basin;
tuyetmai1295@gmail.com; vovanhoa80@yahoo.com
[2] Control Automation Production Institute of Technology (CAPIT);
ledanhtuan@gmail.com

*Corresponding author: vovanhoa80@yahoo.com; Tel.: +84–912509932

Abstract: The research applies self-organizing maps (SOM) technique in combination with K–Means method to objectively classify weather patterns that cause summer heat wave in Viet Nam based on the dataset from 1998 to 2018. The pressure of mean sea level (PMSL) and geopotential height at 500hpa (H500) of JRA25 reanalysis data are used. The heat wave is defined to occur if the daily maximum temperature of at least 2/3 of surface synoptic stations in research area was greater than 35°C. According to above mentioned criteria, 156 summer heat waves were subjectively found at Northern region in period of 1998–2018. In central and southern regions, the summer heat waves were respectively found 204 and 69. By applying SOM and K–Means, there were 4, 3 and 2 key weather patterns that caused summer heat waves in Northern, Central and Sothern region respectively. In fact, the weather pattern caused summer heat waves at research region is usually related to activities of the western hot depression pattern and Northwest Pacific Subtropical High Pressure. The combination of 2 weather patterns or more was usually found Northern and Central region. However, the number of heat wave detected by SOM is smaller than number of heat wave was subjectively determined by forecaster (there are 109, 171 and 62 heat waves detected by SOM for the Northern, Central and Sothern region respectively). The reason for this result is that SOM method has not been able to identify heat waves caused by the combination of many weather patterns or by small and meso–scale weather patterns.

Keywords: SOM; Classification; Summer heat waves; Heat wave weather patterns.

1. Introduction

In recent years, under the impact of climate change, heat wave has been occurring in Vietnam with an increasing trend in both frequency and intensity. Damage statistics show that heat waves are also a type of natural disaster that causes a lot of damage to people and property. Therefore, it is very necessary to increase the understanding of the mechanism that causes heat wave to improve forecast quality. In Vietnam, there have been many studies on summer heat waves in which refer to many aspects from the statistics of heat wave frequency based on past data sets [1–2], causes and patterns of heat wave [3–5], forecasting heat wave is based on statistical methods [6] and NWP models from short term to seasonal scale [7], predicting the change of heat in the future according to climate change scenarios, etc.

In the classification of weather patterns, the SOM method is widely applied, especially related to heavy rain problems [8–11]. The SOM method has been applied in the

classification of heat wave weather patterns, specifically, [12] used ERA–Interim reanalysis data of the ECMWF from 1979 to 2016 to classify patterns for heat waves where the daily maximum temperature on that day is greater than the 90th percentile of the data series. The results show that the synoptic patterns caused heat wave are classified into 6 clusters based on PMSL anomalies in East Asia. Recently, SOM method has been applied in the classification of abnormal heat wave weather patterns in winter for Northern part of Viet Nam [7].

According to studying the causes and weather patterns that cause heat wave, most studies have partially shown the causes and statistics of typical weather patterns. However, these studies are mainly based on synoptic analysis methods and are analyzed subjectively by forecasters, so the obtained results are still subjective and difficult to apply in operational prediction. To contribute additionally to the results of classification of weather patterns that cause heat wave in Vietnam in an objective way, the research applies the SOM (Self–Organizing Map) method in combination with JMA's JRA25 reanalysis data set to identify groups of weather patterns that cause large–scale heat wave events in some areas of Vietnam. The research mainly focuses on summer heat wave in Viet Nam that occurred in large scale and weather patterns that significantly caused these summer heat waves. The daily maximum temperature data at surface synoptic stations, pressure of mean sea level and geopotential height at 500hpa of JRA25 reanalysis data will be used. The next of paper will give out the dataset and methodology in SOM application. The results present in third part of paper. It finally is some conclusions and remarks.

2. Materials and Methods

2.1. Dataset

To be able to find out statistics of heat waves occurring over Vietnam in the 21 years from 1998 to 2018, we collected maximum temperature data (Tx) at 183 surface synoptic stations. As is known, the heat wave cannot be directly observed, but it is determined from the observed quantities based on the given criteria. In the operational forecast, according to the intensity of heat wave, it can be divided into 3 types including heat wave ($35^oC \leq Tx < 37^oC$), strong heat wave ($37^oC \leq Tx < 39^oC$) and extreme heat wave ($Tx \geq 39^oC$). According to influenced area, heat wave can be divided into large–scale heat wave and local heat wave. To simplify the determination and ensure that there is enough sample size for SOM method, in this study we use the criterion $Tx \geq 35^oC$ and have at least 1/2 of the surface synoptic stations in the study area satisfy the condition $Tx \geq 35^oC$ at the same time. In addition, the heat wave in this study is mainly considered under the concept of a "*wave*", which satisfies the condition that at least 1 day occurs or there are 2 or more days as sequency. In case if the satisfactory days are interleaved with the unsatisfactory days, it is also defined as a heat wave.

Heat wave occurs every year in Vietnam and there are differences between climate in terms of origin, causes, intensity, scope, etc. The research aims at a large–scale heat wave as mentioned above and only aims to identify the dominant weather patterns, so there will be many heat waves occurring in many regions at the same time. Therefore, to ensure that the classification by SOM is objective, not duplicated and clearly shows the dominant weather patterns, we divide the study area across the country into 3 regions including the North region, the Central Region and the South region. Basing on the criteria for determining the heat wave, we have subjectively determined 156 heat waves in the northern region for period of 1998–2018. For Central Region and the South region, number of found heat waves is respectively 204 and 69 (Table 1).

To have data on the grid as input for the SOM method, we collected the JRA25 reanalysis data of JMA corresponding to the period of the aforementioned heat wave dataset. NOAA's NNRP2 and ECMWF's ERA–Interim reanalysis data sources were not collected due to

JRA25 has the higher resolution. Because it was developed by JMA, the quality of the JRA25 data source has higher accuracy in Asia area. Since the atmospheric fields are correlated, using all fields in the identification is unnecessary and may give undesirable results. Therefore, the selection of characteristic quantities will limit the amount of work and computation time. Based on the knowledge of weather patterns and previous studies, in this study, we only collect pressure of mean sea level (PMSL) and geopotential height at 500hpa (H500). The PMSL used to characterize the surface hot depression and the H500 will characterize the Northwest Pacific Subtropical High Pressure.

Table 1. Subjective statistics of heat waves for each of research are in Viet Nam by using criteria (Tx ≥ 35°C and have at least 1/2 of the surface synoptic stations in the study area satisfy the condition Tx ≥ 35°C at the same time) for period of 1998–2018.

Year	Northern Area	Central Area	Southern Area	Viet Nam
1998	5	6	1	12
1999	4	8	1	13
2000	3	5	0	8
2001	2	7	0	9
2002	1	4	2	7
2003	8	11	4	23
2004	5	9	2	16
2005	4	9	1	14
2006	8	10	2	20
2007	6	7	5	18
2008	8	10	2	20
2009	8	9	2	19
2010	15	14	7	36
2011	6	8	2	16
2012	11	14	4	29
2013	7	12	7	26
2014	11	11	6	28
2015	16	11	5	32
2016	13	12	11	36
2017	9	15	2	26
2018	6	12	3	21
Sum	**156**	**204**	**69**	**429**

2.2. Methodology

The research flowchart is shown on Figure 1. Specifically, basing on known heat waves that are subjectively determined according to the given criteria, PMSL and H500 data of JRA25 will be normalized and selected principal components based on PCA analysis. These components are then fed into the SOM for analysis and the creation of a U–Matrix map. Next, the K–Means method is applied to find the data clusters. Finally, from the given data clusters, weather pattern maps for each cluster are used to find out the main weather patterns that cause the heat wave. Because the meteorological elements in the JRA25 dataset vary in dimension as well as range of variation. Therefore, the predictors must be normalized before training the SOM network. The set of predictors will be normalized to a new set of factors according to the following formula:

$$x' = \frac{x - \mu_X}{\sigma_X} \qquad (1)$$

where x$^{'}$ is normalized variable of x predictor, μ_X and σ_X respectively are simple average and standard deviation of x predictor that is calculated based on past dataset. After normalizing, x$^{'}$ variable is non–dimension.
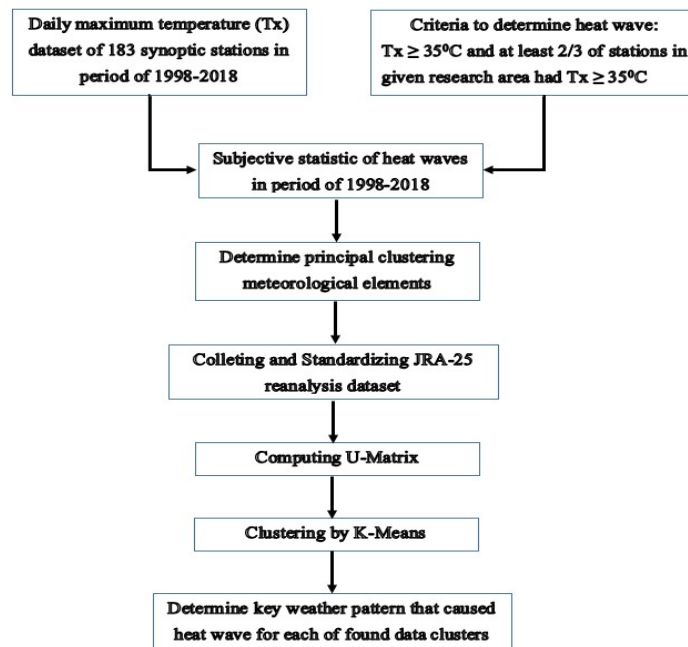
Figure 1. Flowchart of using SOM and K–Means methods to classify weather patterns that cause summer heat wave in Viet Nam based on JRA25 reanalysis dataset.

After normalizing the PMSL and H500 data, in order to reduce data redundancy and computing cost, we will not directly use the PMSL and H500 data on the grid but will analyze them into a series of principal components by applying PCA method. The input data for the PCA method is also normalized according to formula (1) as mentioned above. As a result, instead of including $2 \times 40 \times 29$ variables in clustering (40 and 29 is number of JRA25 grid point in meridional and zonal direction), we only have to cluster the number of 15–20 variables depending on the region.

Basing on principal components are selected, SOM training is implemented according to following step by step:

Step 1: Initializing weight vector $w_j^{(0)}$ with j = 1,2, …, d* by random select in input dataset (D).

Step 2: Iterating solve

Step 3: Assign x in D with given probability value

Step 4: Finding best fit noron i(x) in Kohonen class basing on Euclidean distance between vector w$_j$ and x: $i(x) = \arg \min_{1 \le j \le d^*} \left\| x - w_j^{(s)} \right\|$

Step 5: Updating weight for all norons in out layer as following formular:
$w_j^{(s+1)} = w_j^{(s)} + \eta(s)h_{j,i(x)}(s)(x - w_j^{(s)})$

Step 6: Iterating Step 2 if there is no significant change of SOM feature map (reduce radius of topological neighborhood at specified time)

Step 7: Finishing and give out final SOM feature map

The result of training the SOM network is to create two–dimensional matrix of Kohonen neurons in which each main neuron is a vector whose size is equal to the number of input neurons. Next, to create cluster boundaries in the study, we conduct use the U–Matrix technique combined with the K–Means algorithm. Specifically, basing on the U–Matrix map,

the K–Means method is applied to classify the number of possible heat wave clusters based on the SOM characteristics. After determining the number of possible heat wave clusters, the past heat wave data that subjectively determined as mentioned above will be used to classify each heat wave to which given clusters. After this classification step, the process of determining the dominant weather pattern is performed by displaying atmospheric field maps from the JRA25 data of each heat wave in the given cluster. The all steps to find dominant weather pattern that caused heat wave based on the SOM and K–Means method are shown in Figure 1.

3. Results and Discussion

Figures 2 to 4 show the origin and clustered maps of the U–matrix by applying the K–Means method for the Northern, Central and Southern regions, respectively. Specifically, for the Northern region, the results have 4 clusters found corresponding to 4 groups of weather patterns (Figure 3). However, the degree of clustering (separation between groups) is not really clear. For the Central region, there are 3 data clusters found after applying the K–Means method. For the Southern region, only 2 data clusters were found. Regarding the degree of distinction between clusters, the Southern region shows more clearly than the Northern and Central regions. Figures 5 to 7 respectively show typical weather patterns for each element of the matrix U with size 8 × 8 when applied to PMSL and H500 fields for the North, Central and Southern region, respectively. From these weather pattern maps it is clearer to see the degree of separation in the ground and upper–air weather pattern for each molecule of the U–matrix. Subjectively, the use of the weather pattern matrix as in the Figures 5 to 7 can also analyze and make judgments about the number of groups of weather patterns included in the dataset used to put into SOM analysis.
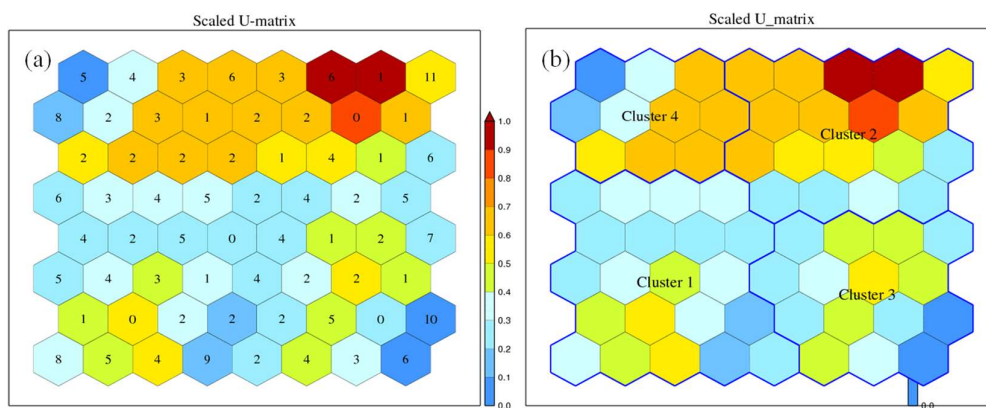

Figure 2. The origin U–Matrix (left) and cluttered U–Matrix by applying K–Means method (right, the cluster boundary line is bold blue) for Northern region basing on JRA25 dataset in period of 1998–2018 (the number means the cases belong the given U–Matrix element, the color palette means the distance of wight vector for each U–Matrix element).

Table 2 gives the results of determining the number of heat waves in each year in the period 1998–2018 for each study region and compares it with the number of heat waves subjectively determined basing on the above criteria. The number of heat waves identified from SOM is less than that determined by forecaster. Specifically, for the Northern region, the number of heat waves detected by the SOM method accounts for about 70% of the subjective determination. Meanwhile in the Central and Southern regions, it is 84 and 90% respectively. Thus, the determination of the number of heat waves by SOM method in the Central and Southern regions is better than in the North. The reason for the difference is the number of weather pattern clusters found by K–Means. Specifically, except for the Southern region (the number of clusters found is equal to the number of weather patterns found

according to the forecaster's analysis), the number of clusters found by K–Mean in the North and Central regions are less than 1 compared to the forecaster's analysis. This result may lead to under–estimate the number of heat waves in these regions. For the Southern region, it is possible that the input data will be in the interference area of the two data clusters.
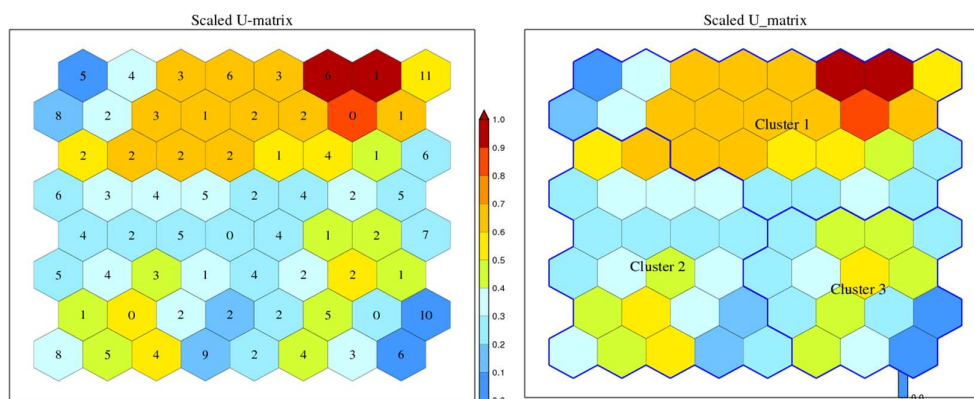
Figure 3. Similar to Figure 2 but for Central region.
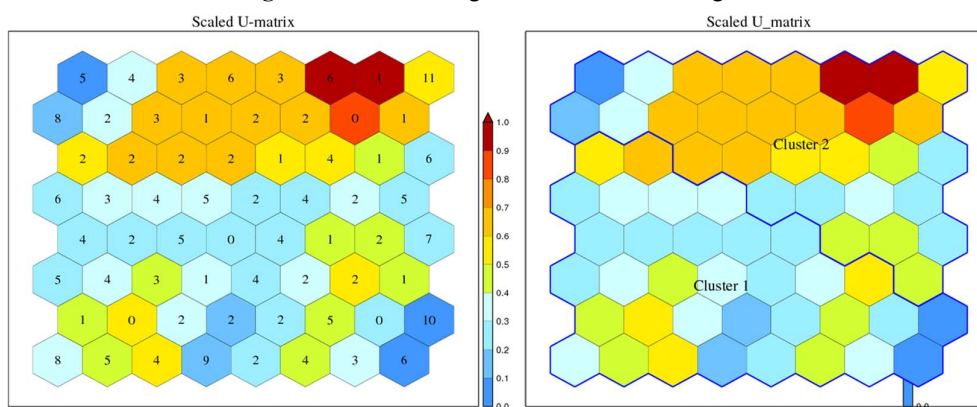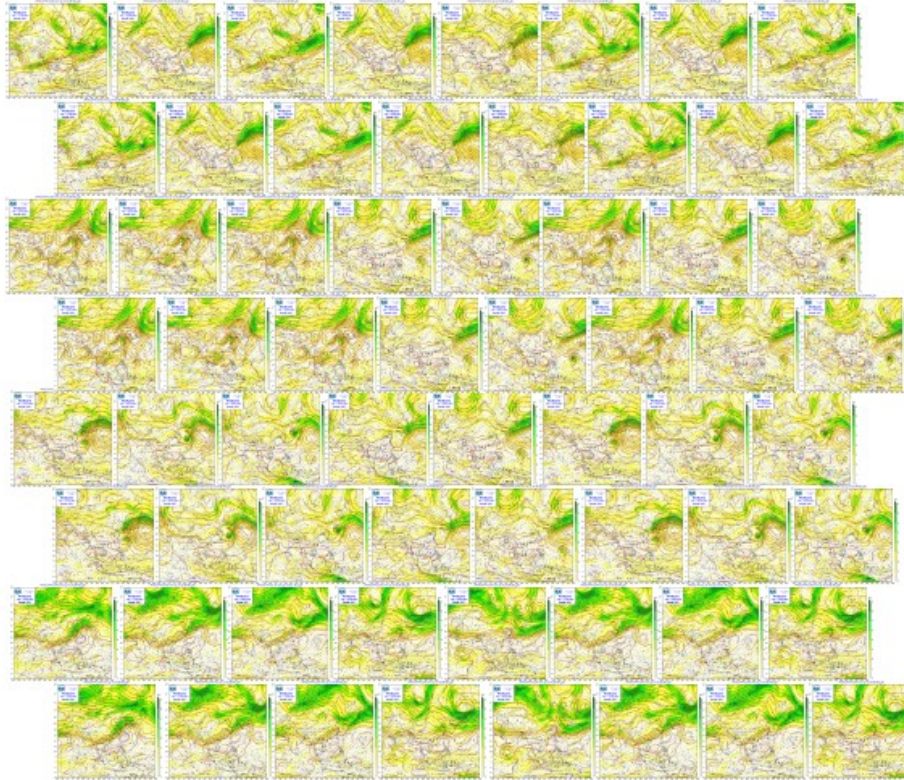

Figure 4. Similar to Figure 2 but for Southern region.
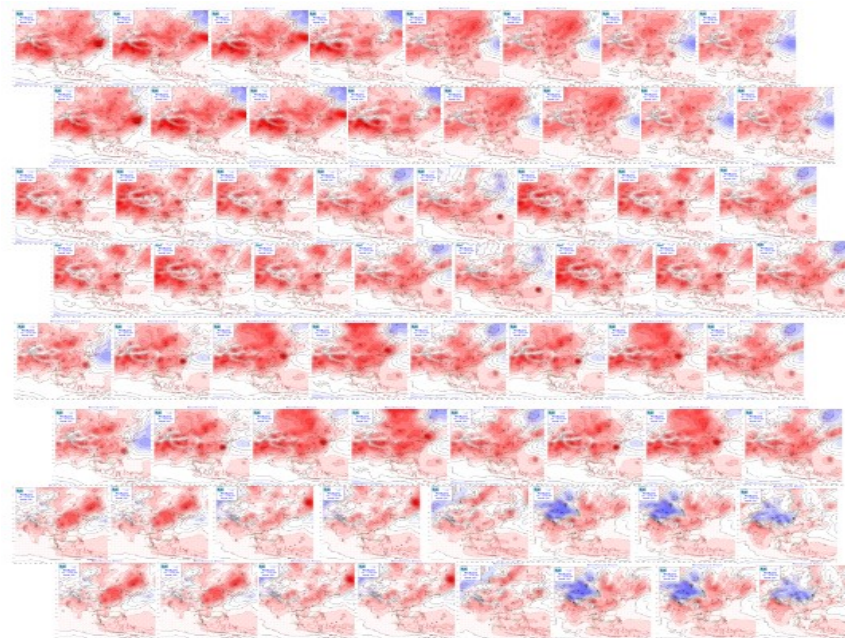
Figure 5a. The significant pressure of mean sea level maps of 8 × 8 U–Matrix for Northern region.

Figure 5b. The significant geopotential height at 500hpa level of 8 × 8 U–Matrix for Northern region.

Figure 6a. The significant pressure of mean sea level maps of 8 × 8 U–Matrix for Central region.

Figure 6b. The significant geopotential height at 500hpa level of 8 × 8 U–Matrix for Central region.
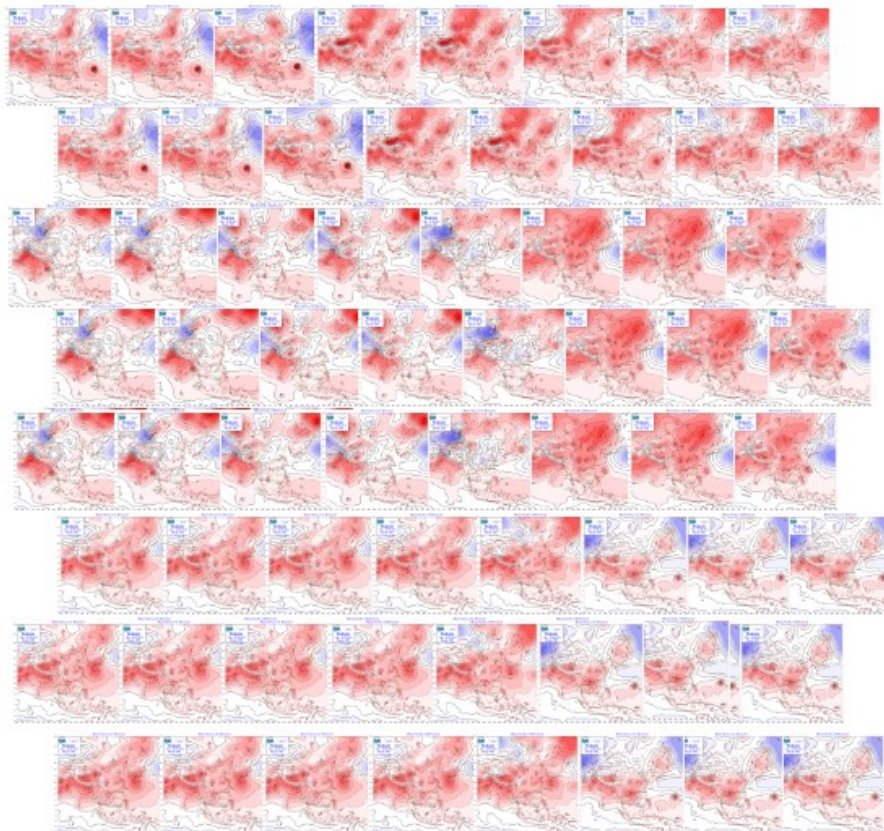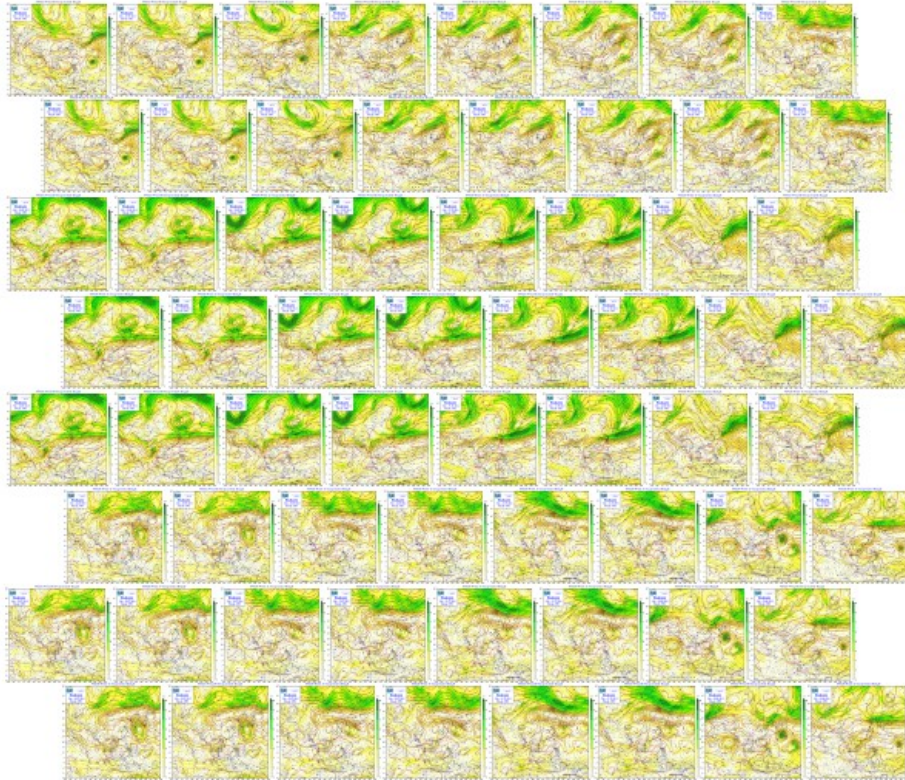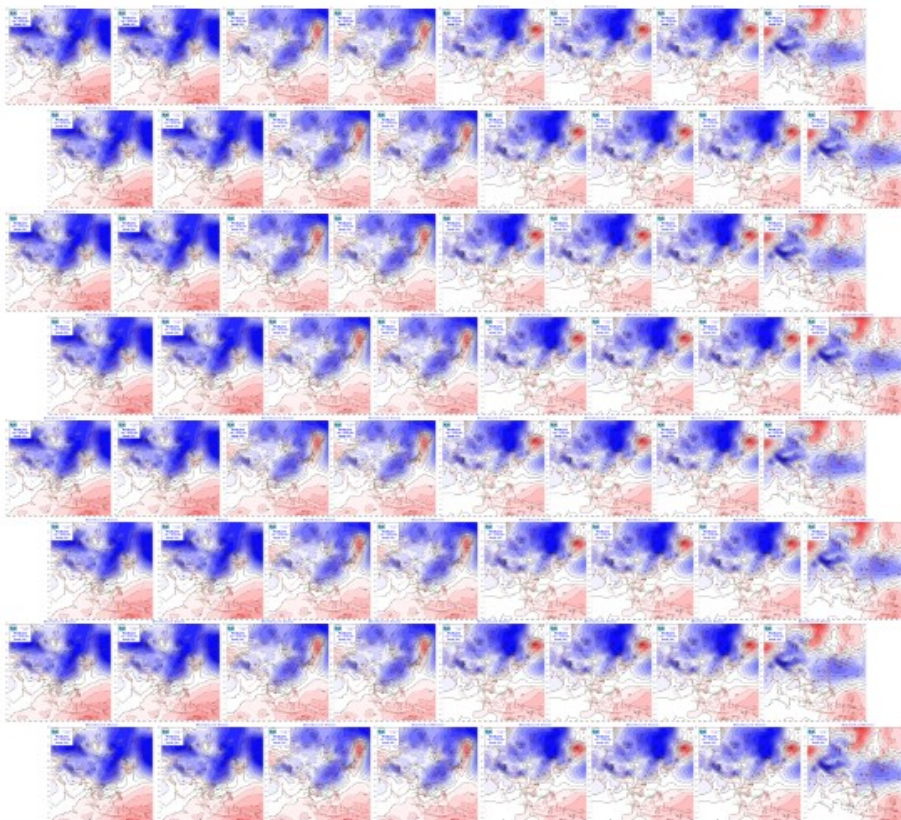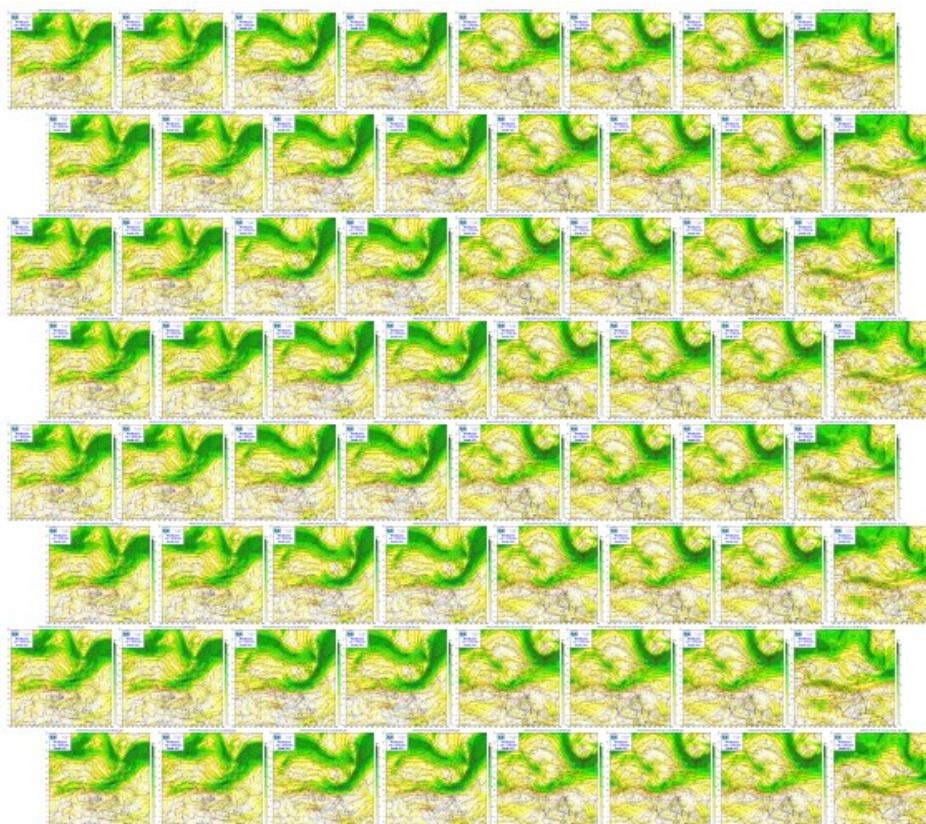

Figure 7a. The significant pressure of mean sea level maps of 8 × 8 U–Matrix for Southern region.

Figure 7b. The significant geopotential height at 500hpa level of 8 × 8 U–Matrix for Southern region.

By reanalyzing significant weather pattern maps of given clusters, we determine dominant weather pattern that caused heat waves in Viet Nam as following:

1. For Northern region: Western hot depression, Northwest Pacific Subtropical High Pressure, South China depression and Northern cold air mass.

2. For Central region: Western hot depression, Northwest Pacific Subtropical High Pressure and South–west monsoon.

3. For southern region: Western hot depression and Northwest Pacific Subtropical High Pressure.

Compared with the clusters of weather patterns determined by the forecaster, the weather patterns found by SOM and K–Means have similarities when only has a large–scale pattern caused heat wave. However, in case of association of 2 or more weather patterns with spatial scale differences, SOM could not be captured all cases, specially has activities of small–scale patterns. This is also the reason why the number of heat waves detected from the SOM is less than that determined by the forecaster's experience.

Table 2. The statistics of heat wave number in period of 1998–2018 for each research region by forecaster and SOM method.

Year	Northern Region		Central Region		Southern Region	
	Forecaster	SOM	Forecaster	SOM	Forecaster	SOM
1998	5	3	6	5	1	1
1999	4	3	8	7	1	1
2000	3	2	5	4	0	0
2001	2	2	7	5	0	0
2002	1	1	4	4	2	1

Year	Northern Region		Central Region		Southern Region	
	Forecaster	SOM	Forecaster	SOM	Forecaster	SOM
2003	8	5	11	9	4	4
2004	5	4	9	7	2	2
2005	4	2	9	8	1	1
2006	8	5	10	8	2	2
2007	6	3	7	5	5	4
2008	8	6	10	8	2	2
2009	8	6	9	7	2	2
2010	15	9	14	12	7	7
2011	6	4	8	6	2	2
2012	11	7	14	12	4	4
2013	7	6	12	12	7	6
2014	11	8	11	10	6	5
2015	16	12	11	10	5	4
2016	13	10	12	10	11	9
2017	9	7	15	12	2	2
2018	6	4	12	10	3	3
Sum	**156**	**109**	**204**	**171**	**69**	**62**

4. Conclusion

The paper has been studied and applied the SOM method combined with K–Means to objectively classify the weather patterns that cause heat wave in Vietnam based on dataset of pressure of mean sea level and 500hpa geopotential height of JRA25 reanalysis in the period 1998–2018. By using the criteria Tx ≥ 35°C and taking a large scale, there respectively were 156 heat waves occurred in the Northern region, 204 in the Central region and 69 in the South in the period 1998–2018. Subjective analysis has shown that there are 5 main weather patterns causing heat wave in the North. For the Central and Southern regions, there are 4 and 2 main weather patterns, respectively. In general, the heat wave occurring in all 3 areas is related to the activity of the western hot depression and the northwest Pacific subtropical high pressure. The heat wave is caused by combination of 2 or more weather patterns mainly occur in the North and Central regions.

The objective classification results based on SOM and K–Means methods are most appropriate in the Southern region, followed by the Central and Northern regions. The number of heat waves detected by the SOM method is less than that determined by subjective methods. The reason is that the number of weather pattern clusters classified from SOM and K–Means is less than in subjective analysis (except for the Southern region). The SOM method classifies well when there is only one dominant weather pattern in large–scale. When there are combinations of 2 or more weather patterns and the impact of these patterns are the same (especially with spatial differences), classification by SOM is difficult because the data will be in the intersection of the clusters. To improve the research results and overcome the above shortcomings, we suggest that it is necessary to optimize the U–matrix size to further enhance the ability to capture the small–scale weather patterns when there is a combination of many weather patterns at different spatial scales.

Author Contribution statement: Conceived and designed the experiments: M.T.T.T., T.L.D.; Analyzed and interpreted the data: H.V.V., T.L.D.; contributed reagents, materials, analysis tools or data: T.L.D.; manuscript editing: H.V.V., M.T.T.T.; Performed the experiments: M.T.T.T.; contributed reagents, materials, analyzed and interpreted the data, wrote the draft manuscript: H.V.V.

Competing interest statement: The authors declare no conflict of interest.

References
1. Ha, H.T.M.; Tan, P.V. The trend and variation of extreme temperatures in Viet Nam in the period of 1961–2007. *Sci. Technol. J. Ha Noi Univ. Sc.* **2009**, 412–422.
2. Lanh, N.V.; Thai, D.V. Prediction of maximum temperature and heat wave in Ha Noi in May–August months under the influence of western hot depression. *VN J. Hydrometeorol.* **2002**, *502*, 14–19.
3. Lanh, N.V. The impact of the depression patterns to summer weather in Viet Nam. *VN J. Hydrometeorol.* **2010**, *593*, 43–53.
4. Lanh, N.V. 2010. Heat wave and key weather patterns that cause heat wave in Viet Nam. *VN J. Hydrometeorol.* **2010**, *597*, 1–8.
5. Hang, P.M.; Dung, T.T.; Quang, N.D. The impact of western hot depression and Northwest Pacific Subtropical High Pressure to heat wave activities over Northern Central region during 2010–2015. *VN J. Hydrometeorol.* **2017**, *674*, 44–52.
6. Lan, H.P.; Dien, N.H. Prediction of maximum temperature by using ANN for northern delta region. *VN J. Hydrometeorol.* **2008**, *571*, 20–23.
7. Hoa, V.V.; Tien, D.D.; Duc, T.A.; Hung, M.K.; Quan, D.D.; Khiem, N.V.; An, N.V. Research on classifying typical synoptic patterns causing abnormal warm spells in early winter in northern area of Viet Nam by a Self–Organizing Map. *VN J. Hydrometeorol.* **2019**, *703*, 51–59.
8. Nishiyama, K.; Endo, S.; Jinno, K.; Uvo, C.B.; Olsson, J.; Bertndtsson, R. Idenfication of typical synoptic patterns causing heavy rainfall in the rainy season in Japan by a Self–Organizing Map. *Atmos. Res.* **2007**, *83*, 185–200.
9. Liu, Y.; Weisberg, RH. Sea surface temperature patterns on the West Florida shelf using growing hierarchical Self–Organizing Maps. *J. Atmos. Ocean Tech.* **2005**, *23*, 325–338.
10. Duc, T.A. Research on classifying heavy rainfall patterns in Viet Nam using Self–Organizing Map method. Climatology and Meteorology Thesis in master degree, Ha Noi University of Science, 2014, pp. 80.
11. Tuan, V.A. etc. Research on objectively classifying heavy rainfall patterns in Viet Nam. *Sci. Rep.* **2015**, pp. 179.
12. Seung–Yoon, B.; Kim, S.W.; Jung, M.I.; Roh, J.W.; Son, S.W. Classification of Heat Wave Events in Seoul using Self–Organizing Map. *J. Clim. Change Res.* **2018**, *9*, 209–221. Doi:10.15531/ksccr.2018.9.3.209.

Research on brewery wastewater treatment by lab scale constructed wetlands

Ton That Lang[1*], Ho Ngoc Han[1]

[1] Ho Chi Minh University for Natural Resources and Environment; ttlang@hcmunre.edu.vn; hongochanhn133@gmail.com

*Correspondence: ttlang@hcmunre.edu.vn; Tel.: +84–903983932

Abstract: Along with the development of the industry in Vietnam, the beery industry has contributed greatly to the state budget and creating jobs for labours. However, beer wastewater contains high organic matter, some wastewater discharged parameters such as BOD, COD, Nitrogen, Phosphorus, etc. usually exceed the permitted standards. Based on the above practice, the research was conducted to evaluate the effectiveness of the horizontal flow wetland model, laboratory scale with organic loading rate varying from 20, 40, 60, 80, 100, 120 kg COD/ha.day. The results show that the concentration of pollutants in the effluent can reach QCVN 40: 2011/BTNMT, column A, before discharged into the environment.

Keywords: Brewery wastewater; Constructed wetland; Horizontal flow wetland.

1. Introduction

Nowadays, the demand for beer is increasing day by day. The growth of the beer industry contributes to increasing products for society, serving people's lives, but on the other hand, it also entails the problem of production waste, especially, wastewater with high concentration of pollution poses a serious threat to the environment [1]. The beer wastewater contains a large number of suspended substances, COD and BOD and high acid content that needs to be treated before being discharged into the receiving water source [2–9]. All organics present in brewery wastewater are easy biodegradable. High BOD/COD ratio ranging from 1.5–2.0 is suitable for biological treatment [9–11].

Biological treatment by microalgae (*Dunalilla*) has reduced the organic matter content in wastewater [12]. The treatment of brewery wastewater through anaerobic AHR tank and aerobic system removed COD with 75%–97% efficiency [13–14]. A brewery (in NSW) [15] will undergo a number of pre–treatments, then enter the GWE ANUBIX–B anaerobic tank, where the COD content is treated by bacteria to reduce the organic matter concentration. At a traditional brewery in the city of Harare [16], the research evaluated the performance of the UASB tank in beer wastewater treatment. After 2 years of research, it was found that the use of the UASB can make the treated wastewater quality meet local requirements. Indicator of COD decreased by 57%, TSS and SS also decreased by 50% and 90%, respectively.

In Vietnam, breweries often use anaerobic, aerobic and reverse osmosis treatment methods to provide better removal efficiency [1–3, 11]. The combination of anaerobic treatment with aerobic treatment can remove 90–98% the high concentration of COD in brewery wastewater [3, 11].

In the process of industrialization and modernization, sustainable economic development and environmental protection are always concerned. The environmentally friendly treatment of industrial wastewater contributes directly or indirectly to the

sustainable development of a country. To that end, biological processes are considered to play a major role. Bioremediation uses beneficial microorganisms [17–19] and plants [20–23] to degrade, reduce or detoxify pollutants.

Wetland wastewater treatment model is a method with many advantages, especially it is very suitable for Vietnamese conditions due to low construction and operating costs [20–23]. Mechanisms of wastewater treatment of wetlands include sedimentation, precipitation, chemisorption, microbial metabolism, and plant uptake [11]. Pollutants can be removed by multiple mechanisms simultaneously in the system. Many studies show that wetlands have the ability to treat wastewater with high concentration of nutrients such as domestic wastewater, industrial wastewater, agricultural wastewater, leachate... [24–28]. A study [25] in 2014 treated turbidity, DO, BOD_5, TP, TN very effectively with the average treatment efficiency of 94%, 86%, 80%, 88%, and 94%, respectyvly The research [26] conducted in 2018 with leachate after being biologically treated with a COD concentration of 575 mg/l passing through the wetland system, the BOD_5 treatment efficiency reached 96.48%, COD 83.24%, total N 91.43%, total P 77.84%, NH_4^+ 86.47%, color 87.91%. The wetland model of [28] was carried out to evaluate the growth and absorption of nitrogen and phosphorus added in the wastewater of intensive farming ponds of catfish using *Hymenachne acutigluma* with treatment efficiency TN, TP reached 80–84.8% and 93.3–95.6%, respectively.

Brewery wastewater contains high organic composition, wide pH range, remarkable TSS, Nitrogen, Phosphorus, Coliform content. When using traditional wastewater treatment technologies, it often costs a lot to invest in construction, chemicals as well as operation and maintenance. Many research works on constructed wetland technology show that this technology is completely in harmony with nature, does not generate a lot of secondary waste, has low investment costs, treat many types of wastewater from low to high loads, using a variety of available, easy–to–find natural plants. This technology can also handle nutrients (Nitrogen, Phosphorus). Plants after treatment from constructed wetlands can be used for different purposes such as forage, improving the local landscape.

In addition, *Colocasia esculenta* and *Caladium bicolor* are easy to find plants and has suitable characteristics for water purification such as growing well in flooded, polluted environments and having strong roots [29]. Therefore, this study evaluates the treatment efficiency of COD, TN, TP, TSS, Coliform of the horizontal-flow constructed wetland model at different loads with *Colocasia esculenta* and *Caladium bicolor*.

2. Materials and Methods

2.1. Objects, scope of research

Objects: Brewery wastewater of Heineken, laboratory model of constructed wetland with *Colocasia esculenta* and *Caladium bicolor*.

Scope: Laboratory scale.

2.2. Model

Laboratory scale of horizontal flow wetland model is made of 10mm thick glass. There are 3 models including:
- Blank model (without growing aquatic plants) – (BL)
- Model of planting *Colocasia esculenta* – (CE)
- Model of planting *Caladium bicolor* – (CB)

All 3 models are designed with the same tank size: length x width x height = 2.0m × 0.6m × 0.5m, length: width ratio is 3: 1 [30]. The bottom of the tank is arranged with a slope of 1% to ensure that the treated wastewater is completely collected [31]. The model

has an external stainless-steel frame to increase the bearing capacity of the tank. On the side of the tank, there is an outlet pipe with a diameter of 27 mm that brings the effluent out.

Each model consists of 3 layers of materials arranged in a certain sequence [24, 29]:

- The top layer is a layer of fine sand with a diameter of 1–2 mm and a layer of soil with a total height of 150 mm. This layer of material is responsible for the main environment for plant roots to attach and develop.

- The middle layer is a layer of round gravel with a diameter of 5–10 mm, the height of this layer is 150 mm. The main task of gravel is to serve as a substrate for microorganisms to attach and grow, and a support layer for the fine sand layer and the upper soil layer.

- The bottom layer is a 10×20 mm rock layer, 100 mm high. The main task of the rock layer is to serve as a substrate for microorganisms to attach and grow, and at the same time a support layer for the gravel, fine sand and the upper soil layer.

The model of planting *Colocasia esculenta* and the model of planting *Caladium bicolor* are arranged in 2 different tanks, in continuous process, in natural conditions and started at the same time, following the operation parameters in Table 1.
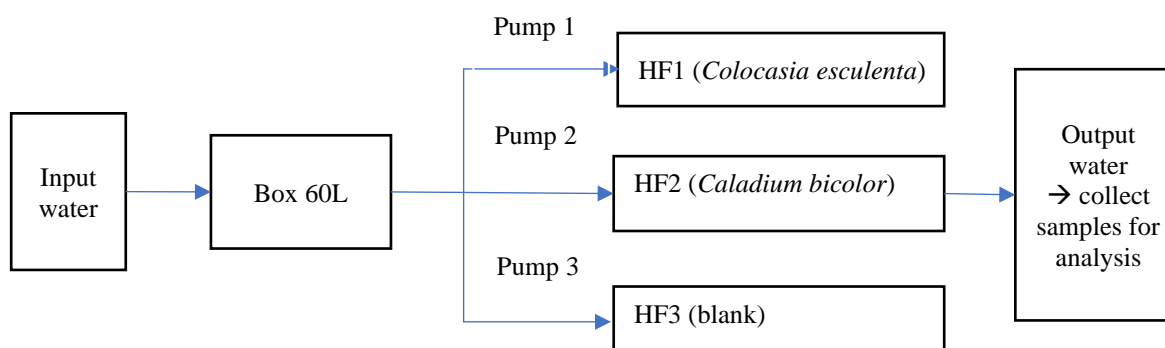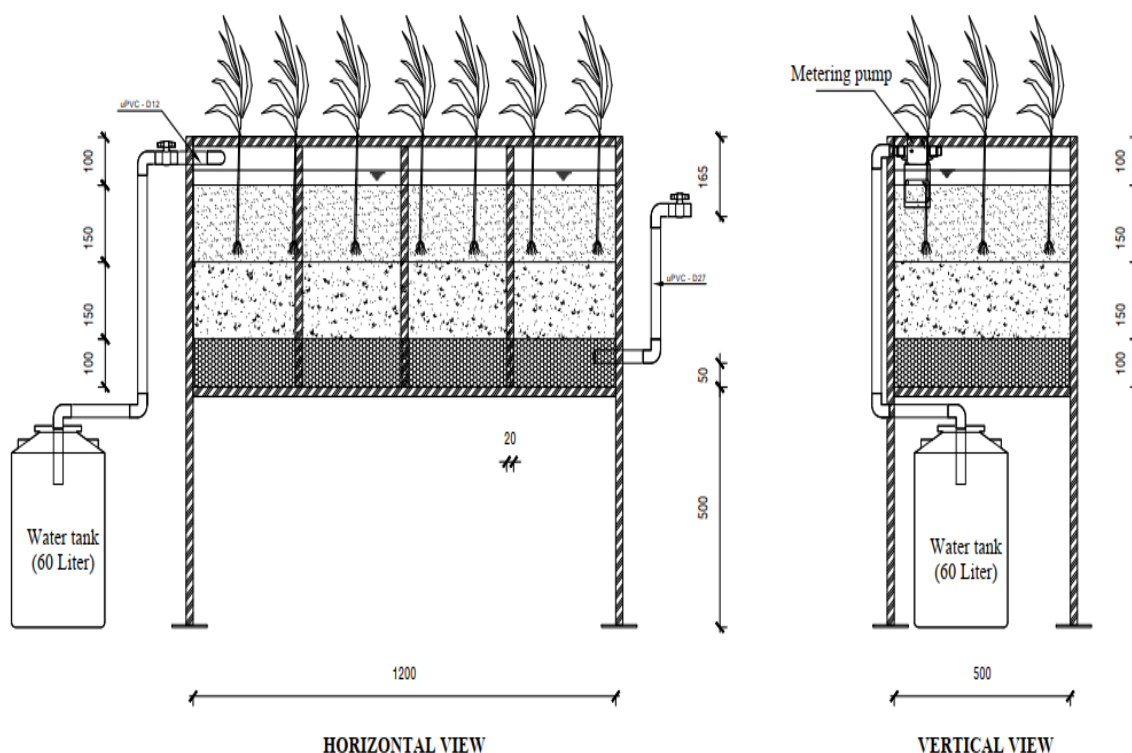

Figure 1. Experimental layout diagram.


Figure 2. Research models.

Table 1. The operation parameters of 2 model.

No.	Organic loading rate (OLR) (kgCOD/ha.day)	Influent flowrate Q (L/day)	Hydraulic Retention time (day)
1	20	7,6	18
2	40	15,2	9
3	60	22,8	6
4	80	30,4	4.5
5	100	38,0	3.6
6	120	45,6	3

3. Results and Discussion

3.1. The variation of pH value in models

The pH value in the effluent of all 3 models ranged from 6.8–7.6 (7.1 ± 0.2) for the blank model, 7.0–7.8 (7.4 ± 0.19) for the CB model and 7.0–7.7 (7.3 ± 0.2) for CE model. In general, the pH value fluctuated around neutral values (7.1–7.4) that can meet column A of QCVN 40:2011/BTNMT (pH = 6–9) and suitable for plant growth (Figure 3).
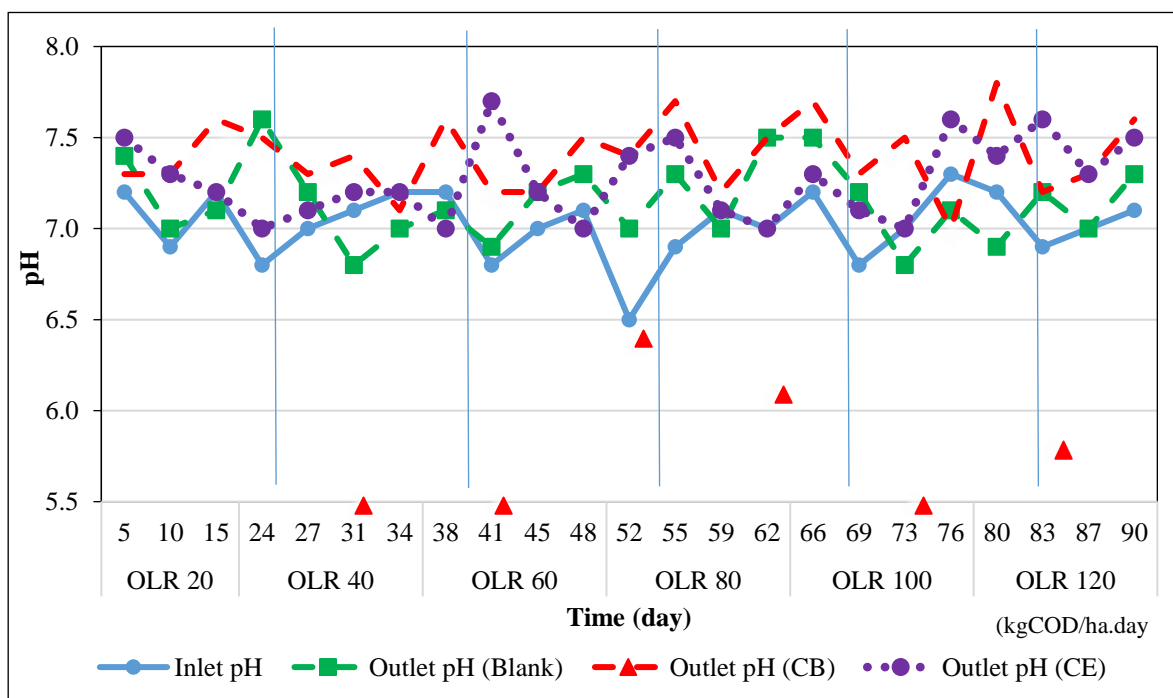
Figure 3. Variation of pH value during experiment.
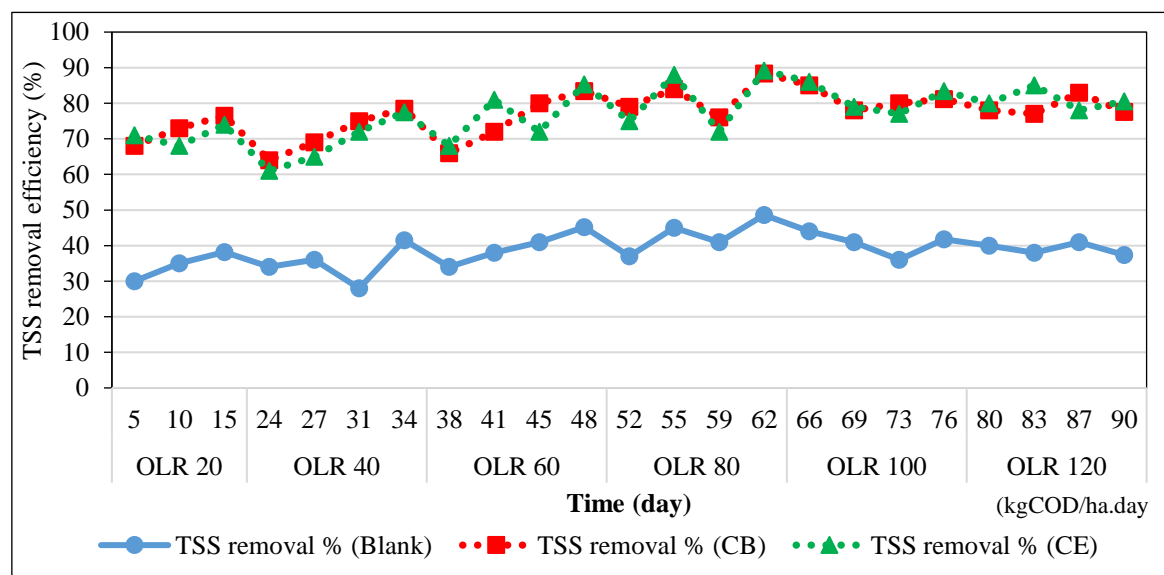
3.2. The variation of COD removal efficiency

Figure 4 shows that the highest COD treatment efficiency of the model is 85% (CB model), 84% (CE model) compared to the blank model, only 29%. COD treatment efficiency increase steadily through the loads and reached at the highest efficiency at organic load of 80 kg COD/ha.day for *Caladium bicolor* and at 100 kg COD/ha.day for *Colocasia esculenta*. The reason are quickly growth of plants and microorganisms leading to increase the absorption organic matter. Compared with QCVN 40:2008/MONRE, the effluent COD concentrations of both planting models can meet the standard and the blank model is not. This result is quite similar to the research results of [29] but is lower than the research of [31], the COD removal efficiency can reach 95% when using *Canna hybrids*.
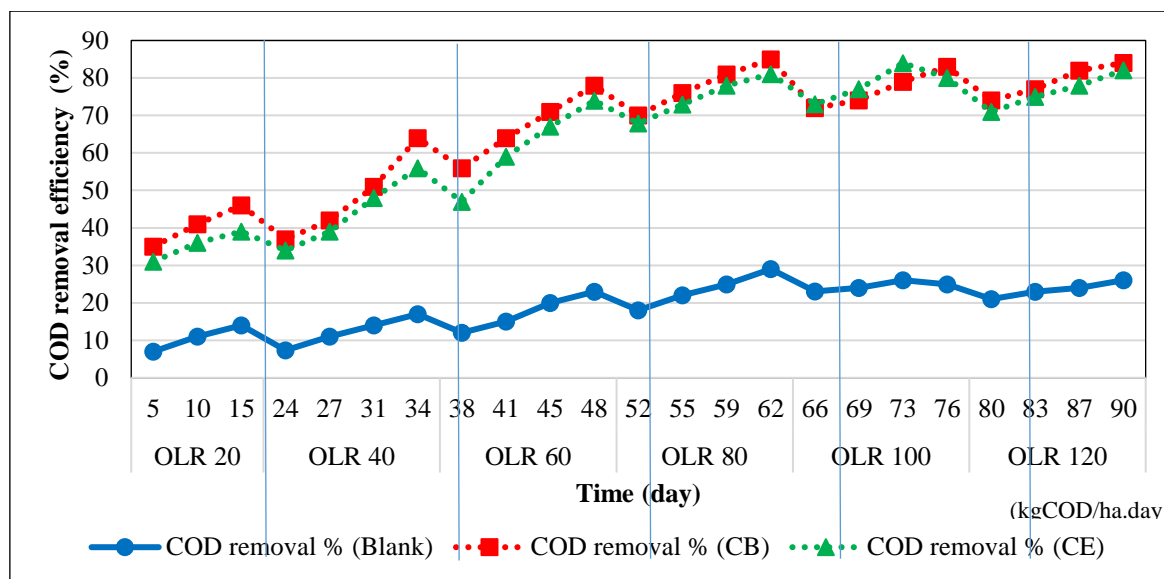
Figure 4. Variation of COD treatment efficiency at different loads.

3.3. The variation of TSS values

In the loading period of 60–80 kg COD/ha.day, TSS concentration in the influent fluctuates in the range of 119.2–143.8 mg/L, but TSS concentration of effluent ranges from 14–47 mg/L (CB model) and 13–44 mg/L (CE model) compared with 61–97 mg/L (blank model). The results of TSS values of the CB, CE model can meet the standards of QCVN 40: 2011/MONRE. The TSS removal efficiency was highest for the CB model (89.2%) and the CE model (88.4%) and compared with the blank model (48.6%). This result is quite similar to the other results, the TSS removal efficiency is from 47%–96% [29] and 88–92% [31], respectively.

Figure 5. Variation of TSS values in wastewater at different loads.

3.4. The variation of Total Nitrogen (TN) removal efficiency

After 90 days of model operation, *Caladium bicolor* and *Colocasia esculenta* grew in height, body size and especially the roots, so that the microflora developed strongly. The highest TN treatment efficiency is 75% (*Colocasia esculenta*); 72% (*Caladium bicolor*), compared with the blank model 29%. For the blank model, due to the concentration of

microorganisms attached to the filter material, the TN removal efficiency is lower. At loads of 80 kg COD/ha.day and 100 kg COD/ha.day those are the period of strong growth plant combined with more nitrifying and denitrifying microorganisms, the concentration of TN is mainly reduced due to nitrification and denitrification process. The TN values in the effluent of two planting models reached at QCVN 14:2008/MONRE, column A. This removal efficiency is lower than the results of [29], but is higher than the results of [31]: the TN removal efficiency is around 70%.
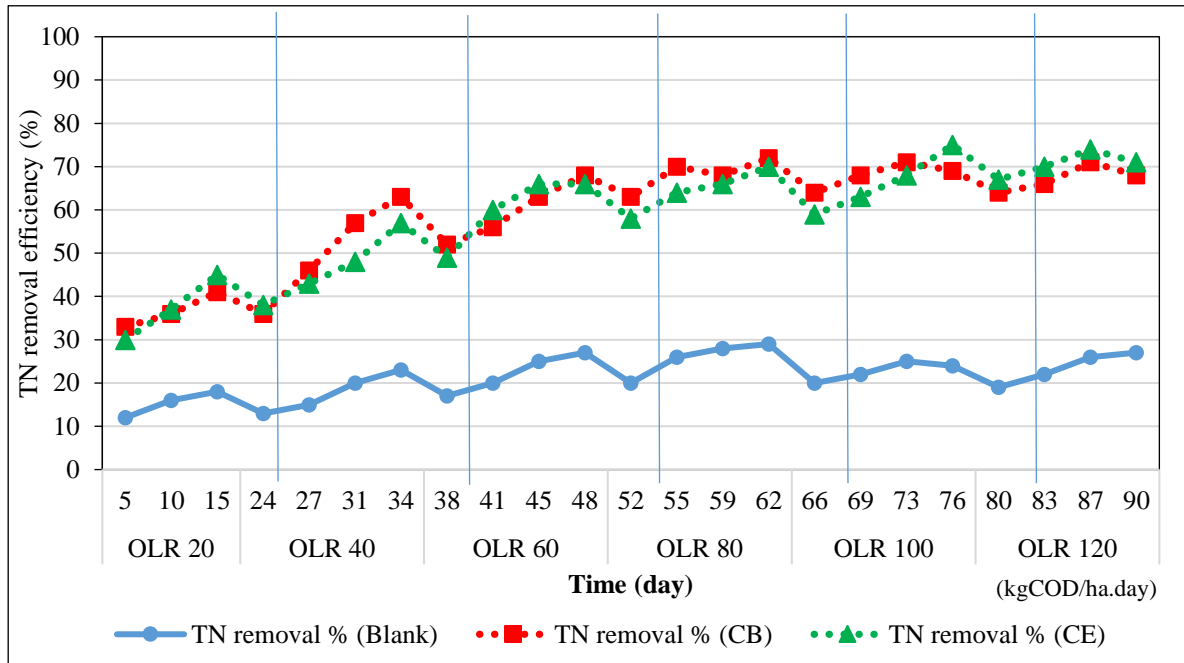

Figure 6. Variation of TN treatment efficiency at different loads.

3.5. The variation of Total Phosphorous (TP) removal efficiency
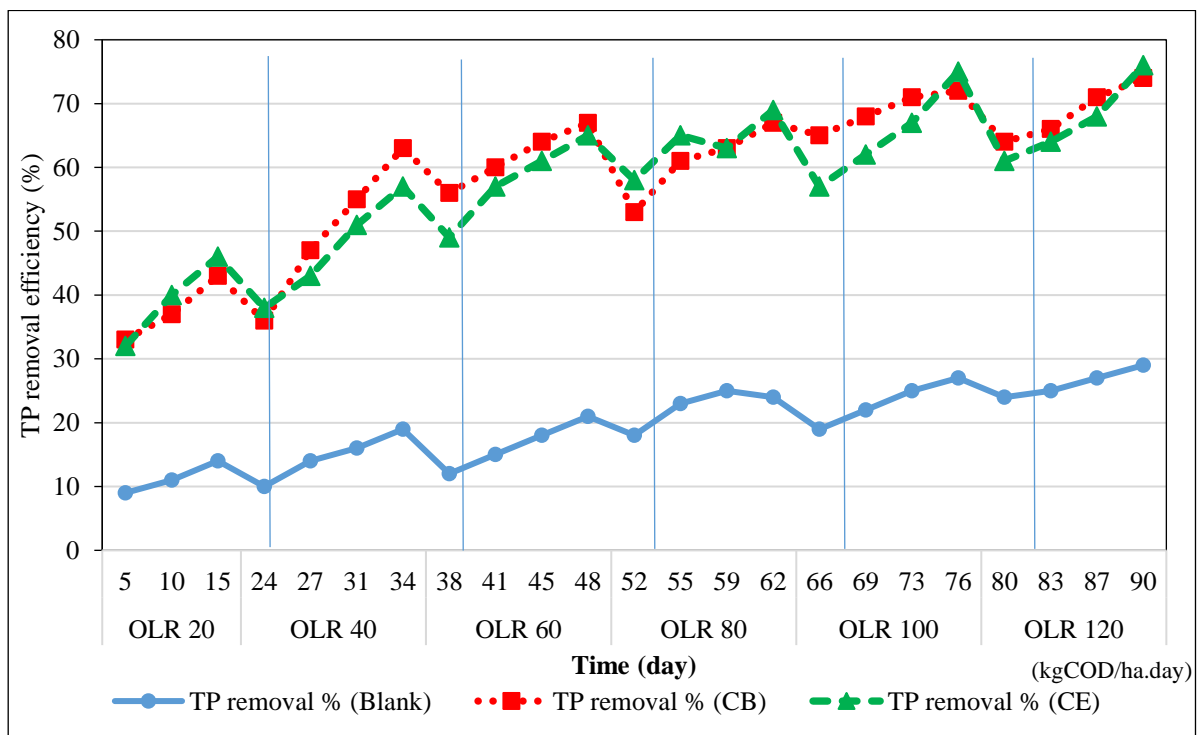

Figure 7. Variation of TP removal efficiency at different loads.

In general, TP treatment efficiency tends to increase with sampling times, the highest efficiency is 76% (*Colocasia esculenta*) and 74% (*Caladium bicolor*) and only about 29% compared to the blank model. The results are similar with the result of [21] treatment brewery wastewater with *Typha latifolia.*

Phosphorus is well reduced in the model by adsorbed by the roots and by adsorbent media. At the same time, the absorption and settling mechanism is enhanced by the mineral content in the filter material. They also use phosphorus as a substrate in the system.

Comparing the total phosphorus concentration in effluent using *Caladium bicolor* and *Colocasia esculenta* with those of QCVN 40:2011/MONRE, both achieved column A.

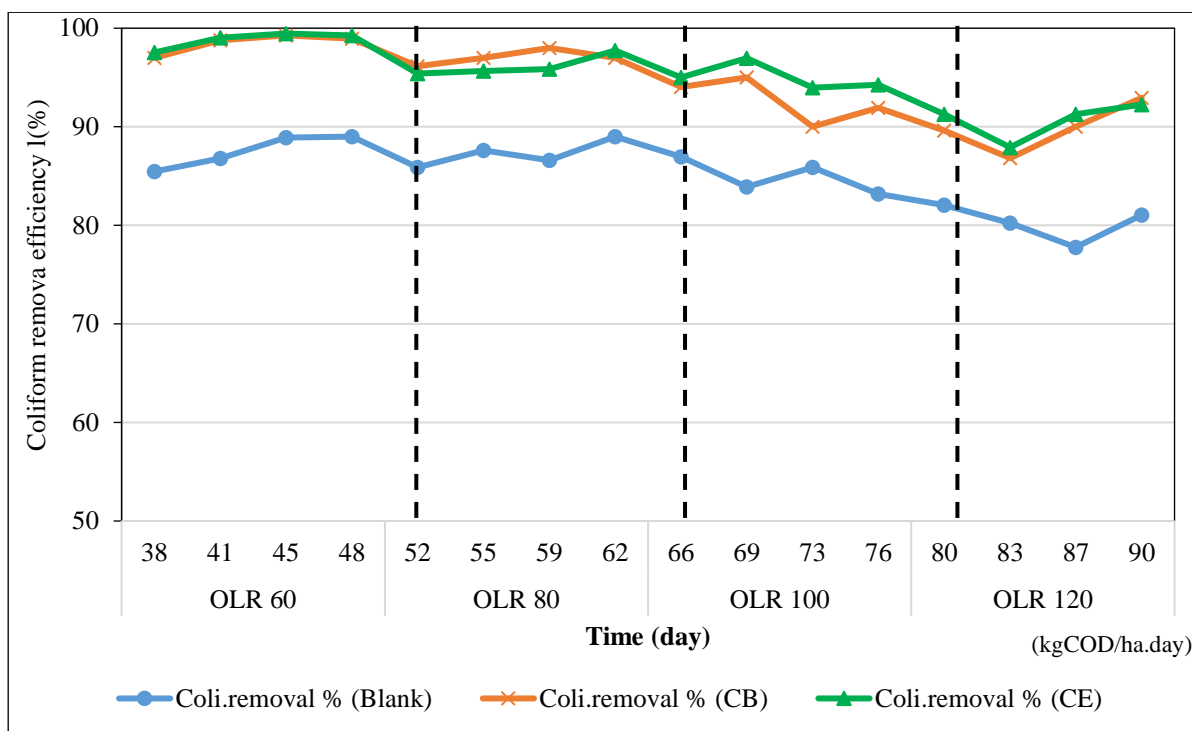### 3.6. The variation of Coliform value

Figure 8. Variation of Coliform treatment efficiency at different loads.

The total Coliform removal efficiency of the wetland model for the *Caladium bicolor* model is 87–99%, the *Colocasia esculenta* model is 88–99% compared to the blank model is 78–89%. In general, the total Coliform removal efficiency of the wetland model gradually decreased when operating at higher organic loads because the retention time at high loads was shorter than before. The effluent Coliform concentration of both planting models can meet the QCVN 40: 2011/MONRE (column A).

The average treatment efficiency (95%) to remove Coliform in this experiment is also consistent with the study [29], the total Coliform efficiency is high and low volatility, stable in the range of 94% to 99%.

3.7. Plant growth process

During operation time, the growth indexes such as the number of branches, tree height, tree body size, etc. all increased fairly at different loads. This is a good sign that shows the plant's adaptation and growth in the actual wastewater and has good treatment capacity.
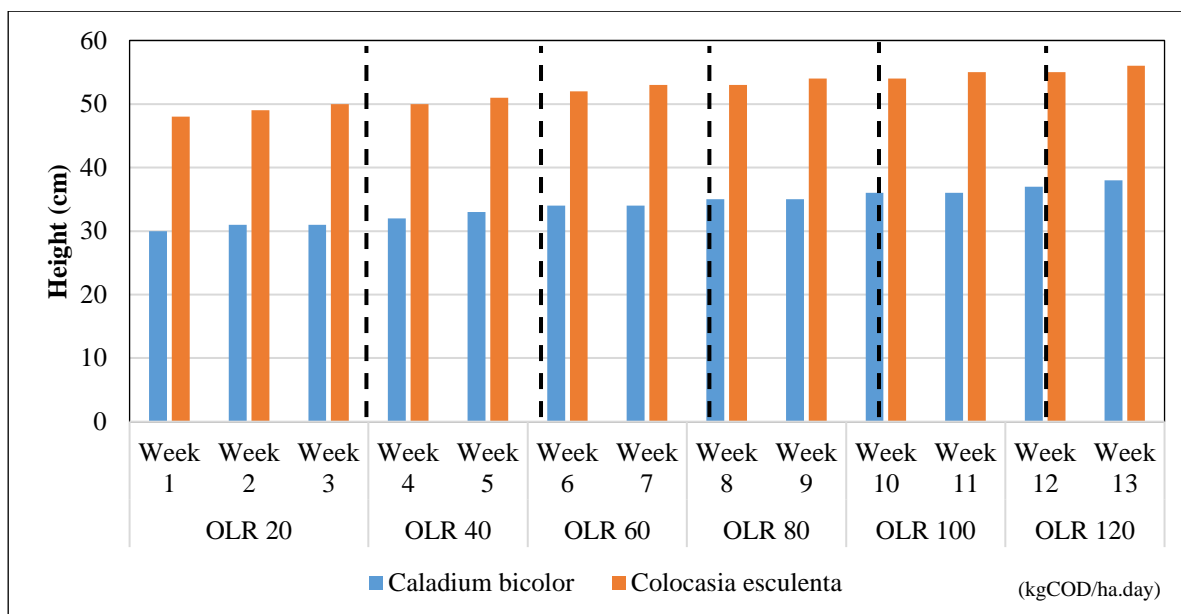
Figure 9. Tree height growth chart
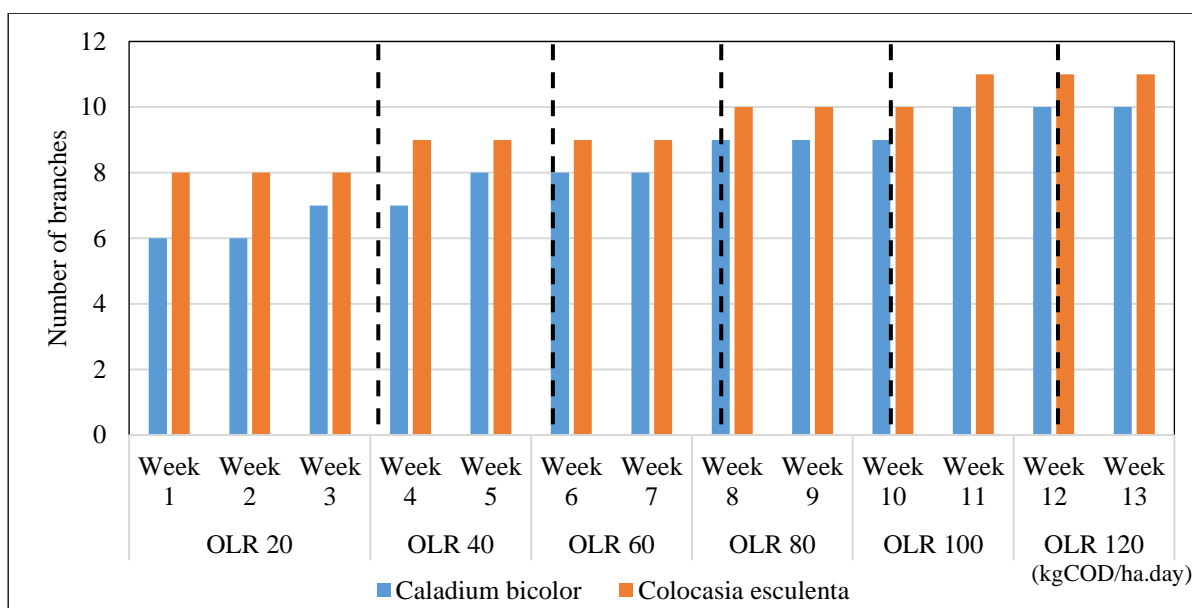
Figure 10. The growth chart of the number of branches

4. Conclusions

The number of pollutants in the brewery wastewater after treatment by constructed wetland has been significantly reduced. Treatment efficiency was highest at 80 kg COD/ha.day for the *Caladium bicolor* and 100 kg COD/ha.day for the *Colocasia esculenta*. Specific treatment efficiency is as follow: COD 84–85%, TSS 88–89%, TN 72–75%, TP 74–76% and total coliform > 90%. The difference between the treatment efficiency of the *Colocasia esculenta* and the *Caladium bicolor* is not significant (from 2–3%).

- It is necessary to continue to study the effectiveness of brewery wastewater treatment with higher loads, with other plant species and build a pilot model to apply this research in practice.

- It is necessary for further study the *Caladium bicolor* and the *Colocasia esculenta* in the treatment of other types of wastewaters to find out the other pollutant removal capabilities of these two plants.

Author contribution statement: Determine content, research methods, write reports: T.T.L; get sample, parsing: H.N.H.

Author Commitment Statement: The paper submitted with the full knowledge and consent of the author (if any), without any prior publishment or copy from other previous studies; There is no dispute of interest in the authors group.

References

1. Hien, N.T. Research on technology model for wastewater treatment of breweries, Master of Science Thesis, Hanoi, 2007.
2. Ha, T.D.; Tin, N.V. Wastewater treatment of breweries according to the model of anaerobic reverse filtration aerotank with intermittent operation. Proceeding of 14th Conference of Science and Technology of Construction University, Hanoi, Vietnam, Construction Publishing House, 2006, pp. 85–93.
3. Lai, T.X.; Duong, N.T. Industrial wastewater treatment, Construction Publishing House, Hanoi, 2009.
4. Azad, H.S. Industrial Wastewater Management Handbook Mc. Graw–Hill. New York, 1976.
5. Beyene, D.; Rao, P.V.V. The effectiveness of waste stabilization ponds in the treatment of brewery effluent the case of META ABO BREWERY waste stabilization ponds, Sebeta, Ethiopia. *J. Res. Sci. Technol.* **2013**, *2(8)*, 10–21.
6. Brewers Association Water and Wastewater: Treatment/Volume Reduction Manual, India, pp. 1–47.
7. Inyang, U.E.; Bassey, E.N.; Inyang, J.D. Characterization of Brewery Effluent Fluid. *J. Eng. Appl. Sci.* **2012**, *4*, 67–77.
8. Simate, G.; Cluett, J.; Iyuke, S.; Musapatika, E.; Ndlovu, S. et al. The treatment of brewery wastewater for reuse: State of the art. *Desalination* **2011**, *273*, 235–247.
9. Enitan, A.M.; Adeyemo, J.; Kumari, S.; Swalaha, F.; Bux, F. Characterization of Brewery Wastewater Composition. *Int. J. Environ. Chem. Ecol. Geo. Geophys. Eng.* **2015**, *9(9)*, 1036–1038.
10. Bodike, R.; Thatikonda, S. Biotreatment of Brewery Effluent Using Pseudomonas Species. *IOSR J. Environ. Sci. Toxicol. Food Technol.* **2014**, *8(6)*, 08–12.
11. Phuoc, N.V. Textbook of domestic and industrial wastewater treatment by biological methods, Construction Publishing House, Hanoi, 2010.
12. Al– Rajhia, S.; Raut, N.; Al–Qasmi, F.; Qasmi, M.; Al Saadi, A. Treatment of Industrials Wastewater by Using Microalgae. Paper presented in 2012 International Conference on Environmental, Biomedical and Biotechnology, IACSIT Press, Singapore **2012**, *41*, 217– 221.
13. World Bank. Industrial Pollution Prevention and Abatement: Breweries. Draft Technical Background Document. Environment Department, Washington, D.C., USA, 1997, pp. 272–274.
14. Kothiyal, M.; Semwal, G.N. Performance Evaluation of Brewery Biological Wastewater Treatment Plant. *MOJ Ecol. Environ. Sci.* **2018**, *3(1)*, 1–4.
15. Brewers Association. Water and Wastewater: Treatment/Volume Reduction Manual, 2011. Available online: BrewersAssociation.org.
16. Parawira, W.; Kudita, I.; Nyandoroh, M.G.; Zvauya, R. A study of industrial anaerobic treatment of opaque beer brewery wastewater in a tropical climate using a full–scale UASB reactor seeded with activated sludge. *Process Biochem.* **2005**, *40*, 593–599.

17. Bodike, R.; Thatikonda, S.R. Biotreatment of Brewary Effluent Using Pseudomonas Species. *IOSR J. Environ. Sci. Toxicol. Food Technol.* **2014**, *8(6)*, 08–12.

18. Amenorfenyo, D.K. et al. Microalgae Brewery Wastewater Treatment: Potentials, Benefits and the Challenges. *Int. J. Environ. Res. Public Health* **2019**, *16(11)*, 1910.

19. Obasi, A.; Agwu, O. Bioremoval of Heavy Metals from a Nigerian Brewery Wastewater by Bacterial Application. *Food Appl. Biosci. J.* **2017**, *5(3)*, 165–175.

20. Lutzu, G.A.; Zhang, W.; Liu, T. Feasibility of using brewery wastewater for biodiesel production and nutrient removal by Scenedesmus dimorphus. *Environ. Technol.* **2016**, *37*, 1568–1581.

21. Gebeyehu, A.; Shebeshe, N.; Kloos, H.; Belay, S. Suitability of nutrients removal from brewery wastewater using a hydroponic technology with *Typha latifolia*. *BMC Biotechnol.* **2018**, 74(2018).

22. Masi, F.; Rizzo, A.; Bresciani, R. Treatment of Wineries and Breweries Effluents using Constructed Wetlands. John Wiley & Sons Ltd, In book: Constructed Wetlands for Industrial Wastewater Treatment 2018, pp. 95–104.

23. Angassa, K.; Assefa, B. Constructed Wetland for the Treatment of Brewery Wastewater Performance evaluation, Lambert Academic Publishing, 2012.

24. El Hamouri, B. et al. A Hybrid Wetland for Small Community Wastewater Treatment in Morocco. Sustainable Sanitation Practice, 2012, pp. 39.

25. Kha, B.T. 2014, The efficiency of domestic wastewater treatment of the tectonic wetland system with clams. *Sci. Technol. Magazine* **2014**, 14.

26. Le, N.A.; Trinh, L.T.M. Application of an artificial wetland model with vetiver and reed grass to treat leachate. *Sci. Technol. Dev. Magazine Nat. Sci. J.* 2018, 2 (5).

27. Pérez, A.G. et al. Sunflowers (*Helianthus annuus L.*) on Top of a Constructed Wetland as an Engineered Ecosystem to Clean Sewage Onsite. *Open J. Water Pollut. Treat.* **2014**, 2374–6351.

28. Kieu, L.D.; Paul, H.T.; Loc, N.X.; Nguyen, P.Q.; Cong, N.V.; Trang, N.T.D. The ability to treat wastewater from ponds of intensive culture of catfish (*pangasianodon hypophthamus*) by *hymenachne acutigluma* in an aerated tectonic wetland system. *J. Agric. Rural Dev.* 2017.

29. Cuong, N.X.; Loan, N.T. The domestic wastewater treatment efficiency of the integrated artificial wetland system. Scientific Journal of Vietnam National University, Hanoi, 2015.

30. Saeed, T.; Sun, G. A lab–scale study of constructed wetlands with sugarcane bagasse and sand media for the treatment of textile wastewater. *Bioresour. Technol.* **2013**, *128*, 438–447.

31. Tu, N.T.; Lang, T.T. Study on treatment of brewery wastewater by horizontal flow constructed wetlands. Natural Resources and Environment Journal, 2022.

Research Article

Measuring land subsidence in Tra Vinh by SAR interferometry Sentinel–1

Ho Tong Minh Dinh[1], Huynh Quyen[2], Bui Trong Vinh[3], Tran Van Hung[4], Le Trung Chon[2*]

[1] UMR TETIS, INRAE, University of Montpellier, France; Dinh.Ho-Tong-Minh@Irstea.fr
[2] Ho Chi Minh City University of Natural Resources and Environment;
Viet Nam; hquyen@hcmunre.edu.vn; chonlt@hcmunre.edu.vn
[3] Ho Chi Minh City University of Technology, VNU-HCM, Viet Nam;
btvinh@hcmut.edu.vn
[4] Department of Natural Resources and Environment, Tra Vinh Province, Viet Nam;
hunggeotv@gmail.com

***Corresponding author: ltchon@hcmut.edu.vn; Tel.: +84–909122367

Abstract: This paper presents the results of analyzing land subsidence in Tra Vinh province in the period of 2015–2019 by using InSAR interference from Big Data of Sentinel–1 satellite radar and ground measurement data to understand the impact of urbanization and groundwater exploitation on land subsidence in the area research. In the period from 2015 to 2019 in Tra Vinh city, there are areas with annual subsidence rate of about 1 cm/year while in Dan Thanh commune, Duyen Hai district subsidence rate is up to 3 cm/year. One of the reasons for such rapid land subsidence is the excessive exploitation of underground water in these areas, especially in Dan Thanh commune, Duyen Hai district.

Keywords: Mekong; Tra Vinh; Subsidence; InSAR; Sentinel–1.

1. Introduction

The countries in the Mekong region, and specifically Vietnam, are among the most affected by climate change. Consequently, the mean temperature is raised and the sea level has risen. One of the big environment impacts is the regional water–pumping induced land subsidence. Combined with the sea level rise due to global climate change, the land subsidence directly impacts a variety of hazards which can be associated with subsurface saline intrusion [1–2], and increases in the depth and duration of annual flooding [3]. In response to these challenges, besides climate change adaptation, the knowledge of the ground subsidence such as their spatial extent and their temporal evolution is essential.

Although traditional ground–based measurement methods such as Global Navigation Satellites System (GNSS such as GPS) and Geodetic Levelling can be made locally, remote sensing observations are essential for mapping spatial extent and temporal evolution of land subsidence over large regions. Large scale land subsidence can be measured using satellite–based SAR imagery processed by interferometry (InSAR) with high accuracy from space [3–7]. Since 2014, the Sentinel–1 satellite provides open access and systematic data (in Terrain Observation with Progressive Scan (TOPS) mode) with 6-day revisit and 20 m spatial resolution [8]. However, TOPS Sentinel–1A/1B phase is very sensitive to geometric errors. In case of a small misregistration error between a pair of images, this residual term leads to a phase jump in the interferometric phase. Due to this limitation, doing TOPS interferometry needs extremely high co–registration requirements (e.g., an accuracy of 0.001 pixel in

azimuth direction) have to be met [9]. Thus, Although Sentinel–1 data offers the best opportunity for land subsidence monitoring, it is challenge to do InSAR processing.

In this paper, we will study on using InSAR techniques to determinate the land subsidence in the Mekong Delta. The work will be focused on the Tra Vinh, a typical coastal province in the delta, to demonstrate the feasibility of Sentinel–1 data.

2. Materials and Methods

2.1. Description of study site

Tra Vinh province, in the Mekong Delta, is one of the most biologically diverse and agricultural regions in the world, but sea level rise, land subsidence, upstream hydropower dam construction and rapid urbanization are the main causes affecting the hydrological, hydraulic, and sediment transport regimes of the whole area. Tra Vinh is located at the end of the island sandwiched between the Tien Giang and Hau Giang rivers. The terrain is mainly flat land with elevation of about 1m above sea level. In the coastal plain, there should be sand dunes, running continuously in an arc and parallel to the coast. The further to the sea, the higher and wider these mounds are. Due to the division by the hills and the system of roads and canals, the terrain of the whole region is quite complicated. Low-lying areas are interspersed with high hills, the slope trend is only shown in each field. Particularly, the southern part of the province is lowland, divided by bow-shaped sand varieties into many local low-lying areas, many places are only at an altitude of 0.5-0.8 m, so it is often flooded with salt water by 0.4 m per year -0.8 m for a period of 3-5 months. The average tidal amplitude of the east coast of Tra Vinh province fluctuates around 2m. Therefore, the delta of Tra Vinh is very vulnerable to land subsidence due to many reasons, in which special attention is paid to the causes of groundwater extraction, urbanization and change of shoreline due to the influence of the sea. The change of hydraulic regime, hydrology and sea level rise due to global climate change. More importantly, these two factors are directly responsible for many hazards such as saltwater intrusion on the surface and underground [1], increased altitude and timing of annual floods [3] and natural arsenic contamination [2]. This really causes concern for policy makers, managers and people living in Tra Vinh Province in particular and the Mekong Delta in general.

91% of Tra Vinh's water source for daily life and production is exploited from groundwater. Groundwater in the Mekong Delta including Tra Vinh is extracted from several aquifers ranging from Holocene to Miocene, in which deep aquifers are exploited the most [10]. When groundwater is extracted, the pore pressure decreases and the sediments undergo compression, which causes land subsidence. In order to determine the subsidence phenomenon as well as the subsidence speed, direct measurement methods such as geometric leveling, GNSS are used in small area, besides remote sensing methods such as: use LiDAR data from unmanned aerial vehicle (UAV) sensors or use radar remote sensing images to determine subsidence and its evolution over time for areas wide.

2.2. Methodology

Let $\varphi^n = \frac{4\pi}{\lambda} R_n$ represent the unwrapped interferometric phase, where R_n is the distance between the target and the n–th orbit acquisition, λ is the carrier wavelength. Then, φ^n is composed of the phase components related to deformation, residual topography, atmosphere, and noise [11]:

$$\varphi^n = \varphi^n_{defo} + \varphi^n_{topo} + \varphi^n_{atmo} + \varphi^n_{noise} + 2k\pi \qquad (1)$$

where φ^n_{defo} is the deformation phase, φ^n_{topo} is the residual topographic phase, φ^n_{atmo} is the atmospheric phase, φ^n_{noise} is the phase noise, and k is an integer ambiguity number.

The goal is to estimate the deformation phase, which can be written as follows (assuming a constant velocity model):

$$\varphi^n_{defo} = \frac{4\pi\tau_n}{\lambda}v \qquad (2)$$

where v is the mean deformation light of sight velocity of the target, and tn is the temporal baseline. The residual topographic phase is given as follows:

$$\varphi^n_{topo} = \frac{4\pi b_n}{\lambda \sin \theta R}\Delta h \qquad (3)$$

where Δh is the residual topography, and θ is the local incidence angle, bn is the normal baseline. The atmospheric phase is the delay of the signal due to weather conditions. The phase noise is due to temporal decorrelation, mis–coregistration, uncompensated spectral shift decorrelation, orbital errors, and thermal noise.

In the conventional spaceborne InSAR, which uses two SAR acquisitions to calculate the interferometric phase, the technique has issues relative to atmospheric, spatial and temporal decorrelations [12]. These parameters can only be overcome by a specific analysis considering phase changes in a series of SAR images acquired at different times over the same region. In fact, multi–temporal Interferometry SAR approach [3–4, 6, 13,] is well–known for its ability to measure subsidence. Particular in [13], he proposed a Maximum Likelihood Estimator (MLE) approach which can jointly exploits all N(N–1)/2 interferograms available from N images, in order to squeeze the best estimates of the N–1 interferometric phases. In this fashion, we can exploit not only persistent scatters (PS) but also distributed scatters (DS) information for estimating the deformation. Such increased number of identified PS/DS points on the ground results at an increased confidence of the ground motion, compared to the previous PS algorithm [3]. The reader is referred to [3–4] for the full descriptions of the processing chain. The results are processed by the TomoSAR platform which offers SAR, InSAR and tomography processing [14].

2.3. Data collection

Tra Vinh is located in the delta area of the Mekong. The water for domestic and industrial use in Tra Vinh comes from wells located within and around the city. The heavy pumping of groundwater has produced a serious settlement problem, which in turn has affected surface structures in the city of Tra Vinh. Figure 1 reports our study area.
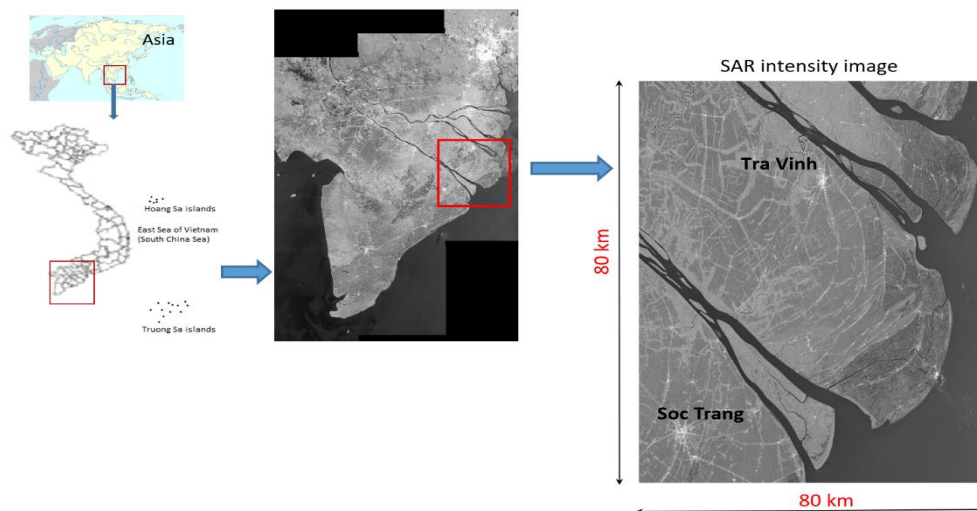
Figure 1. Tra Vinh study area coveraged about 80 km × 80 km. The background image is the SAR intensity.

The SAR stack is from Copernicus Sentinel–1 C–band 2015–2019. To reduce the dimensional of the computational data, we selected a good image for each month, resulting 55 images for PS/DS processing. The baseline distribution is shown in Figure 2.
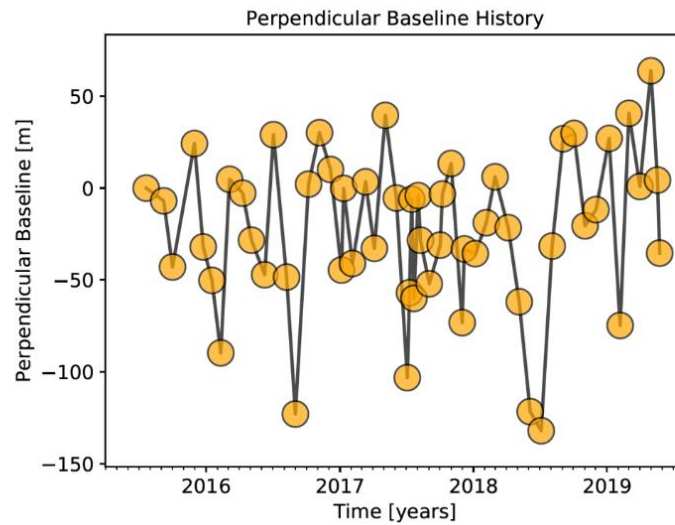
**Figure 2.** Baseline history of the 55 Sentinel–1 SAR images.

3. Results and Discussions

First of all, to provide better understand the presence of the stable points, we calculate the coherence of the stack and report in Figure 3.
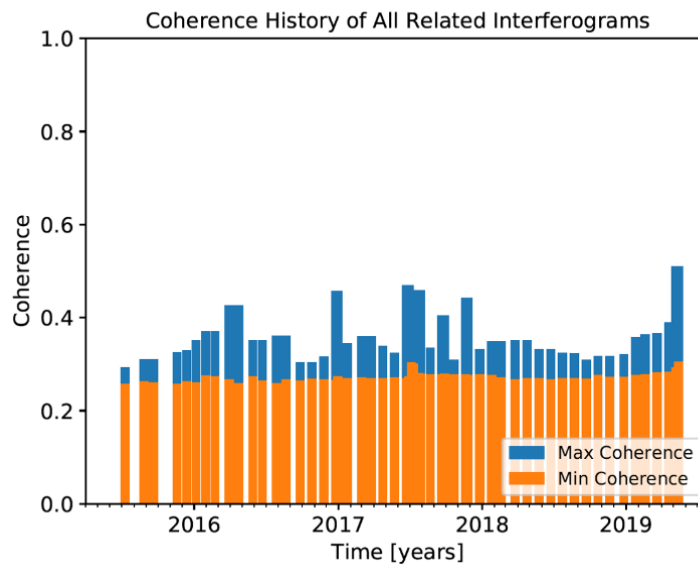
Figure 3. The average velocity trend: Positive velocities (green colors) represent movement uplift; negative velocities (red colors) represent movement subsidence.

The SAR data stack was then processed by using the MLE PS/DS approach. By exploiting the phase information at the PS/DS only, we are able to unwrap all the interferograms. In Figure 4, the unwrapped phase mostly varies from -2π to $+2\pi$. By inversion all the unwrapped interferogram, the average velocity (mm/yr) can be determined as in Figure 5. Positive velocities (blue colors) represent movement uplift; negative velocities (red colors) represent movement subsidence.
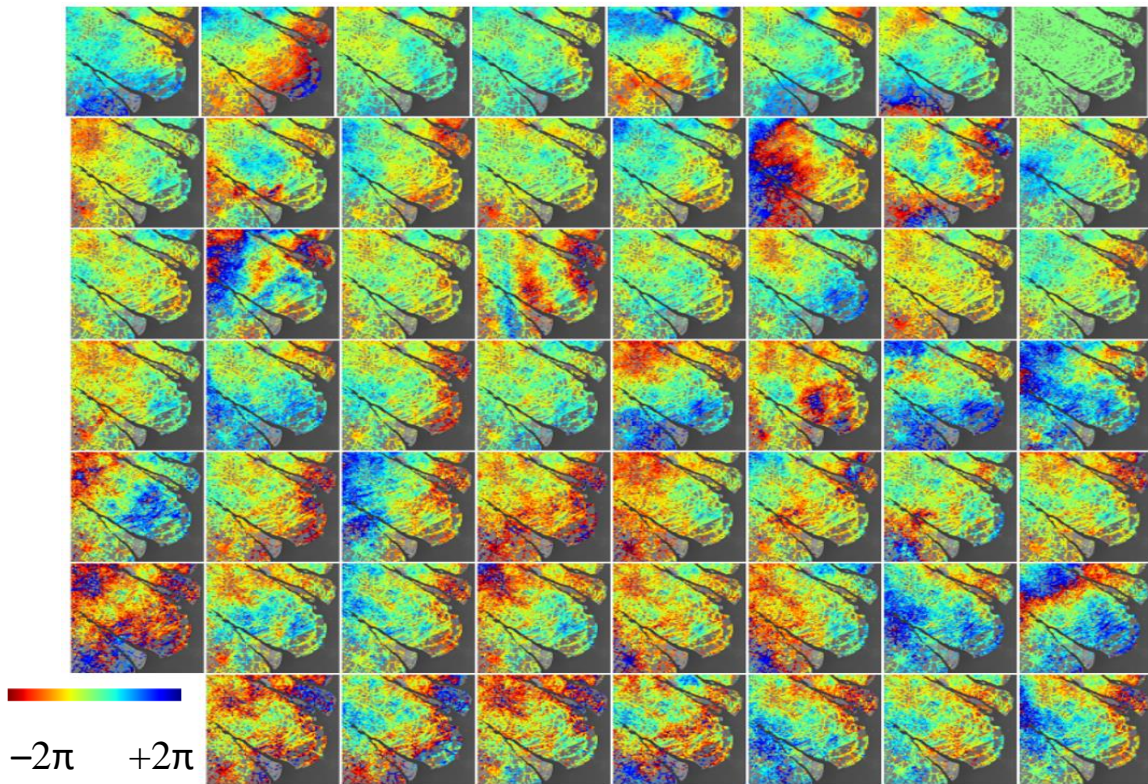
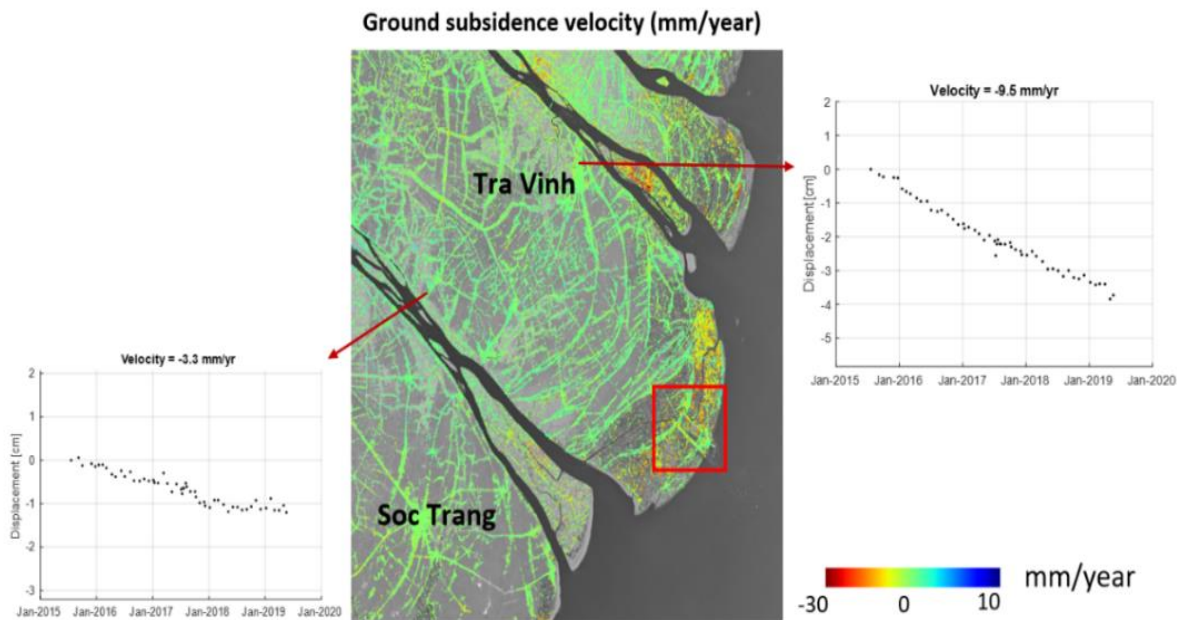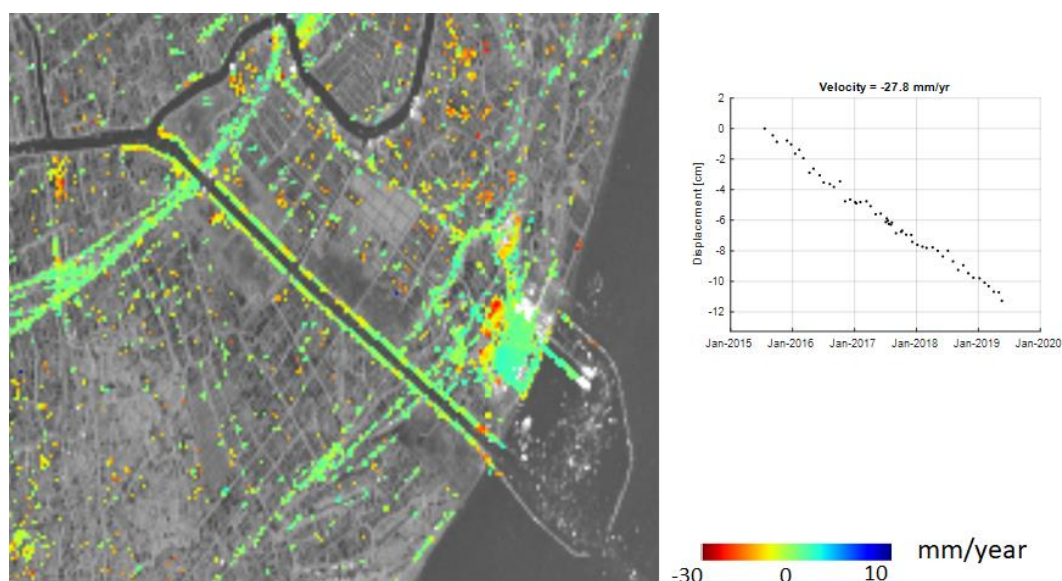**Figure 4.** The unwrapped interferogram. The reference date of calculation is 21 July 2015.

Figure 5. The average velocity trend. Positive velocities (blue colors) represent movement uplift; negative velocities (red colors) represent movement subsidence. The subsidence history is also provide to appreciation the displacement information.

Figure 6. A zoom version of Figure 5 at the Quan Chanh Bo channel.

4. Conclusions

First of all, in figure 3, we can observe that the average coherence of each image ranges from 2.5 to less than 0.5. This indicates the areas is quite challenge for InSAR processing due to less stable measurement points and low coherent signals. However, as the culture of the Mekong, people habitant along the road and river channels, resulting in many scatters can be detected for InSAR processing. This phenomenon is very visible in Figure 5. Thus, we were able to map the subsidence for the whole Tra Vinh.

The average velocity varied from –30 to 10 mm/year. The subsidence phenomena were found mostly in the costal of the sea (the east of the province). Particularly, red pixels correspond to subsidence of nearly 12 cm, from 2015 to 2019, averaging 27.8 mm/yr (see the dots in Figure 6). For comparison, pixels near Soc Trang exhibit an average of 3.3 mm/yr (bottom left plot Figure 5).

The main reason of subsidence of Tra Vinh can be divided into two main kinds. The first kind is natural reason and the other one is artificial reason. Natural reasons include compact sediment, seasonal fluctuation of groundwater and soft soil layers. The artificial reasons include exploitation groundwater; loading capacity by building or levelling; dynamic loading by transportation or under construction building; and suffusion, quicksand negative skin friction. Related to artificial reason, exploitation groundwater pumping wells is most affected.

Groundwater in the Tra Vinh delta is widely pumped from several aquifers ranging from Holocene to Miocene age, where deep aquifers are the most heavily exploited [10]. When groundwater is extracted, pore pressures are reduced and sedimentary layers undergo compaction that can be measured as land subsidence. One of the reasons for such rapid land subsidence is the excessive exploitation of underground water in these areas, especially in Dan Thanh commune, Duyen Hai district.

To conclude, the ground deformation result from SAR of the period 2015–2019 in Tra Vinh describes exactly the subsidence area. This can help to identify hotpot subsidence areas. By using PS/DS processing, it can provide not only the average velocity of ground subsidence but also the subsidence history information. Future works need to focus on validation and also to evaluate the subsidence effects to sea–level rise due to global climate change. Finally, this work has a proof concept on the feasibility of measurement of subsidence for the Delta–wide based on Sentinel–1 data.

Author contribution statement: Conceived and designed the experiments: H.T.M.D., L.T.C.; Analyzed and interpreted the data: H.T.M.D.; Manuscript editing: L.T.C.; Performed the experiments: H.Q., B.T.V., T.V.H; contributed reagents, materials, analyzed and interpreted the data: H.T.M.D., L.T.C.; wrote the draft manuscript: H.T.M.D., L.T.C.

Competing interest statement: The authors declare no conflict of interest.

References

1. Bear, J.; Cheng, H.D.; Sorek, S.; Ouazar, D.; Herrera, I. (Eds) Seawater Intrusion in Coastal Aquifers: Concepts, Methods, and Practices. 1999.
2. Erban, L.E.; Gorelick, S.M.; Zebker, H.A.; Fendorf, S. Release of arsenic to deep groundwater in the Mekong Delta, Vietnam, linked to pumping–induced land subsidence. Proc. Natl. Acad. Sci. U.S.A. **2013**, *110*, 13751–13756.
3. Minh, D.H.T.; Trung, L.V.; Le Toan, L.T. Mapping Ground Subsidence Phenomena in Ho Chi Minh City through the Radar Interferometry Technique Using ALOS PALSAR Data. *Remote Sens.* **2015**, *7*, 8543–8562.
4. Ferretti, A.; Prati, C.; Rocca F. Permanent Scatterers in SAR Interferometry. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39 (1)*, 8–20.
5. Berardino, P.; Fornaro, G.; Lanari, R.; Sansosti, E. A New Algorithm for Surface Deformation Monitoring Based on Small Baseline Differential SAR Interferograms. *IEEE Trans. Geosci. Remote Sens.* **2002**, 2375–2383.
6. Ferretti, A.; Fumagalli, A.; Novali, F.; Prati, C.; Rocca, F.; Rucci, A. A New Algorithm for Processing Interferometric Data–Stacks: SqueeSAR. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49 (9)*, 3460–3470.
7. Minh, D.H.T.; Cuong, T.Q.; Pham, Q.N.; Dang, T.T.; Nguyen, D.A.; El–Moussawi, I.; Toan, T.L. Measuring Ground Subsidence in Ha Noi Through the Radar Interferometry Technique Using TerraSAR–X and Cosmos SkyMed Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12 (10)*, 3874–3884.
8. Torres, R.; Snoeij, P.; Geudtner, D.; Bibby, D.; Davidson, M.; Attema, E.; Potin, P.; Rommen, B.; Floury, N.; Brown, M. et al. GMES Sentinel–1 mission. *Remote Sens. Environ.* **2012**, *120*, 9–24.
9. Prats–Iraola, P.; Scheiber, R.; Marotti, L.; Wollstadt, S.; Reigber, A. TOPS Interferometry with TerraSAR–X. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3179–3188.
10. Minderhoud, P.S.J.; Erkens, G.; Pham, V.H.; Vuong, B.T.; Stouthamer, E. Assessing the potential of the multi–aquifer subsurface of the Mekong Delta (Vietnam) for land subsidence due to groundwater extraction. *Proc. Int. Assoc. Hydrol. Sci.* **2015**, *372*, 73–76.
11. Bamler, R.; Hartl, P. Synthetic aperture radar interferometry. *Inverse Problems* **1998**, R1–R54.
12. Hanssen, R.F. Radar Interferometry: Data Interpretation and Error Analysis. 2001.
13. Rocca, F. Modeling Interferogram Stacks, Geoscience and Remote Sensing. *IEEE Trans.* **2007**, *45 (10)*, 3289–3299.
14. Minh, D.H.T.; Ngo, Y.N. TomoSAR platform supports for sentinel–1 TOPS persistent scatterers interferometry. *IEEE Trans. Geosci. Remote Sens. Symp.* **2017**, 1680–1683.

Soil topography index and SCS – curve number approach in identifying hot spots of runoff potential areas

Anh Bui Khanh Van[1]*

[1] Faculty of Environmental Department – Ho Chi Minh City University of Natural Resources and Environment; bkvanh2020@gmail.com, bkvanh@hcmunre.edu.vn

*Corresponding author: bkvanh2020@gmail.com; bkvanh@hcmunre.edu.vn; Tel.: +84–908836115

Abstract: Runoff reduction is the goal of soil and water conservation in agricultural watersheds. Through the runoff, many substances of soil such as sediment, nutrients have been eroded to end up in streams, rivers, and lakes. In decades, studies have revealed various mitigation, including structure and non–structure conservation ranging from field scale to watershed scale. However, the challenges for effectiveness improvement have increased in recent years within the impacts of anthropogenic activities such as land use land cover change and fluctuation in weather conditions. As a result, the runoff generation has been changing in both terms of quantitative and variable sources areas of runoff generation. From the understanding of runoff generation mechanisms, including infiltration excess and saturation excess, this study was conducted with the objective to propose an application of the Soil Topographic Index (STI) and the Soil Conservation Service Curve Number (SCS–CN) in identifying the areas with high runoff propensity. The method utilized GIS–based indices to indicate the high runoff potential areas. The ranking maps were evaluated by Wilcoxon rank sum test and Getis–Ord Gi* spatial statistics. Results demonstrated that there was a statistical significance of the greater STI in inundated cultivation than STI in cultivation areas. However, STI values were not statistically significant in pasture areas. Alternatively, the combination of STI and SCS–CN detected the statistical significance between calculated indices and inundated observed areas. In conclusion, the combination between STI and SCS–CN values is a potential method in redefining runoff generation hot spots.

Keywords: Runoff generation mechanism; SCS–Curve number; Soil topographic index; Ranking approaches; Hot spots and cold spots.

---

1. Introduction

Runoff and agricultural best management practices (Agricultural BMPs) in soil and water conservation has been a research topic for decades. In 1979, agricultural BMPs controlling runoff resulted in effectiveness of agricultural BMPs was pointed out [1]. Accordingly, the appropriate agricultural BMPs is the lining up between types and purposes of BMPs, which is relevant to the term of targeted conservation, recently. For this reason, the misleading in runoff generation which may cause inaccuracy in identifying high runoff areas has been mentioned in some researches. Since then, many of researches with the purposes to fill the deficiency between BMPs design and the runoff generation mechanisms have been conducted [2–3]. Most recently, 7Rs – Right product, right conservation practices, right

place, right scale, right rate, right method, and right time – again plays an important role in precision of soil and water conservation [4].

Subsequently, there has been many studies concerning either differentiation or combination between infiltration excess and/or saturation excess in runoff generation mechanisms, which influence the results of runoff generation in term of temporal scales and variable sources areas of runoff proneness [2, 5–10]. Importantly, the implication that the Soil Conservation Service Curve Number approach (SCS–CN method) should not be applied in the manner of only infiltration excess and excluding of saturation excess amongst many debates about the application of SCS–CN in rainfall–runoff model [8, 11].

Therefore, in this study, quantitative indices inferring qualitative rank of runoff generation were proposed with the approach of hot spots emergence mechanisms. Hot spot definition was initially proposed in 2003 [12]. The concept of hot spots are areas that show disproportionately high reaction rates related to the surrounding area (or matrix). Hot moments are short periods of time that show disproportionately high rates relative to longer intervention time periods. Emergence of hot spot hot moments highlighted the heterogeneity characteristics of the phenomenon. Hot spot means the spatial intensive concentration of phenomenon at high rate, and hot moment refers to the temporal dimension, during periods of time the phenomenon was enhanced. Hot spots and hot moments may overlap or separate. Most importantly, hot spots identification strongly depends on generating mechanisms. In other word, the meaning of understanding the mechanism is that it can be utilized to predict hot spots in the future [12].

In this research, runoff occurrence was considered a hot spot–hot moment approach due to similarity in disproportionately insightful emerged mechanism. Therefore, utilizing the hot spots emergence mechanism to contribute the research methodology is a potential approach. Also, hot spot theory can apply to emerge the hot spots of critical source areas of sediment, nitrogen and phosphorus. The second condition is the spatial scale and temporal scale which are the considerable factors in hot spots identification. For instance, the forming of surface runoff depends on the integrated impacts of topology, rainfall, soil profile and crop–scape in agricultural watersheds. The high runoff propensity area is the area that satisfies all the high conditions of four features. Therefore, the ultimate distribution of high runoff areas is defined as the areas within the overlapping of these characteristics.

GIS – based indices application has not been a new approach but there is still a lack of using GIS – based indices to propose the hot spots of runoff generation mechanisms in combination of SCS – Curve number approach and the Soil Topographic Index (STI). From the perspective of runoff generation mechanisms, including infiltration excess and saturation excess, this study is conducted with the objective to propose an application of GIS – based indices in identifying the areas with high runoff tendency. The study focuses on answering the question of how to precisely define the high runoff areas in order to propose a suitable soil and water conservation practices and explicit placement of agricultural BMPs. In order to answer the research question, the analogy of combination SCS–CN and STI was proposed.

2. Methodology

Study area was Callahan Creek watershed in Boone County in Missouri with the area approximately 21,960 acres (89 km^2). Location of the basin was as in figure 1. The land use land cover (LULC) types mainly comprise forestry and agricultural areas such as corn, soybean, winter wheat, hay, grass, pasture, and deciduous forest. Also, this area is one of PL–566 watershed projects – The Watershed Protection and Flood Prevention Act.

The Digital Elevation Model (DEM) was downloaded from the U.S. Geological Survey (USGS) database and soil survey data, land use land cover from the U.S. Department of Agriculture (USDA) cropland database at 30m × 30m resolution. The record of inundated areas for Missouri and Illinois in May and June 2019 from National Agricultural Statistic

Services (USDA–NASS) was employed as observed data. The observed inundated cultivation and inundated pasture in this area was extracted; afterward, this observed data was converted to point vector data by GIS toolboxes in order to analyze attribute data and execute statistical hypotheses test by R packages.
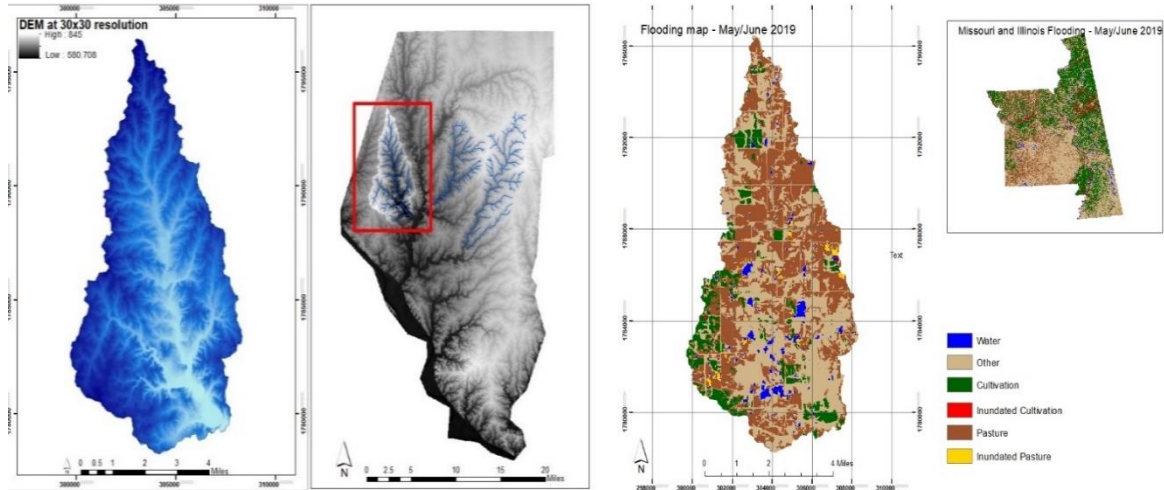
Figure 1. Location of Callahan basin (on the left) and inundated map of Callahan basin (on the right), Boone County, Missouri.

In this study, a proposed approach is that the curve number method can contribute to qualitative evaluation in the linkage of hydrologic soil groups, land–use types, and the condition of land cover. SCS–CN method was applied by the U.S. Department of Agriculture (USDA) in 1972, SCS–CN values indicate the linkage of soil types, antecedent moisture condition, land use types, and surface conditions [13]. Employing SCS–CN to produce runoff map was mentioned in previous study [14]. Subsequently, the STI was originally developed based on the Topographic Wetness Index (TWI) which was respectively published in 1979, 2000 and 2002 [9, 10, 15]. This index demonstrated interaction between topography and soil physical features such as depth of soil and saturated hydraulic conductivity. Applying TWI to produce runoff ranking map was proposed in studies [3] and other indexes for targeted agricultural BMPs was also mentioned [16]. Most significant, the Soil Topographic index (STI) was explained and applied in previous studies [9, 17, 18]. According to, STI was calculated as in formulation (1):

$$STI = \ln\left(\frac{\alpha_i}{\tan(\beta_i)k_s D}\right) \tag{1}$$

where STI is soil topography index; α_i is upslope contributing area per unit contour length (m); $\tan(\beta_i)$ is the local surface topographic slope; k_s is the mean saturated hydraulic conductivity of the soil (m/day); and D is the soil depth to restrictive layer (m).

After that, the calculated raster data was executed spatial join with the SCS–CN and converted to point vector data, employing the explanation and the guidelines from USDA in 1986 to identify suitable curve number values [13]. Accordingly, there are four different types of SCS–CN values depending on Hydrologic Soil Group (HSGs) in drained conditions and LULC conditions. The first letter in the dual HSGs applies to the drained condition, the second letter in the dual HSG applies to the undrained condition. As a result, there were four different scenarios of SCS–CN values according to the combination of draining conditions and LULC conditions. However, the study compared the observed inundated maps at the time in which the extreme inundated event occurred in May of 2019. Therefore, the others

scenarios were eliminated, only SCS–CN values of undrained condition and LULC in poor condition were taken into evaluation.

Notably, there were inconsistency between LULC map and observed inundated map. There were 25 LULC types, while only 6 reclassified types of inundated areas were in the inundated areas map. For example, the deciduous forest in LULC data were reclassified into 6 different types in inundated map which were inundated cultivation, cultivation, inundated pasture, pasture, other, and water areas. Beside the uncertainty of LULC and flooding map, the assumption which were HSGs, LULC, saturated hydraulic conductivity, depth to restrictive layer of the soil in mean values and depth to restrictive layer as in the soil survey database were assigned to calculate STI indices and determine SCS–CN values. Ideally, these parameters should be as realistic as possible to reduce uncertainty in calibration.

In spatial statistic, three distinctions of ranking approaches were proposed and assessed by Getis–Ord Gi* hotspots z scores developed by Getis and Ord to analyze spatial patterns [19–21]. Spatial autocorrelation was developed based on the first law of geography which is "everything is related to everything else, but near things are more related than distant things" [22]. Getis–Ord Gi* scores were calculated to illustrate the spatial autocorrelation, though this statistic cannot interpret the reason why locations that have statistically significant hot spots or cold spots. In other words, this method cannot identify the mechanism which causes hot spots or cold spots. In this study, this approach was utilized to clarify the number and the location of hot spots. Above all, hot spots of three distinct ranking approaches infer the three STI and SCS–CN combinations. Finally, the three distinguished distributions of calculated hot spots were compared with the distribution of observed inundated data.

3. Results and Discussions

Subsequently, the STI index of Callahan Creek ranged from 3.4 to 27.5 (Figure 2). The interpretation is that the higher STI and SCS–CN values represent the higher runoff potential. STI distribution was highly skewed to the right which is not normal distribution with the density curve is not symmetric and bell–shaped. Quantile plot of SCS–CN values describing poor and undrained condition in inundated areas and other types in flooding map evidenced a non–normal distribution. The plots indicated the systematic deviations from a straight line. Outliers appeared as points that were far away from the overall pattern of the plot. Therefore, the statistical method of non–parametric approach was appropriate in these conditions.

The attribute data obtained two types of variables, including numerical variables and categories variables. SCS–CN and STI are numerical variables. The inundated areas of flooding map such as cultivation and pasture are categories variables. According to the quantile plot of the variables, variable distributions are not normal distributions. Therefore, the non–parametric test which are Kruskal Wallis rank sum test, Wilcoxon rank sum test, bootstrapping interval confidence calculation were applied to compare between inundated areas and out of inundated areas. Spatial point data using Getis – Ord Gi* hot spot was applied to identify significant hot spots. The principles of combination between SCS–CN and STI were summarized as in figure 3.

In this proposed method, observation data of inundated areas was important since it was utilized to assess the accuracy of the proposed indices. In observed data, each pixel represented the condition of inundated areas. In this study area, the inundated map recorded the historical inundated of Missouri in May 2019. Accordingly, the inundated areas mainly distributed in the cultivated areas and pasture areas. This distribution not only occurred in nearly streamflow location. The higher SCS–CN and STI inferred the higher potential of runoff. In this study, the condition of LULC at the recorded extreme event was dramatically poor condition without cover crops and low capacity of drainage condition.
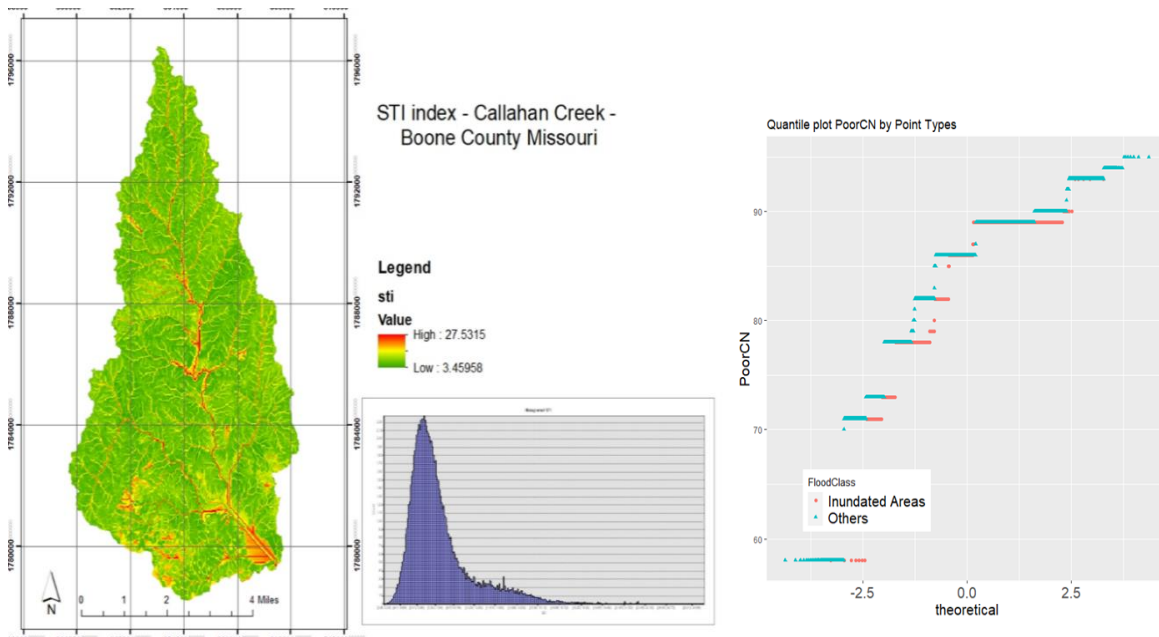
Figure 2. Narrative description of STI and SCS–CN values in poor and undrained conditions.
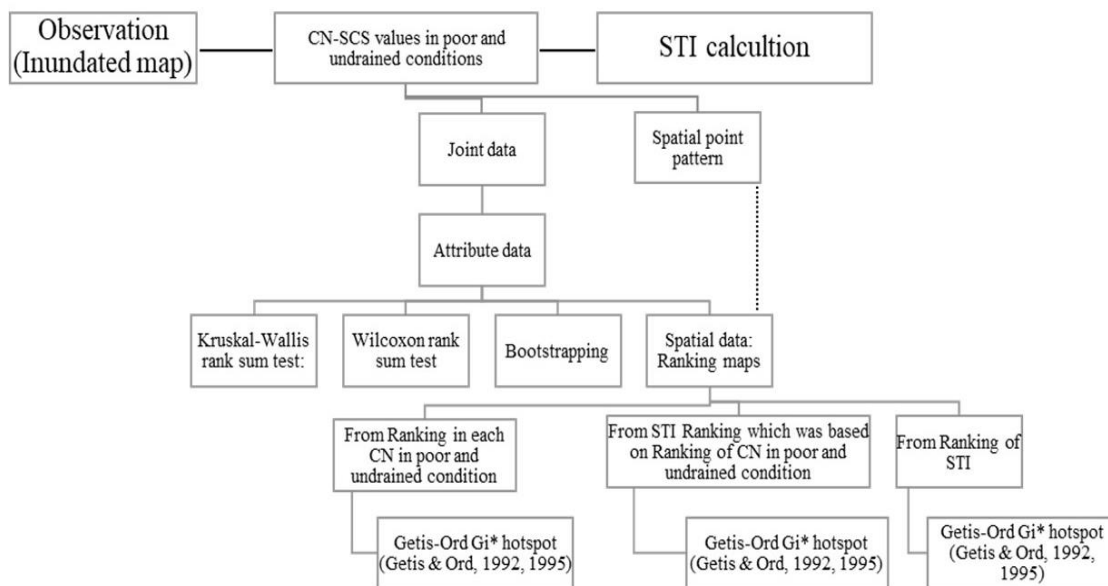
Figure 3. The proposed method of combination between SCS–CN and STI.

Comparing median values between STI inundated areas and STI out of inundated areas illustrated that there was a significant difference between STI and SCS–CN in poor and undrained condition of all groups. The Null hypothesis was that median of STI in inundated areas and median out of inundated areas were equal. The alternative hypothesis was that median in inundated areas was different from median out of inundated areas. Figure 4a revealed that p value proved statistically significant difference in pair comparisons of STI between inundated cultivation and cultivation, inundated pasture and inundated cultivation, cultivation and pasture. However, STI between inundated pasture and pasture were not different. Figure 4b revealed that p value indicated statistically significant difference in pair comparisons of all SCS–CN in poor and undrained condition, but SCS–CN in inundated areas were less than those of drained areas. Figure 4c, density distribution highlighted the higher values of STI in inundated cultivation in comparison to STI in cultivation. In contrast, figure 4d revealed the lower values of SCS–CN in inundated cultivation.
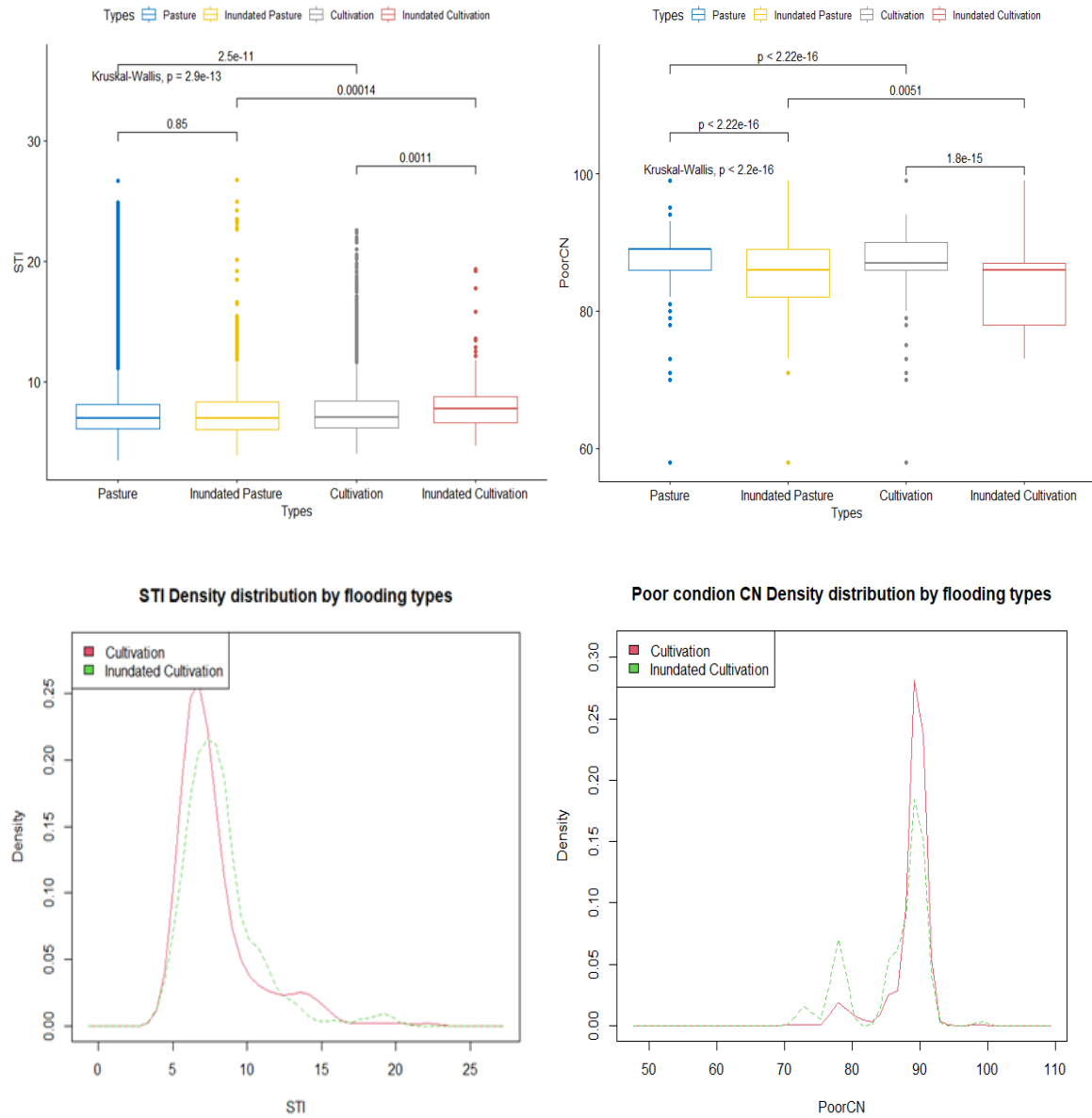
Figure 4. Kruskal–Wallis rank sum test: (a) Kruskal–Wallis test of STI in 4 types of flooding maps; (b) Kruskal–Wallis test of SCS–CN in poor and undrained condition in 4 types of flooding maps; (c) Density distribution of STI in cultivation and inundated cultivation; (d) Density distribution of SCS–CN in cultivation and inundated cultivation.

Table 1. Wilcoxon rank sum test in pair comparison.

Variables	Group x	Group y	W	p–value	Alternative Hypothesis
Wilcoxon rank sum test in the undrained condition					
STI	Inundated_Cultivation	Cultivation	721189	0.0005622	Greater
PoorCN	Inundated_Cultivation	Cultivation	375646	9.25E–16	Less
PoorCN	Inundated_Pasture	Pasture	19321152	2.20E–16	Less

In order to evaluate in pair comparison, Wilcoxon rank sum test computed the value of p as in Table 1. In STI comparison between inundated cultivation and cultivation, p value was 0.0005622, the null hypotheses were rejected and the alternative hypotheses that medians of STI distributions for inundated cultivation were statistically significant and greater than

that of cultivation areas. It means that the STI distributions for inundated cultivation are likely shifted to the right of the STI distributions for cultivation. In SCS–CN evaluation, p values are < 2.2e–16, the null hypotheses were rejected and the alternative hypotheses that medians of SCS–CN values distributions for inundated cultivation were statistically significant and less than those of cultivation and pasture. It means that the SCS–CN values distributions for inundated cultivation and inundated pasture are likely shifted to the left of the SCS–CN distributions for cultivation and pasture. It infers that SCS–CN values in all cases of inundated areas were less than those of drained areas.

Bootstrap confidence interval calculations in table 2 emphasized the similarity to Wilcoxon rank sum test. Only confidence interval of STI between inundated cultivation and cultivation were less than 0, bootstrap confidence interval based on 10,000 bootstrap replicate times were –0.73851. The others results were greater than or equal to 0. These results reinforced the hypothesis that STI in inundated cultivation were higher than STI values in cultivation areas.

Table 2. Bootstrap confidence interval based on 10000 bootstrap replicates.

| Variables	Group 1	Group 2	Resample	Original	BootBias	BootSE	Method	BootMed
STI	Inundated_Cultivation	Cultivation	10000	–0.73623	0.003979	0.21386	Med diff	–0.73851
STI	Inundated_Pasture	Pasture	10000	0.0188	0.002326	0.063783	Med diff	0.01735
PoorCN	Inundated_Cultivation	Cultivation	10000	1	0.0524	0.38369	Med diff	1
PoorCN	Inundated_Pasture	Pasture	10000	0	0.56395	1.1	Med diff	0

The previous results clarified that separately applied STI and SCS–CN values insufficiently reflected the complex mechanisms of runoff generation in a basin. It is crucial to appropriately assemble indices to expose the underlying dynamic processes. Thus, STI and SCS–CN in poor and undrained condition were unified to highlight the trend. In figure 5, the boxplot comparison of STI and SCS–CN in poor condition between inundated cultivation and cultivation in undrained condition indicated that the p value of Wilcoxon rank sum test at SCS–CN 90 was significant. For the inundated pasture, p values of Wilcoxon rank sum test at SCS–CN 78 and 79 were significant. At SCS–CN 71, 82, 86, STI of pasture were higher than STI of inundated pasture.
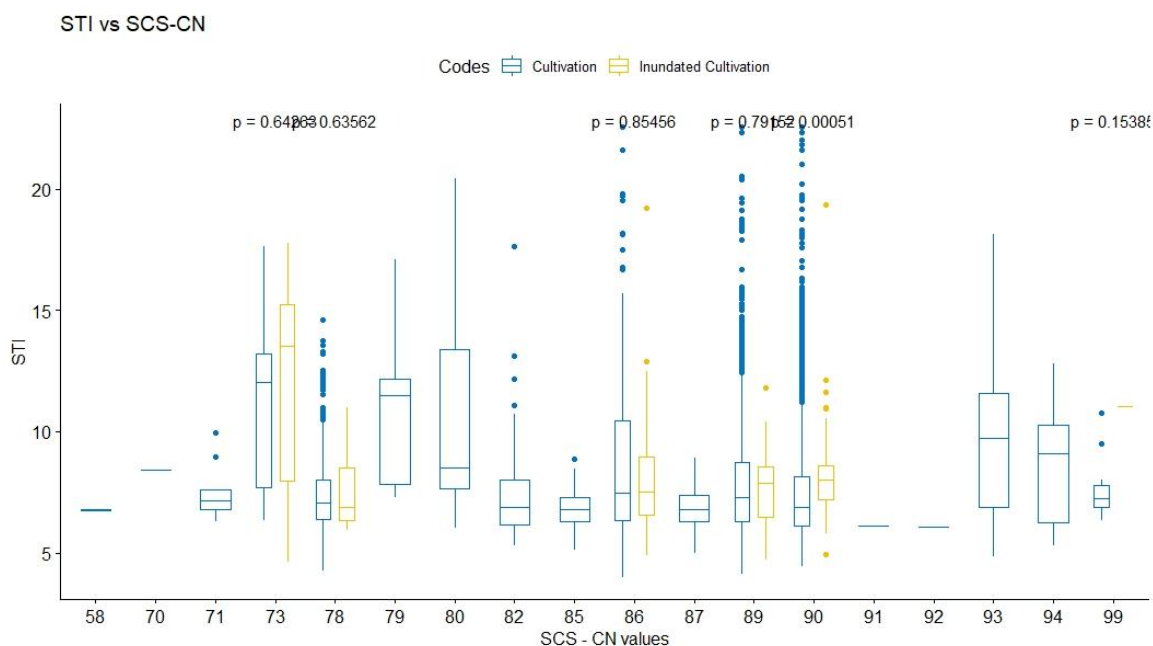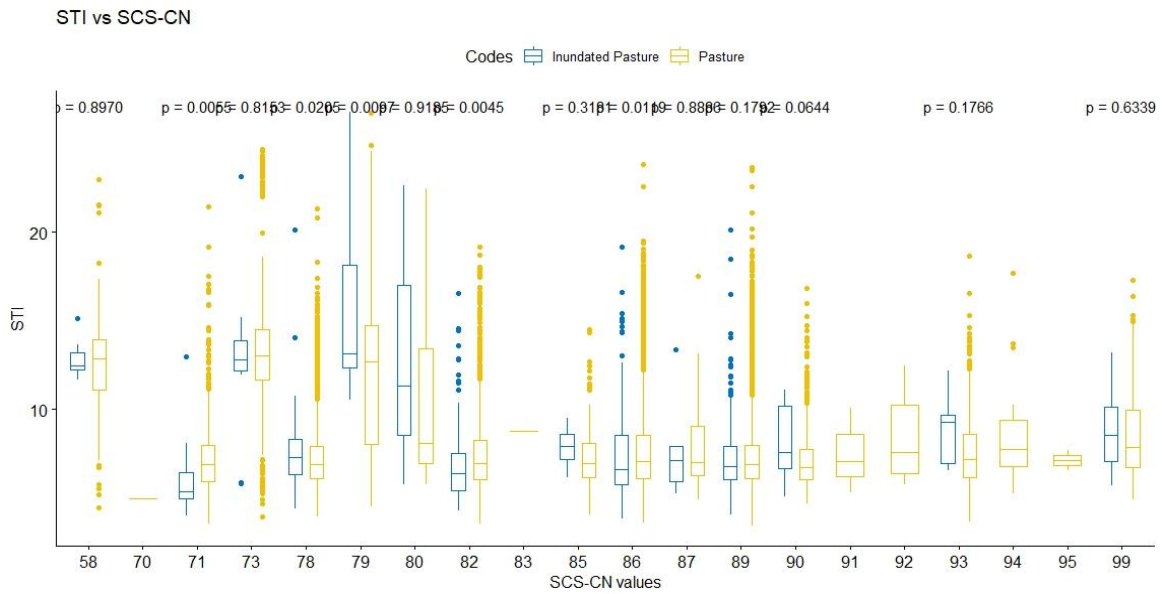
Figure 5. Parallel comparison of STI and SCS–CN in poor and undrained condition; (a) Wilcoxon rank sum test of STI in pairs inundated cultivation and cultivation in each SCS–CN value; (b) Wilcoxon rank sum test of STI in pairs inundated pasture and pasture in each SCS–CN value.

According to the validated results, there were three different ranking approaches. The first ranking was solely based on STI values, the second approach was from STI ranking which was based on the rank of SCS–CN in poor and undrained condition, the third approach was from the ranking in each SCS–CN in poor and undrained condition. Subsequently, Getis–Ord Gi* hotspots were calculated in each ranking map utilizing GIS toolboxes. In the first approach (Figure 6), correlation coefficient between STI and STI ranking was 1 because the ranking was only based on STI values. This ranking simplified that a point with a larger STI represented a higher STI ranking. Overall, the STI ranking changed exponentially with the STI values from SCS–CN 58 to SCS–CN 99. Curve number values did not affect this type of ranking.
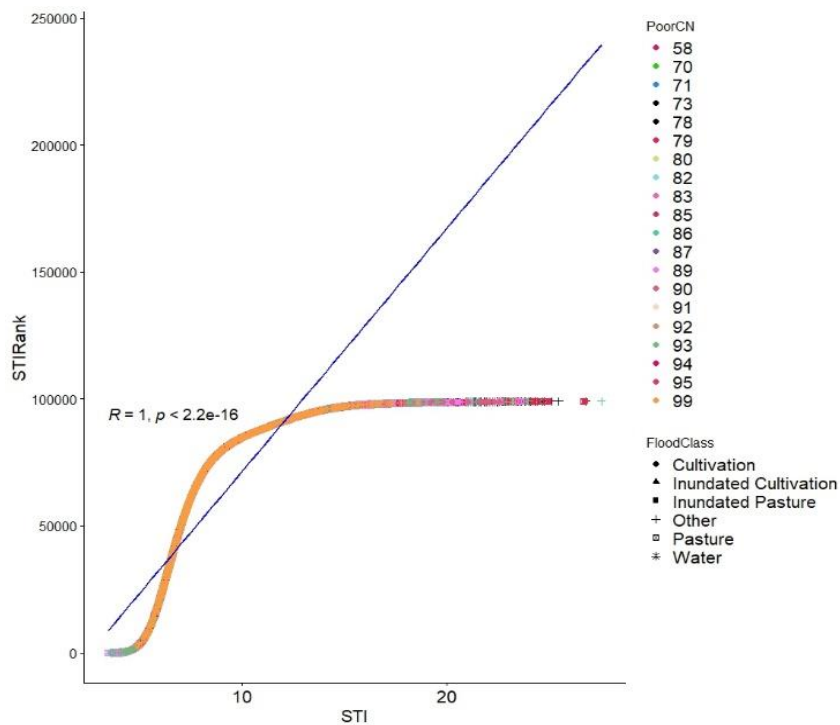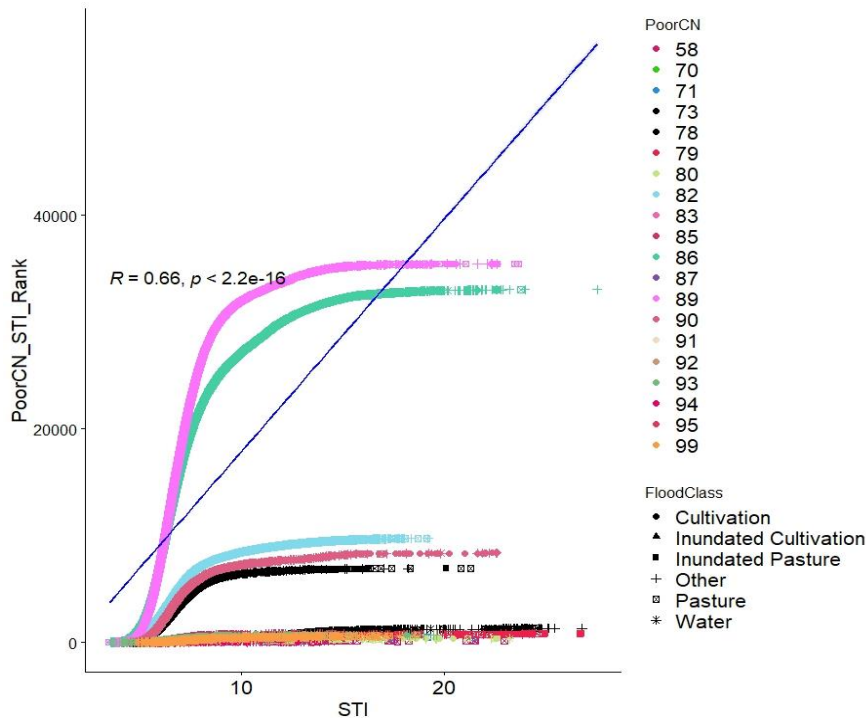

Figure 6a. STI values and Ranking of STI.
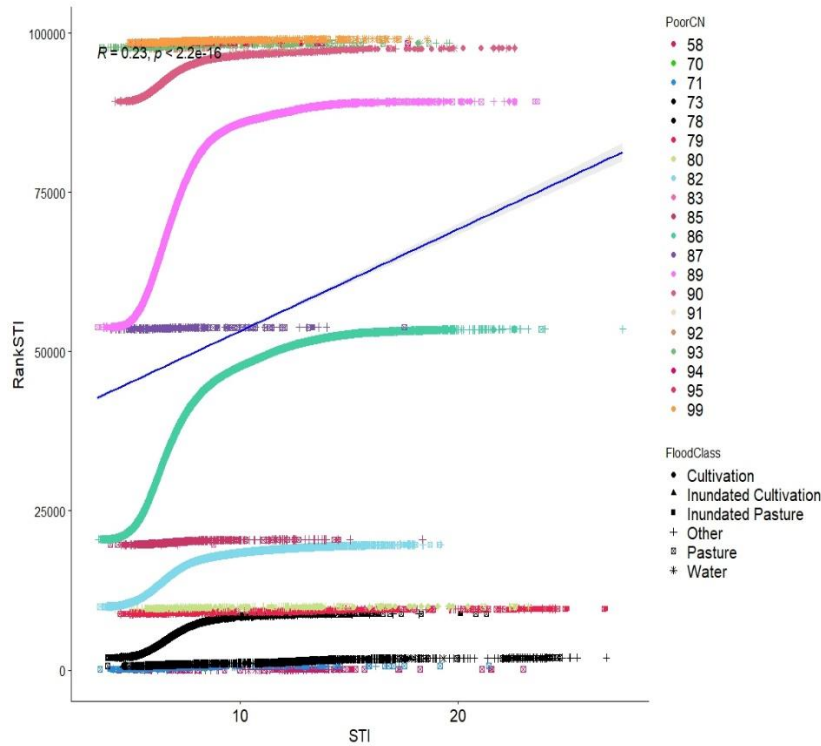
Figure 6b. STI values and Ranking based on Ranking of SCS–CN in poor and undrained condition.

Figure 7. STI based on each CN values ranking map.

The ranking based on the increase of the STI ranking highly depends on the increase of SCS–CN values (Figure 6b). The idea is the ranking of STI depends on the increase of SCS–CN values rather than the increase of STI. Thus, correlation coefficient between STI values and ranking of STI was 0.28. First, SCS–CN values were increased from 58 to 99. Then The STI were sorted from smallest to largest. The lower SCS–CN values led to the lower ranking of STI. In the third ranking approach (Figure 7), STI ranking was based on each SCS–CN values ranking. First, SCS–CN values were arranged from smallest to largest. Second, in each

SCS–CN value, STI were arranged from 1 to largest. The correlation coefficient was 0.66 because the ranking depended on the increase of both SCS–CN values and STI values. In this case, ranking depend on STI values rather than SCS–CN values.

The Getis–Ord Gi* statistic computed a z–score in each feature in the dataset. For statistically significant positive z–scores, the larger the z–score is, the more intense the clustering of high values, named hot–spot. For statistically significant negative z–scores, the smaller the z–score is, the more intense the clustering of low values, named cold–spots. From three different ranking approaches, three different hot spot maps were generated, including a hot spot map of STI ranking, a hot spot map of STI based on SCS–CN values ranking, and a hot spot map of STI ranking based on STI in each SCS–CN values. Table 3 and figure 8 analyzed the distribution of hot spots and cold spots of the third ranking approach, in which STI ranking was arranged in each CN values, had the nearest distribution to the distribution of inundated areas in each LULC types.

Table 3. Total number of hotspots and cold spots in the third approach ranking map.

| Hot spots and Cold spots in comparison with flooding map areas	LULC types	Number of Cold spots			Number of Hot spots			Total cold spots and hot spots in each LULC types
		–3	–2	–1	1	2	3	
Inundated Pasture	Corn						7	7
	Deciduous Forest	51	38	19	2			110
	Developed/Open Space		3		2	4		9
	Grass/Pasture	31	31	28	89	72	9	260
	Open Water				1	6	9	16
	Other Hay/Non–Alfalfa	52	37	4				93
	Soybeans				1	2	2	5
	Shrubland	3	1		1	1		6
Total cold spots and hot spots of inundated pasture in all LULC types		137	110	51	96	85	27	506
Inundated cultivation	Corn					2	30	32
	Deciduous Forest	4	2	1	1			8
	Developed/Open Space				2			2
	Grass/Pasture		1		2	7	3	13
	Open Water						1	1
	Other Hay/Non –Alfalfa	17	5	1				23
	Soybeans					1	8	9
Total cold spots and hot spots of inundated cultivation in all LULC types		21	8	2	2	12	34	88
Total cold spots and hot spots of inundated areas in all LULC types		158	118	53	98	97	61	594

Figure 8 illustrated the comparison between observed inundated areas and hot spot and cold spot distribution from the third approach ranking map. In table 3, total number of cold spots and hot spots in corn LULC areas in inundated cultivation and inundated pasture was 39 points, while the number in the observed inundated map was 40 points in corn LULC type (Figure 8). Generally, the number of hot spots and cold spots in the other LULC types declined in comparison to observed inundated areas. However, in three approaches of the ranking maps, the third ranking approach turned in the best result in comparison with observed inundated data.

Table 3 emerged distribution of hot spots and cold spots in each LULC type and compared to the distribution of inundated points in observed inundated areas. Accordingly, distribution in the hot spots map and inundated areas distribution in observed inundated areas had a similar tendency. The number of hot spots and cold spots decreased from corn to soybean. The highest number of hot spots and cold spots were in grass/pasture in both

distributions. However, the number of points were neither hot spots nor cold spots, which turned in 0 value from Getis–Ord Gi* spatial statistics, and were not analyzed in this description. Figure 9 illustrated the distribution of hot spots and cold spots of inundated cultivation added in google map of Callahan creek basin.
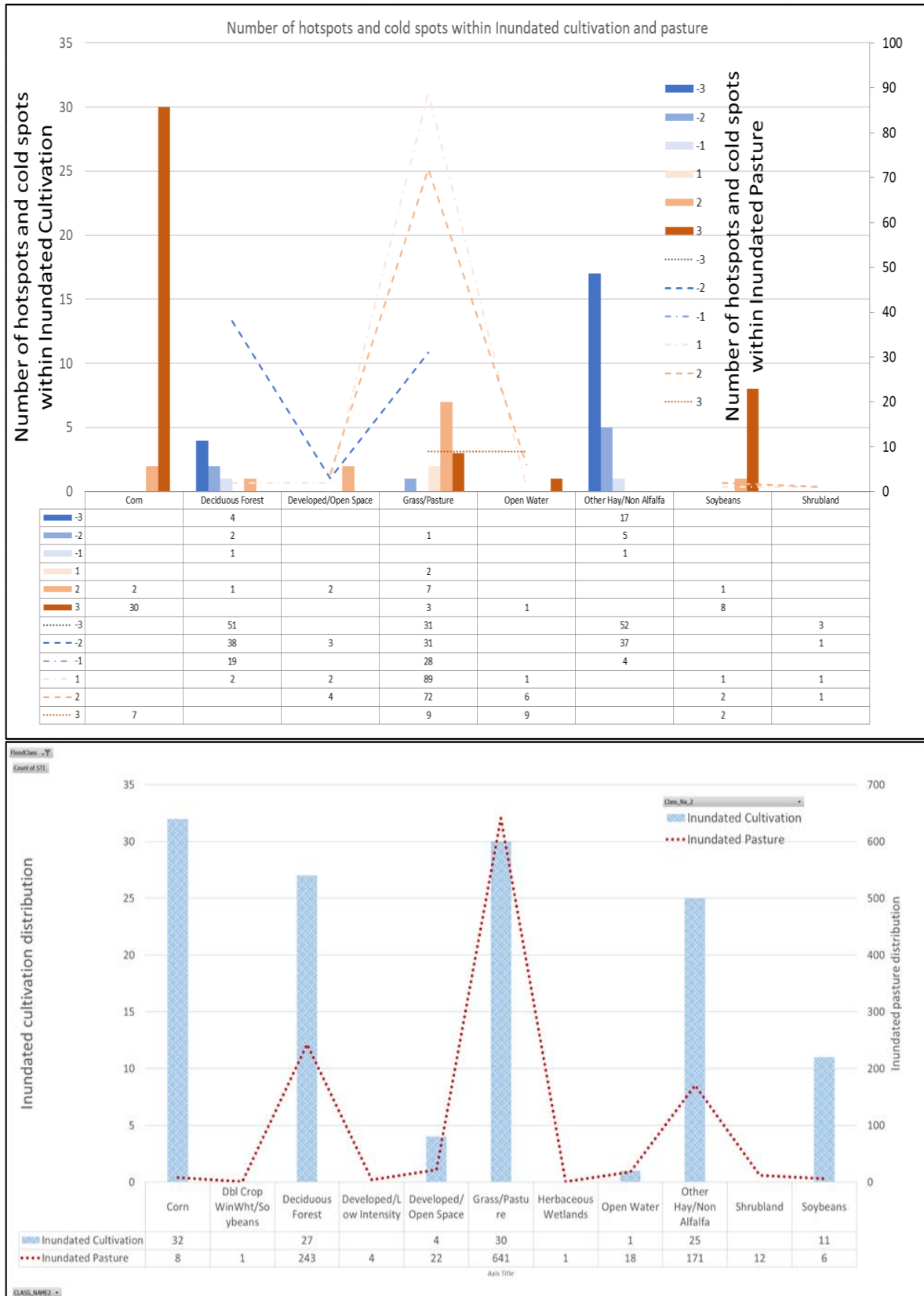
Figure 8. Comparison between observed inundated areas and hot spot and cold spot distribution from the third approach ranking map.
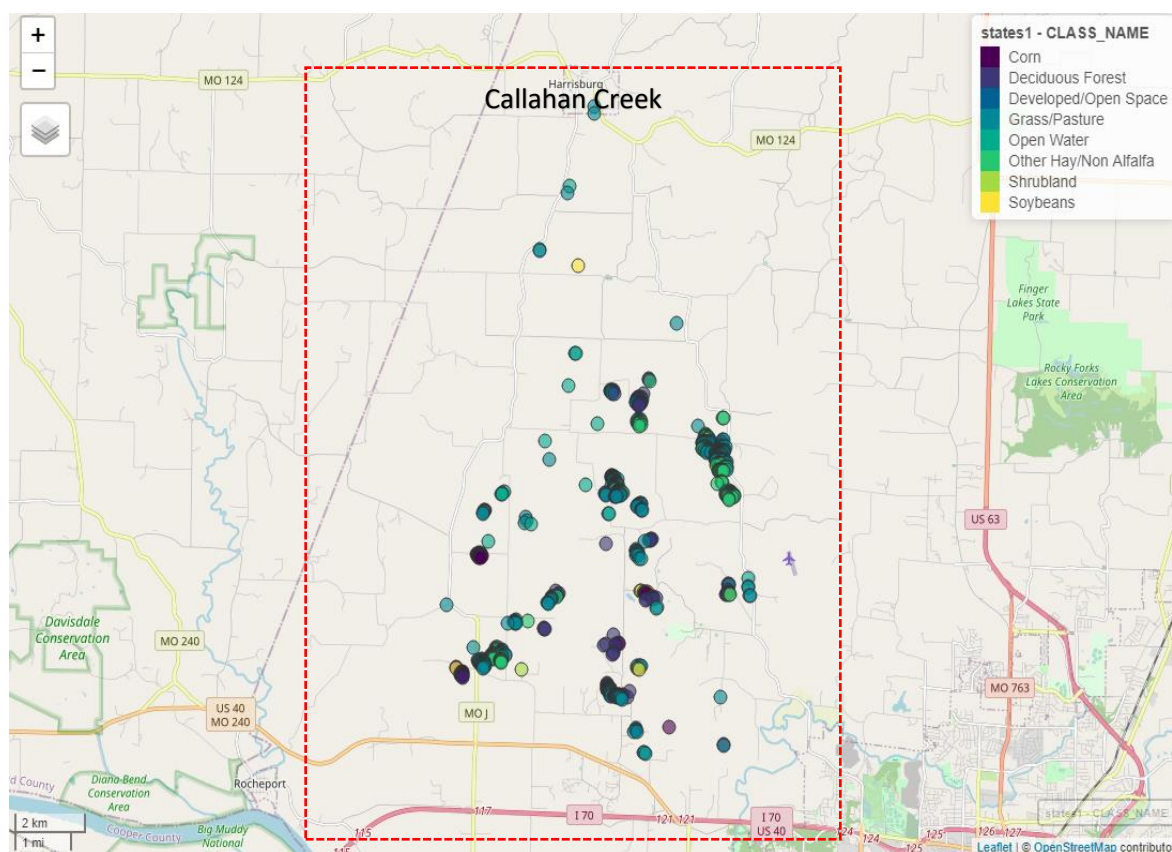
Figure 9. Distribution of hot spots and cold spots of inundated cultivation on google earth map.

4. Conclusions

There was a statistical significance of the greater STI in inundated cultivation than STI in cultivation areas. However, STI values were not significant in STI comparisons of pasture areas. This result requires further observation to analyze the complexity of hydrology in the study areas. In other words, saturated excess mechanism is insufficient to explain runoff generation in pasture areas in this watershed. Therefore, the combination between STI and SCS–CN values demonstrated by the statistical significance of inundated pasture STI refined the explanation of runoff generation mechanisms. At some SCS–CN value, STI in inundated pasture is statistically higher than those of drained pasture. Thus, strongly recommend that STI and SCS–CN values should be combined to produce a ranking map of high runoff potential areas. However, the results need to be validated by soil moisture monitoring or by field studies.

Hot spots and cold spots of runoff generation can emerge either within inundated areas or out of inundated areas. This challenge demands an appropriate scale analysis and hot spot generation mechanism since spatial and temporal resolutions of input data highly affect the hot spots of runoff generations. The distributions of hot spot and cold spots varied in each LULC type. This fluctuation requires high accuracy of land use land cover data as well as distribution of inundated areas. Ranking maps can be highly uncertain in many different runoff generation mechanisms. Therefore, it is necessary to be validated based on field research.

The saturated hydraulic conductivity values should be as precise as possible so that the values can represent the variation of spatial and temporal scale to emerge the hot spots and cold spots. Wilcoxon rank sum test highly depends on the rank sum value of observed data, and thus in this study, the quality of LULC data and observed inundated map significantly influences the results.

Author contribution statement: Data sources from USGS, USDA: seek and opt by Anh Bui KV; STI calculation following Soil and Water Lab document, Biological and Environmental Engineering, Cornell University. SCS–CN following USDA. Validation, Visualization: Application of GIS tools and R software. Literature review, methodology, research design, conducting calculation, data analysis, writing–original draft, writing–review & editing: Anh Bui KV.

Conflicts of Interest: The author declares that there is no conflict of interest.

References

1. Walter, M.F.; Steenhuis, T.S.; Haith, D.A. Nonpoint Source Pollution Control by Soil and Water Conservation Practices. *Trans. Am. Soc. Agric. Eng.* **1979**, *22(4)*, 0834–0840.
2. Schneiderman, E.M. et al. Incorporating variable source area hydrology into a curve–number–based watershed model. *Hydrol. Process.* **2007**, *21*, 25, 3420–3430. doi: 10.1002/hyp.6556.
3. Tomer, M.D.; Dosskey, M.G.; Burkart, M.R.; James, D.E.; Helmers, M.J.; Eisenhauer, D.E. Methods to prioritize placement of riparian buffers for improved water quality. *Agrofor. Syst.* **2009**, *75(1)*, 17–25. doi: 10.1007/s10457-008-9134-5.
4. Delgado, J.; Gantzer, C.; Sassenrath, G. Soil and water conservation A celebration of 75 years. 2020.
5. Agnew, L.J. et al. Identifying hydrologically sensitive areas: Bridging the gap between science and application. *J. Environ. Manage.* **2006**, *78(1)*, 63–76. doi: 10.1016/j.jenvman.2005.04.021.
6. Buchanan, B. et al. Estimating dominant runoff modes across the conterminous United States. *Hydrol. Process.* **2018**, *32(26)*, 3881–3890. doi: 10.1002/hyp.13296.
7. Easton, Z.M.; Fuka, D.R.; Walter, M.T.; Cowan, D.M.; Schneiderman, E.M.; Steenhuis, T.S. Re–conceptualizing the soil and water assessment tool (SWAT) model to predict runoff from variable source areas. *J. Hydrol.* **2008**, *348(3–4)*, 279–291. doi: 10.1016/j.jhydrol.2007.10.008.
8. Lyon, S.W.; Walter, T.M.; Gerard–Marchant, P.; Steenhuis, T.S. Using a topographic index to distribute variable source area runoff predicted with the SCS curve–number equation. *Hydrol. Process.* **2004**, *18(15)*, 2757–2771. doi: 10.1002/hyp.1494.
9. Todd Walter, M.; Steenhuis, T.S.; Mehta, V.K.; Thongs, D.; Zion, M.; Schneiderman, E. Refined conceptualization of TOPMODEL for shallow subsurface flows. *Hydrol. Process.* **2002**, *16(10)*, 2041–2046. doi: 10.1002/hyp.5030.
10. Walter, M.T.; Walter, M.F.; Brooks, E.S.; Steenhuis, T.S.; Boll, J.; Weiler, K. Hydrologically sensitive areas: Variable source area hydrology implications for water quality risk assessment. *J. Soil Water Conserv.* **2000**, *55(3)*, 277–284.
11. Garen, D.C.; Moore, D.S. Curve Number (CN) hydrology in Water Quality Modeling. *JAWRA J. Am. Water Resour. Assoc.* **2005**, *3224(03127)*, 377–388.
12. McClain, M.E. et al. Biogeochemical Hot Spots and Hot Moments at the Interface of Terrestrial and Aquatic Ecosystems. *Ecosystems* **2003**, *6(4)*, 301–312. doi: 10.1007/s10021-003-0161-9.
13. USDA Conservation Engineering Division and NRCS. Urban Hydrology for Small.

Soil Conserv. No. Technical Release 55 (TR–55), 1986, pp. 164.

14. Zhan, X.; Huang, M.L. ArcCN–Runoff: An ArcGIS tool for generating curve number and runoff maps. *Environ. Model. Softw.* **2004**, *19(10)*, 875–879. doi: 10.1016/j.envsoft.2004.03.001.

15. Beven, K.J.; Kirkby, M.J. A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* **1979**, *24(1)*, 43–69. doi: 10.1080/02626667909491834.

16. Dosskey, M.G.; Qiu, Z.; Helmers, M.J.; Eisenhauer, D.E. Improved indexes for targeting placement of buffers of Hortonian runoff. *J. Soil Water Conserv.* **2011**, *66(6)*, 362–372. doi:10.2489/jswc.66.6.362.

17. Buchanan, B.P. et al. Evaluating topographic wetness indices across central New York agricultural landscapes. *Hydrol. Earth Syst. Sci.* **2014**, *18(8)*, 3279–3299. doi: 10.5194/hess-18-3279-2014.

18. Qiu, Z.; Lyon, S.W.; Creveling, E. Defining a Topographic Index Threshold to Delineate Hydrologically Sensitive Areas for Water Resources Planning and Management. *Water Resour. Manag.* **2020**, *34*, 3675–3688. doi:10.1007/s11269-020-02643-z.

19. Anselin, L. Local Indicators of Spatial Association–LISA. *Geogr. Anal.* **1995**, *27(2)*, 93–115. doi: 10.1111/j.1538-4632.1995.tb00338.x.

20. Getis, A.; Ord, J.K. The Analysis of Spatial Association by Use of Distance Statistics. *Geogr. Anal.* **1992**, *24(3)*, 189–206. doi: 10.1111/j.1538-4632.1992.tb00261.x.

21. Ord, J.K.; Getis, A. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geogr. Anal.* **1995**, *27(4)*, 286–306. doi:10.1111/j.1538-4632.1995.tb00912.x.

22. Tobler, W. On the first law of geography: A reply. *Ann. Assoc. Am. Geogr.* **2004**, *94(2)*, 304–310. doi:10.1111/j.1467-8306.2004.09402009.x.

Research Article

Forecasting the track and intensity Damrey storm in 2017 by the multi–physical ensemble Kalman filter

Minh Thi Pham[1*], Trung Le Doan[2], Hang Thi Nguyen[3], Tuong Hong Thi Tran[4], Thuy Kim Pham[3]

[1] Department of Meteorology, Hydrology and Climate change, Ho Chi Minh University of Natural Resources and Environment; minhpt201@gmail.com
[2] Student of Department of Meteorology, Hydrology and Climate change, Ho Chi Minh University of Natural Resources and Environment; dltrungphuhoi@gmail.com
[3] Department of General Science Ho Chi Minh University of Natural Resources and Environment; hang.nguyen687@gmail.com; pkthuy.math@gmail.com
[4] Department of Information Systems and Remote Sensing, Ho Chi Minh University of Natural Resources and Environment; tthtuong@hcmunre.edu.vn

*Corresponding author: minhpt201@gmail.com; Tel.: +84–936069249
Received: 8 April 2022; Accepted: 22 June 2022; Published: 25 June 2022

Abstract: This study applicates the multi–physical method in ensemble Kalman filter determining error of WRF models to forecast the track and intensity of storm Damrey in 2017. The study run three experiments with assimilation of satellite data to forecast Damrey in 2017 at the beginning 00 UTC and 12 UTC November 1st and 2nd: (1) 21 ensemble members which are combined from 11 physics options, no increase in error correlation (MP); (2) Using single set of physical model, 21 ensemble members, inflation factor $\lambda = 6.5$ (MI); (3) Using single set of physical model, 21 ensemble members without increase in error correlation (PF). Statistical results of track errors in MP test at the 24, 48, 72–hour is 12–32% reduction in compared with tests MI and PF. For storm intensity, absolute error of P_{min} in the MP test at 24 and 72–hour is decreased from 30–47% in compared to the other two tests. And the absolute error of Vmax in the MP test at all forecasting terms is 13–26% reduction in compared with tests MI and PF. Thus, the multi–physical ensemble Kalman filter can forecast the track and intensity of storms affecting Vietnam.

Keywords: Ensemble forecasting; Error model; Typhoon; The Kalman filter.

_____

1. Introduction

Basically, data assimilation is a process which the observed data and a background guess field are statistically combined to obtain the best possible initial conditions for the numerical model [1–2]. The goal of assimilation is to find the best possible analysis field for the model input. However, this work depends heavily on the quality of the observed data (related to the error of the observed data) and the quality of the model's background guess data (related to the model's intrinsic error). The error related to the monitoring data belongs to the problem of quality control of professional monitoring; while the background field error is related to the model's internal errors – errors caused mainly by physical processes that are not fully understood [2–4].

Currently, the model error handling techniques in modern data assimilation algorithms include multiplicative inflation techniques [5], additive increasing techniques [6], or

systematic error correction method [7–8], multiple physics technique [9]. The multiple physics technique is the method of using different combinations of physical parameterization diagrams in the WRF model to calculate the parameters related to the error of the model in the ensemble Kalman filter [9]. This method is based on the assumption that the source of the model error is mainly due to the incomplete representation of physical processes [2–3]. The multiple physics has been applied in some previous studies [3, 9–11], and show significant improvements in track and intensity storm prediction results compared with other methods, such as the multiplicative inflation [12] and compared with the case where the model error is zero [3, 9–11]. Accordingly, in the study [9], it was shown that the multiplicative inflation factor 6.5 is the best compared to the multiplicative inflation factor that varies from 1.0 to 6.5, and the multiple physics technique is good choice for short–term forecasting problem, in addition, the study also shows that the optimal number of combinations in combination prediction ranges from 21–24 components which are different combinations of physical parameterization schemes in the WRF model. Therefore, in this study, we will use the multiplicative inflation technique and consider the model to be perfect to compare with the multiple physics' technique, and the number of ensemble components is 21 components for a forecasting session [9].

In addition, storm Damrey in 2017 was a strong storm that directly hit Ninh Hoa–Khanh Hoa At 6:30 am on 4 November 2017 with wind strength increased by 1 level to level 13, level 15, 16 [13]. At 10 o'clock on the same day, the center of the storm was on the mainland of Dak Lak–Lam Dong, the wind strength near the center of the storm decreased to level 10–11, level 13. After that, the storm weakened into a tropical depression. By noon on November 5, the center of the tropical depression in the southern region of Cambodia, the wind dropped below 40 km/h. By November 8, at least 106 people had been killed in Vietnam by the storm, with 197 others injured and 25 missing. It is reported that more than 116,000 homes were destroyed after flooded. The United Nations Children's Fund (UNICEF) estimates that at least four million people have been directly impacted by the storm and need support. Nha Trang beach resort was one of the hardest hit areas, 30,000 residents and tourists had to evacuate the area. A number of previous studies used storm Damrey as in the initial study of vortex chemistry by Nguyen Binh Phong and Associates 2020 to predict the intensity of storm Damrey during the landfall stage [14]. Research results show that storm intensity with vortex initial is improved more clearly in the absence of vortex initialization. Another study related to the forecast of the storm Damrey's trajectory by the method of correcting the forecast of the storm's trajectory from the product of the combined forecasting system through the selection of the optimal composite component of the author Tran Quang Nang and Tran Tan Tien 2020 [15]. The results show that the correction method can only improve the error of trajectory prediction in short–term forecasting terms. Another study by Kulaya Keawsang–in and colleagues 2021 examines the sensitivity of different physical schemes to simulate Typhoon Damrey. The results show that the Belts–Millers–Janjic convection diagram and the WSM6 microphysics diagrams are suitable in the simulation of storm Damrey [16]. Therefore, in this study, we apply multiphysics technique in combinatorial Kalman filter to determine the error of WRF model predicting the trajectory and intensity of storm Damrey 2017.

2. Methods and data

2.1. Ensemble Kalman Filter algorithm

The idea of the LETKF algorithm is to use the background ensemble matrix as a transformation operator from the model space spanned by the grid points within a selected local patch to the ensemble space spanned by the ensemble members, and perform the analysis in this ensemble space at each grid point. For a quick summary of the LETKF

algorithm, assume that a background ensemble $\{\mathbf{x}^{b(i)} i = 1, 2..., k\}$ are given, where k is the number of ensemble members (assuming that we are doing analysis at one instant of time, so no time index is written explicitly). Following [17], an ensemble means $\bar{\mathbf{x}}^b$ and an ensemble perturbation matrix $\mathbf{X}^b$ are defined respectively as:

$$\bar{\mathbf{x}}^b = \frac{1}{k}\sum_{i=1}^{k}\mathbf{x}^{b(i)} ; \ \mathbf{X}^b = \mathbf{x}^{b(i)} - \bar{\mathbf{x}}^b \tag{1}$$

Let $\mathbf{x} = \bar{\mathbf{x}}^b + \mathbf{X}^b\mathbf{w}$, where $\mathbf{w}$ is a local vector in the ensemble space, the local cost function to be minimized in the ensemble space is given by:

$$\widehat{J}(\mathbf{w}) = (k-1)\mathbf{w}^T\{\mathbf{I} - (\mathbf{X}^b)^T[\mathbf{X}^b(\mathbf{X}^b)^T]^{-1}\mathbf{X}^b\}\mathbf{w} + J[\mathbf{x}^b + \mathbf{X}^b\mathbf{w}] \tag{2}$$

where $J[\mathbf{x}^b + \mathbf{X}^b\mathbf{w}]$ is the cost function in the model space. If one defines the null space of $\mathbf{X}^b$ as $N = \{\mathbf{v} \mid \mathbf{X}^b\mathbf{v} = 0\}$, then it is easy to see that the cost function $\widehat{J}(\mathbf{w})$ is composed of two parts: one containing the component of $\mathbf{w}$ in N (the first term in Eq. 2), and the second depending on the component of $\mathbf{w}$ that is orthogonal to N. By requiring that the mean analysis state $\bar{\mathbf{w}}^a$ is orthogonal to N such that the cost function $\widehat{J}(\mathbf{w})$ is minimized, the mean analysis state and its corresponding analysis error covariance matrix in the ensemble space can be found as:

$$\bar{\mathbf{w}}^a = \widehat{\mathbf{P}}^a(\mathbf{Y}^b)^T\mathbf{R}^{-1}[\mathbf{y}^0 - H(\bar{\mathbf{x}}^b)] \tag{3}$$

$$\widehat{\mathbf{P}}^a = [(k-1)\mathbf{I} + (\mathbf{Y}^b)^T\mathbf{R}^{-1}\mathbf{Y}^b]^{-1} \tag{4}$$

where $\mathbf{Y}^b \equiv H(\mathbf{x}^{b(i)} - \bar{\mathbf{x}}^b)$ is the ensemble matrix of background perturbations valid at the observation locations, and $\mathbf{R}$ is the observational error covariance matrix. By noting that the analysis error covariance matrix $\mathbf{P}^a$ in the model space and $\widehat{\mathbf{P}}^a$ in the ensemble space have a simple connection of $\mathbf{P}^a = \mathbf{X}^b\widehat{\mathbf{P}}^a(\mathbf{X}^b)^T$, the analysis ensemble perturbation matrix $\mathbf{X}^a$ can be chosen as follows:

$$\mathbf{X}^a = \mathbf{X}^b[(k-1)\widehat{\mathbf{P}}^a]^{1/2} \tag{5}$$

The analysis ensemble $\mathbf{x}^a$ is finally obtained as:

$$\mathbf{x}^{a(i)} = \bar{\mathbf{x}}^b + \mathbf{X}^b\{\bar{\mathbf{w}}^a + [(k-1)\widehat{\mathbf{P}}^a]^{1/2}\} \tag{6}$$

Detailed handling of more general nonlinear and synchronous observations in LETKF can be found in [17]. It should be mentioned that the above formation is only valid in the absence of model errors. To take into account the model errors, [17] suggested that a multiplicative factor should be introduced in Eq. (4) (specifically, the first factor on the rhs bracket in Eq. 4). Although one could also use the additive inflation, this study focuses only on the multiplicative inflation for the ease of implementation and comparison.

2.2. Models and Study area

Using WRF model version V3.9.1 with 31 levels (Sigma) in the vertical with the highest-pressure level (the upper boundary of the model) is 10hPa [18]. The WRF model is selected with two nested computational domains using the Mercator projection. The mesh area designed for the simulation test of Damrey storm is a nested grid consisting of 2 regions with horizontal resolutions of 36 km and 12 km respectively, grid domain 1 consists of 151×151 grid points and grid domain 2 consists of 151×151 grid points with the domain center. immobility 11.2°N & 112.3°E (Figure 2).
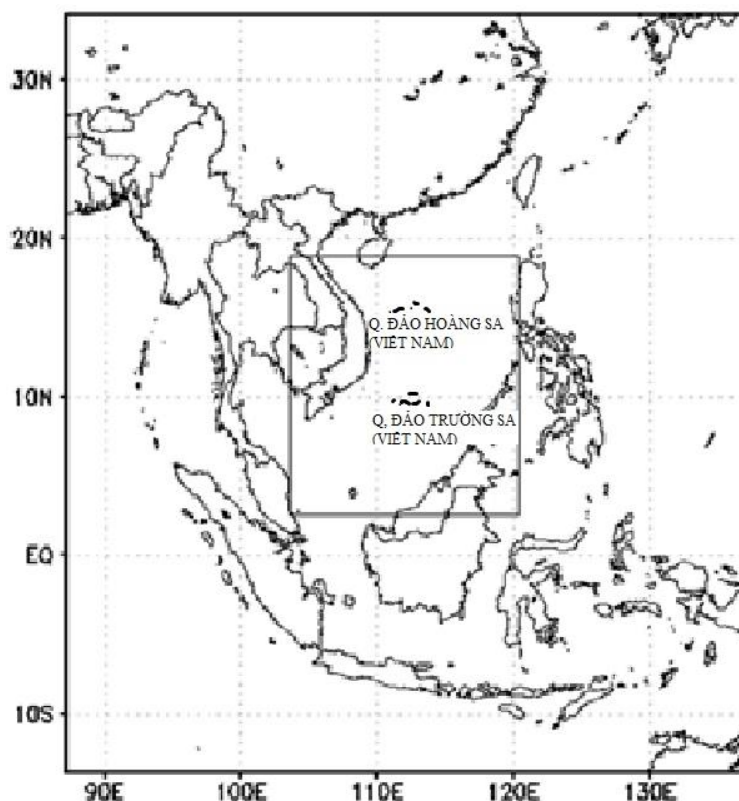
Figure 1. The study area.

2.2. *Experiment descriptions*

In this study, the authors tested the forecast for typhoon Damrey 2017 with a term of 3 days 2017 at the beginning 00 UTC and 12 UTC November 1st and 2nd with experiments are listed in table 1. All the above tests are assimilated satellite wind data by the ensemble Kalman filter.

Table 1. List of experiments with the WRF–LETKF configuration.

Experiments	Description	Boundary Condition
MP	The Combination of 11 options physical model, 21 ensemble members, no increase in error correlation	To ensure that each member has its own lateral boundary condition consistent with its updated nalysis,
MI	Using a set of physical model, 21 ensemble members, inflation factor $\lambda = 6.5$	the WRFDA boundary routine is used to generate boundaries for each
PF	Using a set of physical model, 21 ensemble members without increase in error correlation	ensemble member after the ensemble analysis step is finished for every cycle.

In the first experiment (MP), 21 sets of physical models (Table 3) consisted of the set of combinatorial components of the parameterization schemes in Table 2. In the second experiment (MI), one the set of specific physical models in the WRF model include (a) the WSM3 microphysics diagram, (b) the radiation rapid transmission scheme (RRTM) for both long and short-wave radiation, and (c) BMJ convective parameterization scheme (component 11) is applied to all combinatorial components with a multiplier = 6.5 added in the variable error correlation matrix change P^a in expression (4). However, this coefficient λ does not change in all cycles of the experiment so that the effectiveness of the MI method in handling model error can be compared with that of the MP method. In the third experiment (PF), we consider the model to be perfect with the background error unchanged over all the running cycles of the experiment. Similar to the multiplication method, this

experiment uses the same set of physical models as in the MI experiment so that the effectiveness of the method in handling model errors can be compared with the MP and MI tests.

Table 2. Options table of physical parameterization schemes in WRF model [19].

Schemes	Symbol	Options
Longwave Radiation	**ra_lw_physics**	1. RRTM scheme
Shortwave Radiation	**ra_sw_physics**	1. Dudhia scheme 2. Goddard shortwave
Microphysics	**mp_physics**	1. Kessler scheme 2. Lin et al. scheme 3. WSM 3–class simple ice scheme 4. WSM 5–class scheme 5. Ferrier (new Eta) microphysics 6. WSM 6–class graupel scheme
Cumulus Parameterization	**cu_physics**	1. Kain–Fritsh scheme 2. Betts–Miller–Janjic scheme

Table 3. Encryption of multi–physical ensembles from multiple physics options in WRF model [19].

Complex	Ra_lw_ physics	Ra_sw_ physics	mp_ physics	cu_physics
001	1	2	1	1
002	1	1	1	2
003	1	2	1	2
004	1	1	2	1
005	1	2	2	1
006	1	1	2	2
007	1	2	2	2
008	1	1	3	1
009	1	2	3	1
010	1	1	3	2
011	1	2	3	2
012	1	1	4	1
013	1	2	4	1
014	1	1	4	2
015	1	2	4	2
016	1	1	5	1
017	1	2	5	1
018	1	1	5	2
019	1	2	5	2
020	1	1	6	1
021	1	2	6	1

2.3. Data

The initial and boundary conditions used NCEP/NCAR (NCEP–The National Center for Environmental Prediction/NCAR–The National Center for Atmospheric Research) GFS forecast data with a horizontal resolution of 0.5×0.5 degrees and grib2 format. GFS data were obtained from the website: https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs. The best track data of storm position and intensity are collected from the website: https://www.metoc.navy.mil/jtwc/jtwc.html?western-pacific.

Wind monitoring data from satellites is a particularly important data source for forecasting models running around the world with global coverage and data collection time within 3–6 hours. depending on the characteristics of each satellite. Satellite wind data allows to know the dynamic state of the atmosphere, contributing to the information of the initial field of the forecast model by data assimilation. Currently, satellite wind data are

preprocessed by the University of Wisconsin satellite atmospheric motion vector CIMSS–AMV (Cooperative Institute for Meteorological Satellite Studies – University of Wisconsin satellite atmospheric motion vector CIMSS–AMV) in the same time period. selected. Several studies with CIMSS–AMV data have shown that this data can help improve the predictive quality of various medium–sized systems. The advantage of the CIMSS–AMV data is that the error has been tested for high quality and is determined by a recursive filtering algorithm. Each metric is checked for the best fit with the surrounding data using quality index techniques. Most of the CIMSS–AMV data is distributed in different regions and is currently stored in a variety of formats including ASCII and/or BUFR. In this study, satellite wind data were collected over the Indian, Northwest Pacific region (Figure 2) and downloaded from the website http://tropic.ssec.wisc.edu in ASCII format.
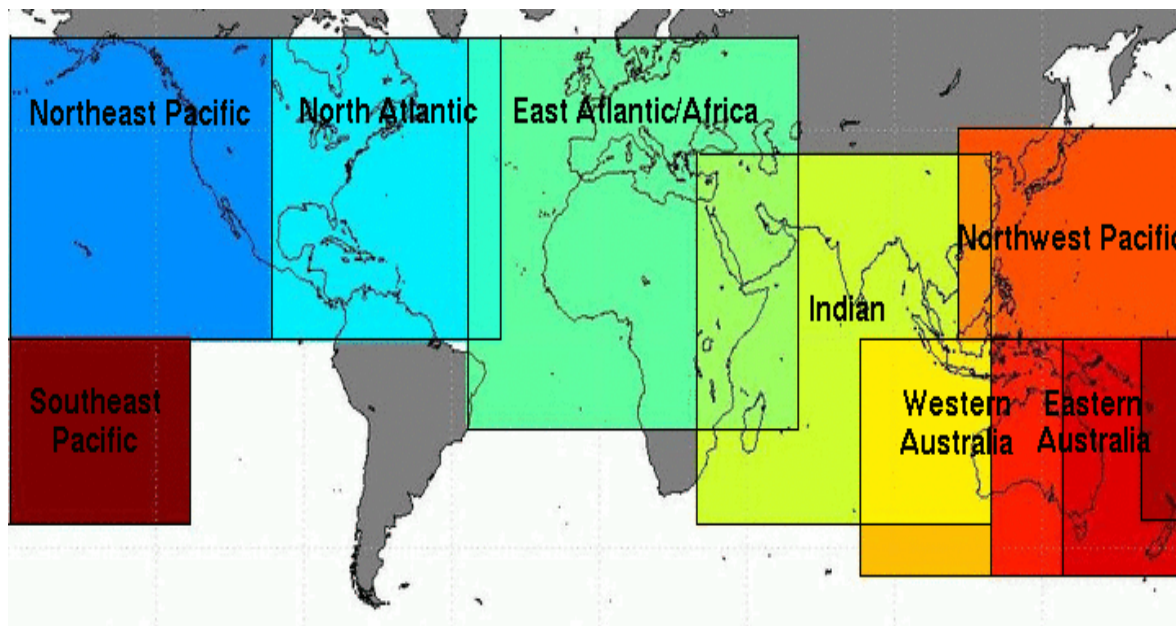
Figure 2. Area is covered by satellite wind data in this study (source: http://tropic.ssec.wisc.edu).

2.4. Evaluation methods

2.4.1. Absolute mean method

According to [20], MAE error is used to evaluate the predictions of continuous atmospheric variables. Therefore, MAE is applied as an index to evaluate the error of storm intensity (minimum sea level pressure at the center – PMIN and maximum wind speed near the center of VMAX). With MAE–mean absolute error is calculated by the formula:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|F_i - O_i| \tag{7}$$

where MAE is the mean absolute error; F_i is the predicted value; O_i is the observed value; and N is the length of the data series.

2.4.1. Storm center distance method

Track error calculated by formula (8).

$$\text{PE} = R_e * \arccos\left[\sin(\alpha_1) * \sin(\alpha_2) + \cos(\alpha_1) * \cos(\alpha_2) * \cos(\beta_1 - \beta_2)\right] \tag{8}$$

where R_e is the radius of the earth (6378.16 km); α_1, α_2 is the latitude of the actual of the storm and the center of the storm predicted by the model (in radians); β_1, β_2 is the longitude of the actual center of the storm and the predicted center of the storm (in radians). The distance mean error is calculated as follows:

$$MPE_j = \frac{\sum_{i=1}^{n} PE_{i,j}}{n} \qquad (9)$$

where PE is the distance error of each forecasting case; n is the number of test cases; j is the forecast term.

3. Results and discursion

3.1. Stream simulation

Basically, when large–scale circulations change and control the storm's active area, it will directly affect the direction of the storm's movement. In the case of cyclone Damrey, large–scale circulation dominated the storm's area of activity, including the northwest Pacific subtropical high and cold high at north of the storm. For the purpose of investigating the applicability of multiphysics techniques in determining the model error in ensemble Kalman filter to prediction storm trajectory, the study compares the stream field in the MP, MI and PF tests in levels 850, 700, and 500 hPa at 12h00 UTC on 3rd November 2017 is the time when the storm begins to make landfall in the forecast session that begins at 12h00 UTC on 1st November 2017 (Figure 3).

Synop
[https://www.tmd.go.th/en/weather_map.php]   MP   MI   PF
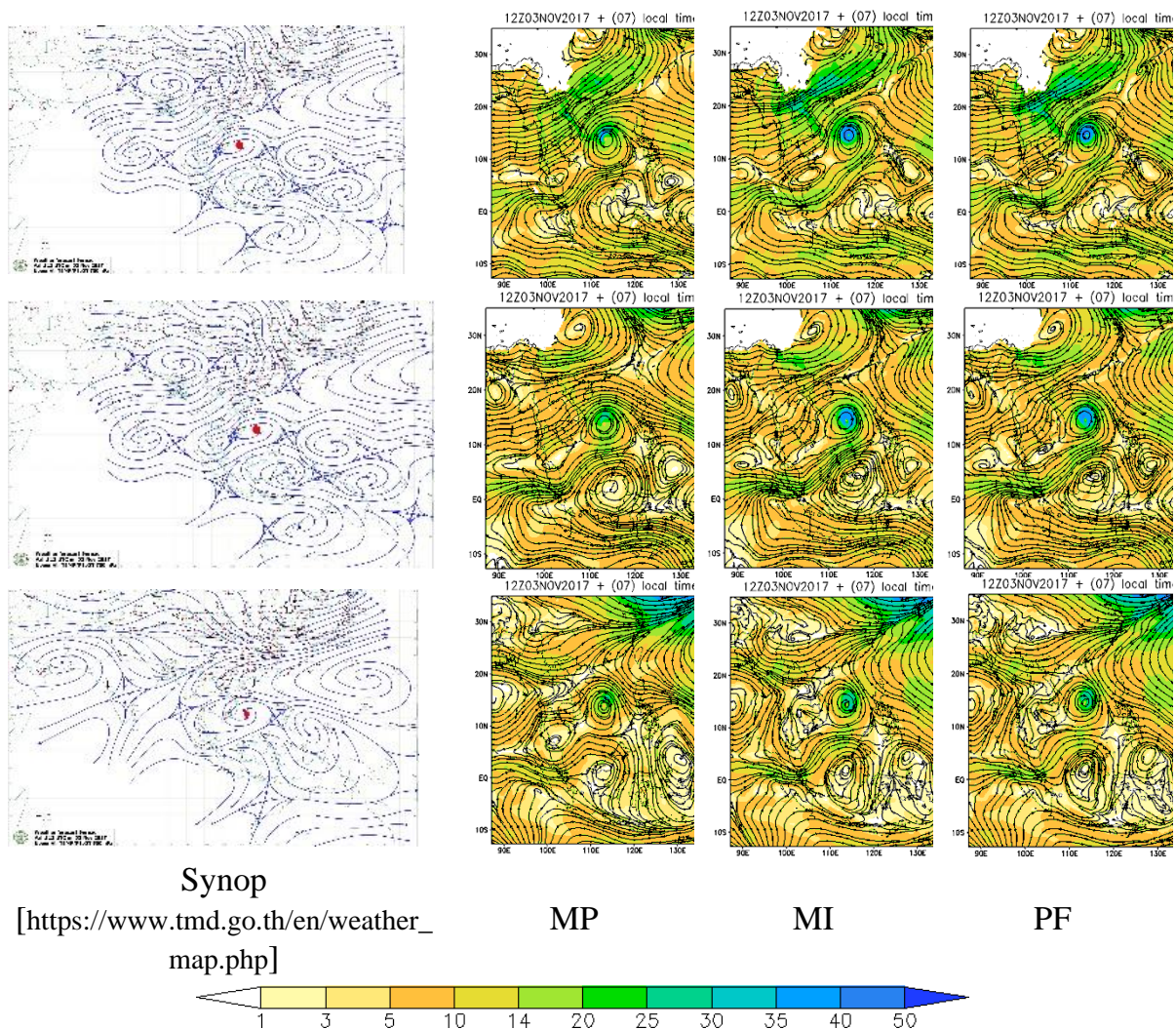
Figure 3. The stream map of 850 hPa (top), 700 hPa (middle) and 500 (bottom) hPa levels.

At 850 hPa and 700 hPa, the MP test simulates a cold high that mixes southward and extends to the east more than the cold high which is simulated in the MI and PF tests. In particular, at 700 hPa and north of the storm there is a fairly barometric saddle in the synop map, and this barometric saddle is also simulated in the MP test, while the MI and PF tests

don't simulate this barometric saddle (Figure 3). At 500 hPa, the MP test simulates the northwest Pacific subtropical high more westward than the MI and PF trials, and has a similar morphology to the synop map (Figure 3). The simulation results show that the direction of the storm in the MP test is closer to the real trajectory than the other two tests [11].

Thus, the multiphysics technique seems to have some impact in the flowline simulation at the levels of 850, 700 and 500 hPa. This result is a consequence of previous work that demonstrated that the volumetric least squares error of the multiphysics technique is smaller than the volumetric least squares error of the multiplicative inflation technique. Meaning predicted error of U, V and T in the MP test improved markedly compared with the MI test [9]. As a result, the forecast results of the trajectory and intensity of storm Damrey in the MP test are also improved relative to the MI and PF tests. In the next section, the paper will examine the ability of the tests to predict the trajectory and intensity of storms.

3.2. Forecasting intensity and track

3.2.1. Track storm

Figure 4 shows that the true trajectory of Typhoon Damrey is moving to the west (Figure 4) and made landfall around 12h00 UTC on 3rd November 2017. Meanwhile, the ensemble components in the MP, MI and PF tests all predict the trajectory of storm Damrey to move to the northwest. At 12h00 UTC on 4th November 2017, the storm moved on the sea in MP, MI and PF tests (Figures 4a–4c). In addition, a few ensemble components of the MP test predicted the location of the storm Damrey's landfall, but it is quite far from the actual location. On the other hand, the dispersion of composite components in the MP test is wider than that of the composite components in the MI and PF tests. This result is similar to the results of the previous study when concluding that the composite dispersion in MP was wider than the composite dispersion in the MI test [9, 11].

Figure 4d is the ensemble mean trajectory in the MP, MI and PF tests, the observed trajectory in black and the trajectory of the GFS data (green). From 00–hour to 30–hour forecast period, it shows that the storm trajectory is not much different between the 3 tests MP, MI and PF, after the 30–hour forecast period to 72 hours, the forecast trajectory of the tests is in the north of the true trajectory, where the predicted trajectory of the MP test is closer to the true orbit than the MI and PF tests. And the Damrey orbit in the GFS data is located south of the true trajectory.

The forecast results of storm Damrey's trajectory in the tests are consistent with the results of the stream simulations in all three tests (section 3.1). Specifically at 12h00 UTC on 3 November 2017 At 850 hPa and 700 hPa, the MP test simulates a cold high which moves down to the south and extends to the east more than with cold high simulated in the MI and PF tests. So the cold high in the MP test limited the direction of the storm's movement to the north. In addition, at 500 hPa, the MP test also simulated the northwest Pacific subtropical high developing to the west, thereby also limiting the northward movement of the storm (figure 4d). However, the speed of storm movement in all 3 trials was slower than observed and GFS data. And to quantify the accuracy of each trial's hurricane trajectory prediction, the study calculated the trajectory prediction error of the trials.

From the graph showing the trajectory prediction error of the MP, MI and PF tests together with the trajectory prediction error of the GFS data (Figure 5), it shows that the trajectory prediction error of the MP test is lower than the forecast error of the storm trajectory of the MI and PF tests at most forecasting periods. Meanwhile, the forecast error of storm trajectory in the MP test did not improve much compared with the track forecast error of the GFS data. This result is also clearly seen in the statistics of the orbital forecast error of the 3 forecasting sessions (Figure 6).

To evaluate the effectiveness of the MP test in predicting the trajectory, we calculated the relative trajectory prediction error between the MI and MP tests, and between the PF and MP tests (table 4). The relative trajectory prediction error results show that the orbital error in the MP test improves from 9% to 32% compared to the trajectory prediction error in the MI test, and improves from 4% to 30% compared to the trajectory prediction error in the PF test at most forecasting terms. This result may be due to the multiphysics technique (determining the error of the model due to the incomplete understanding of physical processes [4, 9] has partly corrected the error of the model. So that the received background field has a significantly reduced error, and leads to a more accurate analysis field for the input of the model than the multiplicative inflation technique and considers the model perfect.
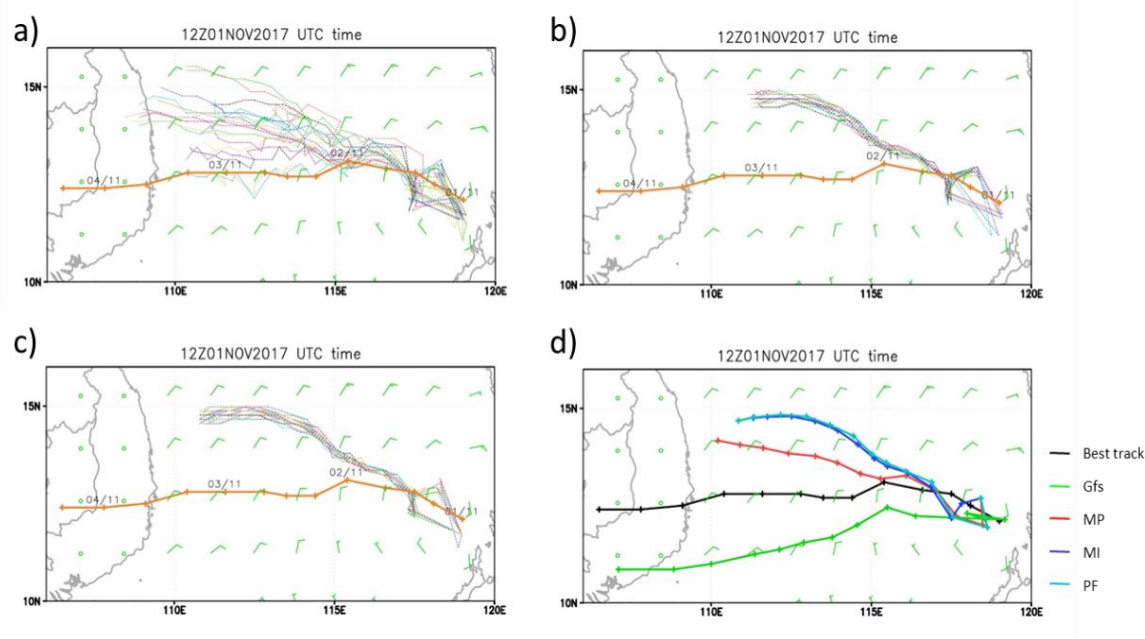
Figure 4. The predicted trajectory of Hurricane Damrey in the MP test (a), the MI test (b) and the PF test (c); observed (orange), composite components are thin lines; Figure d is the observed trajectory, the combined average trajectory of the test MP, MI, PF and GFS data. The forecast start time is 12 o'clock on November 1, 2017.
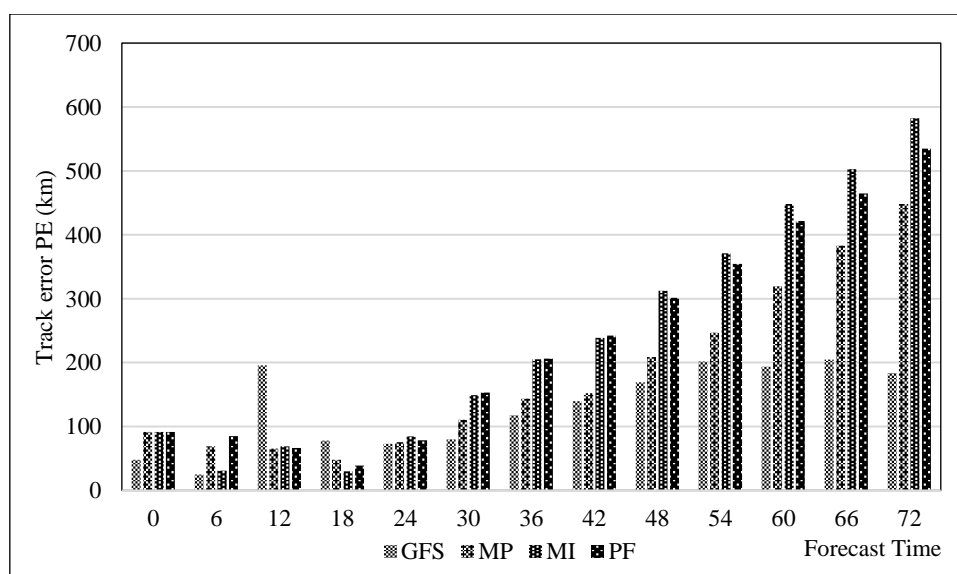
Figure 5. The ensemble means track errors in the MP, MI and PF test. With the forecast session starting at 12h00 UTC on 1st November 2017.
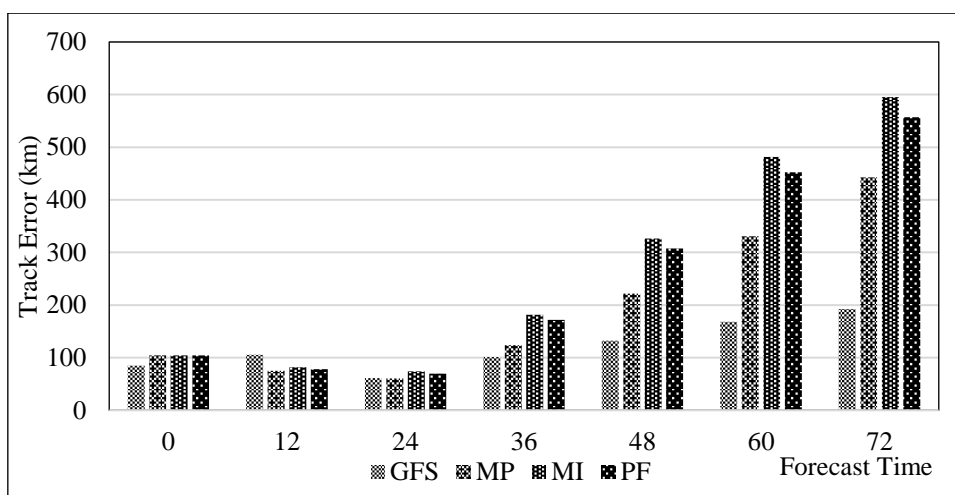
Figure 6. Average track error of 3 forecasting sessions (00 UTC on 1 st November 2017, 12UTC on 1st November 2017 and 00 UTC on 2 nd November 2017) MP, MI, PF test and GFS data.

Table 4. Relative error of MI and PF relative to MP in trajectory simulation, P_{min} and V_{max}.

Periods	Track (km)		Pmin (hPa)		Vmax (ms⁻¹)	
	MI–MP	PF–MP	MI–MP	PF–MP	MI–MP	PF–MP
0	0.00	0.00	0.00	0.00	0.00	0.00
6	–0.23	0.05	0.08	0.03	–0.03	0.14
12	0.08	0.04	0.08	0.10	–0.03	0.08
18	0.09	0.12	–0.13	–0.19	–0.06	–0.13
24	0.18	0.13	0.31	0.27	0.13	0.18
30	0.32	0.28	–0.30	–0.32	0.49	0.49
36	0.32	0.28	–0.16	–0.17	0.36	0.35
42	0.32	0.28	0.68	0.62	0.58	0.52
48	0.32	0.28	–1.56	–1.32	0.27	0.18
54	0.34	0.30	–0.05	–0.05	0.07	0.08
60	0.31	0.27	0.25	0.18	0.30	0.29
66	0.31	0.26	0.38	0.35	0.24	0.19
72	0.26	0.21	0.47	0.43	0.26	0.21

3.2.2. Damrey storm intensity

Storm intensity is usually expressed through minimum pressure (P_{min}) and maximum wind speed (V_{max}). The P_{min} and V_{max} values used in this section are the mean of the 21 ensemble members in each trial.

a) The minimum pressure

The observed P_{min} data in Figure 7 shows that the storm gradually became stronger from 12h00 UTC on 1st November 2017 and the strongest storm at 12h00 UTC on 3rd November 2017 (Figure 7) – expressed through the value of P_{min} down to the lowest. After that, the P_{min} value increases gradually, meaning the storm is getting weaker. The MP, MI and PF tests all predict the P_{min} process which has a decreasing trend similar to the observed P_{min} value. However, after 12h00 UTC on 3rd November 2017, these tests did not predict the changing trend of the P_{min} process as observed (Figure 7). Particularly, the P_{min} of GFS data has a variable similar to the observed P_{min}, but value P_{min} is much larger than the observed P_{min} value, or in other words, the storm intensity in the GFS data is weaker than in reality. At 12h00 on 3rd November 2017 is also the time when the storm makes landfall (according to monitoring data), but in the tests, P_{min} decreased little or not, so the storm still existed at sea. This result is statistically consistent, when the storm makes

landfall, the intensity of the storm decreases. Meaning the P_{min} value increases more than on the sea surface.

The P_{min} absolute error results will indicate the effectiveness of each trial in predicting P_{min}. Figure 8 shows that the P_{min} absolute error in the MP test did not improve much compared to the MI and PF tests at the 60–hour advance term, while at the 66–hour and 72–hour forecast period, the MP absolute error improved significantly. Statistically, all 3 forecasting sessions showed that P_{min} prediction results in the MP test were better than the MI and PF tests at most of the forecasting term (Figure 9).

In addition, similar to the evaluation of the predictability of the Damrey storm trajectory of the multiphysics technique in determining the error of the model in the ensemble Kalman filter, we also calculate the relative error P_{min} between the tests. The results in Table 4 show that, at the 24–hour forecast period and the 2–day prediction term, the MP test improves the P_{min} error by 18% to 47% compared with the MI and PF tests. This result may be due to the effect of the multiphysics technique in forecasting meteorological variables (U, V and T) which is significantly improved compared with the techniques in the MI test [9] and PF test. On the other hand, the P_{min} statistical error results also show that the MP test significantly improves the P_{min} error compared with the GFS data (Figure 9) at most of the forecasting term. However, at the time when the storm was about to make landfall, the P_{min} error in the MP test did not improve compared to the initialization vortex method [14]. In the next section, the article analyzes the predictive ability of Vmax of multiphysics techniques.
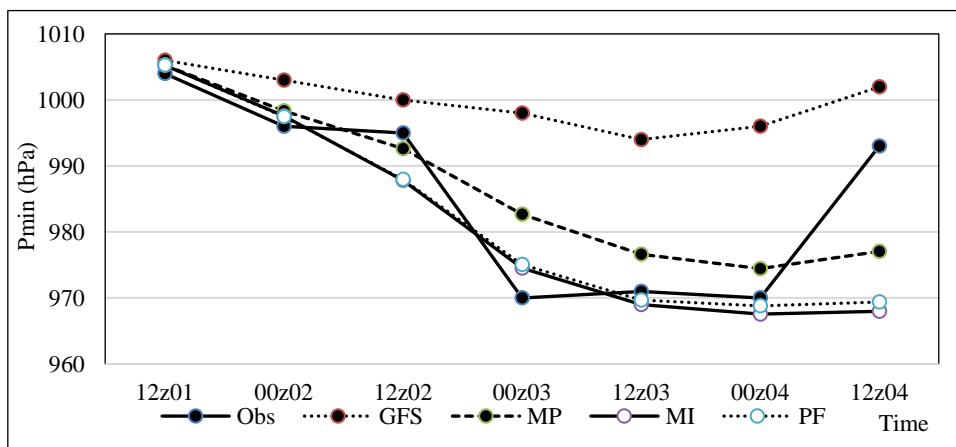
Figure 7. The mean P_{min} process variable in the trials. With the forecast start time at 12h00 UTC on 1st November 2017.
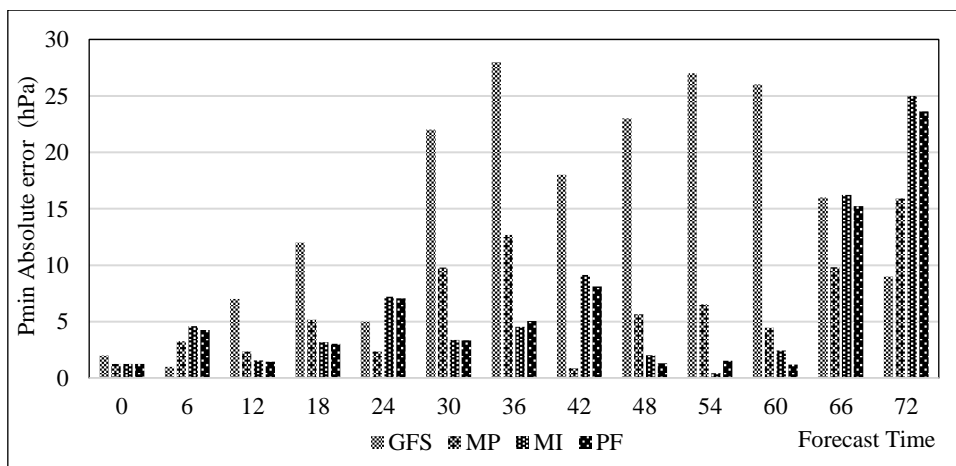
Figure 8. P_{min} Absolute error (hPa) in the MP, MI, PF test and GFS data. Forecast at 12h00 UTC on 1st November 2017.
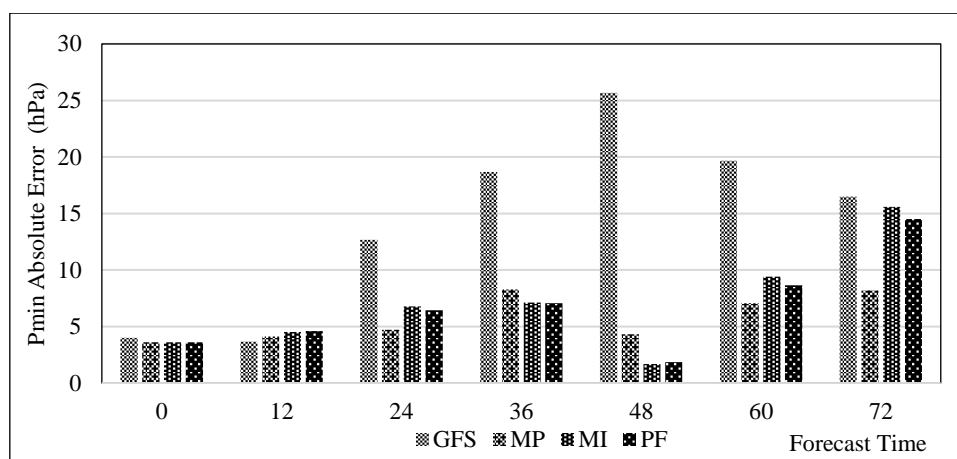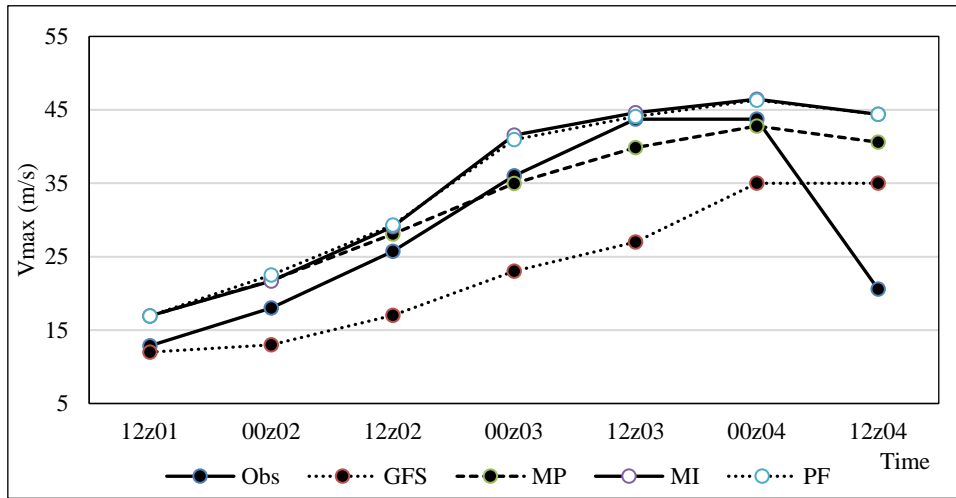
Figure 9. Average absolute error P$_{min}$ in forecast sessions (00 UTC on 11/01/2017, 12 hours UTC on 11/01/2017, 00 UTC on 11/2/2017) in MP, MI, PF tests and GFS data.

b) Maximum wind speed

The maximum wind speed (V$_{max}$) is the second factor representing the intensity of any storm. The change in Vmax during the storm's activity also indicates that the storm is weakening or strengthening. Therefore, the V$_{max}$ variable is a visual image describing the strength or weakening of the storm. Figure 10 is the Vmax variable of storm Damrey with the forecast time at 12h00 UTC on 1st November 2017. Similar to the Pmin variable, the observed Vmax data shows that the storm is getting stronger from 12h00 UTC on 1st November 2017 and the strongest at 12h00 on 3rd November 2017 – this is shown by the value of Vmax reaching the minimum at this time. After that, the storm gradually weakened when it made landfall, due to the influence of surface friction (Figure 8). Meanwhile, the MP, MI and PF tests all predict that the Vmax process tends to increase similarly to the observed Vmax value from 1200h UTC on 1st November 2017 to 12h00 UTC on 3rd November 2017. After 12h00 on 3rd November 2017, the variable Vmax in the tests is different from the observed Vmax variable (Figure 10). Particularly, the Vmax variable of GFS has a variable similar to the observed V$_{max}$, but V$_{max}$ value is much smaller than the observed V$_{max}$ value or in other words, the storm intensity in the GFS data is weaker than in reality. This is consistent with previous studies that simulate the Vmax magnitude of GFS data biased lower than reality. In addition, at 12h00 on 3rd November 2017 is also the time when the storm makes landfall (according to observational data – figure 10, figure 4 – trajectory), but in the tests, V$_{max}$ decreased slightly, so the storm still exists at sea. This result is statistically consistent, when the storm makes landfall, the intensity of the storm decreases sharply – the V$_{max}$ value decreases more than on the sea surface. The results of V$_{max}$ absolute error show the effectiveness of each test in predicting V$_{max}$. Figure 11 shows that the Vmax absolute error of the MP test is significantly improved compared with the MI and PF tests at most of the forecasting terms. Statistically, all 3 forecasting sessions showed that V$_{max}$ prediction results in MP test were better than MI and PF tests at most of the forecast periods except for the 24 hours period (Figure 11).

In addition, we calculate the V$_{max}$ relative error betweenthe tests. The results shown in Table 4 show that the Vmax error value in the MP test is improved by 6% to 58% compared tothe Vmax error in the MI and PF test. This result may be due to the effect ofthe multiphysics technique in forecasting meteorological variables (U, V and T) which is significantly improved compared with the techniques in the MI test [9] and PF test. In addition, the statistical Vmax error also showsthat the MP test significantly improves the Vmax error compared with the GFSdata (Figure 12). However, at the time when the storm was about to makelandfall, the V$_{max}$ error in the MP test did not improve compared to theinitialization method [14]. Thus, for intensity forecasting of Damrey storm, multiphysics
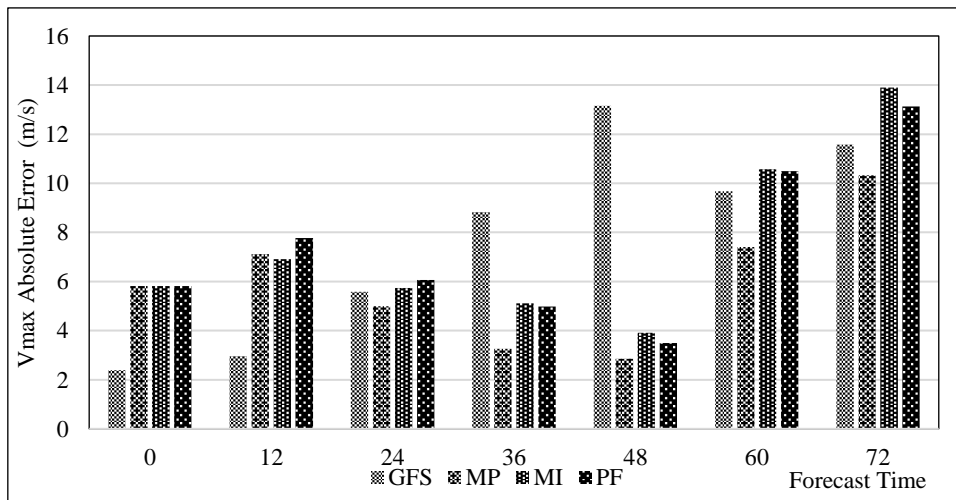
techniques show certain advantages – forecasting storm intensityis more effective than other techniques.

Figure 10. Variable V_{max} means in the trials. With forecast start time at 12h00 UTC on 1st November 2017.
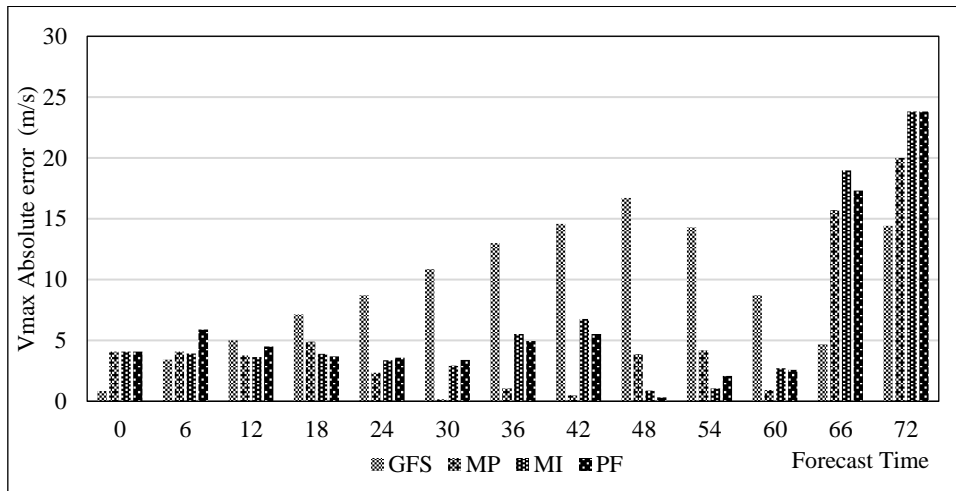
Figure 11. Absolute error V_{max} (m/s) in the MP, MI, PF and GFS tests. Forecast at 12h00 UTC on 1st November 2017.

Figure 12. Average absolute error V_{max} in forecasting sessions (00 UTC on 11/01/2017, 12 hours UTC on 11/01/2017, 00 UTC on 11/2/2017) in MP, MI tests, PF and GFS.

4. Conclusion

In this study, we applied multi–body and multiplier technique and considered the model to be perfect to perform 3 sessions of forecasting the trajectory and intensity of Typhoon Damrey 2017 with 3–day forecast period with input field. is the analysis data which is generated from the Kalman filter combination assimilation of wind data observed from satellites. The results of the comparative analysis and comparison between the above methods show that:

Regarding the trajectory prediction, the MP test (multiphysics technique) showed an improved trajectory error of 9% to 32% in compared with the trajectory prediction error in the MI test, and an improvement of 4% up to 30% of the orbital prediction error in the PF test at most forecasting terms. These results are the consequence of correcting the error of the model because the object processes are not fully represented by the multibody technique. With a specific case at the 12h00 session on 1st November 2017, the multi–physical techniques (MP test) simulation results of the general atmospheric circulation – Cold continental high pressure and subtropical high pressure. The north–west Pacific temperature is quite similar to the synaptic topology, so that the forecasted storm trajectory is closer to the true trajectory than the other techniques (MI and PF). However, at the stage when the storm was about to make landfall, the multiphysics technique did not improve the orbital error compared with the GFS data. This may be due to the storm's interaction with land (topography), so the model error at this point does not seem to be simply due to the physical processes in the model not being fully represented enough. Therefore, the study proposes for the next research direction of applying multiphysics technique in combinational Kalman filter to determine model error for simple cases of storms moving at sea and for hurricanes. landed on land. From there, there is a plan to overcome the error of the trajectory for the case of storms that are about to land on the mainland.

In predicting the intensity (P_{min} and V_{max}), the multiphysics technique also shows the certain advantages over the other two techniques at each forecasting term. Specifically, for P_{min}, at the 24-hour forecast period and the 2 days larger forecast period, the MP test improved 18% to 47% of the P_{min} error in compared with the MI and PF tests. For Vmax, the Vmax error value in the MP test improved by 6% to 58% in compared with the Vmax error in the MI and PF test. These results may be due to the effect of the multiphysics technique in predicting meteorological variables (U, V and T) which is significantly improved compared with the multiplier techniques in the MI test [9] and PF. In compared with using GFS data, the multiphysics technique significantly improved the forecast error of storm intensity (Figure 14 and Figure 15). Especially when the storm was about to make landfall, the strongest storm was at 12h00 UTC on 3rd November 2017, the MP test predicted the storm's value as well as the strengthening trend. In terms of errors, P_{min} and V_{max} are improved compared with GFS data, MI and PF tests, but cannot be improved compared to P_{min} and V_{max} prediction results in the initial study of vortex chemistry [14]. Therefore, the research direction that applies both the multiphysics method and the initial application of vortex chemistry to predict the intensity of storms affecting Vietnam is the next research direction of the authors' group.

Author contribution statement: Conceived and designed the experiments; Analyzed and interpreted the data; contributed reagents, materials, analysis tools or data; manuscript editing: M.P.T.; T.D.L.; H.N.T.; T.H.T.T.; T.P.K.; Performed the experiments; contributed reagents, materials, analyzed and interpreted the data, wrote the draft manuscript: M.P.T.; T.D.L.; H.N.T.; T.H.T.T.

Competing interest statement: The authors declare no conflict of interest.

References

1. Holton, J.R. An introduction to dynamic meteorology, 2004.

2. Chanh, K.Q. Estimation of Model Error in the Kalman Filter by Perturbed Forcing. *VNU J. Sci.: Nat. Sci. Tehnol.* **2010**, *26(3S)*, 310–316.

3. Thomas, M.H.; Jeffrey, S.W. Accounting for the Error due to Unresolved Scales in Ensemble Data Assimilation: A Comparison of Different Approaches. *Mon. Weather Rev.* **2005**, *133*, 3132–3147.

4. Pu, Z. Improving Hurricane Intensity Forecasting through Data Assimilation: Environmental Conditions Versus the Vortex Initialization, Recent Hurricane Research – Climate, Dynamics, and Societal Impacts. Anthony Lupo (Ed.), **2011,** ISBN: 978-953-307-238-8.

5. Anderson, J.L.; Anderson, S.L. A Monte Carlo implementation of the non–linear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.* **1999**, *127*, 27–41.

6. Mitchell, H.L.; Houtekamer, P.L. An adaptive ensemble Kalman filter. *Mon. Wea. Rev.* **2000**, *128*, 416.

7. Dee, D.P.; Silva, A.M. Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.* **1998**, *124*, 269.

8. Li, H. Local ensemble transform Kalman filter with realistic observations. Ph.D. dissertation. University of Maryland. 2007, pp. 131.

9. Chanh, K.; Minh, P.T.; Mai, H.T. An Application of the Multi–Physics Ensemble Kalman Filter to Typhoon Forecast. *Pure Appl. Geophys.* **2013**, *170*, 745–954.

10. Jun, L.I.; Jun, D.U.; Yu, L.I.U. A comparison of initial condition-, multi-physics- and stochastic physics-based ensembles in predicting Beijing "7.21" excessive storm rain event. *Acta Meteorol. Sinica* **2014**. doi:10.11676/qxxb2015.008.

11. Minh, P.T.; Hang, N.T.; Thuy, P.K.; Gia, C.N.H. An application of the multi-physical method of determining error of WRF models to simulate the track and intensity Usagi storm in 2018. *Sci Technol. Dev. J.: Sci. Earth Environ.* **2021**, *5(1)*, 298–311.

12. Anderson, J.L.; Anderson, S.L. A Monte Carlo implementation of the non–linear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.* **1999**, *127*, 2741–2758.

13. https://en.wikipedia.org/wiki/Typhoon_Damrey_(2017).

14. Phong, N.B.; Hiep, N.V.; Thang, N.V. Application of dynamical vortex initialization scheme on intensity forecast and structure study of typhoon Damrey (2017) during near-shore and landfalling period. *J. Clim. Change Sci.* **2020**, *16,* 23–35.

15. Nang, T.Q.; Tien, T.T. Skill validation of probability tropical cyclone track forecast in Bien Dong. *VN J. Hydrometeorol.* **2020**, *717*, 11–19.

16. Kulaya, K.; Sujittra, R.; Pakpoom, R. Sensitivity of different physics schemes using WRF model in Typhoon Damrey (2017) over the Indochina region. *J. Phys.: Conf. Ser.* **2021**, *2145*, 012046.

17. Hunt, B.R.; Kostelich, E.; Szunyogh, I. Efficient data assimilation for spatiotemporal chaos: a local ensemble transforms Kalman filter. *Physica D*. **2005**, *230*, 112–126.

18. https://www2.mmm.ucar.edu/wrf/OnLineTutorial/compilation_tutorial.php.

19. http://homepages.see.leeds.ac.uk/~lecrrb/wrf/aRWUsersGuide.pdf.

20. Wilks, D.S. Statistical Methods in the Atmospheric Sciences, Ithaca New York. **1997**, *59*, 255.

Check for updates

Research Article

Evaluation of rainfall characteristics over Vietnam simulated by the Non–Hydrostatic Regional Climate Model – NHRCM during the 1981–2001 period

Ha Pham-Thanh[1]*, Hang Vu-Thanh[1], Truong Nguyen-Minh [1]

[1] VNU Hanoi University of Science, 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam; phamthanhha.k56@hus.edu.vn; hangvt@vnu.edu.vn; truongnm@vnu.edu.vn

*Correspondence: phamthanhha.k56@hus.edu.vn; Tel.: +84–972355393

Abstract: This study aims to evaluating characteristics of rainfall simulated by Non–Hydrostatic Regional Climate Model (NHRCM) over seven sub–regions of Vietnam during the 1981–2001 period. Features such as seasonal cycle of monthly average daily rainfall, maximum daily rainfall, and frequencies at different thresholds are compared with observations. Statistical evaluations of errors and correlation coefficient are also examined to see the differences between model results and observations. The results show that NHRCM captured well the seasonal cycle of the simulated monthly average daily rainfall, but magnitudes are underestimated in all sub–regions except South Central Vietnam (N1). Generally, underestimations of the simulated daily rainfall are observed in almost all months and sub–regions, higher differences are found in rainy months and the most underestimation of rainfall is observed in South Vietnam (N3). Correlation coefficients over 0.6 are found in North West Vietnam (B1) and South Vietnam (N3). Monthly absolute maxima of the observed daily rainfall are found in North Central (B4) in transition months when the model usually underestimated significantly. In addition, NHRCM tends to simulate cases with rainfall amounts below 16 mm/day or above 50 mm/day with higher frequencies compared with the observations. In contrast, frequencies detected by NHRCM seem to be lower than those of the observations for rainfall amount in 16 – 50 mm/day, especially in sub–regions N2 and N3. These results are supportive for applying the Regional Climate Model in simulating rainfall characteristics over Vietnam, especially for Non–Hydrostatic version.

Keywords: NHRCM model; Rainfall; Vietnam.

1. Introduction

Southeast Asia including Vietnam is a monsoon region where extreme events frequently occur, especially associated with precipitation. Precipitation is an essential component of the earth's climate system affecting the eco–system and socio–economy activities [1]. Numerical experiments of Regional Climate Models (RCM) have been implemented so far by many authors such as [2–7]. Along with the development of science and technology in recent years, super computers have been able to run much finer grid RCMs that help to well reproduce extreme events such as heavy rain and intense heat [8]. It is well known that precipitation simulations of RCMs are biased due to limited understanding of atmospheric physical processes or insufficient spatial resolution [9]. Therefore, a major important thing is that output of RCMs should be evaluated with historical observations [10]. There are several studies [8–13] investigated the skill of the Non–Hydrostatic Regional Climate Model

(NHRCM) of the Meteorological Research Institute (MRI) in simulating the present climate, including heavy precipitation as well as in producing climate projection over Japan and Southeast Asia. Five–year integration from 2001 to 2006 was carried out continuously in the study [8] to evaluate reproducibility of the present climate over Japan using the double–nesting NHRCM with the analysis data. The simulated results of the 4–km grid spacing NHRCM showed that the average annual precipitation was overestimated by only 8% compared to the Automated Meteorological Data Acquisition System (AMeDAS) observations. NHRCM simulated reasonable frequency distribution of precipitation intensity, with a little higher value for the frequency of heavy precipitation than that observed by the AMeDAS [8]. A 20–year–integration of NHRCM of the present climate using an inner nested grid with a spacing of 5 km showed that the model performance fluctuated significantly among regions over Japan. In particular, the model tended to give underestimated precipitation from 20% to 40% along the coast of the Japan Sea and the Nansei Archipelago and overestimated precipitation by more than 40% at some inland locations when compared with observed amounts. However, the model well reproduced the spatial and temporal distributions of the annual mean temperature and precipitation [11]. The study of [12] evaluated the accuracy of four RCMs, including NHRCM, through daily precipitation indices during the period 1985–2004 over Japan at 20–km grid interval. The results indicated that for most indices, such as mean precipitation, number of wet days, mean amount per wet day, 90th percentile of daily precipitation, number of days with precipitation ≥90th percentile of daily precipitation, are often produced well by NHRCM model. However, the accuracy varies depending on the indices, seasons and aspects. The study also highlighted that higher–resolution dynamical models derive a better representation of most daily precipitation indices than reanalysis data, especially in areas with complex terrain and land–sea distribution. The present climate over Southeast Asia, from 1989 to 2008, was simulated by NHRCM with 25–km resolution and 50 vertical levels [13]. It is shown that the model can capture well the topographic effect on rainfall, but the simulated values can be wet (dry) biases in the windward (leeward) side of mountains compared to APHRODITE. In terms of both seasonality and daily distribution of rainfall, the NHRCM model tends to underestimate the number of wet days during the respective wet season of the sub–regions and to overestimate daily rainfall intensity. There are many researches on RCM performance assessment such as RegCM, MM5, WRF… over Vietnam especially on examining rainfall simulation. The study of [14] used RegCM3 to analyze the simulations of surface temperature and precipitation over the Red River Delta of Vietnam. Model results are compared with observations at the 17 meteorological sites in Red River Delta during the 1980–1999 baseline period. They conclude that temperature is underestimated systematically as well as precipitation has cold bias during summer and autumn while is well reproduced in winter and spring. Besides that, the ability simulations of the climate regional non–hydrostatic NHRCM and hydrostatic RegCM4.2 models from 1985–2007 showed in the study of [15]. It can be found that surface temperatures simulated by NHRCM have warm biases while vice versa of RegCM compared with observations. Rainfall simulations of NHRCM have a better agreement with observations than those of RegCM4.2, especially NHRCM can reproduce some heavy rainfall centers. In the study [16], clWRF model is used to simulate the rainfall amount in the period of (1981–2000). The results show that the model simulations are often overestimated comparing to observations especially in rainy season and in southern surface meteorological stations. The performance of simulation and projection of rainfall and tropical cyclone activity over Vietnam was investigated in [17]. The simulated results showed that climatic heavy rainfall centers are well captured in the seasonal cycle. However, the model overestimates rainfall in comparison with the APHRODITE data. In addition, NHRCM underestimates rainfall in North Vietnam but overestimates rainfall in South Vietnam in June–July–August compared to the rain gauge data. The model also

overestimates rainfall in September but underestimates in October–November in Central Vietnam.

The purpose of this paper is to evaluate rainfall simulation of NHRCM over Vietnam as well as to analyze its characteristics in the present climate from 1981 to 2001. Main features including seasonal cycle of monthly average daily rainfall, maximum daily rainfall, and frequencies at different thresholds are compared directly with observations. The rest of this work is organized as follows. Section 2 will provide details of model experiment and data. Section 3 describes results and discussions. Concluding remarks are given in Section 4.

2. Data and Methods

2.1. Data

In this study, daily rain gauge data observed at 127 selected meteorological stations, as shown in Figure 1, are used to estimate the seasonal march of the simulated rainfall in seven sub–regions of Vietnam during the 1981–2001 period. The ERA–Interim reanalysis (hereafter denoted as ERA) is a global atmospheric reanalysis dataset developed by the European Centre for Medium–Range Weather Forecasts (ECMWF), which covers from January 1989 onward and is continuously updated in near–real time [18]. In this study, the ERA data for precipitation and the wind field on the 850mb pressure level for the period 1981–2001 are used to evaluate the performance of the NHRCM model in capturing the features of wind across a year.
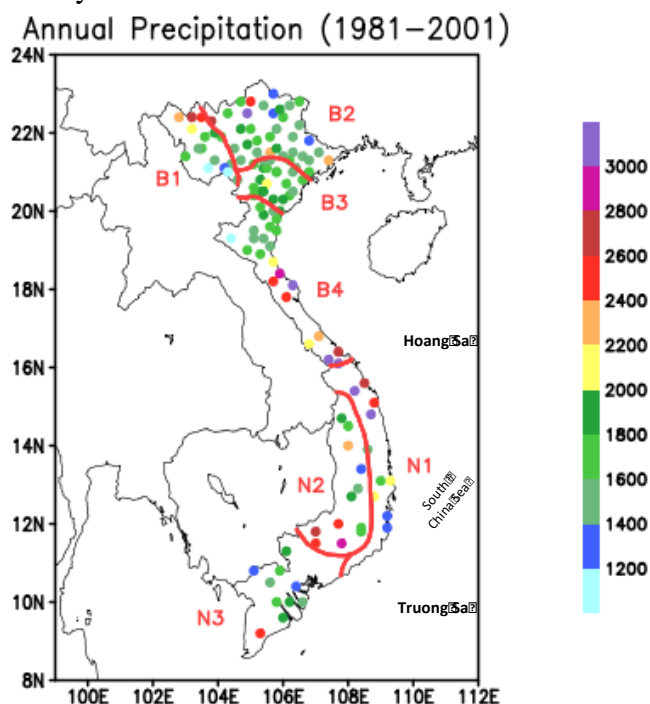
Figure 1. Colors indicate average annual precipitation (mm) over seven sub–regions of Vietnam during the 1981–2001 period. Circles show meteorological observation sites and lines display the boundaries of seven sub–regions.

2.2. Methods

The regional climate model NHRCM is the extended version of an operational Non–Hydrostatic Model (NHM) developed by the Meteorological Research Institute (MRI) and the Numerical Prediction Division of the Japan Meteorological Agency (NPD/JMA). The detail descriptions of NHM can be found in [19]. The Kain–Fritsch scheme [20] is used in this study for cumulus convective parameterization. The soil model is replaced by

MRI/JMA– Simple Biosphere model (MJ–SiB, [21]) and lateral boundary conditions are replaced by spectral nudging boundary conditions.

Simulations have been done over a domain of (6.4° – 25.2°N; 98.4° – 112.5°E) as shown in Figure 2. The time steps for the model configured with 5–km horizontal resolution and 50 vertical levels. Initial and boundary conditions for NHRCM are provided by a simulation performed by an atmospheric general circulation model with a 20–km horizontal grid spacing (AGCM20; [22]). Lateral boundary conditions are updated every 6 hours. The sea surface temperature is given by the Hadley Centre Sea surface temperature data set version 1 (HadISST1) four times a day [23]. Time integration was implemented continuously from January 1981 to December 2000. For each year, the simulation began at 00 UTC 20 November and ended 00 UTC 31 December. The first 42 days of the simulation were regarded as the model spin–up and discarded.
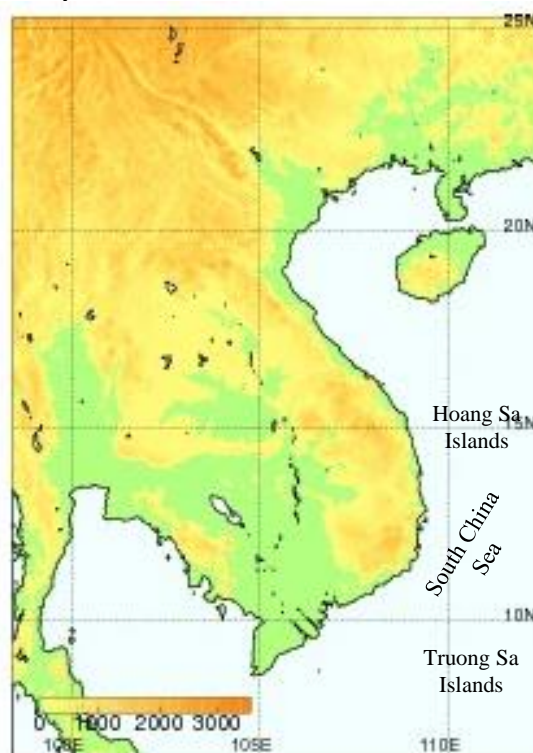
Figure 2. Model domain and topography.

3. Results and discussions

Figure 3 shows the mean 850–hPa wind patterns in March–April–May (MAM), June–July–August (JJA), September–October–November (SON) and December–January–February (DJF) of NHRCM and ERA during 1981–2001. In general, the spatial and temporal variability of the prevailing 850–hPa winds are simulated well by NHRCM compared with ERA for the whole Vietnam region. However, the values of NHRCM are a little higher than those of ERA. During MAM, the mean 850–hPa winds are southwesterly over North and North Central Vietnam, while they are easterly and northeasterly over South Central and South Vietnam. In JJA, the prevailing 850–hPa winds are westerly almost all over Vietnam, however, they change to northeasterly and easterly in Central and South Vietnam while southeasterly and southerly in North Vietnam in SON and DJF. The effects of interactions, especially the interaction between circulations and topography, result in the seasonal cycle of rainfall over each sub–region as illustrated in Figure 4.
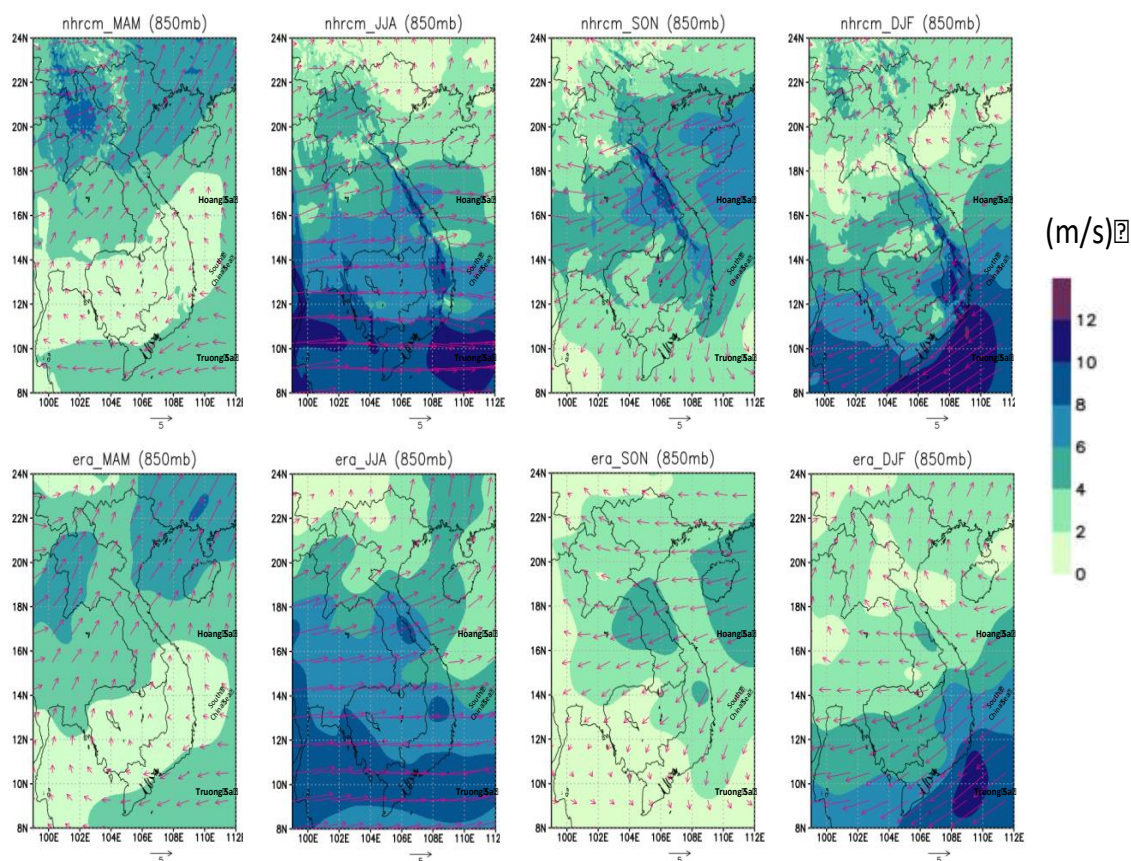
Figure 3. Mean winds at 850–hPa during the 1981–2001 period in (a) MAM, (b) JJA, (c) SON and (d) DJF of NHRCM (upper) and ERA (lower).

Figure 4 describes the monthly average daily rainfall in seven sub–regions of Vietnam of the reanalysis (ERA), model (NHRCM), and observation (OBS) data. It is clear that the maximum average daily rainfall occurs in autumn in B4, N1, and N3 sub–regions while in summer in other sub–regions. The delay of the rainy season in B4 and N1 is due to the barrier effect of topography on prevailing winds during SON, causing heavy rainfall on the windward side of Truong Son Mountains (Figure 3c). According to the observations, the daily rainfall is commonly from 10 mm/day to 15 mm/day in almost all sub–regions, however, rainfall amounts can frequently reach 20 mm/day in B4 and N1 sub–regions (Figure 4a). The average daily rainfall is well captured in terms of the evolution but a little underestimated by NHRCM in all sub–regions except for N1 (Figure 4b). Figure 4c shows that the ERA average daily rainfall also has a good agreement with the observations in the annual march but is much overestimated in all sub–regions.
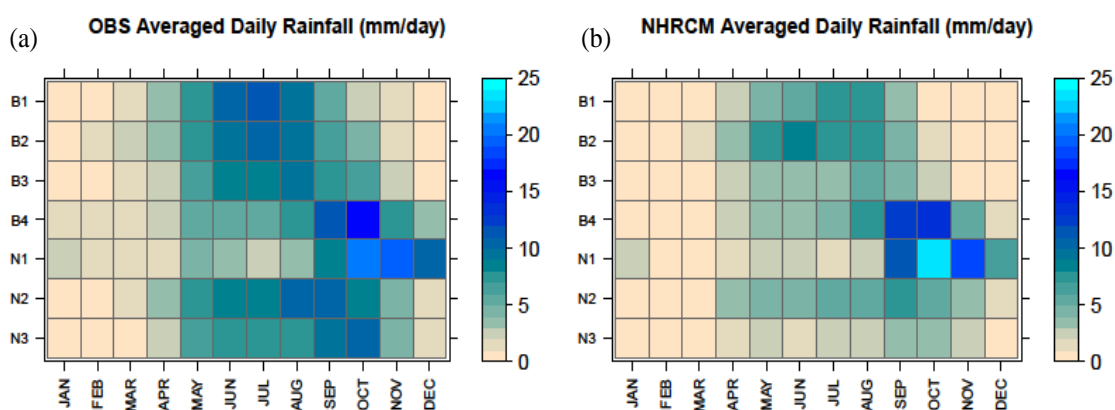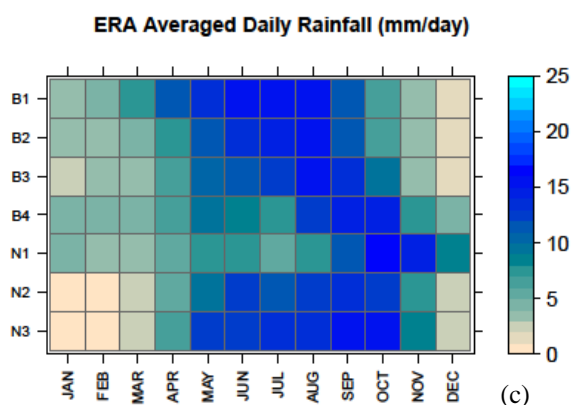
Figure 4. Monthly average daily rainfall in seven sub–regions of Vietnam of (a) observation, (b) NHRCM and (c) ERA during the 1981–2001 period.

The differences in the monthly average daily rainfall between the ERA and NHRCM data and the observations in seven sub–regions are displayed in Figure 5. Accordingly, positive differences between ERA and the observations are found in almost all sub–regions except in October–November for B4, and in October–December for N1. In general, these differences tend to be higher in the rainy season, and the highest ones of 8 mm/day between ERA and the observations can be seen in B1. Conversely, negative differences are detected in almost all sub–regions when comparing the NHRCM and observation data. It is clear that higher differences occurred in higher rainfall months, and the most underestimated rainfall area is N3.
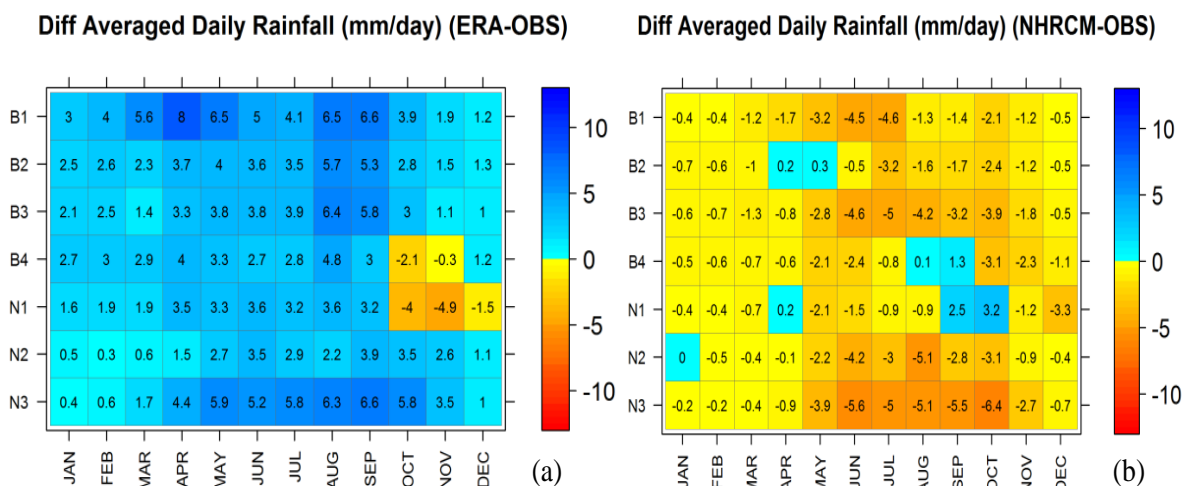
Figure 5. Monthly average daily rainfall differences between (a) ERA and (b) NHRCM and the observations in seven sub–regions of Vietnam during the 1981–2001 period.

Scatter plots of the average daily rainfall of the observations and NHRCM in seven sub–regions and the whole of Vietnam during 1981–2001 are depicted in Figure 6. Accordingly, the correlations of B1 and N3 sub–regions are over 0.6 while the others are over 0.4. The correlation coefficient of the whole of Vietnam is 0.46. It is clear that rainfall amounts of 25–50 mm/day are often observed in B2 sub–region, however, such amounts are normally underestimated by NHRCM. The daily rainfall values of NHRCM in other sub–regions are overestimated compared with the observations.

Maxima and means of the monthly maximum daily rainfall of the observations and NHRCM over seven sub–regions of Vietnam from 1981 to 2001 are shown in Fig. 7. In general, the monthly mean variations of NHRCM well follow those of the observations, however, a clear underestimation is found in N3. Besides, the maxima of the NHRCM maximum daily rainfall are underestimated compared with the observations in late autumn and early winter and vice versa in late spring and summer in almost all sub–regions.
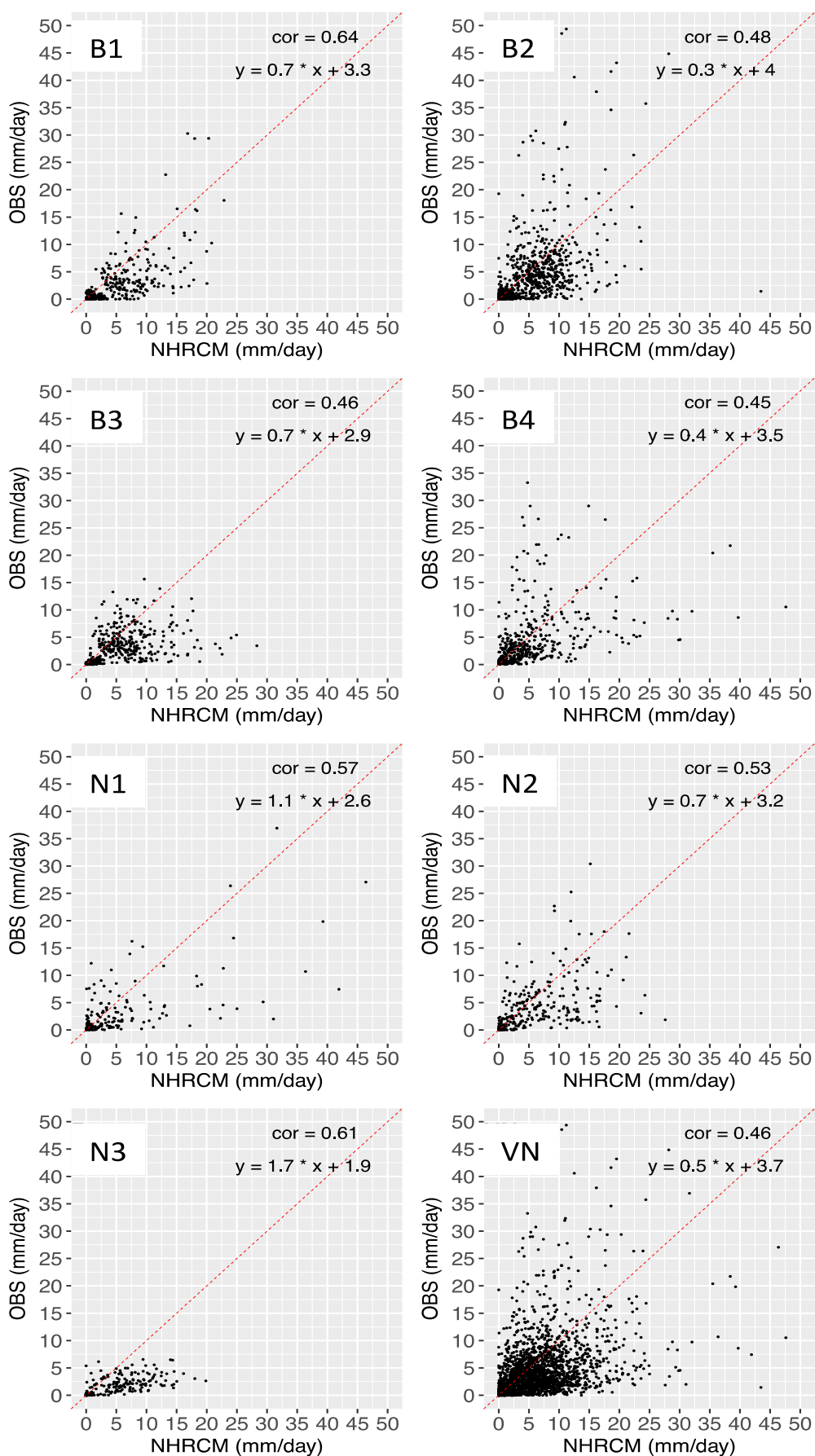
Figure 6. Scatter plots of the average daily rainfall (mm/day) of the observations and NHRCM in seven sub–regions and the whole of Vietnam during the 1981–2001 period.
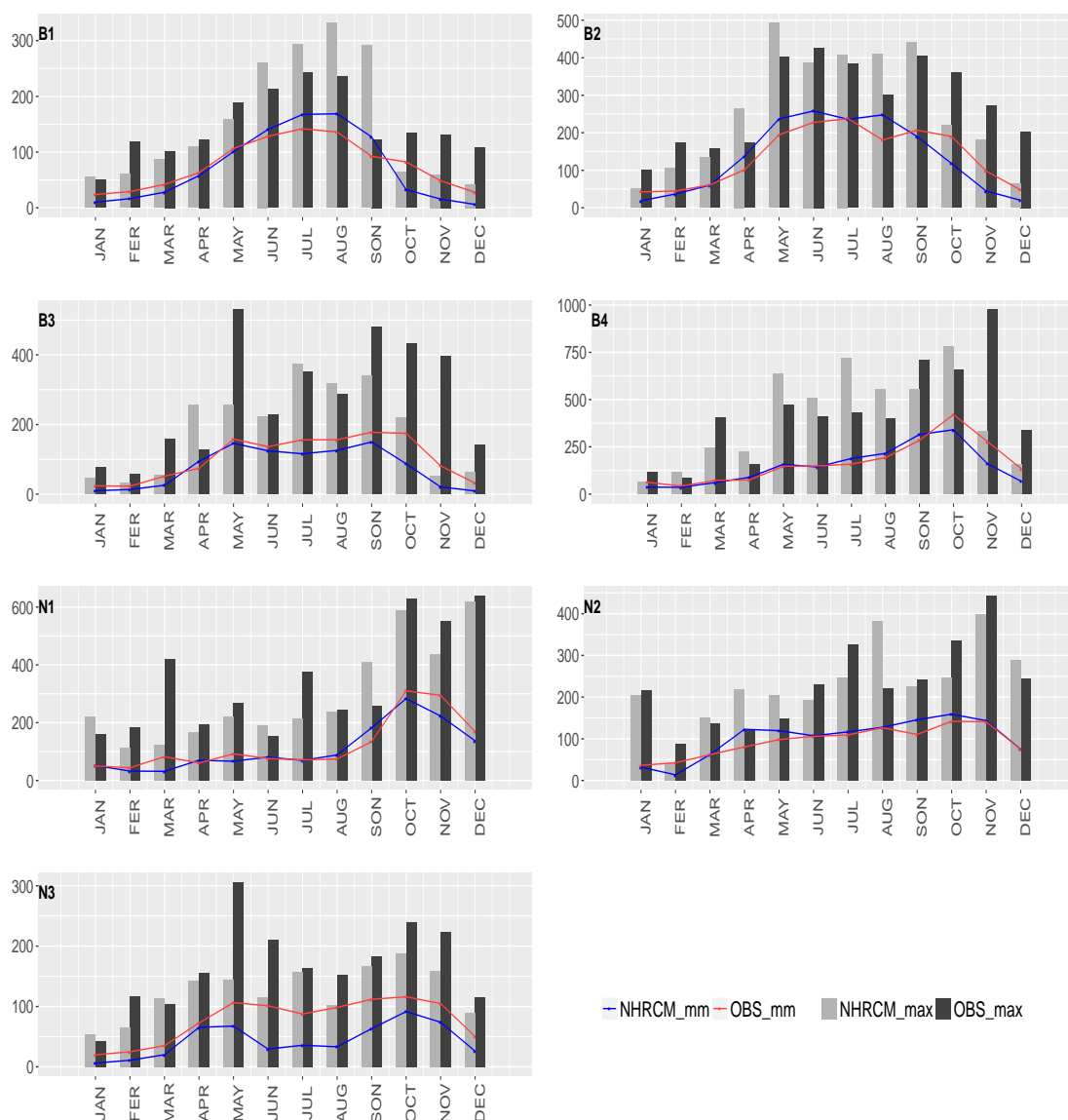
Figure 7. Maxima (columns) and means (lines) of the monthly maximum daily rainfall of the observations and NHRCM over seven sub–regions of Vietnam during the 1981–2001 period.

Frequencies of the daily rainfall at different thresholds of the observations and NHRCM over seven sub–regions of Vietnam during the 1981–2001 period are illustrated in Fig. 8. Accordingly, frequency distribution of the observed daily rainfall is captured well by NHRCM. However, NHRCM tends to simulate cases with rainfall thresholds of below 16 mm/day or above 50 mm/day with higher frequencies compared with the observations. In contrast, frequencies given by NHRCM seem to be lower than the observations for rainfall thresholds from 16 mm/day to 50 mm/day, especially in sub–regions N2 and N3.

Frequencies of the daily rainfall at different thresholds of the observations and NHRCM over seven sub–regions of Vietnam during the 1981–2001 period are illustrated in Fig. 8. Accordingly, frequency distribution of the observed daily rainfall is captured well by NHRCM. However, NHRCM tends to simulate cases with rainfall thresholds of below 16 mm/day or above 50 mm/day with higher frequencies compared with the observations. In contrast, frequencies given by NHRCM seem to be lower than the observations for rainfall thresholds from 16 mm/day to 50 mm/day, especially in sub–regions N2 and N3.
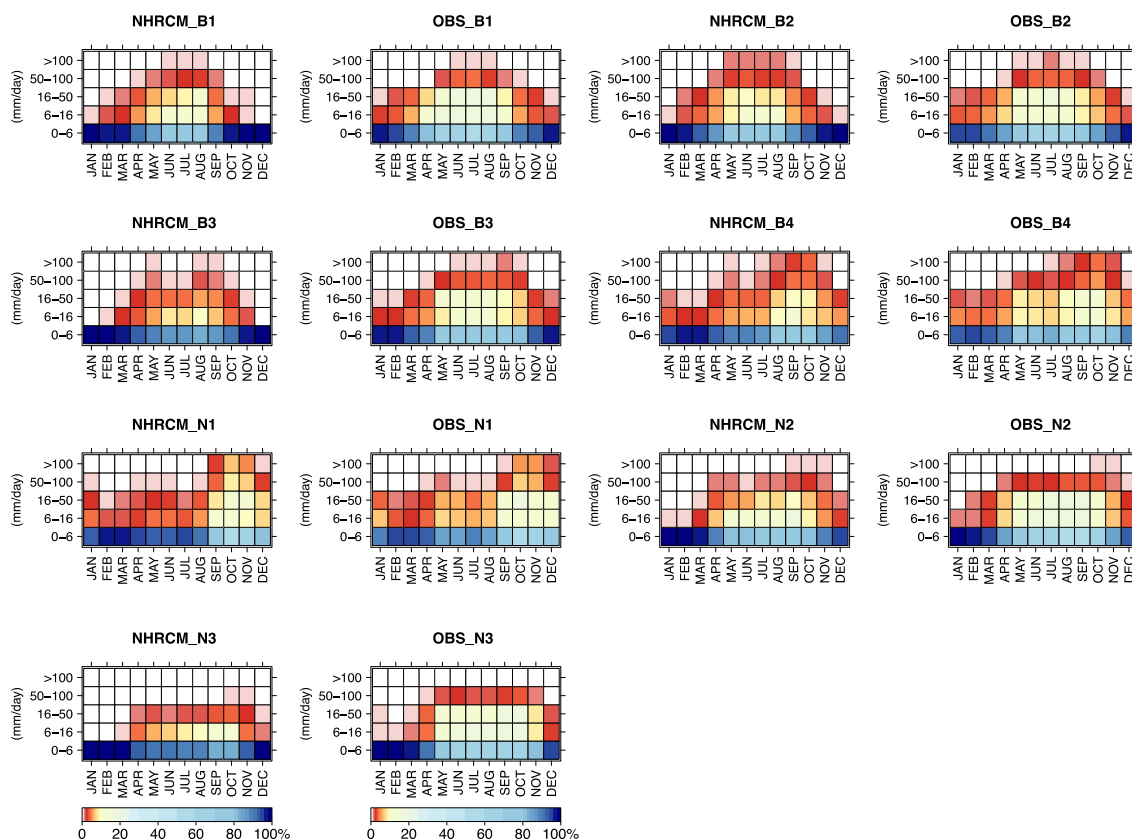
Figure 8. Frequencies of the daily rainfall at different thresholds of the observations and NHRCM over seven sub–regions of Vietnam during the 1981–2001 period.

4. Conclusions

The main goal of this paper is to evaluate the rainfall simulation of NHRCM over Vietnam and analyze its characteristics in the present climate from 1981 to 2001. The results show that the observed seasonal cycle of rainfall is captured well by NHRCM for all sub–regions, but magnitudes are underestimated in all sub–regions except South Central Vietnam (N1). Generally, higher differences are found in rainy months, and the most underestimation of rainfall is observed in South Vietnam (N3). However, correlation coefficients of the average daily rainfall between the observations and NHRCM are over 0.4 for all sub–regions and are higher in North West Vietnam (B1, about 0.64) and South Vietnam (N3, about 0.61). The monthly absolute maxima of the observed daily rainfall are found in North Central (B4) in transition months when the model is underestimated significantly. Besides, NHRCM tends to simulate cases with rainfall thresholds of below 16 mm/day or above 50 mm/day with higher frequencies compared with the observations. Conversely, frequencies found in the NHRCM rainfall data seem to be lower than the observations for cases with rainfall thresholds ranging from 16 mm/day to 50 mm/day, especially in sub–regions N2 and N3. These results have inspired a further study to examine statistical or bias correction for the rainfall simulations and projections of NHRCM over Vietnam.

Author Contributions: Conceptualization; methodology; software; validation; formal analysis; investigation; resources; data curation; writing–original draft preparation; writing–review and editing; visualization: H.P.T.; Conceptualization; methodology; formal analysis; investigation; writing–review and editing; supervision: V.T.H.; Conceptualization; methodology; formal analysis; investigation; writing–review and editing; supervision: N.M.T.

References
1. Wang, B.; Ho, L. Rainy season of the Asian–Pacific summer monsoon. *J. Climate*. **2000**, *15*, 386–398.
2. Dickinson, R.E.; Errico, R.M.; Giorgi, F.; Bates, G.T. A regional climate model for the western United States. *Clim. Change*. **1989**, *15*, 383–422.
3. Giorgi, F.; Bates, G.T. The climatological skill of a regional model over complex terrain. *Mon. Wea. Rev.* **1989**, *117*, 2325–2347.
4. Giorgi, F. On the simulation of regional climate using a limited area model nested in a general circulation model. *J. Climate.* **1990**, *3*, 941–963.
5. Kida, H.; Koide, T.; Sasaki, H.; Chiba, M. A new approach to coupling a limited area model with a GCM for regional climate simulations. *J. Meteor. Soc. Japan.* **1991**, *69*, 723–728.
6. Liang, X.Z.; Li, L.; Kunkel, K. Regional climate model simulation of US precipitation during 1982–2002 – part I: annual cycle. *J. Clim.* **2004**, *17*, 3510–3528.
7. Reboita, M. S.; da Rocha, R. P.; Ambrizzi, T.; Sugahara, S. South Atlantic Ocean cyclogenesis climatology simulated by regional climate model (RegCM3). *Clim. Dynam.* **2010**, *35*, 1331–1347.
8. Sasaki, H.; Kurihara, K.; Takayabu, I.; Uchiyama, T. Preliminary experiments of reproducing the present climate using the Non–hydrostatic Regional Climate Model. *SOLA*. **2008**, *4*, 25–28, doi:10.2151/sola.2008-007.
9. Rauscher, S.; Coppola, E.; Piani, C.; Giorgi, F. Resolution effects on regional climate model simulations of seasonal precipitation over Europe. *Clim. Dynam.* **2010**, *35*, 685–711.
10. Terink, W.; Hurkmans, R.T.W.L.; Torfs, P.J.J.F.; Uijlenhoet, R. Evaluation of a bias correction method applied to downscaled precipitation and temperature reanalysis data for the Rhine basin. *Hydrol. Earth Syst. Sci.* **2010**, *14*, 687–703.
11. Sasaki, H.; Murata, A.; Hanafusa, M.; Oh'izumi, M.; Kurihara, K. Reproducibility of present climate in a Non–Hydrostatic Regional Climate Model nested within an Atmosphere General Circulation Model. *SOLA*. **2011**, *7*, 173–176. doi:10.2151/sola.2011-044.
12. Iizumi, T.; Nishimori, M.; Dairaku, K.; Adachi, S. A.; Yokozawa, M. Evaluation and intercomparison of downscaled daily precipitation indices over Japan in present–day climate: Strengths and weaknesses of dynamical and bias correction–type statistical downscaling methods. *J. Geophys. Res.* **2011**, *116*. doi:10.1029/2010JD014513.
13. Cruz, F.T.; Sasaki, H. Simulation of present climate over Southeast Asia using the Non–hydrostatic Regional Climate Model. *SOLA*. **2017**, *13*, 13–18, doi:10.2151/sola.2017-003.
14. Ngo–Duc, T.; Nguyen, Q.T.; Trinh, T.L.; Vu, T.H.; Phan, V.T.; Pham, V.C. Near future climate projections over the Red River Delta of Vietnam using the Regional Climate Model Version 3. *Sains Malaysiana.* **2012**, *41(11)*, 1325–1334.
15. Xin, K.T.; Hang, V.T.; Duc, L.; Linh, N.M. Climate simulation in Vietnam using regional climate nonhydrostatic NHRCM and hydrostatic RegCM models. *VN J. Nat. Sci. Technol.* **2013**, *29(2S)*, 243–251.
16. Hang, V.T.; Hanh, N.T. Monthly temperature and precipitation seasonal forecast over Vietnam using clWRF model. *VN J. Earth Environ. Sci.* **2014**, *30(1)*, 31–40.

17. Kieu–Thi, X.; Vu–Thanh, H.; Nguyen–Minh, T.; Le, D.; Nguyen–Manh, L.; Takayabu, I.; Sasaki, H.; Kitoh, A. Rainfall and tropical cyclone activity over Vietnam simulated and projected by the Non–Hydrostatic Regional Climate Model – NHRCM. *J. Meteor. Soc. Japan*. **2016**, *94A*, 135–150. doi:10.2151/jmsj.2015-057.

18. Dee, D.P.; et al. The ERA–Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.* **2011**, *137*, 553–597. https://doi.org/10.1002/qj.828.

19. Saito, K.; Fujita, T.; Yamada, Y.; Ishida, J.; Kumagai, Y.; Aramani, K.; Ohmori, A.; Nagasawa, R.; Kumagai, S.; Muroi, C.; Kato, T.; Eito, H.; Yamazaki, Y. The operational JMA Nonhydrostatic Mesoscale Model. *Mon. Wea. Rev.* **2006**, *134*, 1266–1298.

20. Kain, J.; Fritsch, J. Convective parameterization for mesoscale models: The Kain–Fritsch scheme. The Representation of Cumulus Convection in Numerical Models, Meteor. Mongor., *Amer. Meteor. Soc.* **1993**, *46*, 165–170.

21. Hirai, M.; Sakashita T.; Kitagawa, H.; Tsuyuki, T.; Hosaka, M.; Oh'izumi, M. Development and validation of a new land surface model for JMA's operational global model using the CEOP observation dataset. *J. Meteor. Soc. Japan*. **2007**, *85A*, 1–24.

22. Mizuta, R.; Yoshimura, H.; Murakami, H.; Matsueda, M.; Endo, H.; Ose, T.; Kamiguchi, K.; Hosaka, M.; Sugi, M.; Yukimoto, S.; Kusunoki, S.; Kitoh, A. Climate simulations using AGCM3.2 with 20–km grid. *J. Meteor. Soc. Japan*. **2012**, *90A*, 233–258.

23. Rayner, N.A.; Parker, D.E.; Horton, E.B.; Follan, C.K.; Alexander, L.V.; Rowell, D. P.; Kent, E.C.; Kaplan, A. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **2003**, *108*, D14, 4407. doi:10.1029/2002JD002670.

Table of content