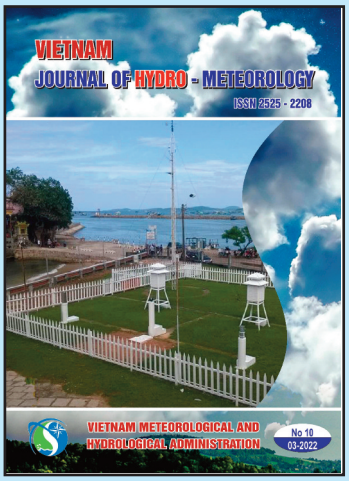
VIETNAM JOURNAL OF HYDRO - METEOROLOGY

ISSN 2525 - 2208

VIETNAM METEOROLOGICAL AND HYDROLOGICAL ADMINISTRATION

No 10
03-2022



Publishing licence
No: 166/GP-BTTTT - Ministry of Information and Communication dated 17/04/2018

Editorial office
No 8 Phao Dai Lang, Dong Da, Ha Noi
Tel: 024.39364963
Email: tapchikttv@gmail.com

Engraving and printing
Vietnam Agriculture Investment Company Limited
Tel: 0243.5624399

VIETNAM JOURNAL OF HYDROMETEOROLOGY
Volume 10 - 03/2022

TABLE OF CONTENT

VIETNAM JOURNAL OF HYDROMETEOROLOGY

Research Article

Characteristics of Heavy Metal Pollution in Industrial Sludge and an Environmental–Friendly Removal Method

Nguyen Thuy Chung[1*], Luong Thi Mai Ly[2], Nguyen Xuan Binh[3], Pham Minh Chinh[4]

[1] School of Environmental Science and Technology, Hanoi University of Science and Technology; chung.nguyenthuy@hust.edu.vn
[2] Faculty of Environmental Sciences, University of Science, Vietnam National University, Hanoi; luongmaily@hus.edu.vn
[3] Institute of Science and Technology – Ministry of Public Security; binh14.163.3@gmail.com
[4] Faculty of Environmental Engineering, National University of Civil Engineering; chinhnkt16@gmail.com

***** Corresponding author: chung.nguyenthuy@hust.edu.vn; Tel.: +84–886114668

Received: 05 December 2021; Accepted: 13 January 2022; Published: 25 March 2022

Abstract: Sludge from wastewater treatment plants (WWTPs) in industrial park is currently a serious problem in Vietnam as well as many countries around the world. Unlike other by–products, sewage sludge from WWTPs contains a lot of toxic components, heavy metals, persistent organic substances and many other hazardous ingredients in high concentrations. Up to now, there has not been a Vietnamese study focusing on systematically assessing the level of toxic pollutants in industrial sludge in Vietnam. Therefore, this study focuses on evaluating the characteristics of industrial wastewater sludge in a specific industrial park, and thereby determining the characteristics and current status of heavy metal pollution in the sludge compared with agricultural soil samples. This study determined the heavy metals enrichments and their possible sources in industrial sludges from different sampling time. The results show that industrial sludge exhibits very high pollution for some typical heavy metals, especially Cu and Cd. The analysis of the correlation relationship between heavy metals also helps to identify the source of emission of heavy metals in the sludge sample. The PI, *Igeo* indexes are also 2–10 times higher than the control soil samples. In addition, the study also used citric acid, GLDA and ascorbic acid solutions as a method of heavy metal extraction from sludge with relatively high efficiency (~80%). Among the chelators, GLDA can be selected as the most effective removal with high capacity to remove Zn and Pb.

Keywords: Industrial sludges; Heavy metal removal; Pollution; Enrichment.

1. Introduction

Industrial sludges from various industries contain trace metals, organic compounds, macronutrients, micronutrients, organic microbial contaminants, microorganisms and eggs of parasitic organisms [1–3]. Previous studies have shown that the annual amount of sludge is constantly increasing due to urbanization and industrialization. Most of the hazardous heavy metals (Zn, Sb, Cr, Ni, Hg, Cd, Sn and V) are found in the sludge with relative concentrations many times higher than the allowable limit [1]. When arbitrarily disposing of industrial sludge into the environment, heavy metals will easily spread to surface water, groundwater and seep into the ground. Heavy metals usually exist in sludge in 5 forms: ionic

form, carbonate bound form, bound form inside or outside the solid particle mass with iron oxide and manganese, bound form with organic compounds, inert form, stable in the environment with mineral grain structure, difficult to be released under natural conditions [3].

Heavy metals such as Cr, Ni, Cu, As, Cd and Pb have been recognized as hazardous elements for the environment. The occurrence of heavy metals in the industrial wastewater and sludge are of interest because they would be often presented at considerable quantities and if leaked into surface waters or arable land, that can have severe effects on the environment and public health [4–5]. Sludge which disposes from different industries contains trace metals, organic compounds, macronutrients, micronutrients, organic micro pollutants [6]. So, the accumulation of industrial sludge poses environment problem and the bioavailable fractions of these wastes may result in secondary environmental pollution. Therefore, contamination of environment by heavy metals from untreated wastewater and sludge of various industries is a worldwide environmental problem [6–8]. Unlike organic wastes, heavy metals are non–biodegradable and thus must be treated to avoid polluting the environment. [5] studied the effect of heavy metals in sewage sludge applied to soil on its metal availability and the growth and yield of crops. Their results indicated that the yields of both cereals and legumes in dressed regions were lower than those of control regions.

In Vietnam, there are clear signs of heavy metal pollution from industrial sludges. The industrial waste problem has become one of the prime concerns in many provinces of Vietnam. Many industries have set up in and around the cities during the last decade, and the number of new industries is continually increasing. In recent times, the rapid development of various industries has created environmental problems that pose a serious threat to the environment [9–12]. A recent study shows that although environmental management has improved in recent years, heavy metal pollution levels are still high in sludge from Hanoi's Kim Nguu River. Most of the sludge samples here have concentrations of Cr, Ni, Cu, As, Cd and Pb exceeding the permissible standards of Vietnam (QCVN 50: 2013/BTNMT), which is caused by industrial wastewater discharged into the Kim Nguu River [12]. This result is in stark contrast to the situation in the Mekong Delta, which is less polluted by industrial activities [10]. The risk of environmental pollution from sludge can be found in Vietnamese statistical reports around Ho Chi Minh City [13–14] where there are many industrial plants but information on composition and volume sludge is still a gap that does not meet the current environmental management needs of Vietnam.

Various surveys of the heavy metal concentrations in sewage sludge have been undertaken to evaluate the suitability of sludge for land application. It is particularly important to study ecological risk assessment to industrial sludges because modern industrial areas are often densely populated due to the presence of industrial and commercial activities as well as easy access to amenities such as transportation, electricity, water, entertainment, and healthcare.

Therefore, this paper studies the characteristics of sludge in an industrial park in the Northern province, thereby assessing the pollution level of heavy metals in the sludge samples. In addition, the study also identifies the emission sources of toxic substances in the analyzed sludge sample, find an appropriate method of recovering heavy metals from the sludges. This study aims to propose an efficient and friendly method of heavy metal extraction from the sludges.

2. Materials and Methods

2.1. Collecting sludge samples

Sludge samples were collected from wastewater treatment plants located in an industrial park (BT, 11 samples) and agricultural soil (control, 4 samples) in the North of Vietnam.

Industrial sludges of the industrial park were collected and managed in the form of ordinary solid waste. Each factory that generates heavy metals in conventional sludge has its own wastewater treatment system, treating heavy metals and some substances before being poured into the general treatment system of the industrial park. However, industrial sludge is generated during wastewater treatment. Industrial wastewater, which is treated in wastewater treatment plants, usually meets column B standards according to QCVN 40:2011/BTNMT before being connected to a common wastewater treatment plant of industrial parks [15]. The industrial park's wastewater treatment plant has been synchronously invested and built by a leading unit in the field of wastewater treatment plant construction for industrial zones. The wastewater treatment system met the standard of grade A wastewater treatment (QCVN 40:2011/BTNMT) with the total capacity according to the design of 4 modules is 10,000 m^3/day. The water collection system in the industrial parks is MMBR system (Moving bed biofilm reactor system) and they had the standard for centralized WWTP. Currently 1 module has been operating with a capacity of 2,500m^3/day. The wastewater treatment system has been fully trained and transferred to the industrial park for operation. Industrial wastewater is treated at each wastewater treatment plant, after partially resolved the heavy metal component, it is brought into the centralized water treatment area. The industrial park which was chosen in this study has 23 manufacturing companies, including 7 production facilities with wastewater containing high heavy metal content, including metallurgical, mechanical and chemical plants.

2.2. Sample pre–treatment

Sludge samples were taken 4 times in a year 2020, total samples were 12, each sample weight was 200 grams and stored in sealed zipper bags. In the sludge storage area, a small shovel was used to scoop up samples at 5 points, then mix well and take 200 gram each. Some indicators such as pH, EC, ORP, COD, T–N, T–P are analyzed at the wastewater treatment station [16]. After collection, the samples were stored in sealed foam containers, in a cool place. Samples were moved to the laboratory according to TCVN 6663–15:2008 (ISO 5667–15:1999). The collected sludge sample is dried in a dark and closed room, then the sample is crushed, removed impurities, sieved through a sieve with a pore size of 0.63 µm and collected samples with a particle size < 0.63 µm to analyze the metal content in the most active sedimentary phase, containing mainly clay and meat particles. Samples were stored in a deep refrigerator waiting for analysis, before analysis the samples were left at room temperature and the drying coefficient was determined according to TCVN 4080:2011. Industrial sludge is collected at the mud drying yard by suitable tools. For the control soil sample, 4 agricultural soil samples were also taken for comparison as the background value for each sampling time. We aim to compare the industrial sludges with agricultural soil samples for risk assessment. The soils is nearby the industrial park and can be shown that they were not contaminated. The morphology and elemental contents on the surface of the investigated sludge samples were observed and analyzed using a scanning electron microscope (SEM) and energy dispersive spectroscopy (EDS) techniques (SEM–EDS, JEOL JSM–7600F using its variable pressure mode and an accelerating voltage of 15 kV).

2.3. Total metal concentration analysis

Total heavy metal analysis was performed in accordance with the sample handling procedure for the analysis of Cd, Cr, Cu, Pb, Ni and Zn metals, which was conducted according to the guidelines of EPA 3050B [17]. Analytical grade (AG) chemicals, procured from E–Merck, India, were used throughout the study without any further purification. The metal standards were prepared from stock certified standard solution of 1000 mg/l (Merck, Germany) by successive dilution with ultra–pure water (TKA Milli–Q Ultra–Pure Water System, Germany). The analytical quality control was assured by standard operating

procedures, repeated analysis of reagent blanks and recovery analysis of several spiked samples. Other parameters were analysed using the standard methods of US. EPA.

2.4. Assessment of heavy metal pollution according to pollution indicators

2.4.1. Pollution Index (PI)

PI: Single pollution index of heavy metals, determined according to the following formula:

$$\mathrm{PI} = \mathrm{TE(industrial\ sludge\ sample)/TE(control\ sample)} \quad (1)$$

where TE (waste sludge) is the average value of heavy metals in the sludge; TE (control soil) is the mean value in the baseline soil sample; PI is classified as follows: low (PI ≤ 1), moderate (1 < PI < 3), high (PI ≥ 3) [13].

2.4.2. Geoaccumulation indexes (Igeo)

Igeo assesses contamination by comparing the total metal content of the sample with the background value of that metal [14].

$$\mathrm{Igeo} = \log 2 \frac{\mathrm{Cn}}{1{,}5\mathrm{Bn}} \quad (2)$$

where Cn: Metal content in the sample; Bn: Base value of metals in the Earth's crust; 1.5: The factor is given to minimize the impact of possible changes to the background value due to lithological changes in the sediment.

where Igeo ≤ 0: Not contaminated, 0 ≤ Igeo ≤ 1: not– average contaminated, 1 ≤ Igeo ≤ 2: average, 2 ≤ Igeo ≤ 3: average– heavily contaminated, 3 ≤ Igeo ≤ 4: heavily contaminated, 4 ≤ Igeo ≤ 5: heavily and seriously contaminated, 5 ≥ Igeo: seriously contaminated.

2.4.3. Treatment of heavy metals in sludge using environmentally friendly chemicals

In this study, two types of chelator solutions were used: 1. Solution of N, N–Dicarboxymethyl glutamic acid tetrasodium salt (GLDA), ascorbic acid and citric acid. The concentrations of the single washing agent are 200 mM for all chelators. The initial solution of heavy metals (3 metals: Pb, Zn, Cu, with concentration corresponding: 300, 1200 and 150 mg/L) were used to investigate the removal effects on heavy metal washing efficiency. The concentrations were similar with heavy metal concentration in the types of industrial sludges. The reaction time was 24 hours with pH 7.0. A solution of 0.1 N HCl were used as control solution in all experiments. pH values of the washing solution were adjusted using HCl and NaOH. This study was similar with method [5].

2.4.4. Statistical data processing

The classical statistical analyzes were processed using IBM SPSS software version 20. The probability level P < 0.05 was considered to be significant.

3. Results

3.1. Sludge characteristic

After processing and collecting data, a table of hysicochemical properties and SEM images of the sewage sludge samples of the research subjects were presented as follows (Table 1). 4 samples of industrial sludges among 12 samples were chosen to investigate the characteristic. According to QCVN 50:2013/BTNMT, the pH of sludge with pH ≥ 12.5 or pH ≤ 2.0 is defined as hazardous sludge. Looking at the data table, it can be seen that the pH

of the sludge sample ranges from 6.05 ± 0.1 to 6.07 ± 0.12, all of which are neutral for industrial sludge. Total organic carbon fluctuated 634 ± 40.88 mg/L.

Table 1. Summary of sludge characteristic.

Parameters	Sample 1	Sample 2	Sample 3	Sample 4
pH	6.3	6.5	6.4	6.3
ORP (mV)	230	350	402	389
COD (mg/L)	1203	1504	1315	1420
TOC (mg/L)	728	837	204	736
T–N (mg/L)	612.3	936.6	699.4	777.5
T–P (mg/L)	45.1	23.1	32.0	34
Potassium (K)	203	201	210	190
Magnesium (Mg)	102	120	115	113

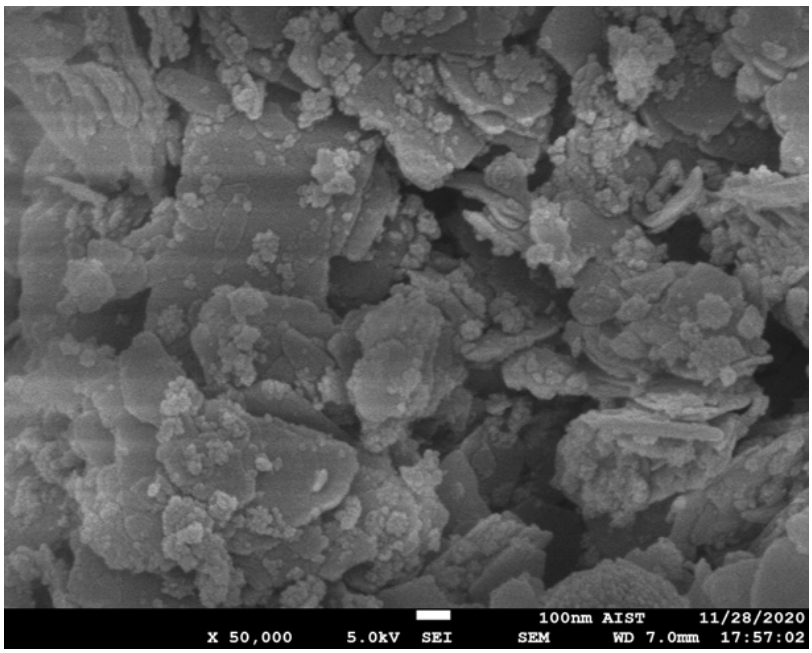


Figure 1. A scanning electron microscope (SEM) of the sludge.

Figure 1 shows microscopic images of representative samples of sludge from the industrial site. In general, the surface morphological characteristics of the sludge are relatively uniform. The high proportions of elemental oxygen and carbon in the sludge samples indicate a large amount of organic matter in the sample, possibly as a result of coagulation of the polymers during the treatment process. The similarity in surface morphology of these samples also supports this conclusion.

3.2. Heavy metal concentrations

The average concentrations of Cd, Cu, Pb, Zn in the sludge samples of the industrial park had the average concentrations of the elements 0.9 mg/kg, 297.2 mg/kg, 164.6 mg/kg and 1177.4 mg/kg, respectively. Most of the elements analyzed were lower than the maximum allowable concentrations for normal sludge (QCVN 50: 2013) (10 mg/kg; 300 mg/kg; 5000 mg/kg, respectively). In some cases, some sludge samples from Ba Thien Industrial Park exceeded the allowable limit for Pb. Therefore, these sludges are considered as hazardous solid wastes, which are not treated and buried according to regulations [15].

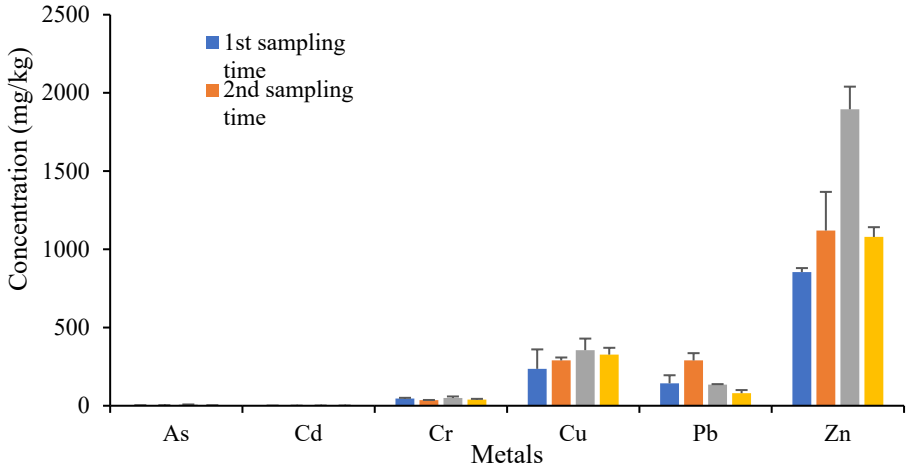


Figure 2. Heavy metal total concentrations in sludges.

3.3. Pollution Index calculation in the industrial sludges

Table 2 describes the pollution index of industrial sludges. The PI (the ratio of TE concentration in sludges to that in soils) was > 3 for Zn, Cu, V and Cr in all samples (Figure 3). Furthermore, the medians of the values were > 5, indicating that these heavy metals were highly enriched in the sludges according to the classification [18].

Table 2. Pollution index of industrial sludges.

	Fe	V	Cd	Cr	Cu	Ni	Mn	Pb	Zn
PI1	64.22	8.54	5.06	9.41	19.87	7.91	3.56	1.65	23.88
PI 2	60.09	11.25	5.52	13.78	35.06	5.12	5.34	1.74	19.83
PI3	95.87	13.96	5.06	16.47	33.26	7.44	6.88	3.97	19.17
PI4	80.28	17.50	4.14	15.13	6.56	3.95	9.61	2.81	17.07
PI5	81.19	19.38	3.22	12.10	8.54	5.12	6.77	2.08	15.38
PI6	130.73	17.50	5.52	19.16	48.54	11.63	5.22	8.24	17.32
PI7	95.87	9.79	4.14	12.77	22.47	5.12	3.68	5.50	14.56
PI8	83.94	20.83	2.76	12.10	26.07	5.58	6.77	11.30	17.28
PI9	87.16	11.88	2.76	11.09	29.66	6.98	7.00	9.77	33.40
PI10	115.14	27.08	3.68	21.51	41.35	10.23	9.97	3.97	40.78
PI11	96.79	12.71	5.52	11.76	22.47	2.51	7.00	4.27	32.82

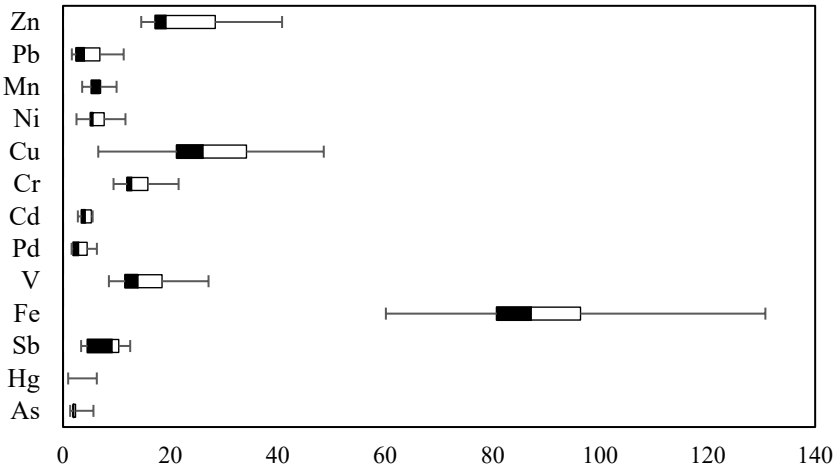


Figure 3. Comparation between industrial sludges and control samples.

In general, compared with the natural soil control sample, most of the PI of metals in the wastewater samples from the industrial park are high. In which, Fe is the metal with the highest PI index, at a high level (PI ≥ 3). Pb has the lowest PI, which is low (PI ≤ 1).

Compared with the control sample of sewage sludge from the groundwater treatment plant, most metals have medium PI, Cd is the metal with the highest PI index, at high level (PI ≥ 3). Fe is the metal with the lowest PI, all PI values < 1, at a low level (PI ≤ 1). This shows that the metal concentration distribution for the ground soil and sludge samples is not the same. For the sludge samples, the metal concentration is distributed more evenly and averagely than in the natural soil environment. Similar to industrial park sludge samples, Fe and Zn are two components that have great influence on the natural soil environment [2, 19, 20].

3.4. Geoaccumulation indexes (Igeo)

There are different indexes generally used to identify metal concentrations of environmental concern like: the metal enrichment factor (EF) and geoaccumulation indexes (Igeo) [8–9]. These indexes identify, numerically, pollution level soils and normally they are calculated on the soil exchangeable fraction because it represents the real bioavailable fraction. The bioavailable metal content in soil exerts a decisive impact on soil quality and it's used in food production. Hence, the assessment of metal contamination is of vital importance in farming areas.

Based on the data in the Table 3, the pollution level based on the Igeo index, it can be seen that: In general, most metals have medium–high Igeo values. In which, Fe has the highest average Igeo value, or in other words, the highest level of Fe contamination. Hg is the metal with the lowest mean Igeo value (all values ≤ 0).

Table 3. Igeo index in industrial park sludge.

Metals / Name	As	Sb	Fe	V	Cd	Cr	Cu	Ni	Mn	Pb	Zn
BT1	–0.17	1.42	5.42	2.51	1.87	2.64	3.73	2.36	1.25	0.13	3.99
BT2	0.37	0.68	12.51	4.58	–0.91	4.19	7.44	2.29	7.64	4.66	8.83
BT3	0.60	1.70	13.18	4.90	–1.03	4.44	7.36	2.83	8.01	5.85	8.78
BT4	0.15	0.26	12.93	5.22	–1.32	4.32	5.02	1.92	8.49	5.35	8.61
BT5	0.42	0.72	12.94	5.37	–1.68	4.00	5.40	2.29	7.98	4.92	8.46
BT6	0.46	1.90	13.63	5.22	–0.91	4.66	7.91	3.47	7.61	6.91	8.63
BT7	0.68	2.09	13.18	4.38	–1.32	4.08	6.80	2.29	7.11	6.32	8.38
BT8	0.37	2.15	12.99	5.47	–1.91	4.00	7.01	2.42	7.98	7.36	8.63
BT9	0.42	1.53	13.04	4.66	–1.91	3.87	7.20	2.74	8.03	7.15	9.58
BT10	1.76	1.77	13.45	5.85	–1.49	4.83	7.68	3.29	8.54	5.85	9.87
BT11	1.92	1.87	13.20	4.76	–0.91	3.96	6.80	1.26	8.03	5.96	9.55

3.5. Heavy metal recovery experiment

In this study, sewage sludge was used as the subject of metal recovery study. Chemical analysis using the ICP–MS method showed that the sludge was heavily contaminated by heavy metals, especially zinc (Zn) with concentrations higher than 1200 mg/kg. Therefore, the study used 3 types of specific chelators to test the ability to recover 3 typical heavy metals in wastewater.

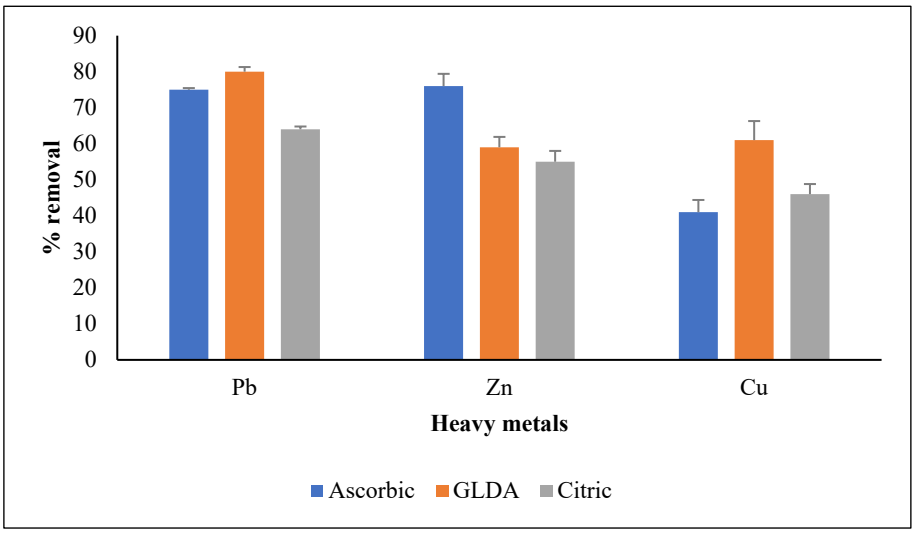


Figure 4. Effect of different extractant concentrations on metal removal.

Overall, among the studied chelators, ascorbic acid was the most effective in extracting Zn from contaminated sewage sludge, followed by GLDA and citric acid (Figure 4). GLDA performed best in extracting Pb and Cu from contaminated sludge although the overall removal efficiency was not really higher than the other 2 chelators for Pb. It can be seen that citric acid has the worst removal efficiency compared with other chelators under the same study conditions.

More importantly, 80.1% Pb, 76,5% Cu, two metals with levels exceeding the national allowable limit were extracted from the sludge using only 200 mM recoverable ascorbic heavy metal. The results of the study open up new application directions for environmentally friendly chemicals used to replace strong acids in cleaning and recovering heavy metals in polluted sludge.

4. Discussion

Waste sludge at Ba Thien 2– Vinh Phuc Industrial Park and industrial sludge collected at Thanh Cong 2 Cement Plant both show very high pollution for some typical heavy metals, especially Cu and Cd. Among the heavy metals, Cadmium has the highest ecological potential risk compared to other elements.

Sludge from the two sampling points above is considered hazardous waste. The results of analysis of pollution index (PI) and ecological risk index (RI) both show that the sludge sample has 2–10 times higher results than the control sample. Industrial sludges have a higher immobilization capacity, with (–COOH) and (–OH) being the typical functional groups present on the sludge surface.

The Pollution Index (PI) in metals and Geoaccumulation Index (Igeo) are indicators studied to calculate the presence and intensity of anthropogenic contaminant deposition on soil or sludges. These indexes of potential contamination are calculated by the normalization of one metal concentration in the research sample respect to the concentration of a reference element. In this study, although the geological accumulation risk indicators are at a moderate level of pollution, the ecological risks of each metal Er and the pollution index of each metal are quite high, potentially causing polluted environment. Therefore, it is necessary to strictly manage as well as take reasonable measures to handle and avoid risks.

The final result shows that most heavy metals in the sludges samples were lower than the maximum allowable concentrations in standard guideline (QCVN 50: 2013) (10 mg/kg; 300 mg/kg; 5000 mg/kg, respectively). But specifically, lead, zinc and copper in some sludge

samples exceeding the national allowable limit so these metals were chosen for removal study. Lead and copper were extracted from the sludge using only 200 mM recoverable ascorbic, GLDA and citric solution that worked effectively. The results of the study open up new application directions for environmentally friendly chemicals used to replace strong acids in cleaning and recovering heavy metals in contaminated industrial sludges.

5. Conclusions

It is essential to determine the heavy metal concentrations in the industrial sludges to select appropriate disposal methods. We conducted a survey of heavy metal concentrations of sludge samples from 11 industrial sludge samples from an industrial park located in Vinh Phuc Province. The average concentrations of Cd, Cu, Pb, Zn in the sludge samples of the industrial park had the average concentrations of the elements 0.9 mg/kg, 297.2 mg/kg, 164.6 mg/kg and 1177.4 mg/kg, respectively. This study also characterizes the physico–chemical characteristic of industrial sludge. Environmental issues related to possible management options are also addressed. Sludge samples from industrial parks were analysed and calculated the PI and Igeo index in comparison with the natural soils. The results indicate that pollution indexes of Cu, Pb, and Zn could be 2–10 times higher than the control natural soil and it may pose a potential threat to the water quality for sludge dumped near water bodies. Therefore, we recommend avoiding uncontrolled upland disposal of such sludge. Some preliminary treatment to remove heavy metals from the sludges should be applied. A recovery of Cu and Pb using GLDA solution from this sludge could be considered with high efficiency.

Supplementary Materials: None

Author Contributions: Conceptualization, N.T.C., L.T.M.L.; methodology, N.X.B.; software, P.M.C.; validation, N.T.C., L.T.M.L., N.X.B.; formal analysis, P.M.C., N.X.B.; investigation, L.T.M.L.; data curation, N.T.C.; writing–original draft preparation, N.T.C., P.M.C.; visualization, L.T.M.L.; project administration, N.T.C.; All authors have read and agreed to the published version of the manuscript".

Funding: This research was funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 105.08–2019.15 with title: "Assessment of toxic chemicals in industrial sludge, their potential impact on ecosystem and an eco–friendly remediation method for those materials in Vietnam".

References

1. Islam, M.S.; Ahmed, M.K.; Raknuzzaman, M.; Habibullah–Al–Mamun, M.; Kundu, G.K. Heavy metals in the industrial sludge and their ecological risk: A case study for a developing country. *J. Geochem. Explor.* **2017**, *172*, 41–49.
2. Dung, T.T.T.; Cappuyns, V.; Vassilieva, E.; Golreihan, A.; Phung, N.K.; Swennen, R. Release of potentially toxic elements from industrial sludge: Implications for land disposal. *Clean – Soil Air Water* **2015**, *43*, 1327–1337.
3. Udayyanga, W.D.C.; Veksha, A.; Giannis, A.; Liang, Y.N.; Lisak, G.; Hu, X.; Lim, T.T. Insights into the speciation of heavy metals during pyrolysis of industrial sludge. *Sci.Tot. Env.* **2019**, *691*, 232–242.
4. Agoro, M.A.; Adeniji, A.O.; Adefisoye, M.A.; Okoh, O.O. Heavy Metals in Wastewater and Sewage Sludge from Selected Municipal Treatment Plants in Eastern Cape Province, South Africa, P2–3. *Water* **2020**, *12(10)*, 2746.
5. Van Thinh, N.; Osanai, Y.; Adachi, T.; Vuong, B.T.S.; Kitano, I.; Chung, N.T.; Thai, P.K. Removal of lead and other toxic metals in heavily contaminated soil using biodegradable chelators: GLDA, citric acid and ascorbic acid. *Chemosphere* **2021** *263*, 127912.

6. McLaughlin, M.J.; Hamon, R.E.; McLaren, R.G.; Speir, T.W.; Rogers, S.L. A bioavailability–based rationale for controlling metal and metalloid contamination of agricultural land in Australia and New Zealand. *Soil Res.* **2000**, *38(6)*, 1037–1086.
7. Tytła, M. Assessment of heavy metal pollution and potential ecological risk in sewage sludge from municipal wastewater treatment plant located in the most industrialized region in Poland – Case study. *Int. J. Environ. Res. Public Health.* **2019**, *16(13)*, 2430.
8. Zhang, X.; Wang, X.Q.; Wang, D.F. Immobilization of Heavy Metals in Sewage Sludge during Land Application Process in China: A Review. *Sustainability* **2020** *9(11)*, 2020.
9. Thai, N.T.K. Hazardous industrial waste management in Vietnam: current status and future direction. *J. Mater. Cycles Waste Manag.* **2009**, *11*, 258–262.
10. Strady, E.; Dinh, Q.T.; Némery, J.; Nguyen, T.N.; Guédron, S.; Nguyen, N.S.; Denis, H.; Phuoc Dan Nguyen, P.D. Spatial variation and risk assessment of trace metals in water and sediment of the Mekong Delta. *Chemosphere* **2017**, *179*, 367–378.
11. Nguyen, T.C.; Loganathan, P.; Nguyen, T.V.; Pham, T.T.N.; Kandasamy, J.; Wu, M.; Naidu, R.; Vigneswaran, S. Trace elements in road–deposited and waterbed sediments in Kogarah Bay, Sydney: enrichment, sources and fractionation. *Soil Res.* **2015**, *53(4)*, 401–411.
12. Hung, C.V.; Cam, B.D.; Mai, P.T.N.; Dzung, B.Q. Heavy metals and polycyclic aromatic hydrocarbons in municipal sewage sludge from a river in highly urbanized metropolitan area in Hanoi, Vietnam: levels, accumulation pattern and assessment of land application. *Env. Geochem. Health.* **2015**, *37*, 133–146.
13. Viet, N.T.; Dieu, T.T.M.; Loan, N.T.P. Current status of sludge collection, transportation and treatment in Ho Chi Minh city. *J. Env. Protect.* **2013**, *4(12)*, 1329-1335.
14. QCVN 50:2013/BTNMT – National technical regulation on hazardous thresholds for sludge from water treatment process.
15. QCVN 40:2011/BTNMT – National technical regulation on industrial wastewater.
16. APHA, AWWA, WEF – Standard Methods for the Examination of Water and Wastewater. 21st edition, Washington DC., USA. 2005, 1070–1072.
17. EPA Method 3050B (SW–846) Acid Digestion of Sediments, Sludges, and Soils. 2019.
18. Barbieri, M.; Nigro, A.; Sappa, G. Soil Contamination evaluation by enrichment factor (EF) and geoaccumulation index (Igeo). *Senses. Sci.* **2015**, *2(3)*, 94–97.
19. Dung, T.T.T.; Golreihan, A.; Vassilieva, E. Insights into solid phase characteristics and release of heavy metals and arsenic from industrial sludge via combined chemical, mineralogical, and microanalysis. *Env. Sci. Pollut. Res.* **2015**, *22*, 2205–2218.
20. Vo, T.H.; Le, T.K.C.; Bui, X.T.; Nguyen, P.D. Assessment of decomposition and leaching ability of heavy metals in composting and aerobic sludge digestion. AUN/SEED – Net 2nd Regional Conference on Global Environment. Proceeding: "Global Environmental Issues for Sustainable Development in the ASEAN Region", Kyoto, Japan, 2010.

VIETNAM JOURNAL OF HYDROMETEOROLOGY

Research Article

Study to assess the impact of saltwater intrusion in Ho Chi Minh City under climate change conditions

Nguyen Van Hong [1*], Nguyen Phuong Dong[1]

[1] Sub Institute of HydroMeteorology and Climate Change; nguyenvanhong79@gmail.com; nguyenphuongdongkttv@gmail.com

*Corresponding author: nguyenvanhong79@gmail.com; Tel.: +84–913613206
Received: 08 December 2021; Accepted: 03 March 2022; Published: 25 March 2022

Abstract: The study on assessing the impacts of saline intrusion on water resources, from building a hydraulic model MIKE 11 including hydraulic model, rain model, and advection–dispersion module (HD + RR + AD) together with the calculation scenarios in consideration of climate change through the high scenarios RCP 8.5 and low scenarios RCP 4.5 in 2025, 2030 and 2050 are highlighted in the paper. In addition, the impacts of salinity, salinity margins on the constructions, and water resources through the cases with or without the saltwater prevention works are considered, thereby assessing the impact of salinity intrusion through scenarios and structural solutions for water resources in HCMC.

Keywords: Saline intrusion; Saline boundary; Saigon–Dong Nai River system; Climate change; Sea level rise; MIKE modeling.

1. Introduction

At present, the saline intrusion is a serious problem for many local governments in coastal areas, in the context of increasingly complex and unpredictable climate change, possibly resulting in the increasing risk of saline intrusion, impacting water resources [1–3]. The saline intrusion from the seas to the rivers is very widespread for the coastal plains, especially in the dry season [4–6]. The extent of saline intrusion depends on many natural factors such as topographical characteristics, flow regime from upstream, the tidal regime of estuary, changes in rainfall, temperature and evaporation, sea–level rise trend, or human activities such as groundwater extraction, change of land use, destruction of mangroves [7–11].

Ho Chi Minh City is surrounded by a very developed network of a total of 3,020 rivers and canals in a total length of 5,075km, including three main river systems, Dong Nai River, Saigon River, and Vam Co River [12]. With relatively flat terrain, the hydrological and hydraulic regimes are not only strongly influenced by the East Sea tide, but also clearly affected by the exploitation of terraces of reservoirs at the upstream at present and in the future (such as the reservoirs of Tri An, Dau Tieng, Thac Mo...) [13]. Despite several abundant water resources thanks to a dense network of rivers and streams, in recent years, water shortages are more and more seriously due to colonization and saline boundary (SB) moving deeply into the field as consequences of climate change. Accordingly, fresh water supply to the city is more and more challenging [14–16]. To deal with the above matter, the study of the saline intrusion on the main rivers of Ho Chi Minh City should be implemented in parallel with the assessment based on the climate change and sea–level rise scenarios in the future context for proper management and policies to ensure the sustainable socio–economic development. Currently, the modeling method is widely applied in simulating saltwater intrusion in many river systems [17–18]. Some case studies have used the HD, RR,

and AD modules of the MIKE 11 model, all developed and supported by the Danish Hydraulic Institute (DHI) to model the effects of river flows and saltwater intrusion [8, 19]. After proper calibration and verification, the HD, RR, and AD modules of the MIKE 11 have been used to estimate the salt profile. Scenarios for salinity intrusion simulation were developed based on the Climate change and sea–level rise scenarios for Vietnam by MONRE [20]. Results of salinity intrusion computation for RCP4.5 and RCP8.5 scenarios up to 2050 are shown for some cross–sections in the main Sai Gon–Dong Nai River systems.

2. Materials and Methods

2.1. Description of the study site

Ho Chi Minh City is the largest metropolis in Vietnam and the main center for most economic activities, not only for the South region but for the whole country in general. It is also a hub for educational, scientific, cultural, and technological activities. The city currently has 24 administrative units, including 19 districts and 5 districts. Ho Chi Minh City has a complex natural drainage system, including many sewer lines discharging water into rivers and canals connecting each other. The main river system is the Dong Nai–Saigon River system in the east and the Vam Co River system in the west.

2.2. Calculation of rainfall-runoff model (MIKE NAM)

- Using Hydrometeorological data at measuring stations in the study area and surrounding areas, collected from the Southern Regional Hydrometeorological Station from January 1, 2017, to December 1, 2017:

+ Daily rainfall data at five stations on the Saigon–Dong Nai river system: Ta Lai, Tri An, Tan Son Hoa, Dong Ban, and Loc Ninh.

+ Evaporation data estimated by Blaney–Crridle method based on temperature data of the Tan Son Hoa station.

+ Average daily flow data to the Tri An and Dau Tieng Reservoirs.

- According to the river system, the river basins in the study area are divided into basins and sub–basins. The basins are demarcated based on the database of the digital elevation model (DEM) (30 m × 30 m). The sub–basins area is considered the basis for the calculation of mean precipitation by the Thiessen method. The precipitation means for each sub–basin corresponding to the rainfall stations will be used as input for the NAM model (Figure 1).

2.3. MIKE 11 model

+ Cross–section data: the section is defined in compliance with the national standard elevation on the topographic map of 1:10,000 scale as a basis for determining the location of the river section on the hydraulic diagram (Cross-sectional data are inherited from the topic of HCMC climate change response plan in 2020).

+ The network consists of 79 large and small tributaries, 674 cross–sections, 68 points of entry and exit. The maximum distance dx on the tributaries is 500–1000 m, and the smallest is 100–200 m.

+ The upstream boundary data are calculated from the NAM model with 5 stations, specifically the water level at Go Dau and Tan An corresponding to Tay Ninh and Long An basins, and discharge flows of Phuoc Hoa, Tri An, and The The Dau Tieng Reservoirs.

+ The downstream boundary data are the water level at 4 stations (Soai Rap, Dinh Ba, Long Tau, Thi Vai) which is correlated from the water level data at Vam Kenh and Vung Tau stations (Figure 2).

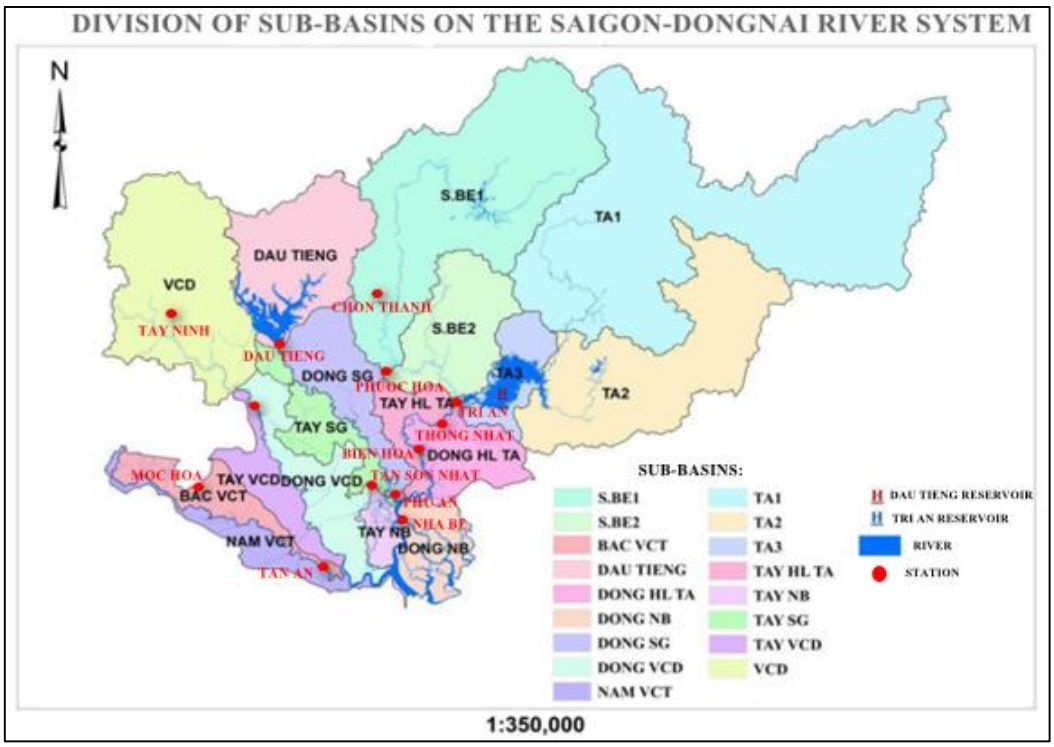


Figure 1. Division of sub–basins on the Saigon–Dong Nai River system.

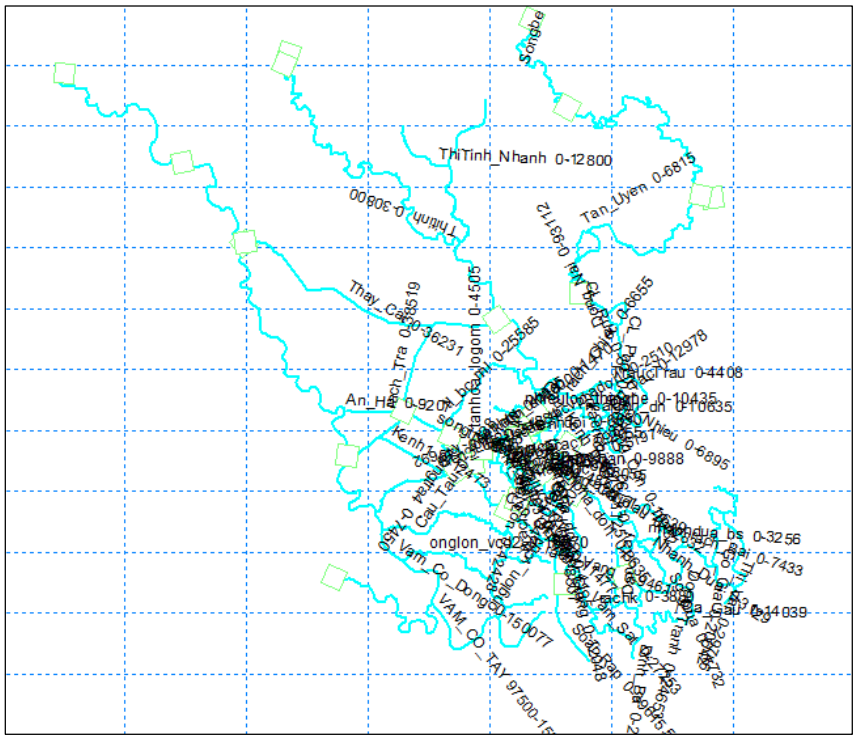


Figure 2. Hydraulic diagram of Saigon–Dong Nai river basin.

3. Results and Discussion

3.1. Results of calibration and validation of NAM model for rainfall–runoff

The calibration and verification are based on the flows to the Tri An Reservoir, the Dau Tieng Reservoir, and Be River "The Q data about the lake was collected at the hydro-meteorological station of Nam Bo province. On the Be river, Q data is taken from Phuoc

Long hydrological station to correct upstream. And Q returned to Phuoc Hoa lake to test the downstream side of the river", with the error between the calculated and measured values being evaluated by the coefficient of determination R^2 and NSE. The results of model calibration and verification are presented in detail in Table 1. In addition, the basin of the Dau Tieng Reservoir is also calibrated and tested further for three more stages with the parameters of the model taken in the year of further calibration in 2016.

The results of the correlation analysis between the measured and calculated flows are pretty good for the basins of Dau Tieng and Be rivers (NSE and R^2 are greater than 0.65) and good for the Tri An Reservoir ($R^2 > 0.8$). This result shows that the discharge process curve calculated from rain by NAM model is quite consistent with the measured flow process curve. The results of calibration and verification are presented representatively at the Tri An Reservoir as in Figure 3.

In general, the above results are consistent with the hydrological conditions of the study area. The set of parameters achieves high reliability, which is qualified to simulate the flow from rainfall, corresponding climate change scenarios. The runoff generated from the rain and the discharge of wastewater generated from economic activities is used as the inflow of rivers.

Table 1. Results of calibration and verification of NAM model.

Basin	Coefficient	Calibration		Validation				
		2016	2017	2011–2015	2007–2010	2000	2010	2018
The The Dau Tieng Reservoir	NSE	0.752		0.657	0.7	0.669		
	R^2		0.65					
Be River	NSE	0.867					0.717	
Tri An Reservoir	R^2	0.82						0.80

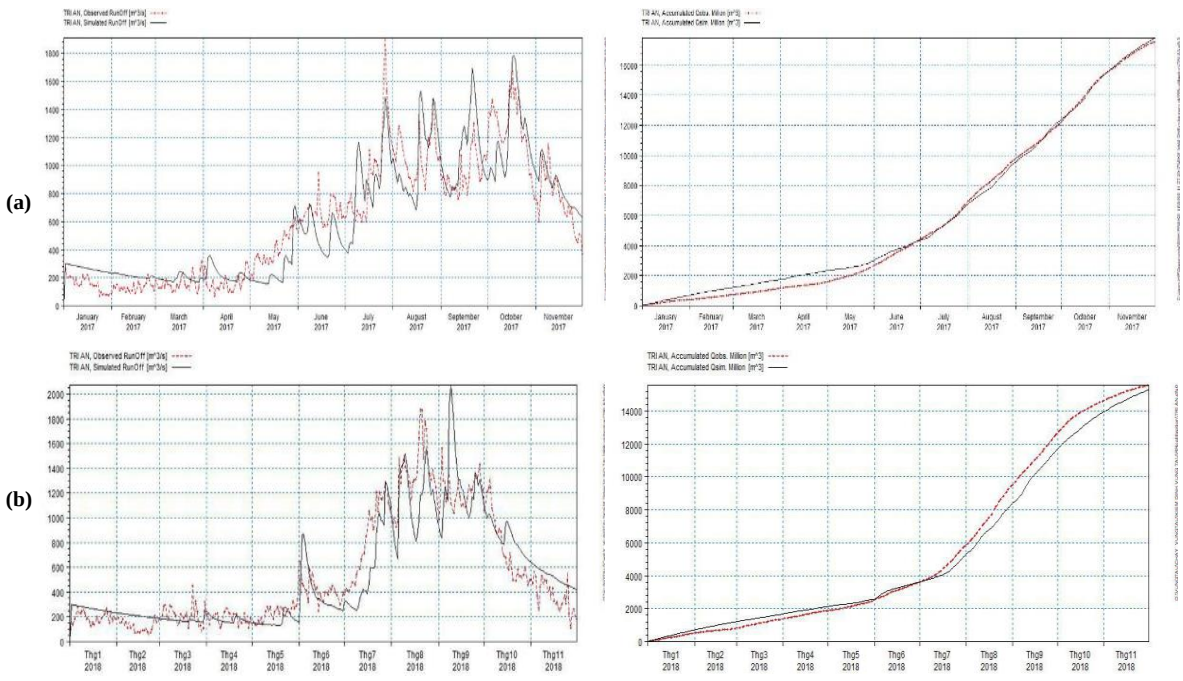


Figure 3. Results of calibration (a) and verification (b) of MIKE NAM model at Tri An reservoir.

Because the article inherits the results from the climate assessment project of HCMC. Therefore, data collection is quite difficult, the authors can only test the dry and rainy seasons in 1 year.

3.2. Calibration and validation results of the MIKE 11 model

+ Time to calibrate the model is taken from 0:00 on April 1, 2017 to 23:00 on April 30, 2017; then data for three days from 9:00 on April 26, 2017 to 8:00 p.m. on April 28, 2017, was extracted for hydraulic testing for the study area; time step $\Delta t = 1$ minute.

+ Initial conditions: At the initial time of taking the static water level, the flow $Q = 0$ m^3/s.

+ Stations for saline calibration and verification: Cat Lai, Nha Be, Phu An, Phu Cuong, Vam Sat, Vam Co, Hoa An, Binh Phuoc, from 9:00 April 26, 2017, to 20:00 April 28, 2017.

+ The time to calibrate the hydraulic model is from April 1-25, 2017. The time to test the hydraulic model is from April 26-28, 2017.

+ The time to test the salinity model is from 9:00 on April 26, 2017 to 20:00 on April 28, 2017.

3.3. Simulation results of saline intrusion according to the scenarios

3.3.1. Saline intrusion in Ho Chi Minh City in the current period

The saline intrusion has been defined by the boundaries of the salinity thresholds (SB) corresponding to the selection criteria for the salinity limit based on the influence on the demand for surface water of water treatment plants in the Saigon–Dong Nai basin, as presented explicitly in Table 2. To analyze the spread of salinity in the river, the displacement of the saline boundary over each period will be considered and evaluated based on the inherited mapping (Figure 5).

Table 2. The boundaries of the salinity thresholds affect the demand for surface water.

No.	Limits of saline zones correspond to salinity thresholds	Criteria for selecting the salinity limit based on the influence on the demand for surface water usage
1	< 0.25‰ SB1 (0.25‰)	Qualified for domestic water supply (after normal treatment)
2	0.25–0.5‰ SB2 (0.5‰)	Qualified for domestic water supply (after normal treatment), conservation of aquatic plants, and other purposes
3	0.5–1‰ SB3 (1‰)	Qualified for irrigation purposes or other uses with equivalent water quality requirements
4	1–2‰ SB4 (2‰)	- Good for brackish water aquaculture - Yield reduction of salt–sensitive crops
5	2–4‰ SB5 (4‰)	- Good for brackish water aquaculture - Yield reduction of different crops
6	4–8‰ SB6 (8‰)	- Cultivating some types of brackish–water aquatic products - Yield reduction of salt–sensitive crops
7	8–18‰ SB7 (18‰)	- Cultivating some types of brackish water aquatic products - Impossible for irrigation.
8	> 18‰	Saline intrusion, unusable.

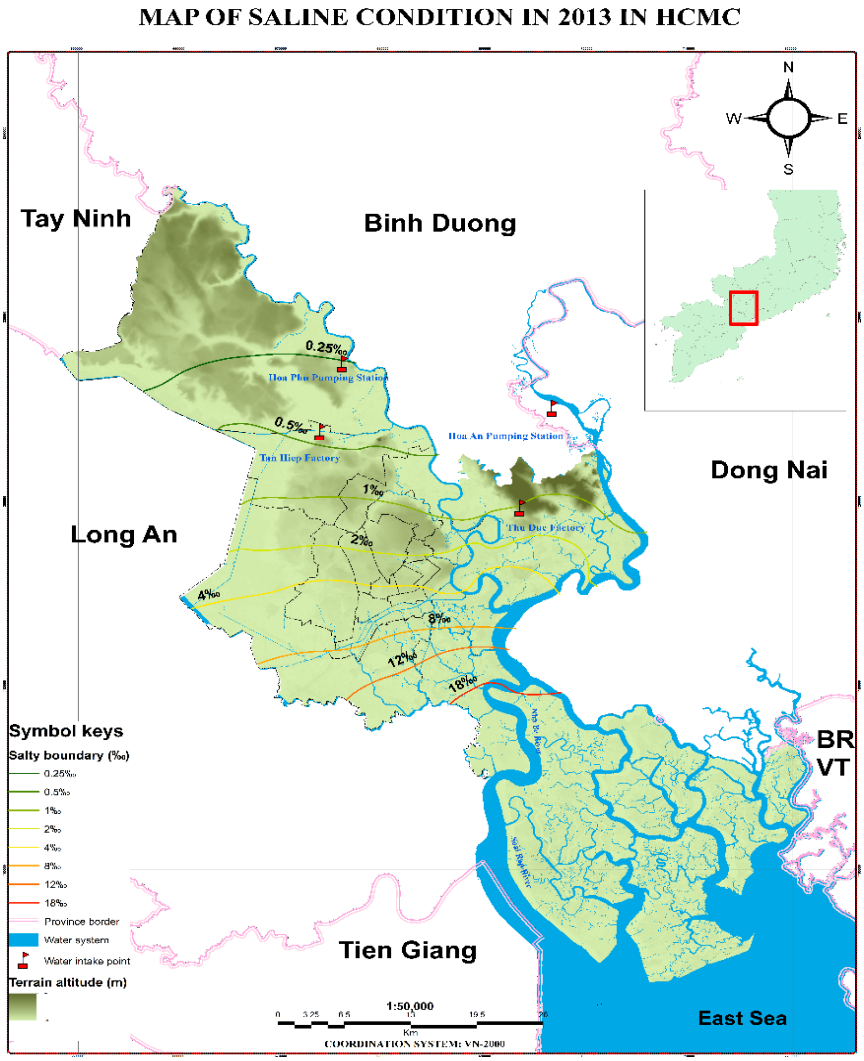


Figure 4. Map of saline condition in 2013 in HCMC.

Using 2013 as the current condition is because the article inherits some results from the project to update the action plan for Ho Chi Minh City in 2020.

a) Analysis of saline boundary on the Saigon River in the current condition

- SB1 < 0.25‰: farthest to the upstream of the Dau Tieng Reservoir, passing Hoa Phu pumping station, but the distance between this SB and the pumping station is negligible from 0.5–1 km.

- SB2 0.25–0.5‰: available at Tan Hiep water plant, from SB2 position, about 10 km from Hoa Phu pumping station to the downstream of the Saigon River.

- SB3 0.5–1‰: about 18 km from Hoa Phu pumping station and Binh Phuoc hydrological station, 5 km toward the downstream of the Dau Tieng Reservoir.

- SB4 1–2‰: SB4 is 5 km from SB5 along the length of the Saigon River from upstream to downstream.

- SB5 2–4‰: SB5 is 5 km from SB4 along the length of the Saigon River from upstream to downstream and 28 km from Hoa Phu pumping station and Tan Hiep water plant, and 18 km toward the upstream of the reservoir; This SB is located near Phu An hydrological station and about 1km from Thu Thiem salinity measurement station toward the upstream of the Dau Tieng Reservoir.

- SB6 4–8‰: SB6 is about 1 km from Cat Lai salinity measurement station toward the upstream.

- SB7 8–18‰: available at the intersection between Saigon River and Dong Nai River.

b) Analysis of saline boundary on the Dong Nai River in the current condition

- SB1 < 0.25‰: toward the upstream of Tri An Reservoir, SB1 goes further about 3 km from Hoa An pumping station.
- SB2 0.25–0.5‰: SB2 is located between Hoa An pumping station and Thu Duc water plant.
- SB3 0.5–1‰: available at the water source area of Thu Duc plant, about 10 km from Hoa An pumping station and about 0.5 km from Long Dai salinity measurement station toward the upstream of Tri An reservoir.
- SB4 1–2‰: SB4 is available at the location of the Cat Lai hydrological station.
- SB5 2–4‰: about 1–1.5 km from SB4.
- SB6, SB7 are evenly distributed along the river.

In general, from the distribution map of SB on the Saigon–Dong Nai river system in the HCMC area: SB1 and SB2 are evaluated to determine the area serving the domestic water supply after only the normal treatment process, conservation of aquatic plants, and activities for other purposes with similar water quality factors with salinity > 0.5‰. SB3 is assessed to determine the area for water use for irrigation or other purposes consistent with the salinity of this saline boundary. Similarly, SB4 and SB5 are suitable in conditions that can be treated by conventional systems and satisfied for most of the water demand for residential areas from SB to the upstream of the reservoir. Particularly for SB6, SB7 with high salinity < 18‰, it is impossible for water supply purposes or crops due to yield reduction, including salt–tolerant crops, but they can mainly be used for some good salt–tolerant aquatic species.

3.3.2. Impact of saline intrusion on water resources in the context of climate change in Ho Chi Minh City in case of no salinity prevention works

Under the impacts of climate change factors such as temperature, heat, rainfall, and sea–level rise, the hydraulic regime in the river is changed, and the saline intrusion is deeper toward the inner field, causing changes compared to the current status. The salinity change and salinity spreading tend to increase gradually in the future, toward the upstream of the Dau Tieng Reservoir in Saigon River and the Tri An Reservoir in Dong Nai River. For each RCP scenario, the movement of saline boundaries in each period (year) is different, corresponding to sea–level rise.

According to the sea level rise RCP4.5 and RCP8.5 with salinity risk atlas, it is shown that there are no significant differences between 2025 and 2030; it seems that salinity is not much intruded into the upstream. Particularly for the RCP scenario for the year 2050, the data about the sea level rise increases significantly, resulting in apparent changes in the risks of saline intrusion and salinity as well as the negative impacts on the domestic water supply of water treatment plants and regulating irrigation of canals and channels in the city. As a result, the study will analyze and assess the salinity changes in 2025 and 2030 in general and analyze separately and, more specifically, the salinity changes in 2050. The evaluation of the process of saline intrusion is based on the current situation and (in the context of) climate change in the Saigon and Dong Nai rivers from Table 3 to Table 6.

Table 3. Risk assessment of the saline intrusion according to the RCP4.5 scenario compared with the current situation in the absence of structures on the Saigon River.

Saline boundary	2025	2030	2050
Saline boundary 0.25‰ (SB1)	SB1 moves further about 0.25 km from Hoa Phu pumping station toward upstream of the		SB1 moves further the Hoa Phu pumping station about 3.5 km compared to the 2025 and 2030 scenarios.

Saline boundary	2025	2030	2050
	Dau Tieng Reservoir compared to the current situation.		
Saline boundary 0,5‰ (SB2)	SB1 moves further into the field, about 3–5 km from Hoa Phu pumping station toward upstream of the Dau Tieng Reservoir compared to the current situation.		From the position from Hoa Phu pumping station to the upstream of Dau Tieng lake, it was moving about 1.5 km compared to the current situation and 5 km compared to the period of 2025–2030.
Saline boundary 1‰ (SB3)	SB3 is about 24–25 km from Hoa Phu pumping station toward downstream.		Compared to 2013, SB3 continues to move upstream, about 20 km from Hoa Phu pumping station toward downstream.
Saline boundary 2‰ (SB4)	It is 35 km from Hoa Phu pumping station toward downstream to Dau Tieng lake.		Compared to 2013, it is 30km from Hoa Phu pumping station, moving about 5km more.
Saline boundary 4‰ (SB5)	It is 46 km from Hoa Phu pumping station toward downstream to Dau Tieng lake.		It is 45 km from Hoa Phu pumping station toward downstream to Dau Tieng lake.
Saline boundary 8‰ (SB6)	It is 60 km from Hoa Phu pumping station toward downstream to Dau Tieng lake.		It is 56 km from Hoa Phu pumping station toward downstream to Dau Tieng lake.
Saline boundary 18‰ (SB7)	It is 70 km from Hoa Phu pumping station toward downstream to Dau Tieng lake.		It is 67 km from Hoa Phu pumping station toward downstream to Dau Tieng lake.

Table 4. Risk assessment of the saline intrusion according to the RCP8.5 scenario compared with the current situation in the absence of structures on the Saigon River.

Saline boundary	2025	2030	2050
Saline boundary 0,25‰ (SB1)	SB1 moves further about 0.25 km from Hoa Phu pumping station toward upstream of the Dau Tieng Reservoir compared to the current situation.		It is 5 km from Hoa Phu pumping station upstream to Dau Tieng lake, increasing 4 km more than in 2013.
Saline boundary 0,5‰ (SB2)	SB1 moves further into the field, about 3–5 km from Hoa Phu pumping station toward upstream of the Dau Tieng Reservoir compared to the current situation.		It is 9.1 km from Hoa Phu pumping station toward Dau Tieng lake downstream, increasing 7 km more than in 2013.
Saline boundary 1‰ (SB3)	SB3 is about 24–25 km from Hoa Phu pumping station toward downstream		It is 19.7 km from Hoa Phu pumping station toward downstream to Dau Tieng lake.
Saline boundary 2‰ (SB4)	It is 35 km from Hoa Phu pumping station toward downstream to Dau Tieng lake.		It is 30 km from Hoa Phu pumping station toward downstream to Dau Tieng lake.
Saline boundary 4‰ (SB5)	It is 46 km from Hoa Phu pumping station toward downstream to Dau Tieng lake.		It is 45 km from Hoa Phu pumping station toward downstream to Dau Tieng lake.
Saline boundary 8‰ (SB6)	It is 60 km from Hoa Phu pumping station toward downstream to Dau Tieng lake.		It is 56 km from Hoa Phu pumping station toward downstream to Dau Tieng lake.
Saline boundary 18‰ (SB7)	It is 70 km from Hoa Phu pumping station toward downstream to Dau Tieng lake.		It is 66 km from Hoa Phu pumping station toward downstream to Dau Tieng lake.

Table 5. Risk assessment of the saline intrusion according to the RCP4.5 scenario compared with the current situation in the absence of structures on the Dong Nai River.

Saline boundary	2025	2030	2050
Saline boundary 0,25‰ (SB1)	SB1 moves further about 3–5 km from Hoa Phu pumping station toward upstream of the Tri An Reservoir compared to the current situation.		Moving further upstream from Hoa An Pumping Station, about 8 km toward upstream, compared to 2013.
Saline boundary 0,5‰ (SB2)	SB1 moves further into the field, about 6–7 km from Hoa An pumping station toward upstream of the Tri An Reservoir compared to the current situation, increasing about 2–3 km.		It is 4 km from Hoa An pumping station toward downstream to Tri An lake, increasing about 5 km.
Saline boundary 1‰ (SB3)	It is 17.5 km from Hoa An pumping station toward downstream to Tri An lake, increasing about 1.5 km.		It is 14.5 km from Hoa An pumping station toward downstream to Tri An lake, increasing about 5 km.
Saline boundary 2‰ (SB4)	It is 30 km from Hoa An pumping station toward downstream to Tri An lake.		It is 27 km from Hoa An pumping station downstream to Tri An lake.
Saline boundary 4‰ (SB5)	It is 34 km from Hoa An pumping station toward the downstream to Tri An lake.		It is 32 km from Hoa An pumping station toward downstream to Tri An lake.
Saline boundary 8‰ (SB6)	It is 40 km from Hoa An pumping station toward downstream to Tri An lake.		It is 37 km from Hoa An pumping station toward the downstream to Tri An lake.
Saline boundary 18‰ (SB7)	It is 52 km from Hoa An pumping station toward the downstream to Tri An lake.		It is 49 km from Hoa An pumping station toward downstream to Tri An lake.

Table 6. Risk assessment of the saline intrusion according to the RCP8.5 scenario compared with the current situation in the absence of structures on the Dong Nai River.

Saline boundary	2025	2030	2050
Saline boundary 0,25‰ (SB1)	SB1 moves further about 3–5 km from Hoa Phu pumping station toward upstream of the Tri An Reservoir compared to the current situation.		SB1 is 8km from the pumping station upstream to Tri An lake.
Saline boundary 0,5‰ (SB2)	SB1 moves further into the field, about 6–7 km from Hoa An pumping station toward upstream of the Tri An Reservoir compared to the current situation, increasing about 2–3 km.		It is 3 km from Hoa An pumping station toward downstream to Tri An lake, increasing about 6 km.
Saline boundary 1‰ (SB3)	It is 17.5 km from Hoa An pumping station toward downstream to Tri An lake, increasing about 1.5 km.		It is 14 km from Hoa An pumping station toward downstream to Tri An lake, increasing about 5 km.
Saline boundary 2‰ (SB4)	It is 30 km from Hoa An pumping station toward downstream to Tri An lake.		It is 27 km from Hoa An pumping station downstream to Tri An lake.
Saline boundary 4‰ (SB5)	It is 34 km from Hoa An pumping station toward downstream to the Tri An Reservoir.		It is 31.5 km from Hoa An pumping station toward downstream to Tri An lake.
Saline boundary 8‰ (SB6)	It is 40 km from Hoa An pumping station toward downstream to Tri An lake.		It is 37 km from Hoa An pumping station toward the downstream to Tri An lake.
Saline boundary 18‰ (SB7)	It is 52 km from Hoa An pumping station toward the downstream to Tri An lake.		It is 48.5 km from Hoa An pumping station toward downstream to Tri An lake.

3.3.3. Impact of saline intrusion on water resources in the context of climate change in Ho Chi Minh City in case of available salinity prevention works

The Saline control gate are mainly located at the intersections of rivers entering the field. If the works are completed and put into operation by 2019, the effectiveness of these gates in salinity prevention is shown in Figures 6a–6b.

When there are no structures to prevent salinity in normal conditions, salinity will increasingly encroach into the upstream area. The saline intrusion will be spread far away in the future, narrowing the safety level of freshwater sources, affecting the daily production and daily life of the local people. Thanks to the operation of 6 Saline control gate in the future, the salinity will be significantly reduced on small river tributaries going deep into the field, such as Ben Luc branch, Doi cannel – Te cannel, Phu Xuan canal (District 7), Cay Kho canal (Nha Be). However, the effectiveness of salinity prevention on the two main rivers (Saigon River and Dong Nai River) is negligible.

In 2025, salinity at a point located 1.3 km from Phu An station, nearly 1 km from Ben Nghe sluice, along the Ben Luc branch (in District 1) will decrease by nearly 36% when the works are in operation. Similarly, the salinity concentration at a point on the Doi cannel branch located 1.4 km from Tan Thuan sluice along the Doi cannel branch will decrease by nearly 29%, showing the difference in salinity concentration in the 2 cases. On the Saigon River, the salinity does not change significantly during the salinity observation at any point, about 100 m from Phu Cuong station upstream and 500 m from Hoa Phu water pumping station downstream (Figure 6c). The salinity in the case of available structures increases by 0.002% compared to the case without structures. Therefore, the Saline control gate can only reduce salinity into the field, reduce salinity in water for production and irrigation, but they are not effective for the two main rivers, Saigon River and Dong Nai River. In short, the effect of the works on the deep saline intrusion upstream of the main river can be negligible.

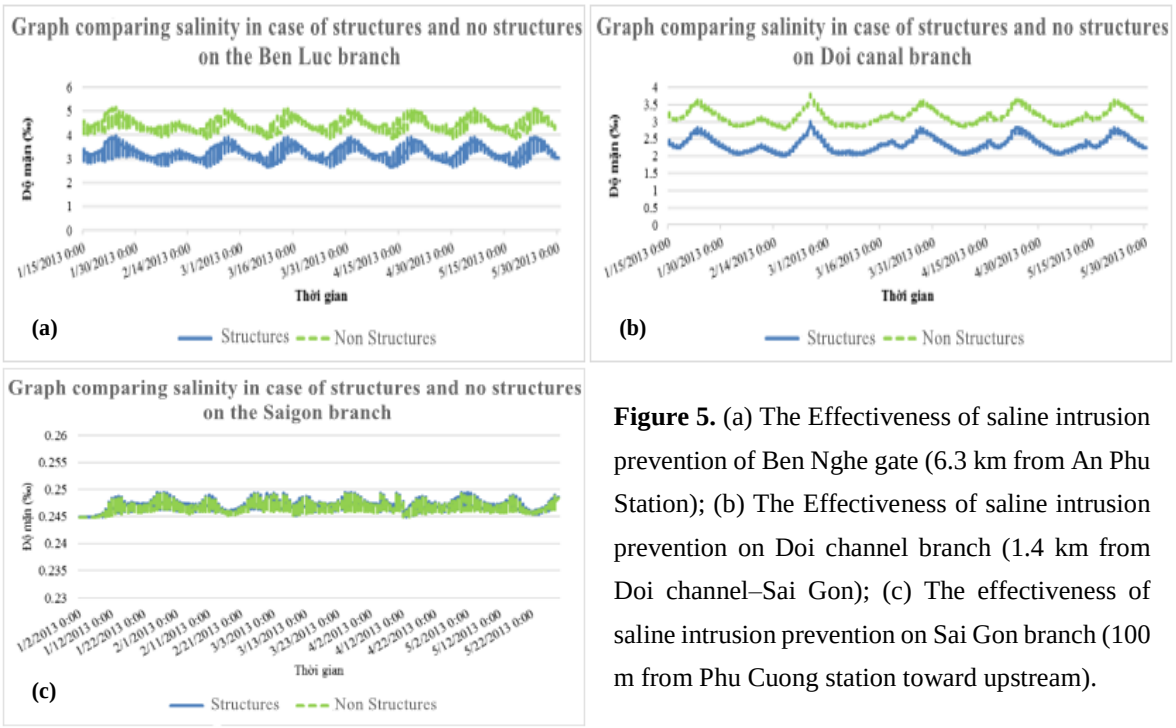


Figure 5. (a) The Effectiveness of saline intrusion prevention of Ben Nghe gate (6.3 km from An Phu Station); (b) The Effectiveness of saline intrusion prevention on Doi channel branch (1.4 km from Doi channel–Sai Gon); (c) The effectiveness of saline intrusion prevention on Sai Gon branch (100 m from Phu Cuong station toward upstream).

The saline intrusion toward inland in the Ho Chi Minh City in the RCP4.5 scenario 2025 in the case of operational Saline control gate is presented in below Figure 7.

Figure 6. Map of saline intrusion in Ho Chi Minh City under the 2025–RCP4.5 scenario.

4. Conclusion

As Saigon River and the Dong Nai River are the two main rivers serving the water supply of the city, the saline boundaries impacting river water quality are analyzed corresponding to the location of the Hoa Phu raw water pumping station (Cu Chi) on the Sai Gon River and Hoa An station (Dong Nai) on the Dong Nai River.

According to the current scenario, SB1 0.25‰ on the Saigon River moves the farthest to the upstream of the Dau Tieng Reservoir, going through Hoa Phu pumping station with a distance of 0.5–1 km. On the Dong Nai River, SB1 moves toward the upstream of the Tri An Reservoir, about 3 km from Hoa An pumping station. The differential in salinity between the two scenarios RCP 4.5 and RCP 8.5 in the absence of salinity compartments, on the Saigon River branch, respectively in 2025–2030 and 2050, according to RCP4.5, the salty border is 0.25 ‰ away Hoa Phu station is 0.25 km in turn; 3.5 km compared to 0.25 km; 3.5 km according to RCP8.5. On Dong Nai river branch, compared with station Hoa An, the

corresponding figures are 3–5 km; 8.0 km under RCP4.5 and 3–5 km; 8.0 km according to RCP8.5.

When the Saline control gate are put into operation, the saline intrusion shall decrease significantly on small tributaries such as Ben Luc, Doi channel–Te channel, Phu Xuan canal (District 7), Cay Kho canal (Nha Be) but these gates do not have many effects on the two main rivers, Saigon River and Dong Nai River. However, the salinity of the Saigon River toward upstream will increase by 0.002% compared to the case without the structure.

Therefore, the proper usage of the surface water should be paid attention to distribute the surface water and rain between the two seasons reasonably to serve the daily life, economic activities, and operation of water treatment plants in the future with the impacts of climate change.

However, the study of factors influencing saline intrusion in Ho Chi Minh city remains limited. Especially the regulation of neighboring reservoirs and the Tri An system from the upstream and the limitted of saline monitoring data. This factor necessitates a long-term research complex because of the climate change of the current period and is influenced by reservoir regulation upstream.

Author contributions: N.V.H., N.P.D. discussed the original idea of the draft. N.V.H. analyzed and designed the input data. N.V.H. established the input data and performed the model simulation. N.V.H., N.P.D. wrote and edited the manuscript. N.V.H. analyzed the output date. All authors were a reviewer and submitted the final version of the manuscript.

Conflict of interest: The authors declare that there are no conflicts of interest.

References

1. 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).
2. IPCC. Climate Change 2007: Synthesis Report Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change; IPCC: Geneva, Switzerland, 2007.
3. IPCC. The Physical Science Basis: Contributing of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, 2013.
4. Ha, N.T.T.; Trang, H.T., Vuong, N.V.; Khoi, D.N. Simulating impacts of sea level rise on salinity intrusion in the Mekong Delta, Vietnam in the period 2015–2100 using MIKE 11. *Naresuan Univ. Eng. J.* **2016**, *11*, 21–24.
5. Nguyen, A.D.; Savenije, H.H.G.; Pham, D.N.P.; Tang, D.T. Using salt intrusion measurements to determine the freshwater discharge distribution over the branches of a multi–channel estuary: The Mekong Delta case. *Estuarine Coastal Shelf Sci.* **2007**, *77*, 433–445.
6. Tri, D.Q.; Nguyen, C.D.; Chen, Y.C.; Pawan, K.M. Modeling the Influence of River Flow and Salinity Intrusion Processing in the Mekong River Estuary, Vietnam. *Lowland Technol. Int.* **2014**, *16(1)*, 14–25.
7. Nguyen, A.D.; Savenije, H.H. Salt intrusion in multi–channel estuaries: A case study in the Mekong Delta, Vietnam. Hydrol. *Earth Syst. Sci*. **2006**, *10*, 743–754.
8. Tai, V.C.; Linh, N.K.; Ngan, V.H.; Quang, D.N.; Anh, T.V. Investigation of saltwater intrusion in Thanh Han river system by MIKE hydron river package. *Transport Commun. Sci. J.* **2021**, *72(1)*, 57–68.
9. Tri, D.Q. Application MIKE 11 model on simulation and calculation for saltwater intrusion in Southern region. *VN J. Hydrometeorol.* **2016**, *11*, 39–46.
10. Tri, D.Q.; Tuyet, Q.T.T. Effect of climate change on salinity intrusion: case study Ca River Basin. *J. Clim. Change* **2016**, *2(1)*, 91–101.

11. Vu, D.T.; Yamada, T.; Ishidaira, H. Assessing the impact of sea level rise due to climate change on seawater intrusion in Mekong Delta, Vietnam. *Water Sci. Technol.* **2018**, *77*, 1632–1639. https://doi.org/10.2166/wst.2018.038.
12. Van, L.C.J.; Dan, N.P.; Dieperink, C. The Challenges of Water Governance in Ho Chi Minh City. *Integr. Environ. Assess. Manage.* **2016**, *12(2)*, 345–352.
13. Thanh, B. Impact of climate change and proposed response orientation for water resources, land use planning, infrastructure, and urban flood control in Ba Ria – Vung Tau province. *VN J. Hydrometeorol.* **2013**, *12*, 12–17.
14. Khoi, D.N.; Thom, V.T.; Linh, D.Q.; Quang, C.N.X.; Phi, H.L. Impact of climate change on water quality in the Upper Dong Nai River Basin, Vietnam. Proceedings of the 36th IAHR World Congress, Netherland, 2015.
15. Kim, T.T.; Son, B.H.; Bay, N.T.; Diem, P.T.M.; Phung, N.K. The calculation of salinization by numerical method: A case study for Sai Gon river. *VN J. Hydrometeorol.* **2019**, *3*, 17–29.
16. Thuy, N.T.D.; Phung, N.K.; Hoan, N.X.; Khoi, D.N. Assessing the impacts of the changes in the upstream flow and sea level rise due to climate change on seawater intrusion in Ho Chi Minh city using the HEC – RAS 1D Model. *VN J. Hydrometeorol.* **2018**, *01*, 64–69.
17. Moriasi, D.N.; Arnold, J.G.; Van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* **2007**, *50(3)*, 885–900.
18. Nielsen, S.A; Hansen, E. Numerical simulation of the rainfall–runoff process on a daily basis. *Hydrol. Res.* **1973**, *4(3)*, 171–190.
19. Lee, H.L.; Tangang, F.; Hamid, M.R.; Benson, Y.; Razali, M.R. Modeling the Influence of River Flow and Salt Water Intrusion in the Terengganu Estuary, Malaysia. *Mater. Sci. Eng.* **2016**, *136*, 012076.
20. MONRE (Ministry of Natural Resources and Environment of Vietnam). Climate change and sea level rise scenarios for Vietnam, 2016.

VIETNAM JOURNAL OF HYDROMETEOROLOGY

Research Article

Evaluate coastal seawater quality and propose sampling frequency for monitoring in the Northeast of Quang Ninh Province, Vietnam

Huu Tuan Do[1*]

[1]Faculty of Environmental Sciences, VNU University of Science, Vietnam National University, Hanoi. Add: 334 Nguyen Trai Street, Thanh Xuan District, Ha Noi, Viet Nam.

*Corresponding author: tuandh@vnu.edu.vn; Tel: +84–2438584995

Received: 21 January 2022; Accepted: 7 March 2022; Published: 25 March 2022

Abstract: Sampling frequency plays important role in water quality monitoring activity. A suitable sampling frequency could save time and cost of monitoring work. In this study, coastal seawater quality of the Northeast of Quang Ninh Province, Vietnam was evaluated by single indicator and statistical analysis of the monitored data from 2016–2019. Then the monitoring frequency was adjusted to match the current pollution status of the study area. The results showed that seawater of the area has good quality. Monitored parameters: pH, DO, TSS, Oil, and grease, Coliform were under the QCVN 10–MT:2015/BTNMT. The manager should pay more attention to NH_4^+ concentration in seawater by controlling the pollution source of $NH4^+$. Seawater sampling frequency should be rearranged. More samples should be taken at potential pollution points, while reduced in low potential pollution points.

Keywords: Coastal seawater; seawater quality; sampling frequency; Quang Ninh.

1. Introduction

Monitoring seawater quality plays important role in managing and controlling pollution in a coastal zone. The monitored data helps managers make right decisions to adjust polluted activities or to expand social economic development. A suitable seawater monitoring stations and frequency will inform right status of seawater quality, track the change, and identify pollution sources. To evaluate seawater quality, many methods could be used such as WQI [1–2], satellite imagery [3–4], grey systems theory [5], modelling [6]. Particularly, statistical method could be used to analyze water quality parameters and identify pollution factors or pollution sources [7–8]. In Vietnam, there are some researches were performed to assess seawater quality by monitoring station data in the South region of Vietnam [9], in Quang Ninh–Hai Phong coastal area [10], in Quang Binh province [11]. These researches mainly focused on assessing the status of water quality. To get more values, it is necessary to further analyze and use monitoring data. Determination of a suitable sampling frequency is important to water quality monitoring. A good sampling frequency could save time and cost. In addition, the number of samples could have influence on the evaluation of water quality status [12] or the usage of data, the prediction reliability of modelling [13]. Some common methods could be used to calculate water sampling frequency: statistical method [14], systems analysis [15], non–parametric tests [16] or cost–effective selection [17]. In which, historical monitoring data was an important parameter to calculate or adjust sampling frequency by statistical analysis [14, 18].

The Northeast region of Quang Ninh plays important role in development of social economic of the province. With the increasing of rapid investment and development in the region, seawater quality may be affected by those activities. Current seawater monitoring points in the region had equal sampling frequency as 4 samples/point/year. This frequency does not consider the current pollution status of the region and may not timely adjust to meet the rapid changing of environmental quality under the impact of increasing development activities. Therefore, it is necessary to analyze the seawater quality of the region, then propose a suitable sampling frequency to match the new management requirement.

In this research, a combination of water quality assessment by statistical analysis with historical monitoring data to analyze and propose a sampling frequency. The objectives of this study included: (1) Assessment of seawater quality of coastal zone in the Northeast region of Quang Ninh province from 2016–2019; (2) Calculate and adjust sampling frequency for monitoring points.

2. Methods

2.1. Study area

This study was conducted in coastal zone in the Northeast of Quang Ninh Province. The data was collected from 2016–2019 in 10 monitoring points (coded as P01–P10) (Table 1 and Figure 1). These points monitored the seawater quality through 6 parameters: pH, Dissolved Oxygen (DO), Total Suspended Solid (TSS), Oil and grease, Ammonium (NH_4^+), and Coliform [19]. These parameters play important role in evaluating coastal seawater for beaches and aquaculture.

Table 1. Coordinates of sampling points.

No.	Sampling points	Code	Coordinates	
			Latitude	Longitude
1.	Mong Duong river mouth	P01	21.07112	107.366827
2.	Cam Hai, Cam Pha City	P02	21.09351	107.371072
3.	Mui Chua Port	P03	21.28557	107.458244
4.	Dam Ha–Dam Ha District	P04	21.31621	107.629111
5.	Quang Phong, Quang Dien–Hai Ha District	P05	21.37457	107.751824
6.	Quang Nghia–Mong Cai City	P06	21.49848	107.817986
7.	To Chim–Hai Hoa–Mong Cai City	P07	21.5111	108.038567
8.	Tra Co Beach–Mong Cai city	P08	21.47793	108.029884
9.	Co To Port	P09	20.96942	107.761974
10.	Vung Cat 2. 3 Group 4, Co To Town	P10	20.95323	107.740462

2.2. Methods

2.2.1. Statistical analysis

The collected data were analyzed using R software – a statistical programming language. The ggplot2 package was used to draw graphs. The monitored data was compared with the national technical regulation on marine water quality QCVN 10–MT:2015/BTNMT.

2.2.2. Frequency calculation

To calculate water sampling frequency for multiple variables and multiple stations, [14] introduced an equation based on comparing the weighting factors of whole monitoring network [14, 18].

$$N_k = W_k \times P \quad (1)$$

where P is the total number of samples obtained through the monitoring network in a year; W_k weighting values of station s.

$$W_k = \frac{\sum_{j=1}^{n} W_i}{n}; \; W_i = \left(\frac{M_i}{\sum_{i=1}^{s} M_i}\right) \quad (2)$$

where W_i is the weighting value of variable i at station s; N_k is the number of samplings at station i; M_i is the historical mean value of variables at station i; s is the number of stations; n is the number of variables.

Figure 1. Coastal seawater monitoring points.

3. Results and Discussion

3.1. Coastal seawater quality

a) pH

The pH values of seawater in the region ranged from 7.4 to 8.4. Average and median were 8.0 and 7.99 respectively. All pH values of 10 points were in compliance with the national technical regulation on marine water quality QCVN 10–MT:2015/BTNMT (range, 6.5 to 8.5). Figure 2 showed that the pH values of points P01, P09 and P10 were stable, while points P06, P07 and P08 were more fluctuated in 4 years.

b) DO

The mean value of DO concentration was 6.95 mg/l (range, 5.9 mg/l to 8.3 mg/l (Figure 3). DO concentration of all points were higher than the QCVN 10–MT:2015/BTNMT (DO ≥ 5 mg/l for beach and water sports, DO ≥ 0.4 mg/l for Aquaculture). Most of DO values concentrated at 6.5 mg/l to 7.5 mg/l. The concentration indicated that DO in seawater is suitable to any seawater using purposes.

c) TSS

Figure 4 showed the distribution of TSS concentrations. TSS values ranged from 2.5 mg/l to 57.2 mg/l. The mean value of TSS was determined to be 13.6 mg/l. About 98.7%

TSS concentration were under the QCVN 10–MT:2015/BTNMT (TSS ≤ 50 mg/l). TSS concentration at P03 and P08 were more fluctuated than the others.

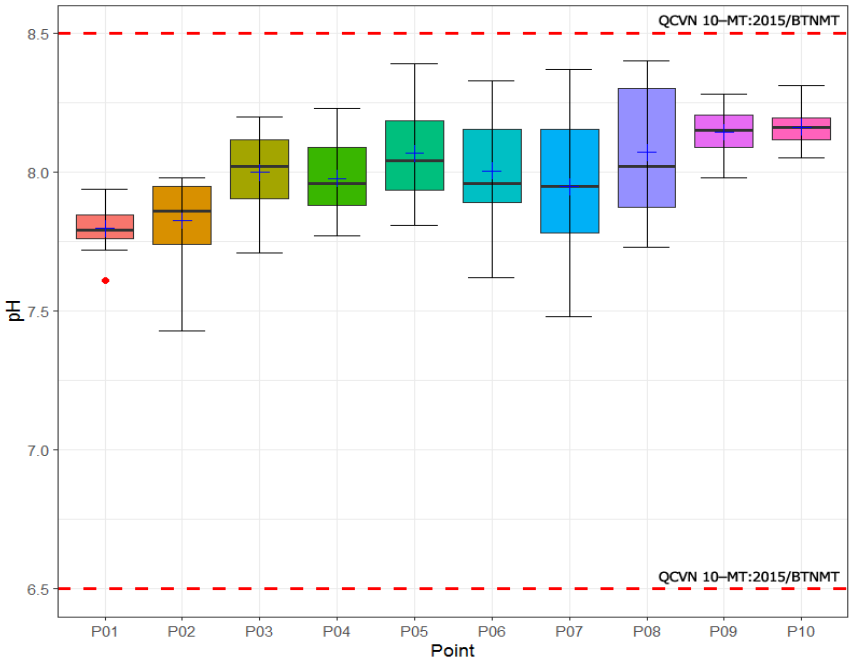


Figure 2. pH values (2016–2019).

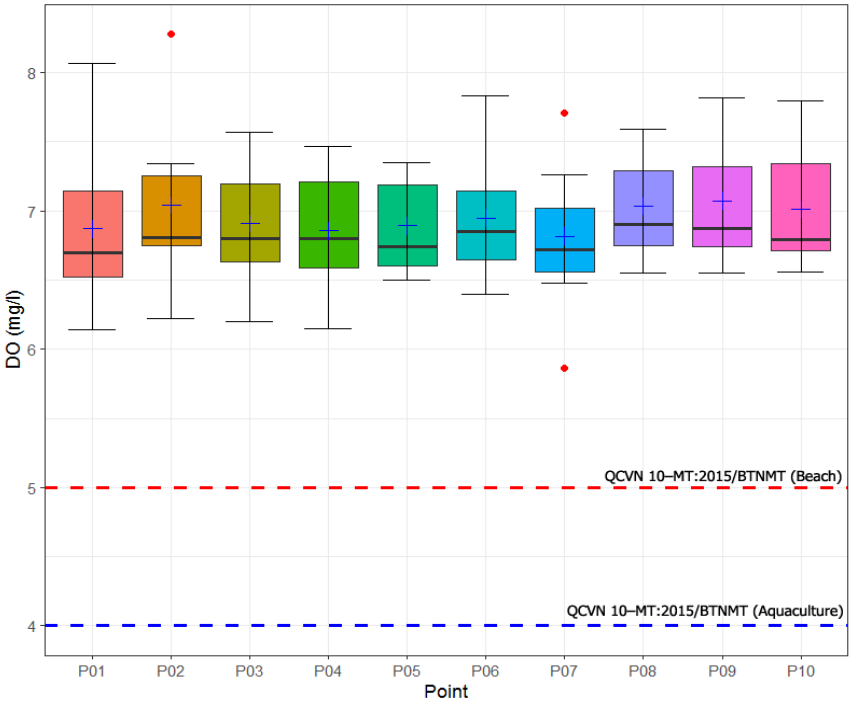


Figure 3. DO concentration (2016–2019).

d) NH_4^+

NH_4^+ concentration was found to be the most polluted parameter in the region. The mean and median values (0.38 mg/l and 0.286 mg/l respectively) were all higher than QCVN 10–MT:2015/BTNMT for marine species protection and aquaculture ($NH_4^+ \leq 0.1$ mg/l) but lower than QCVN 10–MT:2015/BTNMT for beach and watersport ($NH_4^+ \leq 0.5$ mg/l). The TSS

concentration were also highly fluctuate from 0.036 mg/l to 0.952 mg/l (Figure 5). About 24.67 % NH_4^+ concentration was lower than the QCVN 10–MT:2015/BTNMT (≤ 0.1 mg/l) and 62.67% concentration lower than the QCVN 10–MT:2015/BTNMT (≤ 0.5 mg/l), concentrated in 2017 with the mean value of 0.343 mg/l. However, in 4 years, i.e., 2016–2019, the mean value of NH_4^+ concentration was still lower than the standard.

Domestic wastewater released from channel and river flowing through community was identified as pollution sources of NH_4^+. To control NH_4^+ concentration, it is necessary to collect and treat domestic wastewater before discharging to coastal zone.

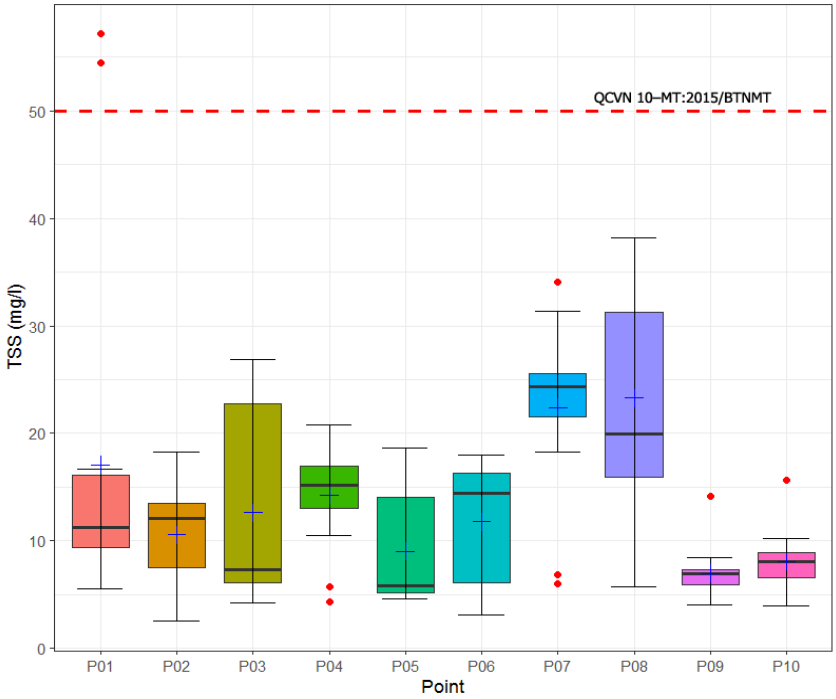


Figure 4. TSS concentration (2016–2019).

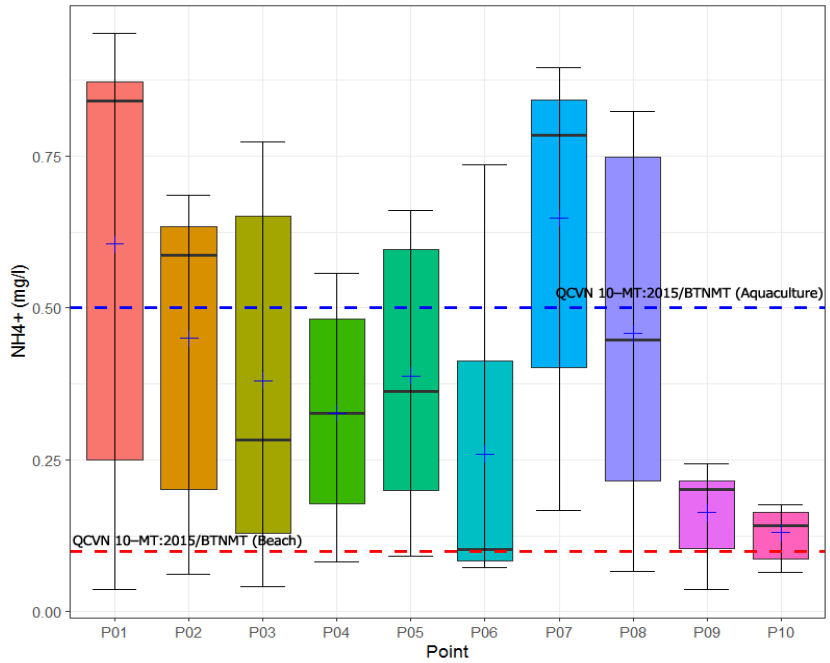


Figure 5. NH_4^+ concentration (2016–2019).

e) Coliform

In 4 years, 2016–2019, 99.32% Coliform samples were lower than the QCVN 10–MT:2015/BTNMT (1000 NPM/100ml). The Coliform concentration were very highly fluctuated from 3 NPM/100ml to 1500 NPM/100ml (Figure 5). Points with high Coliform concentration appeared near river mouth (P02 and P07), while monitoring points located in island had lower Coliform concentration (P05, P09 and P10). Average Coliform concentration in 4 years meet water quality requirement for all using purposes.

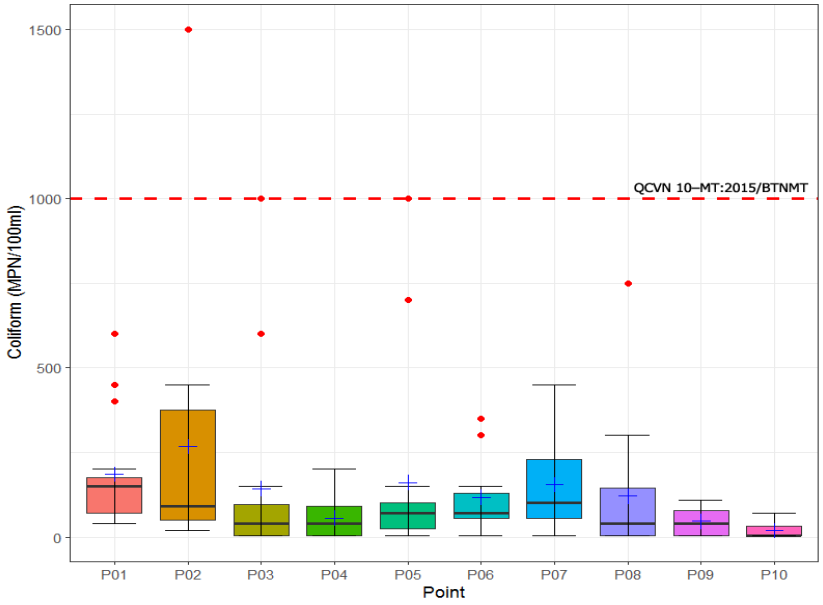


Figure 6. Coliform concentration (2016–2019).

f) Oil and grease

All the monitoring points had lower oil concentration than the QCVN 10–MT:2015/BTNMT (0.5 mg/l). This result indicated that, inspire of high ship activities in the region, the oil and grease concentration still very low (Figure 7). This well supports for the development of the ecosystem in the coastal zone of the region.

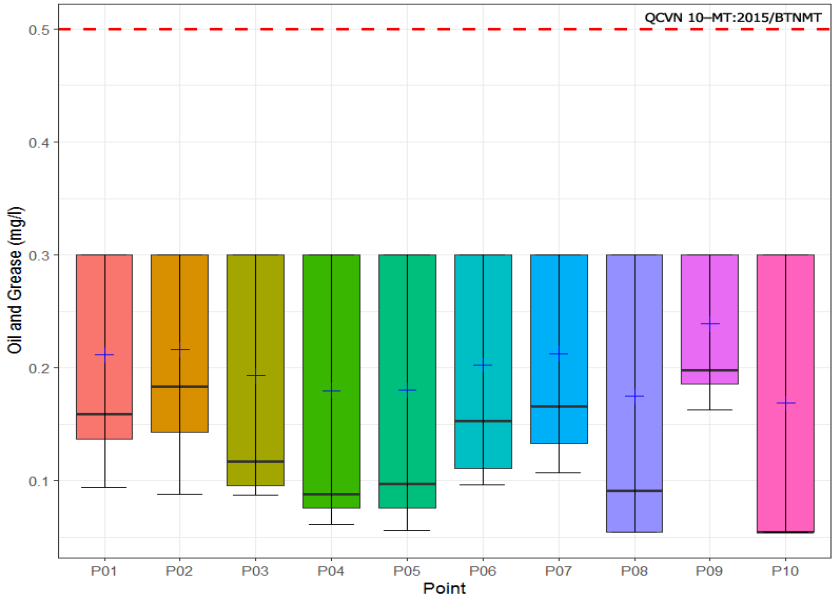


Figure 7. Oil and grease concentration (2016–2019).

3.2. Sampling frequency

a) Weighting values of variables

Comparing weighting values of pH, DO, TSS, $NH4^+$, Oil and Coliform among 10 monitoring points indicated that points P09 had the highest weighting value of pH, DO and Oil, while TSS, NH4+, and Coliform were P08, P07, and P01, respectively (Table 2 and Figure 8). The high weighting value of variables in 10 points (W_k) were P07, P01, P02, and P08. It indicated that sampling frequency of these points should be increased while the P09 and P10 should be decreased.

Table 2. Weighting values of variables.

Point	W_{pH}	W_{DO}	W_{TSS}	W_{NH4^+}	W_{Oil}	$W_{Coliform}$	W_k
P01	0.098	0.099	0.126	0.160	0.107	**0.146**	**0.122**
P02	0.098	0.101	0.077	0.118	0.112	0.210	**0.119**
P03	0.100	0.099	0.093	0.099	0.096	0.113	0.100
P04	0.100	0.099	0.105	0.086	0.091	0.044	0.088
P05	0.101	0.099	0.066	0.102	0.091	0.127	0.098
P06	0.100	0.100	0.087	0.068	0.102	0.092	0.091
P07	0.099	0.098	0.165	**0.170**	0.107	0.122	**0.127**
P08	0.101	0.101	**0.171**	0.120	0.086	0.095	**0.112**
P09	**0.102**	**0.102**	0.051	0.042	**0.122**	0.036	0.076
P10	0.102	0.101	0.059	0.034	0.086	0.015	0.066

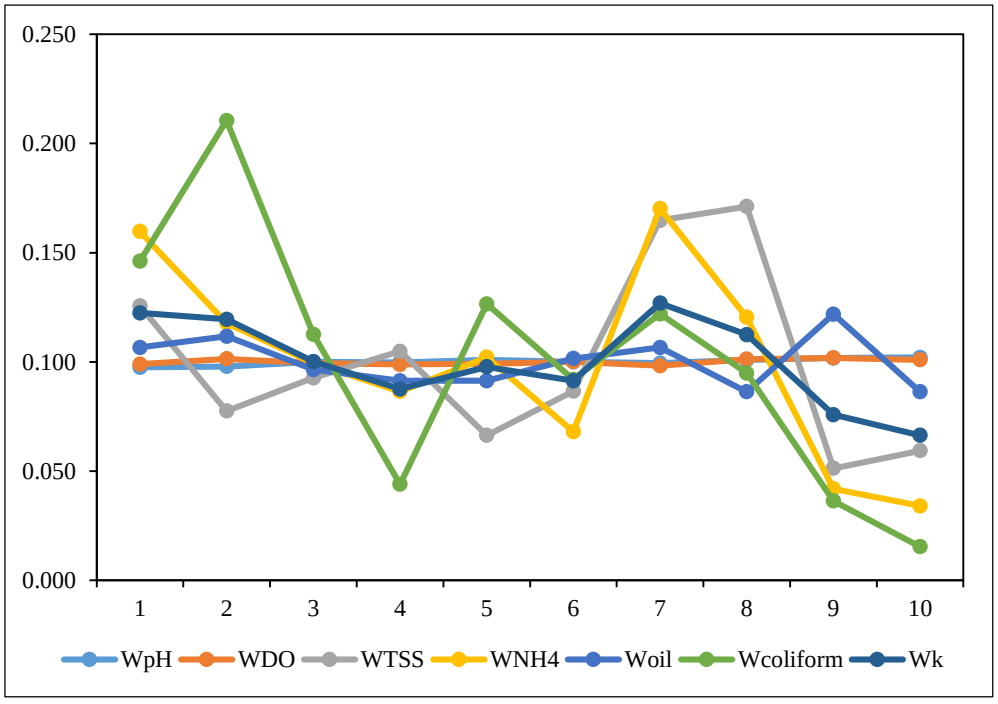


Figure 8. Weighting values of pH, DO, TSS, NH_4, Oil and grease, Coliform and Wk.

b) Sampling frequency

The W_k indicated that point P01, P02, P07, and P08 were more polluted than the others, while point P09 and P10 had better seawater quality. Figure 9 showed the sampling frequency of 10 points in a year, in which the frequency should be increased at point P01, P02, and P07, while decreased at point P09 and P10.

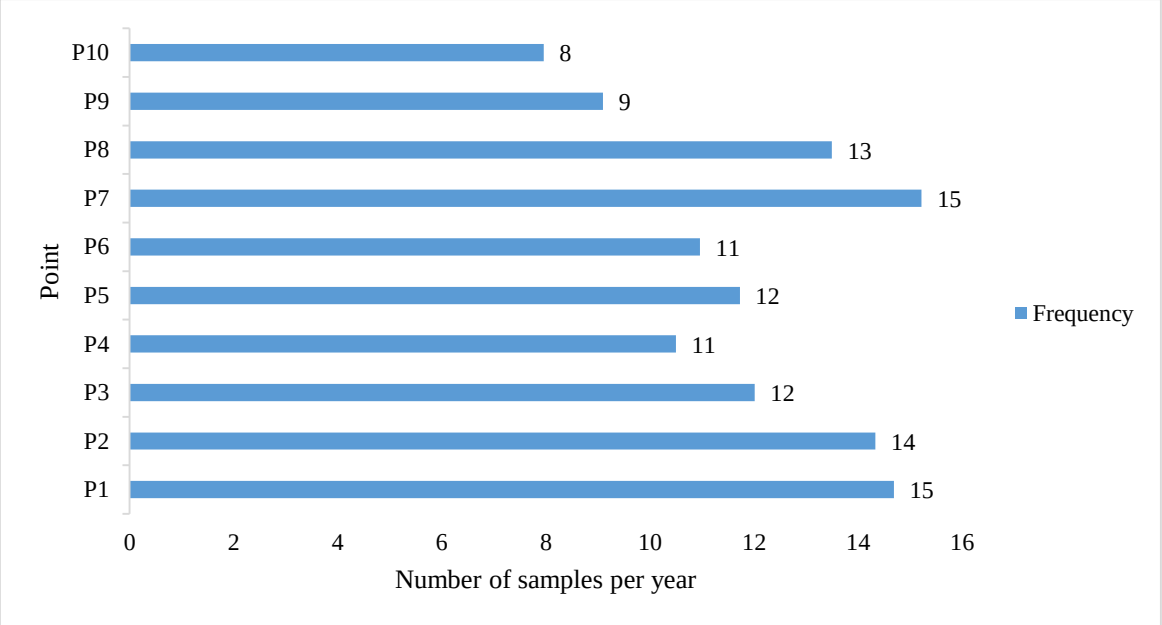


Figure 9. Proposed sampling frequency.

In environmental monitoring and management, the places with more polluted than the others, more actions should be considered to control pollution sources [14, 18]. In which, monitoring activities should be increased to monitor and track the sources as well as the changes of pollution. The calculated sampling frequency indicated that the monitoring activities should be focused on P01, P02 and P07, while reduced in P9 and P10.

4. Conclusion

Seawater in the study area has good quality. In 6 monitored parameters in 4 years, 2016–2019, (pH, DO, TSS, Oil and grease, NH_4^+, Coliform), 5 parameters were under the QCVN 10–MT:2015/BTNMT. However, it had a signal of NH_4^+ pollution. Manager should pay more attention to NH_4^+ concentration in seawater of the region by controlling pollution source of NH_4^+. Sampling frequency should be rearranged. More samples should be taken at potential pollution points P01, P02 and P07, while reduced in P09 and P10. Industrial and domestic wastewater near P01, P02, P08 and P09 must be treated before releasing. Data series of this research were short so the research results were limited.

Acknowledgement

Author specially thanks to Natural Resources and Environmental Monitoring Center, Quang Ninh Province has provided reference materials for the research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ma, Z.; Li, H.; Ye, Z.; Wen, J.; Hu, Y.; Liu, Y. Application of modified water quality index (WQI) in the assessment of coastal water quality in main aquaculture areas of Dalian, China. *Mar. Pollut. Bull.* **2020**, *157*, 111285.
2. Tam, P.H. Application of Water Quality Index to Assess Environmental Quality in Coastal Monitoring Stations in the South Viet Nam in the Last 5 Years (2011–2015). *VNU J. Sci.: Earth Environ. Sci.* **2016**, *4(32)*, 36–45.
3. Caballero, I.; Román, A.; Tovar–Sánchez, A.; Navarro, G. Water quality monitoring with Sentinel–2 and Landsat–8 satellites during the 2021 volcanic eruption in La Palma (Canary Islands). *Sci. Total Environ.* **2022**, *822*, 153433.

4. Ashikur, M.R.; Rupom, R.S.; Sazzad, M.H. A remote sensing approach to ascertain spatial and temporal variations of seawater quality parameters in the coastal area of Bay of Bengal, Bangladesh. *Remote Sens. Appl.: Soc. Environ.* **2021**, *23*, 100593.
5. Zheng, Y.; Zheng, X.; Gao, Z.; Zhang, Y. Prediction of Seawater Quality in Rigs–to–Reefs Area Based on Grey Systems Theory. *Procedia Environ. Sci.* **2013**, *18*, 236–242.
6. Wu, Y.; Xie, P.; Dahlak, A. Utilization of Radial Basis Function Neural Network model for Water production forecasting in Seawater Greenhouse units. *Energy Rep.* **2021**, *7*, 6658–6676.
7. Franklin, J.B.; Sathish, T.; Vinithkumar, N.V.; Kirubagaran, R..Madeswaran, P. Seawater quality conditions of the south Andaman Sea (Bay of Bengal, Indian Ocean) in lustrum during 2010s decade. *Mar. Pollut. Bull.* **2018**, *136*, 424–434.
8. Lušić, D.V.; Kranjčević, L.; Maćešić, S.; Lušić, D.; Jozić, S.; Linšak, Ž.; Bilajac, L.; Grbčić, L..Bilajac, N. Temporal variations analyses and predictive modeling of microbiological seawater quality. *Water Res.* **2017**, *119*, 160–170.
9. Tam, P.H. Coastal Seawater Quality from Data at South Vietnam Monitoring Stations during 2013 – 2017. *VNU J. Sci.: Earth Environ. Sci.* **2018**, *2(34)*, 95–109.
10. Nam, L.V.; Nghi, D.T..Ngan, L.T.K. Using index for assessment of water quality and classification of eutrophication levels in the Quang Ninh – Hai Phong coastal area. *J. Mar. Sci. Technol.* **2018**, *17(4)*, 490–497.
11. Nguyen Dinh Nguyen, N.D.T., Vu Van Tich, Vu Viet Duc, Hoang Van Hiep Characteristics of Seawater Environment and Geochemistry of Surface Sediment in Quang Binh Sea Area (60–100m Depth). *VNU J. Sci.: Earth Environ. Sci.* **2018**, *34(4)*, 89–97.
12. Pekárová, P.; Pekár, J..Miklánek, P. Impact of water sampling frequency on estimating water quality status in the Ondava River. *Ecohydrol. Hydrobiol.* **2006**, *6(1)*, 105–113.
13. Piniewski, M.; Marcinkowski, P.; Koskiaho, J..Tattari, S. The effect of sampling frequency and strategy on water quality modelling driven by high–frequency monitoring data in a boreal catchment. *J. Hydrol.* **2019**, *579*, 124186.
14. Ward, R.C.; Loftis, J.C.; Nielsen, K.S..Anderson, R.D. Statistical evaluation of sampling frequencies in monitoring networks. *J. Water Pollut. Control Fed.* **1979**, *51(9)*, 2292–2300.
15. Chappell, N.A.; Jones, T.D.; Tych, W. Sampling frequency for water quality variables in streams: Systems analysis to quantify minimum monitoring rates. *Water Res.* **2017**, *123*, 49–57.
16. Naddeo, V.; Scannapieco, D.; Zarra, T.; Belgiorno, V. River water quality assessment: Implementation of non–parametric tests for sampling frequency optimization. *Land Use Policy* **2013**, *30(1)*, 197–205.
17. Loftis, J.C.; Ward, R.C. Cost–effective selection of sampling frequencies for regulatory water quality monitoring. *Environ. Int.* **1980**, *3(4)*, 297–302.
18. Sanders, T.G.; Ward, R.C.; Loftis, J.C.; Steele, T.D.; Adrian, D.D..Yevjevich, V. Design of networks for monitoring water quality. *Water Res. Publ.* **1983**, 98–150.
19. EMAC, Q.N., Current Environmental Quality in Quang Ninh Province. 2016–2019: Quang Ninh, Vietnam.

VIETNAM JOURNAL OF HYDROMETEOROLOGY

Research Article

Spatial and Temporal Modeling of Land use/Land cover Change at the Ca River Basin (North Central Viet Nam) Using Markov Chain and Cellular Automata Approach

Bang Nguyen Thanh [1,2,3*], Phong Doan Ha[1]

[1] Hydro–Meteorology and Climate change Technology Research Division, Vietnam Institute of Meteorology, Hydrology and Climate change, No. 23, Nguyen Chi Thanh street, Dong Da district, Hanoi 100000, Vietnam
[2] Geographic Information System Group, Department of Business and IT, University College of Southeast Norway, Gullbringvegen 36, N–3800, Bø i Telemark, Norway
[3] Department of Health and Environmental Studies (INHM), University College of Southeast Norway, Gullbringvegen 36, N–3800, Bø i Telemark, Norway

*Corresponding author: bangnt.imhen@gmail.com; Tel.: +84–838734488
Received: 08 January 2021; Accepted: 08 March 2022; Published: 25 March 2022

Abstract: Simulation of Land use/Land cover (LULC) change has been conducted extensively in the past with varying techniques and methodologies with Markov Chain incorporating Cellular Automata approach among those. The Markov–Cellular Automata (Markov_CA) model has been applied worldwide, however, model parameter calibration is site–specific. In Viet Nam, research on LULC change a pressing issue given the rapid socio–economic development. Research on LULC change is a necessary starting point for impacts assessment on water resources, land resources, ecosystems, environment, etc. However, what we lack is a method for modeling our insights to simulate LULC fluctuations and to project future LULC. Therefore, this article offers a way to combine known problems to produce a new result. The change of LULC for the period 2005–2015 will be simulated and will result in a prediction of the LULC of 2030. In addition, the calibrated Markov_CA model adapted to the study area will also be a valuable reference for employment in similar areas. Finally, the expected results and the calibrated model are validated by the Kappa coefficient and provide a good level of agreement.

Keywords: Land use/Land cover Change; Markov Chain; Cellular Automata; Ca River Basin; Viet Nam.

1. Introduction

Land change science has emerged as a fundamental component of global environmental change and sustainability research [1]. Land use/land cover (LULC) has interaction with soil, water resources, biodiversity, ecosystem, climate [2]. Changes in land use and land cover will consequently result in the changes of the latter. LULC changes are often caused by two influencing factors: anthropogenic and natural [3]. Human–induced land cover change such as deforestation has been a major contributor to increasing CO_2 concentration [4], the rapid expansion of agriculture reduce available freshwater given the intensive water use nature of agriculture (70% of total freshwater used by humankind), land exploitation disrupts the biotic function of soils [2]…. However, these changes can also be caused by natural factors, in particular vegetation cover [3]. Therefore, detecting and projecting LULC dynamics is necessary.

Recent advances in remote sensing data and growing advances in their temporal, spatial, and spectral resolutions provide useful data and tools for the detection of changes on LULC at different scales [5]. Remote sensing and geographical information system (GIS) together can provide an accurate depiction of changes in LULC [6], while remaining cost–effective [7].

A typical approach to simulate and predict LULC changes is to investigate the factors contributing to the land transitions and to provide a probabilistic prediction of where the changes may occur through modeling [5, 8]. There are various modelling approaches for the simulation and exploration of LULC changes [9–10]. According to [11] a set of 19 land–use models were reviewed in detail as representative of the broader set of models identified from the more comprehensive review of the literature. They included Markov models, Cellular Automata models, logistic regression models, econometric models, weights models, etc.

Markov model is stochastic modeling that uses evolution from "*t–1*" to "*t*" to project probabilities of changes for a future date "*t+1*" [12]. However, a stochastic Markov model is not appropriate because it does not consider spatial knowledge distribution within each category and transition probabilities are not constant among landscape states; so it may forecast the right magnitude of change but not the right direction [13]. This deficiency of the Markov model can be offset through the integration with other spatial component models [14]. Hence, the Cellular Automata Markov model combines the concepts of Markov Chain, Cellular Automata, Multi–Criteria Evaluation and Multi–Objective Land Allocation [8] is an interesting approach to modeling both spatial and temporal changes. [8] also determined that Cellular Automata Markov gave the approximate results to Multi–Layer Perceptron Markov [15] and outperformed Stochastic Markov in various validation techniques including: per category method, kappa statistics, components of agreement and disagreement, three map comparison, and fuzzy method.

The article aims to combine the scientific basis of the Markov–Cellular Automata method and the practice of the Ca River basin to find the influencing factors and model parameters to simulate the change of land cover in the Ca River basin. At the same time, the results of the article are the future land cover in 2030 of the Ca River basin will also supplement the data source – which is still lacking, to serve the planning and management of water resources in this area.

2. Study area and data used

2.1. Description of the study area

The Ca River system is a transnational basin originating from the upper 2000 m high mountain range in Xieng Khuang province of Laos, flowing northwest–southeast before entering the North Central of Viet Nam and pouring into the sea at the Hoi estuary. The Ca River system is located between 18°15'50" to 20°10'30" north latitude; and 103°45'20" to 105°15'20" east longitude. The outlet of the basin is at 18°45'27" north latitude; and 105°46'40" east longitude. The starting point of the Ca River system within Viet Nam is at 19°24'59" north latitude; and 104°04'12" east longitude [16].

The mainstream of the Ca River system is approximately 513 km in length, of which the length of the reach within Viet Nam's territory is 361 km long [17]. The mainstream flows through most parts of Nghe An Province, known as the Ca river. In Anh Son District, the river receives tributary is the Hieu River. Downstream of the Ca River is its confluent with the La river flowing from Ha Tinh Province. From this reach to the sea, the river is called the Lam river (Figure 1).

The basin area within Viet Nam is 17,730 km^2 in a total basin area of 27,200 km^2 [18]. Every year, the basin receives an average precipitation of 1100 ÷ 2500 mm. In the large rainfall centers such as upstream of Hieu, La Rivers, average annual rainfall could be as high

as 2000 ÷ 2400 mm. Land cover in the river basin composes of 44% forest, 16% paddy rice, 2% vegetable and crops, 38% others [19].

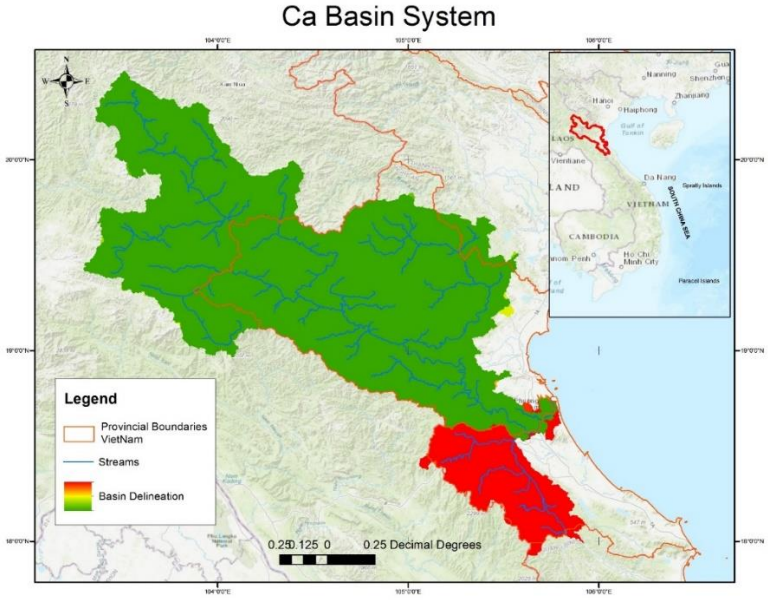


Figure 1. Ca River Basin.

2.2. Data used

2.2.1. Remote sensing data

In order to study how the landscape has changed over the 2005–2015 period, land cover maps of 2005–2010–2015 should be developed. The maps are based on LANDSAT image data from the United States Geological Survey (USGS). The collected images were LANDSAT 5 TM and LANDSAT 8 OLI/ TIRS with the same resolution of 30 m (Table 1), where the image of path 127 rows 47 was the largest covering, approximately 80% of the whole Ca river basin. Selected images based on the criteria: low cloud cover, no "scan line" error, the time of the image is not too far apart, especially the images 127–47.

The Landsat 5 TM images consist of six spectral bands with a spatial resolution of 30 meters for Bands 1–5 and 7. Landsat 8 OLI and TIRS images consist of nine spectral bands with a spatial resolution of 30 meters for Bands 1 to 7 and 9. The ultra–blue Band 1 is useful for coastal and aerosol studies. Band 9 is useful for cirrus cloud detection. The resolution for Band 8 (panchromatic) is 15 meters. Thermal bands 10 and 11 are useful in providing more accurate surface temperatures and are collected at 100 meters [20]. For LULC classification, images are collected as shown in (Table 1). High spatial resolution images from Google Earth and current land use status map published by the Department of Survey, Mapping and Geographic Information Viet Nam are used to validate the results.

Table 1. Description of the data sources and types used in this study.

Year	Data type and resolution	Path–row	Date	Source
2005	Landsat 5 TM	126–47	14th July 2005	https://earthexplorer.usgs.gov/
	30 m	126–48	09th April 2005	
		127–46	18th May 2005	
		127–47	18th May 2005	
		128–46	23rd April 2005	
		128–47	07th April 2005	

Year	Data type and resolution	Path–row	Date	Source
2010	Landsat 5 TM	126–47	12th July 2010	https://earthexplorer.usgs.gov/
	30 m	126–48	12th July 2010	
		127–46	08th Nov 2010	
		127–47	08th Nov 2010	
		128–46	21st April 2010	
		128–47	30th Oct 2010	
2015	Landsat 8 OLI/TIRS	126–47	11th Aug 2015	https://earthexplorer.usgs.gov/
	30 m	126–48	28th Sept 2015	
		127–46	30th May 2015	
		127–47	30th May 2015	
		128–46	28th Oct 2015	
		128–47	28th Oct 2015	

2.2.2. Ancillary data and field data (GPS)

a) Digital Elevation Model–DEM

ASTER Global Digital Elevation Model 2.0 data (GDEM 2.0) is a product of the Ministry of Economy, Trade, and Industry (METI) and National Aeronautics and Space Administration (NASA) collected from the US Geological Survey (USGS). GDEM 2.0 was announced by METI and NASA in mid–October 2011, inheriting almost all the features of GDEM 1.0 with a resolution of 30 m, covering from latitude 83º North to 83º South. But GDEM 2.0 has a higher horizontal resolution by using a 5×5 correlation kernel instead of 9×9 as used for GDEM 1.0. GDEM 2.0 has a total accuracy of 17 m compared to 20 m of GDEM 1.0 along with a 95% certainty [21] (ASTER–GDEM, October 2011).

The DEM data for the study area were collected from latitude 18º to 19º North, longitude 103º to 105º East. The ArcSWAT tool is then used to calculate flow direction, accumulate flow, create sub–basin area, create flow net, discharge outlet, etc [22].

b) Current land use status map

Beginning in 1999, under Directive 24/1999/CT–TTg of the Prime Minister of Viet Nam on land inventory is issued in 2000. Since then, the inventory and mapping of current land use status have been performed in 2005, 2010, 2015 (Figure 2). They are valuable ancillary references for LULC classification.

The current land use status map is a map showing the distribution of land categories according to the regulations on an inventory of land use purposes at the time of land inventory. The current land use status map is drawn up on the basis of the cadastral map, in comparison with the field data and land inventory data; In case no cadastral map is available, use aerial photographs or high–resolution satellite images converted into orthogonal photographs combined with field data and land inventories to make the current land use status map; In case there are no such maps, current land use status map of the previous period is used and also will be checked with field data and land inventory data.

The land cover based on use purpose includes agriculture production, forestry, aquaculture, salt production, other agricultural, built–up, specialized (eg.: State's office, defence, and security, transport, medical, education, etc.), rivers and water surfaces, bare and unused land. They can be regrouped into 5 classes: Forest, Agriculture, Built–up, Waterbody, and Bare area (Figure 2).

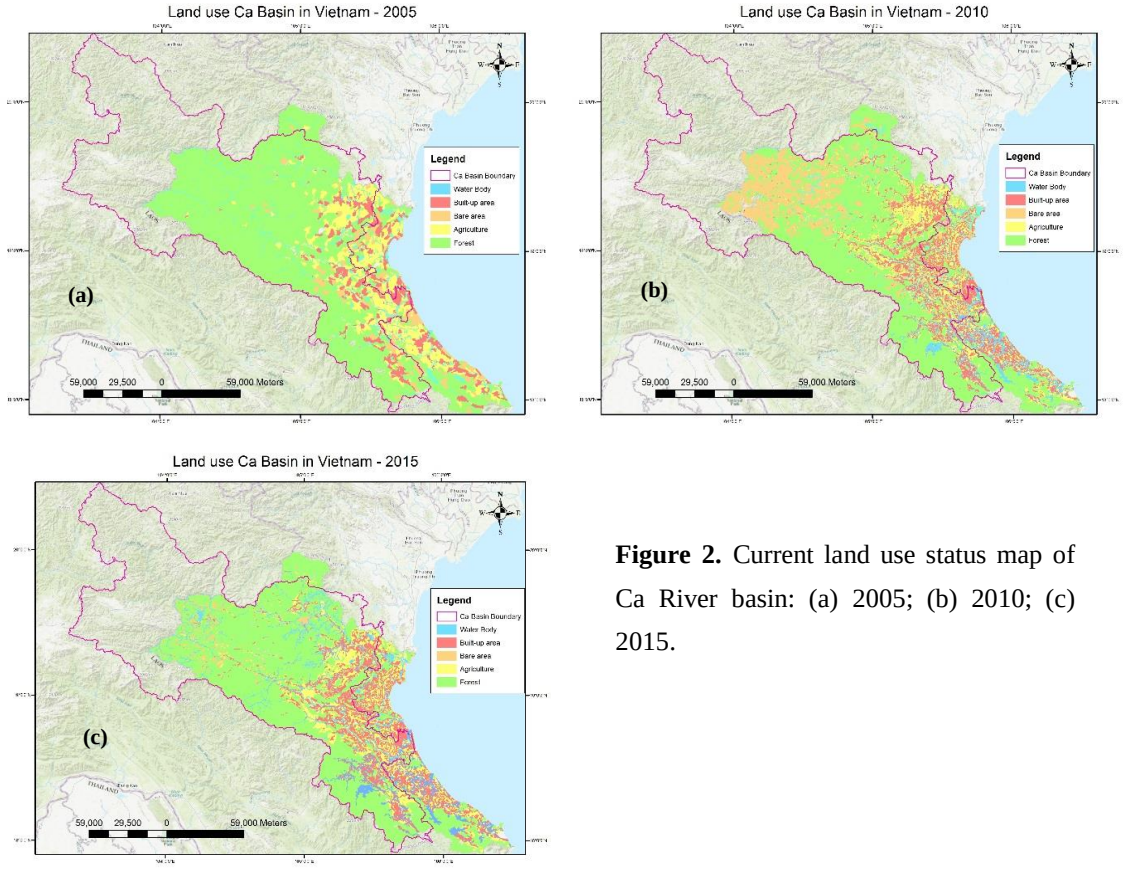


Figure 2. Current land use status map of Ca River basin: (a) 2005; (b) 2010; (c) 2015.

c) Field data

Ground data is collected for classification and verification of classification results. The total number of samples acquired is 120 samples. Because the latter works related to three periods, all the ground data were consolidated; care was taken to ensure that areas that had undergone a change (e.g., burn regeneration) were excluded from the investigation. Through consultation with local people, five classes were sampled – Forest, Agriculture, Built–up, Waterbody, Bare area – with about 20 ground data for each class. Some test data is additionally collected using Google Maps by random points algorithm.

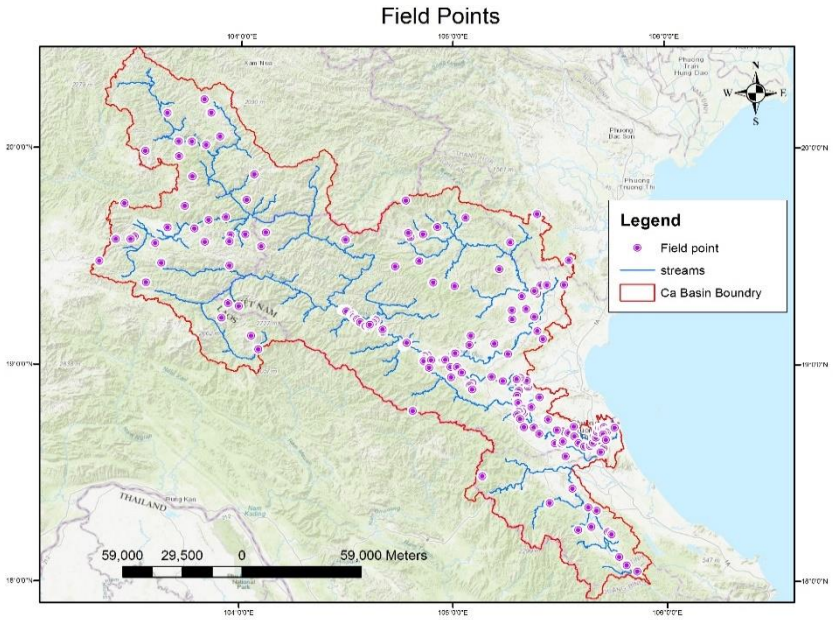


Figure 3. Location of ground data.

3. Theoretical background of the method used

3.1. Maximum Likelihood classification

The classification method used is the Maximum Likelihood Classification (MLC), which is one of the methods of the Supervised Classification [7, 23]. This method is based on a given set of sample pixels and hence identifies pixels with the same spectral characteristics. Next, the estimated (Gaussian) probability density function is used to identify other pixels of the same land use/land cover [24]. The MLC principle also can be found in Foody and Strahler's researches [24–25]. This is a commonly used method of image classification and provides relatively high classification accuracy.

In this study, the Landsat data were classified with the maximum likelihood decision rule and some ancillary data (e.g., DEM, land use data, vegetation index, and textural analysis of the Landsat images) were combined through an expert (or hypothesis testing) system to improve the classification accuracy [26]. Considering the spectral characteristics of the satellite images and existing knowledge of land use/land cover of the study area, five LULC categories were identified and classified for 2005, 2010, and 2015 (Table 2).

Table 2. LULC categories distributed for the classification (Circular 08/2007/TT–BTNMT).

LULC category	Description
Forest	Land with natural forest or planted forests meeting the forest standards (e.g., production forests, protection forests, and special–use forests).
Agriculture	Land for agriculture production including Land for planting annual crops (e.g., paddy land, grassland used for breeding, other annual crops); Land for perennial crops (e.g., orchards, perennial crops).
Built–up	Land for construction of dwelling houses, construction of works, land for offices of agencies and non–business works; land protection, security.
Waterbody	Land for rivers and streams and specialized water surfaces, coastal water surface.
Bare area	Land with no purpose of use including unused plain land, unused hill or mountain land, Rocky Mountains without forests.

3.2. Markov Chain

Markov process is a special random moving from one state to another state at each time step via the use of transition probability matrices [14, 27]. The transition probability matrix is calculated by assuming that probability distribution over the next state only depends on the current state, but not on previous ones [10]. In this study, a probability matrix based on the likelihood of the LULC variations between 2005, 2010, and 2015 was used to predict the LULC map in 2030. The transition matrix can be presented as follows [14]:

$$P = (P_{ij}) = \begin{vmatrix} P_{11} P_{12} \dots P_{1n} \\ P_{21} P_{22} \dots P_{2n} \\ \dots \\ P_{n1} P_{n2} \dots P_{nn} \end{vmatrix} 0 \le P_{ij} \le 1 \quad \sum_{i=1}^{n} P_{ij} = 1 \qquad (1)$$

where P is the transition probability matrix, P_{ij} is the probability of the i^{th} LULC changing to j^{th} LULC from initial year to illation year and n is the number of LULC classes.

The Markov chain model is very powerful to determine the possibility of land–use change between two time periods. However, the Markov chain model cannot provide the spatial distribution of occurrences of land–use change [28].

3.3. Cellular automata

Generally, Cellular automata (CA) models aim to simulate the real nature regulations. Land–use change modeling using the CA technique is one of the preferred methods because it gives explicit spatial modeling results based on defined transition rules [23]. A CA consists of discrete cell space, in which states characterize every cell. [29] define a simple CA to include the following components: (1) a grid space L on which the model operates, (2) cell states Q in the grid space, (3) transition rules f, which determine the spatial dynamic process, (4) status of the neighborhood Δ that influences the central cell. Hence, the spatiotemporal changes of state in a system can be described as [30]:

$$A = [L,Q, \Delta, f] \quad (2)$$

Each Q cell of L grid space will change their state in discrete time steps. The state of a cell Q depends on its neighborhood Δ (the surrounding cells) and the corresponding f transition rules. However, the most important concern in the CA model is defining appropriate transition rules f based on training data that controls the model [31].

3.4. Accuracy assessment

There are many accuracy assessment methods that have been discussed in the remote sensing context e.g., [32–35], but the most widely proposed and used method is confusion matrix or error matrix. A measurement termed "percentage of cases correctly allocated" derived from a confusion matrix has been used to measure classification accuracy [36]. The accuracy of the individual class may be derived from the matrix by relating the number of cases correctly allocated to the class to the total number of cases of that class (Figure 3). This leads to two concepts: user's accuracy and producer's accuracy. The user's accuracy provides the user information on the accuracy of the LCLU data against actual data. Producer's accuracy indicates the percentage of samples of a certain (reference) class that were correctly classified [37]. These accuracies are calculated based upon the confusion matrix's row or column [38].

Actual Class

Predicted Class	A	B	C	D	∑
A	n_{AA}	n_{AB}	n_{AC}	n_{AD}	n_{A+}
B	n_{BA}	n_{BB}	n_{BC}	n_{BD}	n_{B+}
C	n_{CA}	n_{CB}	n_{CC}	n_{CD}	n_{C+}
D	n_{DA}	n_{DB}	n_{DC}	n_{DD}	n_{D+}
∑	n_{+A}	n_{+B}	n_{+C}	n_{+D}	n

$$\text{User's accuracy} = \frac{n_{ii}}{n_{i+}} \quad (3)$$

$$\text{Producer's accuracy} = \frac{n_{ii}}{n_{+i}} \quad (4)$$

Figure 4. Error matrix with n_{ij} representing the proportion of predicted class i and the actual class j.

On the issue of the chance of agreement, Cohen's kappa coefficient has been used and be adopted as a standard measure of classification accuracy [39]. Kappa takes the chance

agreement into account and Kappa adjusts the percentage correct measure by subtracting the estimated contribution of the chance agreement [40]. The definition of Kappa (k) is:

$$k = \frac{p_0 - p_e}{1 - p_e} \quad (5)$$

where p_0 is the observed proportion correct, p_e is the expected proportion correct due to change.

4. Proposed methodology for Spatial and Temporal Modeling of Land Cover Change at the Ca River Basin (North Central Viet Nam)

This study employed an integrated Markov – Cellular Automata (Markov–CA) model to predict LULC changes for the Ca River basin in the target years 2030. Data preprocessing and format unification were achieved using GIS, which provides numerous functions for visualizing and analyzing the data [41]. Markov–CA model is applied by TerrSet, developed by Clark Labs at Clark University, is an integrated geospatial software with the ability to incorporate the IDRISI GIS analysis for monitoring and modeling purposes [42]. In general, the flow chart of the methodology is summarized in (Figure 5).

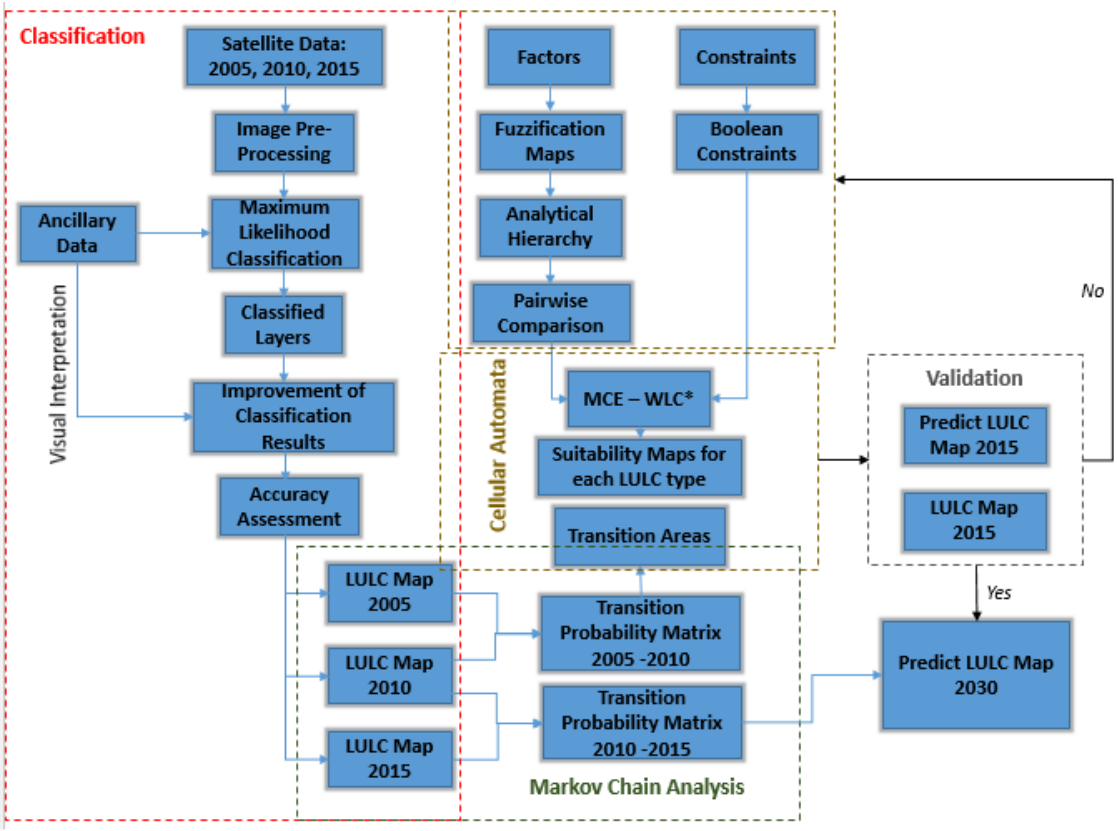


Figure 5. Workflow showing the methodology in the study; *MCE–WLC: Multi–criteria evaluation–Weighted linear combination.

In the first phase, the Landsat images are classified and LULC layers are prepared. In the second phase, the Transition Probability Matrix and Transition Areas are calculated with Markov Chain Analysis. At the same time, factors and constraints are set up for each land–use class and fuzzy functions are applied for each factor and assigned Boolean values (0 or 1) for constraints. Then Analytical Hierarchy Process and Pairwise Comparison are used to assign the weight of each factor. Weighted factors and Boolean constraints are used in the MCE–WLC function to generate suitability maps for each LULC type. In the third phase, all previous components are thrown into the Cellular Automata module and output to the projected LULC map of the next period (2015). In the validation phase, the projected LULC

map is compared to the LULC map on the agreement level by the Kappa coefficient. If the validation results indicate that a bad agreement, the Weighted factors, and Boolean constraints will be reconsidered. Otherwise, the model is ready to predict LULC maps in the future.

5. Result

5.1. Image processing and classification

For an accurate assessment of LULC between 2005 and 2015, atmospherically–corrected surface reflectance Landsat 5 TM and Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) images were collected from the United States Geological Survey (USGS) website. All scenes were verified for geometric accuracy and all data were projected on WGS 1984, UTM zone 48N.

Images were stacked, subset, and analyzed in ENVI, ArcGIS software, and classified using the maximum likelihood algorithm. Supervised approaches using a maximum likelihood classifier algorithm were applied for the extraction of LULC. A modified land–cover classification system was used for remote sensing data as recommended by [43] and 5 classes were identified: built–up, forest, agriculture, waterbody, bare area (Table 2).

Segmentation provides an approach to extracting features from imagery based on objects. These objects are created via an image segmentation process where pixels in close proximity and having similar spectral characteristics are grouped into a segment. Segments exhibiting certain shapes, spectral, and spatial characteristics can be further grouped into objects – meaningful object–oriented feature class. The result is a grouping of image pixels into a segment characterized by an average color.

A supervised classification requires collecting training samples as the basis for the maximum likelihood algorithm speculating other pixels of the same class. The more accurate the data collected, the more accurate the classification. Bands composite method is widely used in remote sensing to support that. Each composite has its advantages in classifying LULC. A composite image using bands 4, 3, 2 in Landsat 5 TM images or bands 7, 6, 5 in Landsat 8 for the red, green, blue channels, respectively will be easier to detect roads, water–body, and agriculture class. Other composite images are also used to expose other land classes e.g. bands 5, 4, 3 or 4, 5, 3 in Landsat 5 TM or 6, 5, 4 in Landsat 8 for forest classification. The results are presented in (Figure 6).

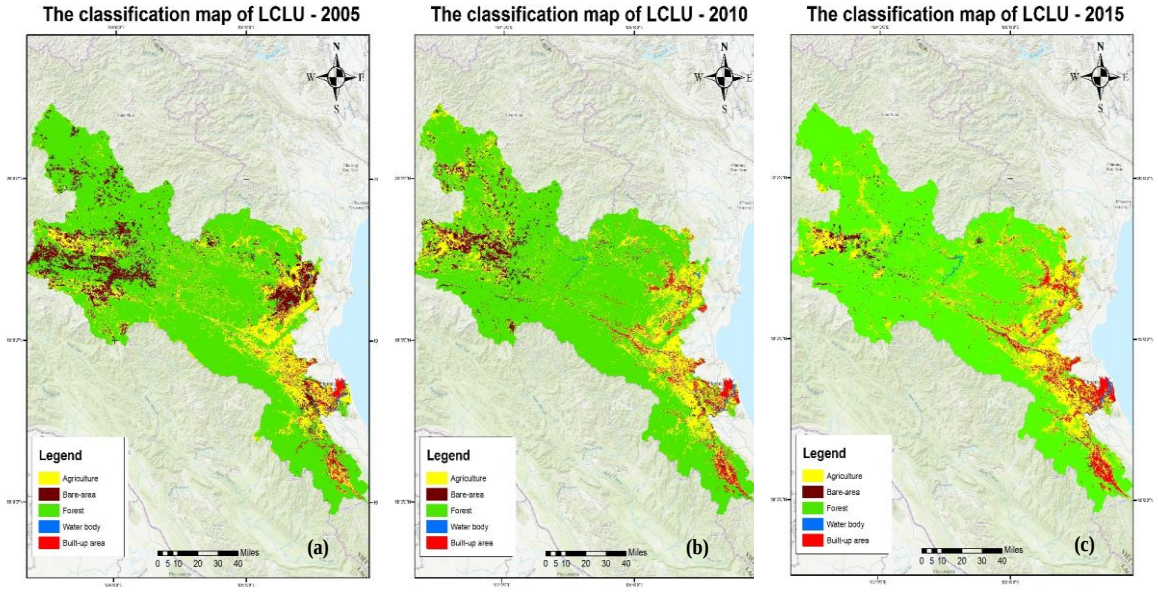


Figure 6. The classification map of LULC (a) 2005, (b) 2010, (c) 2015.

5.2. Landuse/landcover classification and analysis

Accuracy assessment has been used to evaluate the accuracy of classified data. 175 test samples were selected, of which 90 samples were GPS points collected in the field, the rest were randomly selected points from Google Maps. The study calculated and evaluated PA, UA, CA, and kappa index for classification data of 2005, 2010, and 2015. The results are shown in Table 3.

Table 3. Accuracy assessment of the Land use/Land cover classification using the validation dataset PA: Producers Accuracy; UA: Users Accuracy; CA: Classification Accuracy.

Land use/Land cover class	2005		2010		2015	
	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)
Agriculture	68.40	74.28	77.77	80.00	83.33	85.71
Bare area	74.19	65.71	84.85	80.00	86.11	88.57
Forest	78.38	82.86	80.00	91.43	91.67	94.29
Waterbody	87.88	82.86	94.44	97.14	100	100
Built–up	77.78	80.00	93.33	80.00	93.75	85.71
Overall CA (%)	77.14		85.71		90.86	
Kappa index	0.7143		0.8214		0.8857	

Accordingly, the PA and UA of each class are greater than 75% for the classification data of 2010 and 2015. Particularly for 2005, the PA of agriculture class is only 68.40%, the UA of bare area class is only 65.71%, however, PA and UA of other classes still reach over 75% for all 3 years. Overall Classification Accuracy of classification data for 2005, 2010 and 2015 is 77.14%, 85.71%, 90.86%, respectively. The Kappa index for 2005 data is 0.71, 2010 is 0.82, 2015 is 0.88. The CA and Kappa index of 2005 was relatively lower than other years can be explained by the fact that test data was collected in 2018, so there are certain differences compared to 2005.

5.3. Generation of Transition Probability Matrix (TPM), Transition Area Matrix (TAM)

a) TPM

A transition probabilities matrix determines the likelihood that a cell or pixel will move from a land–use category or class to every other category from date 1 to date 2. This matrix is the result of cross–tabulation of the two images adjusted by the proportional error and is translated in a set of probability images, one for each land–use class [12]. As mentioned above, TerrSet software is one of the best platforms to conduct CA–Markov model, which is developed by Clark Labs in the U.S. Hence, transition probabilities matrix are built from the land–use/land–cover images of 2005–2010 and 2010–2015 by ArcGIS and TerrSet software (Table 4).

Table 4. Transition probability matrix of 2005–2010 and 2010–2015 periods (%).

		Agriculture (%)	Bare Area (%)	Forest (%)	Water Body (%)	Built–up (%)
2005–2010	Agriculture	42.98	7.14	38.58	1.76	9.54
	Bare Area	35.74	17.35	43.15	0.39	3.37
	Forest	22.53	5.97	70.52	0.37	0.61
	Water Body	24.07	0.63	7.92	63.22	4.15
	Built–up	18.75	3.15	5.71	1.96	70.43
2010–2015	Agriculture	63.89	4.26	21.69	1.20	8.96

	Agriculture (%)	Bare Area (%)	Forest (%)	Water Body (%)	Built–up (%)
Bare Area	25.76	66.40	4.70	0.38	2.75
Forest	24.81	3.34	70.76	0.47	0.63
Water Body	15.00	0	0	85.00	0
Built–up	0	0	0.01	15.97	84.02

Table 4 shows that the transition probability of agriculture and bare–area is higher than forest, built–up, and water–body in the 2005–2010 period. Built–up and water body has only little probability to change to another type of land cover, about 15% in the 2010–2015 period.

b) TAM

A transition area matrix that records the number of cells or pixels that are expected to change from each land–use class to each other land–use class over the next period. This matrix is produced by the multiplication of each column in the transition probability matrix by the number of cells of corresponding land use in the later image [12]. Overlapping of land cover maps in 2005–2010 and 2010–2015 (Figure 6). Set the time interval between two maps to be five years, the proportional error to 0.15 in case of Maximum Likelihood Classification. The transition area matrix is presented in (Table 5).

Table 5. Transition area matrix of 2005–2010 and 2010–2015 periods (pixel).

		Agriculture (pixel)	Bare Area (pixel)	Forest (pixel)	Water Body (pixel)	Built–up (pixel)
2005–2010	Agriculture	3,511,522	583,006	3,151,858	144,081	779,401
	Bare Area	744,148	361,252	898,504	8,171	70,201
	Forest	4,778,388	1,265,217	14,955,883	78,376	129,558
	Water Body	95,896	2,527	31,554	251,882	16,539
	Built–up	309,583	51,943	94,306	32,381	1,162,909
2010–2015	Agriculture	5,974,382	398,421	2,028,508	112,114	837,973
	Bare Area	579,574	1,494,112	105,851	8,601	61,880
	Forest	4,479,663	602,590	12,778,159	84,471	114,425
	Water Body	81,596	0	0	462,369	0
	Built–up	19	56	169	358,662	1,886,814

Table 5 shows a clearer view of how many pixels have changed from a class to another class. Specifically, the number of pixels in the forest class has the greatest change in both phases, followed immediately by the agriculture class. The class that has the least change is the water body.

5.4. Suitability Map

Suitability maps present the probability of suitability of a pixel belonging to the corresponding LULC class. They range from 0 to 255 with 255 being the most likely and 0 is unlikely. Each suitability map is created by transition rules that are formed by the linkage between socioeconomic, ecological, and spatial variations (e.g. built–up tends to develop near the road). Besides, there are also restrictions on each type of LULC class (e.g. forest areas are planned for conservation). Therefore, factors and constraints are two driving forces of change that determine which lands to be considered for further development.

In this study, slope, digital elevation model (DEM), distance to water bodies, distance to main roads were set as driving factors, but also there were constraints considered (e.g. water and built–up represented constraints for transition to bare–area). They were chosen because of the similarity in their use in many previous studies such as [5, 12, 15, 28, 44–47] and the author's knowledge of the study area. The constraints and factors were standardized into a Boolean (0 and 1) character and a continuous scale of suitability from 0 (least suitable) to 255 (most suitable), respectively. To do that, three types of fuzzy membership function (linear, sigmoidal, and J–shaped) and control points were determined as a necessity to measure the scale of potential suitability for each class (Table 6). Selection of the type of fuzzy membership function and control points is prone to subjectivity and can change according to the knowledge of decision–makers [15]. A fuzzy set theory can be found at [48–50].

Table 6. Standardization of factors by Fuzzy module.

Class	Factors	Functions	Control Points
Agriculture Built–up	Slope	J–shaped	0 degree highest suitability
			0–20 degree decreasing suitability
			>20 degree no suitability
	DEM	J–shaped	0 m highest suitability
			0–350 m decreasing suitability (Agriculture)
			0–150 m decreasing suitability (Built–up)
			> 350 m no suitability (Agriculture)
			> 150 m no suitability (Built–up)
	Distance to rivers	Sigmoidal	< 1.5 km highest suitability
			1.5–5.5 km decreasing suitability
			> 5.5 km no suitability
	Distance to main roads	J–shaped	< 0.2 km highest suitability
			0.2–5 km decreasing suitability
			> 5 km no suitability
Waterbody	Slope	J–shaped	0 degree highest suitability
			0–15 degree decreasing suitability
			> 15 degree no suitability
	DEM	J–shaped	0 m highest suitability
			0–300 m decreasing suitability
			> 300 m no suitability
	Distance to rivers	Sigmoidal	< 1 km highest suitability
			1–5 km decreasing suitability
			> 5 km no suitability
Forest Bare area	Slope	Sigmoidal	< 5 degree no suitability (Forest)
			< 20 degree no suitability (Bare)
			5–18 degree increasing suitability (Forest)
			20–40 degree increasing suitability (Bare)
			> 18 degree highest suitability (Forest)
			> 40 degree highest suitability (Bare)

Class	Factors	Functions	Control Points
	DEM	Sigmoidal	< 150 m no suitability (Forest)
			< 1300 m no suitability (Bare)
			150–700 m increasing suitability (Forest)
			1300–1700 m increasing suitability (Bare)
			> 700 m highest suitability (Forest)
			> 1700 m no suitability (Bare)
	Distance to main roads	Sigmoidal	< 1 km no suitability
			1–10 km increasing suitability
			> 10 km highest suitability

Analytical hierarchy process and pairwise comparison were then applied to develop a set of relative weights for a group of factors in a multi–criteria evaluation. The weights were developed by providing a series of pairwise comparisons of the relative importance of factors to the suitability of pixels for the activity being evaluated. These pairwise comparisons were then analyzed to produce a set of weights that sum to 1 [42]. The procedure by which the weights are produced follows the logic developed by [51–52]. The larger weight denoted a more important criterion in terms of overall factors (Table 7).

Table 7. Factors and their weights used in the construction of suitability maps.

Factors	Forest	Agriculture	Built–up	Water body	Bare area
Slope	0.5917	0.1740	0.5232	0.3874	0.1571
DEM	0.3332	0.2696	0.2976	0.1692	0.2493
Distance to main roads	0.0751	0.0795	0.1222		0.5936
Distance to rivers		0.4768	0.0570	0.4434	
Consistency ratio	0.01	0.02	0.03	0.02	0.05

Then, the Multi–Criteria Evaluation (MCE) module was used to make decisions which is a choice between alternatives. In an MCE, an attempt is made to combine a set of criteria to achieve a single composite basis for a decision according to a specific objective. Through a Multi–Criteria Evaluation, these criteria images representing suitability may be combined to form a single suitability map from which the final choice will be made [42]. Weighted linear combination (WLC) methods are used to include both weighted factors and constraints by the logical AND operation. The intersection of all the criteria leads to obtaining suitable areas for a specific LULC class (Figure 7).

5.5. Simulation of the land use/land cover

The integrated model of Cellular Automata and Markov models can predict LULC changes based on two–time intervals. Therefore, the transition probabilities for the period 2005–2010 along with the basis LULC 2010 were used to simulate LULC in 2015. Each pixel of each LULC type was attributed future suitability by the suitability map for each LULC class. In addition, a standard 5×5 boolean mask filter was used to analyze the neighborhood definition that the suitability weight of the pixels will decrease far from the existing areas and allocate preference to the neighboring suitable areas [41]. The number of iterations i is also the number of time steps that will be used in the model. Choosing this number is also one of the elements that will influence the model's expected results. Indeed, to reach the optimal parameters, the number of iterations was examined in [53]. In the case of this study, the number of iterations was tested as 5, 10, and 20 (Figure 8).

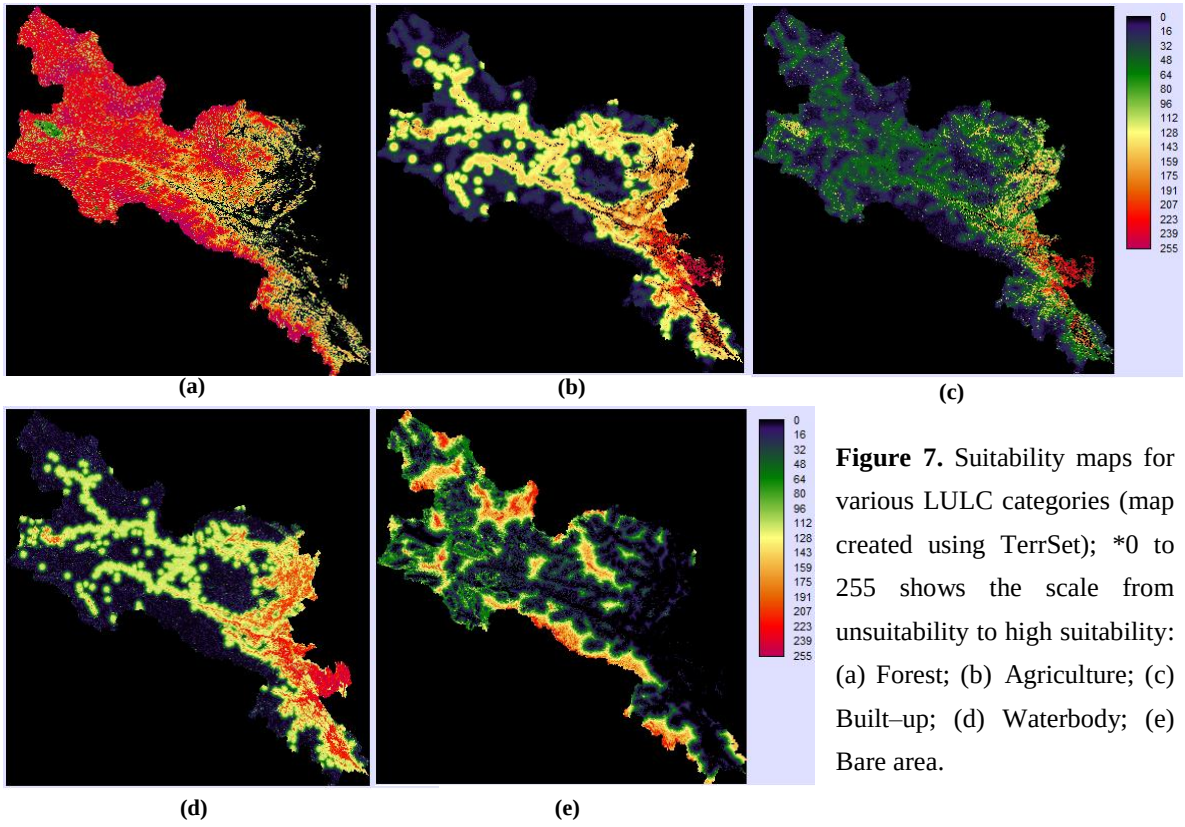


Figure 7. Suitability maps for various LULC categories (map created using TerrSet); *0 to 255 shows the scale from unsuitability to high suitability: (a) Forest; (b) Agriculture; (c) Built–up; (d) Waterbody; (e) Bare area.

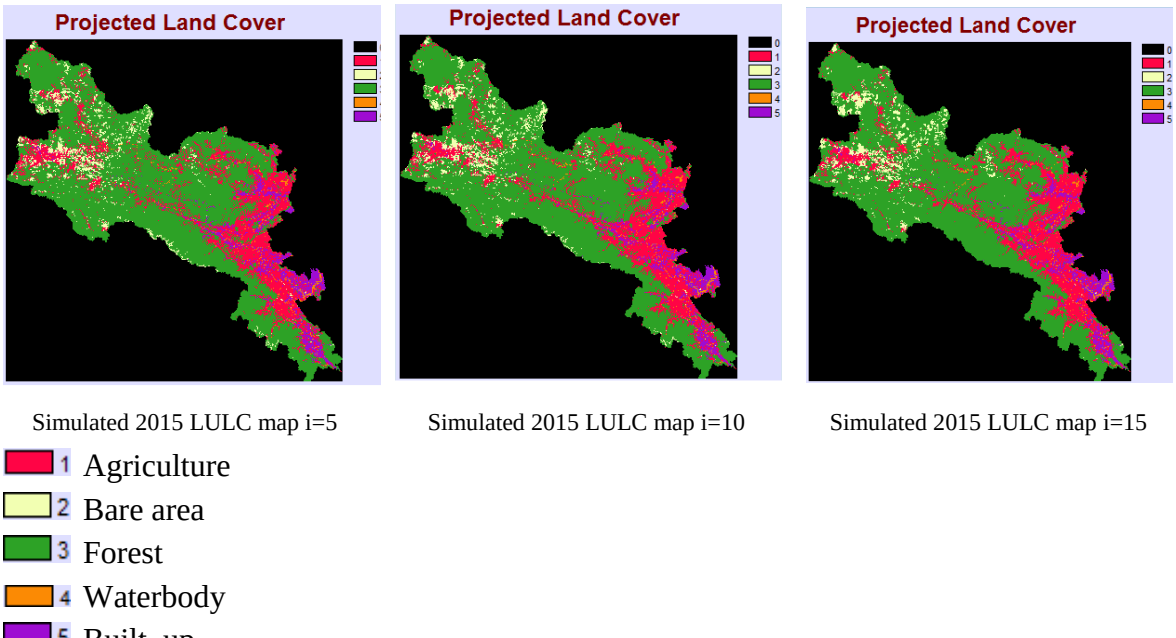


Figure 8. Simulated LULC map of 2015 at different number of iterations through CA–Markov.

5.6. Model validation

Model validation is always an important part to verify and evaluate the accuracy of a model. Nevertheless, there are no consolidation criteria for assessing the feasibility of land change models [54]. To quantify the proficiency of the model, we need to compare the predicted result of the model with a similar and reliable map using the Kappa coefficient [55]. But [54, 56] proved that standard Kappa (Cohen's Kappa) offers almost no useful information because it confounds quantification error with location error. Hence, in addition to Kappa standard ($K_{standard}$), different components of the Kappa index including the Kappa

for no information (K_{no}), Kappa for grid–cell level location ($K_{location}$), and Kappa for stratum–level location ($K_{locationStrata}$) were used to supplement the deficiencies [54]. In short, the simulated 2015 LULC map was validated with the classified 2015 map, and the results are shown below (Table 8).

Table 8. Summary of Kappa statistics for the models on validation data (2015).

Kappa Indices	2015 LULC (i = 5)	2015 LULC (i = 10)	2015 LULC (i = 20)
K_{no}	0.9507	0.9349	0.9119
$K_{location}$	0.9178	0.8887	0.8451
$K_{locationStrata}$	0.9178	0.8887	0.8451
$K_{standard}$	0.9156	0.8865	0.8420

[57] claimed that associations between two variables that both rely on coding schemes with $K < 0.7$ is often impossible and said that content analysis researchers generally think of $K > 0.8$ as good reliability, with $0.67 < K < 0.8$ allowing tentative conclusions to be drawn. Therefore, the simulation provided valid results, then the calibrated model could be applied for the prediction of future patterns – 2030 LULC map.

5.7. Future land use/land cover modeling

After calibration, the CA–Markov model has proven its viability in performing future LULC simulations, 2030. Therefore, this model has continued to be used with parameters that have been demonstrated to be accurate in study area conditions – the Ca river basin – as factors along with its weights and constraints, number of iterations, etc. However, the difference was the input data: 1) satellite–derived LULC maps for 2010–2015 were used to project the LULC for 2030; 2) constituents that generated factors such as main roads, rivers were updated until 2015. The predicted LULC map of 2030 is shown in Figure 9.

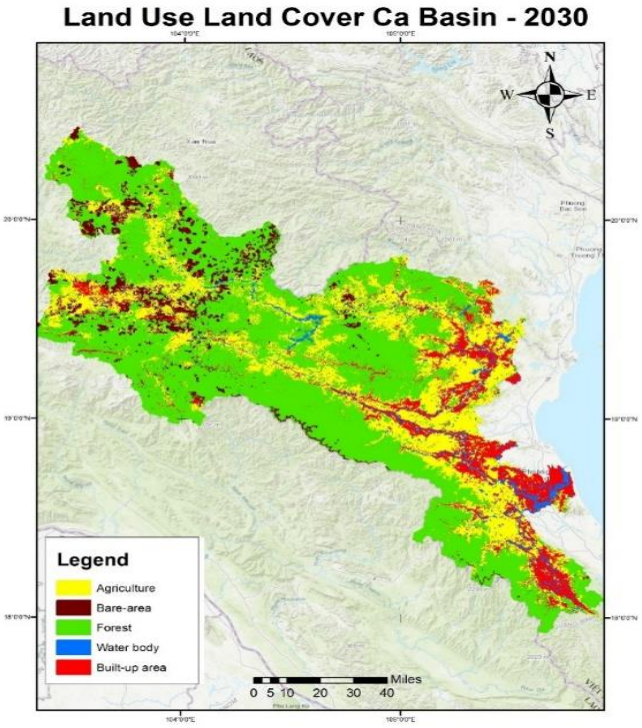


Figure 9. Predicted LULC map of 2030.

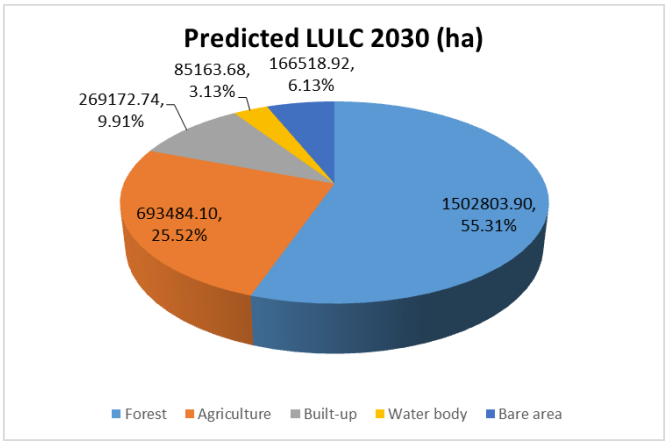


Figure 10. Area (ha) and percentage distribution (%) for each LULC class in 2030.

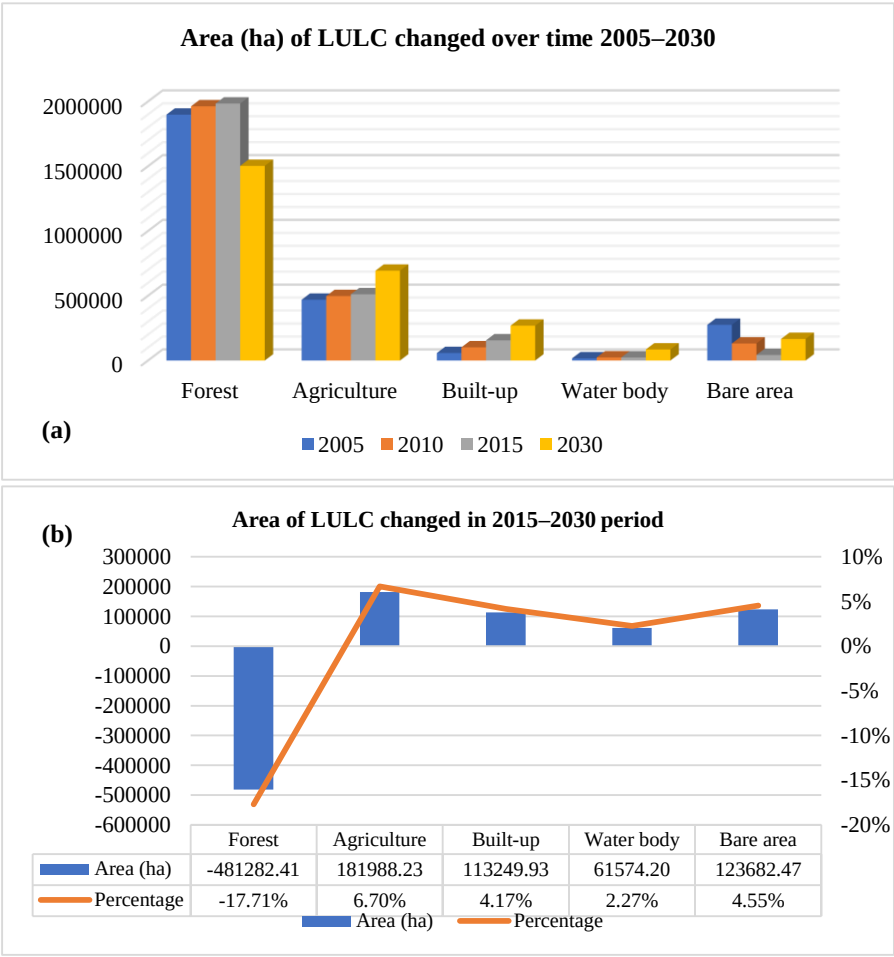


Figure 11. The predicted amount of change from a) 2005 to 2030; b) 2015 to 2030.

Figure 9 demonstrates the spatial distribution of each LULC class in 2030 that predicted by the CA_Markov model. Figure 10 is a quantitative figure for the number of hectares of each type of land: forest, agriculture, built–up, waterbody, bare area corresponding to 55%, 26%, 10%, 3%, 6% of the total land in 2030. We can also easily see the increase of bare areas in the Western uplands by 2030 compared to previous years. In addition, forest areas are projected to decline sharply, especially in areas close to agricultural land and built–up land. Agricultural land in 2030 increased compared to 2005 (Figure 11a). Specifically, in Figure

11b, bare land increased by 4.55%, waterbody increased by 2.27%, built–up land increased by 4.17%, agricultural land increased by 6.70% between 2015 and 2030. The area of forest land has increased slightly in the period of 2005–2010, but in the years after that figure kept going down, especially from 2015 to 2030, the area of forest land decreased by 17.71%. In total, hectares were projected to change from 2015 to 2030 were 961777.24 ha, equivalent to 35.39% of the total studied land area.

6. Conclusion

This study demonstrates the feasibility of the Markov Chain and Cellular Automata approach for modeling the LULC in the Ca river basin, Viet Nam. The research process also clarified the model test and evaluation options, resulting in a calibrated model with appropriate parameters for the conditions of the study area. Validation results with Kappa coefficients of K_{no} = 0.95, $K_{location}$ = 0.91, $K_{locationStrata}$ = 0.91, $K_{standard}$ = 0.91 showing strong agreement between satellite–derived and simulated LULC maps also denote that the model has good reliability.

The results of this study reveal that LULC in the Ca river basin has been, and will continue to change. Specifically, by 2030 the area of forest land will be reduced 17.71% and transformed into other types of land such as agricultural land 6.70%, construction 4.17%, and vacant land 4.55%. This also accurately reflects current socio–economic development trends: urbanization, agricultural land expansion, deforestation, etc. The change will increase the pressure on other natural factors such as soil status, water resources in this area. This still requires further research, but knowing the specific quantitative numbers of the area and type of LULC that will change will greatly assist in the process of assessing the impacts of LULC on other natural factors.

Author contribution: The authors confirm contribution to the paper as follows: study conception and design: B.N.T.; data collection: B.N.T.; analysis and interpretation of results: B.N.T.; draft manuscript preparation: B.N.T., P.D.H. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgements: This research was funded by the Ministry of Natural Resources & Environment (MONRE) of Viet Nam through Project TNMT.2018.02.08 “Study to develop LULC change simulation method and evaluate the impact on water resources for two river basins in North Central Viet Nam”. The authors would like to thank the Department of Business and IT, Department of Health and Environmental Studies, University College of Southeast Norway for providing necessary research facilities to carry out this research work.

Conflict of interest: The authors declare that there are no conflicts of interest.

References

1. Turner, B.L.; Lambin, E.F.; Reenberg, A. The emergence of land change science for global environmental change and sustainability. *Proceedings of the National Academy of Sciences*, **2007**, *104(52)*, 20666–20671.
2. Sciences, Committee on Grand Challenges in Environmental Sciences. Grand challenges in environmental sciences. *National Academy Press*, **2001**.
3. Lambin, E.F. Modeling and monitoring land–cover change processes in tropical regions. *Prog. Phys. Geogr.* **1997**, *21(3)*, 375–393.
4. Kabat, P. The role of biospheric feedbacks in the hydrological cycle; the IGBP–BAHC special issue. *Global Change Newsletter* **1999**, *39*, 1–3.
5. Halmy, M.W.A.; Gessler, P.E.; Hicke, J.A.; Salem, B.B. Land use/land cover change detection and prediction in the north–western coastal desert of Egypt using Markov–CA. *Appl. Geogr.* **2015**, *63*, 101–112.

6. Corner, R.J.; Dewan, A.M.; Chakma, S. Monitoring and prediction of land–use and land–cover (LULC) change. Change. In: Dewan A., Corner R. (eds) Dhaka Megacity. Springer Geography. Springer, Dordrecht, **2014**, 75–97.
7. Rogan, J.; Chen, D. Remote sensing technology for mapping and monitoring land–cover and land–use change. *Prog. Plann.* **2004**, *61(4)*, 301–325.
8. Ahmed, B.; Ahmed, R.; Zhu, X. Evaluation of model validation techniques in land cover dynamics. *ISPRS Int. J. Geo-Inf.* **2013**, *2(3)*, 577–597.
9. Overmars, K.d.; De Koning, G.; Veldkamp, A. Spatial autocorrelation in multi–scale land–use models. *Ecol. Modell.* **2003**, *164(2–3)*, 257–270.
10. Veldkamp, A.; Lambin, E. F. Predicting land–use change. In: *Elsevier,* **2001**.
11. Agarwal, C.; Green, G.M.; Grove, J.M.; Evans, T.P.; Schweik, C.M. A review and assessment of land–use change models: dynamics of space, time, and human choice. Gen. Tech. Rep. NE–297. Newton Square, PA: US Department of Agriculture, Forest Service, Northeastern Research Station, **2002**, *61*, 297.
12. Houet, T.; Hubert–Moy, L. Modeling and projecting land–use and land–cover changes with Cellular Automaton in considering landscape trajectories. *EARSeL eProceedings* **2006**, *5(1)*, 63–76.
13. D BEHERA, M.; Borate, S.N.; Panda, S.N.; Behera, P.R.; Roy, P.S. Modeling and analyzing the watershed dynamics using Cellular Automata (CA)–Markov model–A geo–information–based approach. *J. Earth Syst. Sci.* **2012**, *121(4)*, 1011–1024.
14. Guan, D.; Li, H.; Inohae, T.; Su, W.; Nagaie, T.; Hokao, K. Modeling urban land–use change by the integration of cellular automaton and Markov model. *Ecol. Modell.* **2011**, *222(20–22)*, 3761–3772.
15. Ozturk, D. Urban growth simulation of Atakum (Samsun, Turkey) using cellular automata–Markov chain and multi–layer perceptron–Markov chain models. *Remote Sen.* **2015**, *7(5)*, 5918–5950.
16. Tung, H.T.; Cat, V.M.; Ranzi, R.; Hoa, T.T. Research on medium-term flood forecasting in Ca River basin. *Journal of Water Resources and Environmental Engineering*, **2010**, 28. (In Vietnamese)
17. Ly, N.T.K. Flood Characteristics of Lam River basin. *University of Science - Viet Nam National University*, 2017. In Vietnamese.
18. Giang, P.T. Studying flood characteristics for flood warning in the Lower Lam River basin. (Master Thesis), *Hanoi University of Science – Viet Nam National University*, **2014**. In Vietnamese.
19. Jayawardena, A.; Takahasi, Y.; Tachikawa, Y.; Takeuchi, K. Catalogue of Rivers for Southeast Asia and the Pacific – Volume 6: UNESCO–IHP Regional Steering Committee for Southeast Asia and the Pacific, **2012**.
20. Barsi, J.A.; Lee, K.; Kvaran, G.; Markham, B.L.; Pedelty, J.A. The spectral response of the Landsat–8 operational land imager. *Remote Sen.* **2014**, *6(10)*, 10232–10251.
21. ASTER–GDEM. ASTER GDEM 2 README, **2011**.
22. Olivera, F.; Valenzuela, M.; Srinivasan, R.; Choi, J.; Cho, H.; Koka, S.; Agrawal, A. ARCGIS&SWAT: A geodata model and GIS interface for SWAT 1. *JAWRA J. Am. Water Resour. Assoc.* **2006**, *42(2)*, 295–309.
23. Al–sharif, A.A.; Pradhan, B. Monitoring and predicting land–use change in Tripoli Metropolitan City using an integrated Markov chain and cellular automata models in GIS. *Arabian J. Geosci.* **2014**, *7(10)*, 4291–4301.
24. Foody, G.M.; Campbell, N.; Trodd, N.; Wood, T. Derivation and applications of probabilistic measures of class membership from the maximum–likelihood classification. *Photogramm. Eng. Remote Sen.* **1992**, *58(9)*, 1335–1341.
25. Strahler, A.H. The use of prior probabilities in maximum likelihood classification of remotely sensed data. *Remote Sen. Environ.* **1980**, *10(2)*, 135–163.

26. Manandhar, R.; Odeh, I. O.; Ancev, T. Improving the accuracy of land use and land cover classification of Landsat data using post–classification enhancement. *Remote Sen.* **2009**, *1(3)*, 330–344.
27. López, G.G.I., Hermanns, H.; Katoen, J.P. Beyond memoryless distributions: Model checking semi–Markov chains. In Process Algebra and Probabilistic Methods. Performance Modelling and Verification. Springer, **2001**, 57–70.
28. Araya, Y.H.; Cabral, P. Analysis and modeling of urban land cover change in Setúbal and Sesimbra, Portugal. *Remote Sens.* **2010**, *2(6)*, 1549–1563.
29. Moreno, N.; Wang, F.; Marceau, D. J. Implementation of a dynamic neighborhood in a land–use vector–based cellular automata model. *Comput. Environ. Urban Syst.* **2009**, *33(1)*, 44–54.
30. Balzter, H.; Braun, P.W.; Köhler, W. Cellular automata models for vegetation dynamics. Ecol. Modell. **1998**, *107(2–3)*, 113–125.
31. Al–shalabi, M.; Billa, L.; Pradhan, B.; Mansor, S.; Al–Sharif, A.A. Modeling urban growth evolution and land–use changes using GIS–based cellular automata and SLEUTH models: the case of Sana'a metropolitan city, Yemen. *Environ. Earth Sci.* **2013**, *70(1)*, 425–437.
32. Aronoff, S. Classification accuracy: a user approach. *Photogramm. Eng. Remote Sens.* **1982**, *48(8)*, 1299–1307.
33. Aronoff, S. The minimum accuracy value is an index of classification accuracy. *Photogramm. Eng. Remote Sens.* **1985**, *51(1)*, 99–111.
34. Kalkhan, M.A.; Reich, R.M.; Czaplewski, R.L. Statistical properties of five indices in assessing the accuracy of remotely sensed data using simple random sampling. Paper presented at the Proceedings ACSM/ASPRS Annual Convention and Exposition, **1995**.
35. Koukoulas, S.; Blackburn, G.A. Introducing new indices for accuracy evaluation of classified images representing semi–natural woodland environments. *Photogramm. Eng. Remote Sen.* **2001**, *67(4)*, 499–510.
36. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sen. Environ.* **2002**, *80(1)*, 185–201.
37. Lucas, I.; Janssen, F.; van der Wel, F.J. Accuracy assessment of satellite–derived landcover data: A review. *Photogramm. Eng. Remote Sen.* **1994**, *60(4)*, 479–426.
38. Story, M.; Congalton, R.G. Accuracy assessment: a user's perspective. *Photogramm. Eng. Remote Sens.* **1986**, *52(3)*, 397–399.
39. Smits, P.; Dellepiane, S.; Schowengerdt, R. Quality assessment of image classification algorithms for land–cover mapping: a review and a proposal for a cost–based approach. *Int. J. Remote Sen.* **1999**, *20(8)*, 1461–1486.
40. Campbell, J.B.; Wynne, R.H. Introduction to remote sensing: Guilford Press. **2011**.
41. Lu, Q.; Chang, N.B.; Joyce, J.; Chen, A.S.; Savic, D.A.; Djordjevic, S.; Fu, G. Exploring the potential climate change impact on urban growth in London by a cellular automata–based Markov chain model. *Comput. Environ. Urban Syst.* **2018**, *68*, 121–132. doi:10.1016/j.compenvurbsys.2017.11.006.
42. Eastman, J.R. TerrSet manual. *TerrSet Version* **2015**, *18*, 1–390.
43. Anderson, J.R. Land use and land cover classification system for use with remote sensor data, US Government Printing Office, **1976**, pp. 964.
44. Hadi Memarian, S.K.B.; Talib, J.B.; Sung, C.T.B.; Sood, A.M.; Abbaspour, K. Validation of CA–Markov for Simulation of Land Use and Cover Change in the Langat Basin, Malaysia. *J. Geog. Inf. Sys.* **2012**, *4*, 13. doi:10.4236/jgis.2012.46059.
45. Shafizadeh Moghadam, H.; Helbich, M. Spatiotemporal urbanization processes in the megacity of Mumbai, India: A Markov chains–cellular automata urban growth model. *Appl. Geogr.* **2013**, *40*, 140–149. doi:10.1016/j.apgeog.2013.01.009.

46. Yang, Y.; Zhang, S.; Liu, Y.; Xing, X.; Sherbinin, A. Analyzing historical land–use changes using a Historical Land Use Reconstruction Model: a case study in Zhenlai County, northeastern China. *Sci. Rep.* **2017**, *7*, 41275. doi:10.1038/srep41275.
47. Zabihi, H.; Ahmad, A.; Vogeler, I.; Said, M.N.; Golmohammadi, M.; Golein, B.; Nilashi, M. Land suitability procedure for sustainable citrus planning using the application of the analytical network process approach and GIS. *Comput. Electron. Agric.* **2015,** *117*, 114–126.
48. An, P.; Moon, W.; Rencz, A. Integration of geological, geophysical, and remote sensing data using fuzzy set theory. *Can. J. Explor. Geophys.* **1991**, *27(1)*, 1–11.
49. Luo, X.; Dimitrakopoulos, R. Data–driven fuzzy analysis in quantitative mineral resource assessment. *Comput. Geosci.* **2003**, *29(1)*, 3–13.
50. Ranjbar, H.; Honarmand, M. Integration and analysis of airborne geophysical and ETM+ data for exploration of porphyry type deposits in the Central Iranian Volcanic Belt using fuzzy classification. *Int. J. Remote Sen.* **2004**, *25(21)*, 4729–4741.
51. Saaty, T.L. A scaling method for priorities in hierarchical structures. *J. Math. Psychol.* **1977**, *15(3)*, 234–281.
52. Saaty, T.L. Analytic hierarchy process. In Encyclopedia of operations research and management science. Springer, **2013**, 52–64.
53. Arsanjani, J.J.; Kainz, W.; Mousivand, A.J. Tracking dynamic land–use change using spatially explicit Markov Chain based on cellular automata: the case of Tehran. *Int. J. Image Data Fusion* **2011**, *2(4)*, 329–345.
54. Pontius, R.G. Quantification error versus location error in comparison of categorical maps. *Photogramm. Eng. Remote Sen.* **2000**, *66(8)*, 1011–1016.
55. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **1960**, *20(1)*, 37–46.
56. Pontius Jr, R.G. Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogramm. Eng. Remote Sen.* **2002**, *68(10)*, 1041–1050.
57. Kirppendorff, K. Content analysis: An introduction to its methodology. Beverley Hills: Sage. 1989.

VIETNAM JOURNAL OF HYDROMETEOROLOGY

Research paper

Assessing pesticide fate and transport following modeling approach: A case study of fipronil in the Sakura River watershed, Japan

Le Hoang Tu[1*], Hirozumi Watanabe[2]

[1] Research Center for Climate Change, Nong Lam University Ho Chi Minh city, Ho Chi Minh city, Vietnam; tu.lehoang@hcmuaf.edu.vn
[2] International Environmental and Agricultural Science, Tokyo University of Agriculture and Technology, Tokyo, Japan; pochi@cc.tuat.ac.jp

*****Corresponding author: tu.lehoang@hcmuaf.edu.vn; Tel.: +84–931844631
Received: 5 February 2022; Accepted: 16 March 2022; Published: 25 March 2022

Abstract: Modeling approach has considered as an effective alternative method for environmental risk assessment in recent decades. This work aimed to assess the pesticide fate and transport from rice paddy which has higher potential of pesticide runoff compared to upland fields as reported in previous studies. The study area was the Sakura River watershed, Ibaraki Prefecture, Japan. For modeling rice pesticide, the study applied the PCPF–1@SWAT2012 model. The model was used to simulate concentration of a rice pesticide namely fipronil ($C_{12}H_4Cl_2F_6N_4OS$) in 2009. The simulated streamflow and pesticide concentration were calibrated and validated. The results showed that the maximum pesticide concentrations at the monitored point in the wastershed was 0.008 µg/L in rice paddy cultivation season of 2009. In conclusion, the modeling of the pesitcide was successfully performed in the Sakura River watershed by using the PCPF–1@SWAT2012 model. The fate and transport of the pesticide were assessed. Thus, the modeling can be useful tool for environmental risk assessment.

Keywords: The PCPF–1@SWAT2012 model; Pesticide fate and transport; Rice paddy; Rice pesticide; The Sakura River watershed.

1. Introduction

Rice is main daily meals for nearly half of the world's population especially in Asia [1]. The total global rice consumption is increased from 150 million tons in 1961 to 475 million tons in 2016 and predicted continue to rise in the future [1–2]. Maintaining production of rice is very important task for agriculturists. Due to occurrences of various insects, diseases and weeds, rice farmers have been forced to depend on pesticides [3–4]. However, inappropriate use and management of rice pesticides may adversely affect the aquatic environments. Numerous monitoring studies from Europe and Japan have provided evidence that high pesticide concentrations were usually found in rivers during pesticides application periods of rice cultivation season [5–6]. Because pesticides are applied in the rice paddy where rice is cultivated under the submerged condition, pesticide runoff can occur more frequently via drainage or seepage and percolation [7]. Asian countries produce 90% of rice production in the world [2]. As a result, the aquatic environment of these countries may be at high risk of water contamination due to pesticides loss from rice paddy fields.

Japan is the tenth largest producer of rice in the world. Though pesticides use in Japan has decreased, it is still higher compared to other Asian countries [8]. Some studies reported

that the loss of pesticides from rice paddies is one of the major non–point sources of pesticide pollution of water in streams or rivers in Japan [7, 9–10]. Sakura River watershed is located about 50 km north–east of Tokyo and one of the popularity monitored watershed. Sakura River watershed is an agricultural watershed with 77.6% of the geographical area under forest and agriculture in Ibaraki prefecture, Japan [11]. During the rice cultivation season of the watershed, pesticides loads of the streams which are elevated due to agricultural drainages from rice paddies, have a potential to cause aquatic toxicities. Recent investigations reported that more than 39 kinds of herbicides, insecticides, and fungicides were detected in the watershed [11–12]. Specifically, in 2007 and 2008, concentrations of herbicides such as bromobutide, daimuron, and imazosulfuron were monitored at more than 2 µg/l in early–mid of rice season while simetryn and bentazone were high in mid and late of rice season. The high concentrations of these pesticides may adversely affect aquatic ecosystems by changing water quality and interrupting the aquatic food chain resulting in the loss abundance aquatic species [13]. Due to these reason, pesticides use in the rice production of Sakura River watershed is of great concern. Therefore, the prediction and assessment of their fate and transport in water is required to minimize the adverse impacts in the aquatic environment of the watershed.

In recent decades, computer models have been developed and widely applied in many fields such as graphics, geology, geography, environment and agriculture. For rice paddy, they have become an advantages management tool since the last two decades. Since a rice paddy model in watershed scale has been required for assessing the potential environmental risks in Sakura River watershed, RICEWQ–RIVWQ, PADDY–Large, and PCPF1@SWAT2012 model could be considered as best candidates (REF). However, the RICEWQ–RIVWQ and PADDY–Large algorithms for runoff and pesticide movement have focused only on simulation of paddy hydrology and ignored other types of land uses, which may significantly influence the hydrologic dynamics and pesticide concentrations of river basins [14]. On the other hand, the PCPF–1@SWAT2012 simulates both hydrologic processes and pesticide transports from the watershed at two phases [15]. First, the upland phase controls the amount of surface runoff and pesticide loadings to the main channel from upland fields. Second, the water or routing phase controls the movement of water and pesticide loadings through the channel networks of the watershed into the outlet. Thus, the PCPF–1@SWAT2012 is a more appropriate model for this specific study. Therefore, this study aims to evaluate a rice pesticide transport at the Sakura River watershed by using the PCPF–1@SWAT2012model.

2. Materials and Methods

2.1. Study area

The Sakura River watershed located in Ibaraki Prefecture, Japan (Figure 1). The watershed area is about 335 km^2 and main stream, namely the Sakura River, which flows into Lake Kasumigaura is 53.4 km long [11, 16]. The topography of the watershed is classified into mountain areas in the north, and flat in the west and southeast of the watershed, with average elevation ranging from 8 to 852 m [17]. The land use in the Sakura watershed consists of forest land (32.0%), rice paddies (28.6%), upland agricultural fields (17.0%), residential land (13.9%), and other land use (8.5%) [11, 18]. With respect to soil types, the lower and upper parts of the watershed are mostly Brown forest, Black, and Gray lowland soils while other parts are mostly composed of Gley and Peat soils [19]. The Sakura watershed generally has a temperate climate; with the average annual rainfall of 1,318 mm. The average daily maximum and minimum temperatures are 19.6°C and 10.1°C, respectively [20].

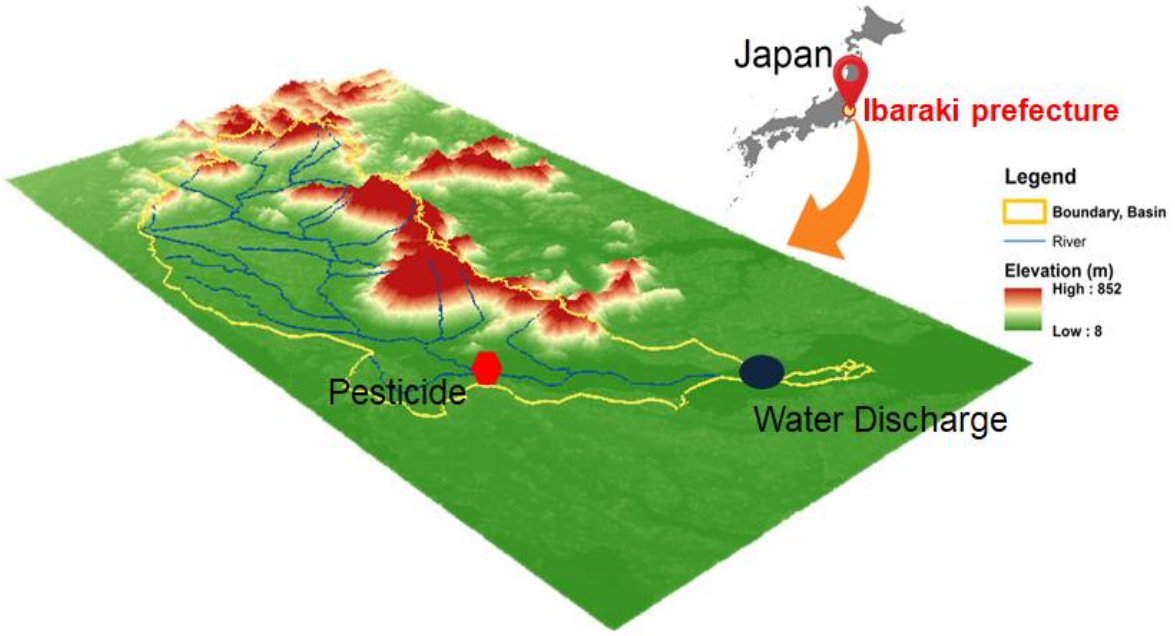


Figure 1. Location and elevation of the Sakura River watershed.

2.2. PCPF–1@SWAT2012 model

2.2.1. Brief model description

PCPF–1@SWAT2012 was updated from the PCPF–1@SWAT which was developed for assessing the impacts of rice pesticides on aquatic environments in watershed scale [21]. Similar to the Soil and Water Assessment Tool (SWAT) model, the PCPF–1@SWAT2012 model also requires topography, land use, soil, weather, crops management practices and pesticide as input data. Figure 2 shows the implementation of Pesticide Concentration in Paddy Field (PCPF–1) model into SWAT model version 2012. In the PCPF–1@SWAT2012 model, rice paddy has been defined as pothole, which is a kind of water bodies for impoundment function in SWAT model. Hence, all performances of the PCPF–1 model are executed inside the pothole of SWAT model. In the PCPF–1@SWAT2012 model, the subbasin can be divided into one or multi hydrologic response units (HRUs). Each subbasin can be set one or multi potholes. When the water is ponding into the pothole, a water balance algorithm is used to calculate the daily amount of runoff. This water balance includes precipitation, water inflow, surface runoff, evapotranspiration, seepage and discharge. In addition, the calculation of water balance components, irrigation process and the pothole variables were redefined. When integrating PCPF–1 model into SWAT model, a procedure to calculate the concentration of pesticide sorbed on sediment was added in the PCPF–1@SWAT2012 model. Because the sediments dissolved in paddy water are not simulated by the PCPF–1 model, pesticide sorbed on soil could not be predicted by overflow [22]. Moreover, recirculation scheme of water was developed. This option aims to calculate the water loss via surface water drainage and tile drainage, which can be collected and re–injected in the field to reduce fresh water requirement.

The PCPF–1@SWAT2012 model was verified and validated in two phases . In the first phase, the pothole algorithms and pesticide mass balance of the model were checked with single and multiple pesticide applications scenarios. The verified results showed that the algorithms used to simulate paddy field water management and pesticide concentrations for single and multiple applications were also correctly implemented into SWAT, and the PCPF–1 was correctly linked to the SWAT model. For the second phase, the

PCPF–1@SWAT2012 model was applied in Sakura River watershed (Ibaraki, Japan) for simulating four herbicide fate and transport. The simulated water flow rate and pesticides concentrations in the Sakura River watershed were good. The model needs to be checked and verified in other watersheds with various pesticides and pollutants.

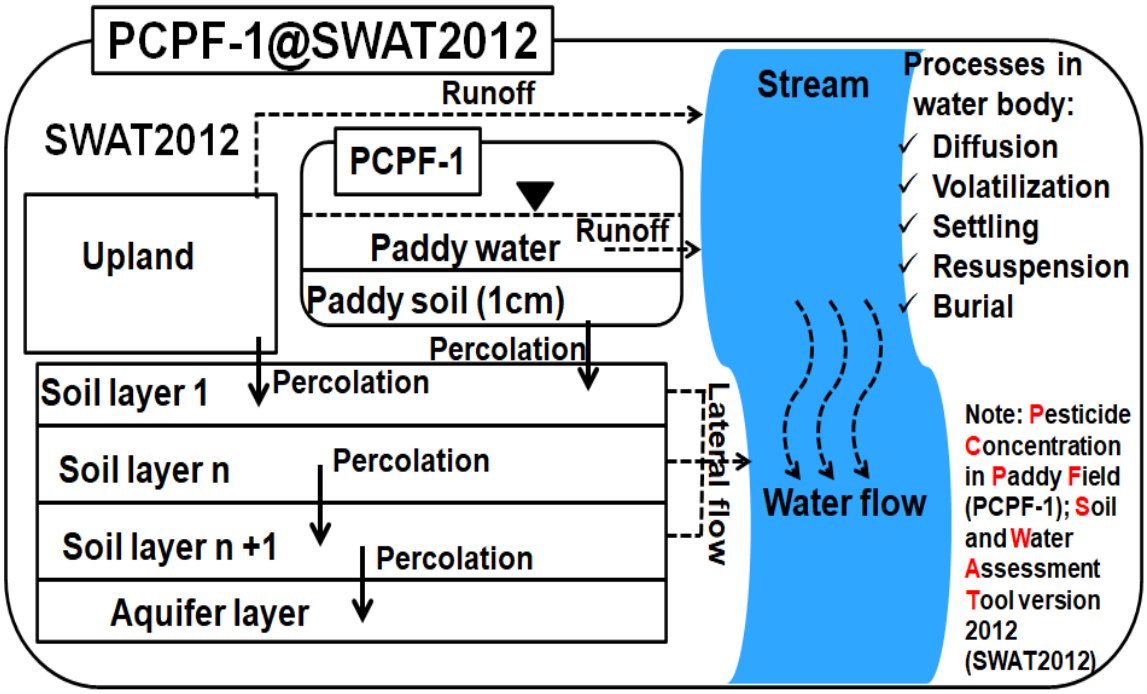


Figure 2. The implementation of PCPF–1 model into SWAT model flowchart [14].

2.2.2. Data collections and processing

The topographic data was obtained from the website of the The Terra Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model (ASTER–GDEM) at resolution of 30 m [17]. Stream network and subbasin boundaries data were downloaded from the National Land numerical information download service [18]. The land use data of the Sakura River watershed used in this study were created in 2008 [18]. The data were downloaded from National Land numerical information download service. The dominant land use types of the watershed were forest (32.01%), paddy fields (28.55%), upland agricultural (17.04%), and residential land (13.92%). Paddy fields predominantly covered the west and south parts of the watershed. Soil types were identified in the catchment based on a 1:25,000 digital cultivated soil data for Ibaraki prefecture in 2007 [19]. The dominant soil types of the lower and upper parts of the Sakura river watershed were mostly Brown forest, Black and Gray lowland soils. The remaining parts of the Sakura river watershed were mostly composed of Gley and Peat soils. Two of the above listed data were provided in the Japan Profile for Geographic Information Standards format (JPGIS) under Geographic Projection (JGD 2000), which need conversion into spatial vector–type GIS (shapefile) and SWAT attribute format under Universal Transverse Mercator Projection (JGD 2000 UTM Zone 54). The observed daily data of precipitation, minimum and maximum temperatures were also collected. Four years weather data (2006–2009) were downloaded from Japan Meteorological Agency– Radar–AMeDAS–analyzed data base [20]. Water flow rates at the outlet were acquired for 2008 and 2009 from the observation data of the Water Information System of the Ministry of Land, Infrastructure and transport, Japan [18].

Regarding pesticide data, the PCPF–1@SWAT2012 model requires two groups of pesticide data including application and pesticide properties. The model demands pesticide

application time, rate, area, and water holding period (WHP) for creating pesticide input table (Table 1). The application rates of the pesticides were obtained based on shipment amount, usage rate, percentage active ingredient and pesticide product information which were extracted from various literatures especially pesticide database of Japan Plant Protection Association (JPPA) [23]. An insecticide namely fipronil ($C_{12}H_4Cl_2F_6N_4OS$) was selected for the model simulation because the required input data of the pesticides were available.

Table 1. Required pesticides application data for writing the pesticide input table of the model.

Parameter	Unit	Definition	Input file
MGT_OP	none	Operation code. MGT_OP=19 for rice pesticide application	.mgt
MONTH/DAY or HUSC	days	Day and month when the rice pesticide is applied in the HRU	.mgt
pcpfipest	none	Integer that identify the pesticide name	.mgt
pst_pcpfkg	g/m^2	Pesticide application rate	.mgt
pcpfarea	%	Percentage of the HRU where the pesticide has been applied	.mgt
pcpfwhp	days	Water holding period	.mgt

2.2.3. The model evaluation

The study used the Nash–Sutcliffe model efficiency coefficient (NSE), the Root Mean Square Error (RMSE) and Percent bias (PBIAS) to evaluate the prediction performance, tendency and model accuracy [24–27]. NSE can range from $-\infty$ to 1 and an NSE of 1 corresponds to a perfect match between estimation and observations. An NSE of 0 indicates that the model estimations are as accurate as the mean of the observed data, whereas an NSE less than zero ($-\infty$ <NSE < 0) occurs when the model prediction of observed mean is not accurate. Similar to NSE, RMSE is also one criterion most widely used for assessment of model output against observed data. The RMSE values can range from 0 to + ∞ with 0 being a perfect prediction. Because NSE is related normalization of the mean squared error (MSE) and RMSE [28] the PBIAS was additionally calculated. The optical value of PBIAS is 0.0, positive and negative PBIAS values indicate model underestimation and overestimation bias, respectively [26, 29].

3. Results and Discussions

3.1. The model calibration and validation

The Sakura River watershed is divided into 36 subbasins based on hydrological characteristics of the watershed. Simulations have been done for the calibration (2008) and validation (2009) of water discharge in daily time step (Figure 3). Since the rice paddy accounted to 28.5% of the Sakura River watershed area, the water management practices in the paddy field have significant effect on the water flow simulation. However, according to [22] no reliable data regarding that kind of activities in the watershed were available. Therefore, the input data related to the water management practices in the paddy field of the watershed were generated based on assumptions, and those data were extracted from the study [14]. The water management input data include (i) the water holding period, which was 7 days after the pesticides application, (ii) the tile flow rate to channel from paddy fields, which was 0.12 cm/day, and (iii) the percolation rate in the paddy fields, which was 1 cm/day. In addition, the study could only calibrate parameters of pesticide simulation in 2009 due to the observed pesticide data limitation. The selected parameters for calibration of the water discharge and pesticide simulations in the Sakura River watershed are shown in Table 2.

Table 2. The calibrated values of the parameters used for the water flow rate simulation.

Parameter	Description	Unit	Calibrated range	Output variables
GWQMN	Threshold depth of water in the shallow aquifer	mm	0–5000	Water discharge
GW_DELAY	Groundwater delay	day	0–500	
GW_REVAP	Groundwater "revap" coefficient	none	0.02–0.2	
REVAPMN	Threshold depth of water in the shallow aquifer for "revap" to occur	mm	0–500	
LAT_TTIME	Lateral flow travel time	day	0–180	
CN2	Initial SCS CN II value	none	–0.2–0.2	
PERCOP	Pesticide percolation coefficient	none	0–1	
CHPST_REA	Pesticide reaction coefficient in reach	1/day	0–0.1	Pesticide
CHPST_KOC	Pesticide partition coefficient between water and sediment in reach	m^3/g	0–0.1	

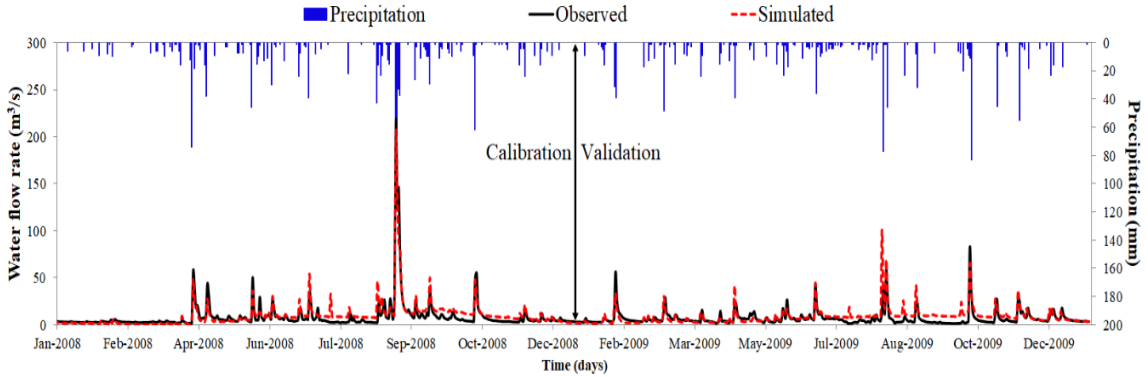


Figure 3. The simulated water flow rate at the outlet of the Sakura River watershed during the 2008–2009 period.

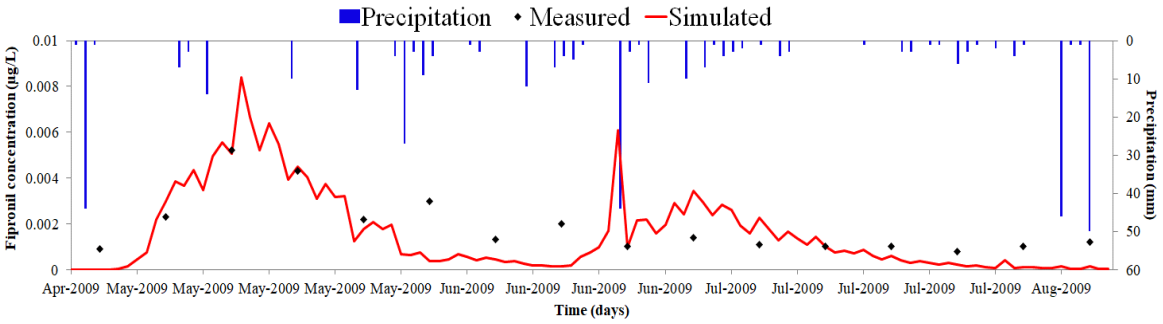


Figure 4. The predicted fipronil concentration for rice season during 2009 at the pesticide monitoring point in the watershed.

Although the simulated baseflow was fluctuated more than that of the observed data, the simulated flow rate showed good response with the observed data and precipitation. Specifically, when the rain came, peak of predicted water flow rate was achieved and vice versa; and the tendency between the predicted flow rate and the observed data was found similar. Table 3 shows the calculated statistical indices for evaluating the model performance of water discharge simulation. The calculated RMSE and PBIAS values, respectively, showed that the predicted water discharge rate had large errors and they were overestimated. Meanwhile, the NSE values indicated that the water discharge rate simulation of the model was good in the calibration year and acceptable in the validation year.

The predicted concentrations of the insecticides were lower compared to the measured insecticides concentrations. The statistical evaluation results of the predicted values of the insecticides concentrations are summarized in Table 3. The NSE and RMSE values were very close to zero. However, the values of PBIAS index indicated that the predicted concentrations of the insecticide in the reach were underestimated possibly because of the insecticide application timing and rate, which did not match with the corresponding actual application timing and rate. In conclusion, the predicted insecticides concentrations were found acceptable.

Table 3. The computed values of model evaluation indices values for the simulated concentrations of the two insecticides.

Variables	Period	NSE	RMSE	PBIAS
Water discharge	Calibration	0.85	6.52	–15.68
	Validation	0.22	7.03	–26.24
Fipronil	Calibration	0.19	0.001	19.12

3.2. Assessment of Fipronil transport

In the Sakura River watershed, the paddy fields are allocated a long with the river in low–land area. So, the applied pesticides in paddy fields are likely to spread to the surrounding aquatic environment. The simulated fipronil concentrations at the monitoring point are displayed in figures 4. The maximum values of the predicted and monitored insecticides concentrations were 0.008 and 0.005 μg/L in 2009 for fipronil. Since insecticides are applied to protect rice against insects throughout the whole growing season, their concentrations in reach increased two times in the 2009 rice season. However, the predicted insecticides concentrations occurred at the beginning of May, rose up in the middle of May, and then decreased.

The pesticides concentrations, which were simulated by the PCPF–1@SWAT2012 model, showed that the rice treated area, application timing, rate and water solubility have strongly affected the prediction. Specifically, the peaks of the insecticides in the middle of May were probably due to nursery–box application upon transplanting. Although fipronil had large rice treated area its peak concentrations of fipronil were low. That can be explained by the application rate of 0.0101 kg/ha, and water solubility values of 3.78 mg/L for fipronil. In addition, the simulated concentrations of the insecticides were low might because of the high Koc (803 ml/g for fipronil) and expected to be mainly applied following the pest forecasting. In other words, the differences in the application rates and methods for the insecticides probably explain why they were detected and simulated at low concentrations. In addition, the pesticide transport was associated with rainfall. When ranfaill eccexed certain amount, it caused loss of the applied pesticide from rice paddy. On the other hand, high rainfall also diluted the pesticide concentration in water bodies.

4. Conclusions

The PCPF–1@SWAT2012 model was applied for predicting transport of a rice insecticide namely fipronil in the Sakura River watershed during 2008–2009 period. The model simulated the observed data with acceptable tendency. The fipronil concentration was increase during rice seasons in the watershed. Strong relationship existed between the increase in the simulated pesticide concentration in the rivers and pesticide application timing and rainfall. However, the model needs to be verified with other pesticides in this watershed as well as in other watersheds. Furthermore, to improve the model accuracy,

detailed information regarding water management and pesticide use in the watershed are required.

Author contribution statement: Collected data: L.H.T., H.W.; Analyzed and processed the data: L.H.T.; Model performing: L.H.T.; wrote the draft manuscript: L.H.T., H.W.; Manuscript editing: L.H.T., H.W.

Acknowledgments: Authors are thankful to MEXT scholarship for support during the study. Authors are also thankful to the Ministry of Economy, Trade, and Industry (METI) of Japan, the United States National Aeronautics and Space Administration (NASA), The Ministry of Land, Infrastructure and Transport (MLIT) of Japan, National Institute for Agro–Environmental Sciences (NIAES) of Japan, Japan Meteorological Agency (JMA) and Japan Plant Protection Association (JPPA) for providing required data for this study.

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

References

1. Mohanty, S.. Trends in global rice consumption. Rice today 2, 2013.
2. FAOSTAT. Global Rice consumption. Food Agric. Organ. 2016. URL <http://www.fao.org/faostat/en/#data> (accessed Apr.10.2017).
3. Gianessi, L. Importance of Pesticides for Growing Rice in South and South East Asia. *Crop. Int.* 2014, pp. 4.
4. Savary, S.; Willocquet, L.; Elazegui, F.A.; Teng, P.S.; Du, P.; Van, Zhu, D.; Tang, Q.; Huang, S.; Lin, X.; Singh, H.M.; Srivastava, R.K. Rice pest constraints in tropical Asia: characterization of injury profiles in relation to production situations. *Am. Phytopathol. Soc.* **2000**, *84*, 341–356.
5. Lamers, M.; Anyusheva, M.; La, N.; Nguyen, V.V.; Streck, T. Pesticide Pollution in Surface– and Groundwater by Paddy Rice Cultivation: A Case Study from Northern Vietnam. *Clean Soil, Air, Water* **2011**, *39*, 356–361.
6. Numabe, A.; Nagahora, S. Estimation of pesticide runoff from paddy fields to rural rivers. *Water Sci. Technol.* **2006**, *53*, 139–146.
7. Nakano, Y.; Miyazaki, A.; Yoshida, T.; Ono, K.; Inoue, T. A study on pesticide runoff from paddy fields to a river in rural region – 1: Field survey of pesticide runoff in the Kozakura River, Japan. *Water Res.* **2004**, *38*, 3017–3022.
8. FAOSTAT. Rice Production Quantity. Food Agric. Organ. 2014. URL <http://www.fao.org/faostat/en/#data/QC> (accessed Apr.10.2017).
9. Inao, K.; Watanabe, H.; Karpouzas, D.G. Simulation Models of Pesticide Fate and Transport in Paddy Environment for Ecological Risk Assessment and Management. *Jpn. Agric. Res. Q.* **2008**, *42*, 13–21.
10. Kawata, K.; Kose, T. Behavior of Pesticides and Their Transformation Products in River Water in Japan. 2012.
11. Iwafune, T.; INao, K.; Horio, T.; Iwasaki, N.; Yokoyama, A.; Nagai, T. Behavior of paddy pesticides and major metabolites in the Sakura River, Ibaraki, Japan. *Pestic. Sci.* **2010**, *35*, 114–123.
12. Vu, S.H.; Watanabe, H.; Ishihara, S. Probabilistic risk assessment through pesticide fate modeling for evaluating management practices to prevent pesticide runoff from paddy fields. In: The 11th IUPAC International Congress on the Chemistry of Crop Protection. Kobe, Japan, 2006, pp. 263.
13. Uddin, H.; Amin, A.K.M.R.; Haque, M.; Islam, A.; Azim, M.E. Impacts of organophosphate pesticide , sumithion on water quality and benthic invertebrates in aquaculture ponds. *Aquac. Reports.* **2016**, *3*, 88–92.

14. Boulange, J.; Watanabe, H.; Inao, K.; Iwafune, T.; Zhang, M.; Luo, Y.; Arnold, J. Development and validation of a basin scale model PCPF–1@SWAT for simulating fate and transport of rice pesticides. *J. Hydrol.* **2014**, *517*, 146–156.
15. Neitsch, S.; Arnold, J.; Kiniry, J.; Williams, J. Soil & Water Assessment Tool Theoretical Documentation Version 2009. Texas Water Resour. Institute, TR–406. 2011.
16. Karpouzas, D.G.; Ribarbelli, C.; Pastori, M.; Capri E. Landscape risk analysis for pesticides applied to rice paddies. *Agron. Sustain. Dev.* **2006**, *26(3)*, 167–177.
17. METI and NASA. The ASTER Global Digital Elevation Model (ASTER GDEM). The Ministry of Economy, Trade, and Industry (METI) of Japan The United States National Aeronautics and Space Administration (NASA). 2012. URL<http://www.jspacesystems.or.jp/ersdac/GDEM/E/4.html> [accessed 16 May 2012].
18. MLIT. Digital national Land Information. Ministry of Land, Infrastructure and Transport, Japan; 2012. URL<http://nlftp.mlit.go.jp/ksj–e/jpgis/jpgis_datalist.html> [accessed 16 May 2012].
19. NIAES. Cultivated soil information system. National Institute for Agro–Environmental Sciences. 2009.
20. Agency, J.M. Japan Meteorological Agency. Japan Meteorological Agency. Japan Meteorological Agency, 2012. URL<http://www.jma.go.jp/jma/index.html> [accessed 28 May 2012].
21. Tu, L.H.; Boulange, J.; Iwafune, T.; Yadav, I.C.; Watanabe, H. Improvement and application of the PCPF–1@SWAT2012 model for predicting pesticide transport: A case study of the Sakura River watershed. *Pest Manage. Sci.* **2018**, 74, 2520–2529. doi:10.1002/ps.4934.
22. Boulange, J. Development and application of the PCPF–1@SWAT model for simulating the fate and transport of rice pesticides in watersheds containing paddy fields. Tokyo University of Agriculture and Technology. 2013.
23. JPPA . Pesticide database. Japan Plant Protection Association, Tokyo, Japan. 2009.
24. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)? –Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250.
25. Gassmann, M.; Stamm, C.; Olsson, O.; Lange, J.; Kümmerer, K.; Weiler, M. Model–based estimation of pesticides and transformation products and their export pathways in a headwater catchment. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 5213–5228.
26. Moriasi, D.N.; Arnold, J.G.; Van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Trans. ASABE.* **2007**, *50*, 885–900.
27. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I – A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290.
28. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **2009**, *377*, 80–91.
29. Gupta, H.V.; Sorooshian, S.; Yapo, P.O. Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration. *J. Hydrol. Eng.* **1999**, *4*, 135–143.

VIETNAM JOURNAL OF HYDROMETEOROLOGY

Research Article

Application of artificial neural network with fine–tuning parameters for forecasting PM2.5 in deep open–pit mines: A case study

Xuan–Nam Bui[1,2*], Hoang Nguyen[1,2], Qui–Thao Le[1,2], Tran Quang Hieu[1,2]

[1] Department of Surface Mining, Mining Faculty, Hanoi University of Mining and Geology, 18 Vien street, Duc Thang ward, Bac Tu Liem district, Hanoi 100000, Vietnam; nguyenhoang@humg.edu.vn (H.N.); lequithao@humg.edu.vn (Q.T.L.); tranquanghieu@humg.edu.vn (T.Q.H.)

[2] Innovations for Sustainable and Responsible Mining (ISRM) Research Group, Hanoi University of Mining and Geology, 18 Vien street, Duc Thang ward, Bac Tu Liem district, Hanoi 100000, Vietnam.

*Corresponding author: buixuannam@humg.edu.vn. Tel.: +84–989583095

Received: 5 February 2022; Accepted: 23 March 2022; Published: 25 March 2022

Abstract: In this paper, an artificial neural network (ANN) model was applied to forecast PM2.5 at the Coc Sau open–pit coal mine (Northern Vietnam) with fine–tuning parameters. It aims to provide the feasibility and insights into controlling air quality in open–pit mines using artificial intelligence techniques. Accordingly, an air quality monitoring system was established to monitor hourly PM2.5 datasets for more than three months. Subsequently, 80% of the whole data was used to design and tune the ANN model, and the remaining 20% was used for testing the PM2.5 predictions. An ANN model with a single hidden layer and ten nodes was developed for this aim. The stochastic gradient descent algorithm was applied to train the ANN model under the learning rate of 0.001 to avoid the overfitting of the model. In addition, 10 time steps (multi–step forecasting model) were applied to forecast the next time step. The results indicated that ANN is a potential model for forecasting PM2.5 in open–pit mines with high accuracy (RMSE = 2.000), and it can be used to control real–time air quality in open–pit mines.

Keywords: Open–pit mine; Air quality controlling system; PM2.5; Artificial neural network; Multi–variate multi–step time series forecasting.

1. Introduction

Surface mining is one of the most popular methods for mineral exploitation; however, its side effects on the surrounding environment are significant. Of the side effects, dust is considered one of the significant concerns in open–pit mines, and it can spread from the surface of open–pit mines to the surrounding areas [1–2]. From the environmental point of view, particular matter (PM) of dust in the atmosphere is an important factor in evaluating the size of dust, as well as its dangerous levels for human health, particularly in the workers in mine sites [3–4]. Of those, PM2.5 is considered more dangerous than PM10 due to its size and the ability to enter the human body, primarily through inhalation [5]. Therefore, this study considered PM2.5 in open–pit mines as the main objective, and the feasibility of artificial neural network (ANN) will be discovered to forecast PM2.5 in open pit mine.

Review of related works shows that artificial intelligence (AI) techniques have been widely applied to forecast air quality and different PM, especially PM2.5. However, studies

with the applications of AI for forecasting PM2.5 in door and out door are popular, but in open–pit mines is very rare. For instance, [6] applied a machine learning model, namely gradient boosting machine (GBM) associated with particle swarm optimization (PSO) to forecast PM2.5 in an open–pit mine with different scenarios. Average accuracy of 0.954 was indicated in their study with the proposed PSO–GBM model. [7] also applied a deep learning algorithm, namely long short–term memory (LSTM) neural network–attention for forecasting PM2.5, and it was then compared to the autoregressive integrated moving average (ARIMA) model. Their results showed that the LSTM–attention model could forecast PM2.5 with better accuracy (from 3–5.6%).

In this study, we tried to discover the feasibility of ANN in forecasting PM2.5 at an open–pit coal mine in Vietnam. The structure of the ANN, as well as its performance, will be diagnosed in this study to interpret whether it is suitable for forecasting and controlling air quality in open–pit mines in Vietnam. It is worth mentioning that the ANN model is developed as the second phase based on the dataset monitored by an air quality monitoring system that was developed by our research group (Innovations for Sustainable and Responsible Mining – ISRM), and they are considered a completed air quality controlling system in open–pit mine.

2. Background of ANN

In this study, ANN is considered and fine–tuned to forecast PM2.5 in deep open–pit mines. The applications of AI in general and ANN have been introduced and developed over the last decade [8–10]. ANN simulates the relationship between input and output variables and recognizes a biological brain function to solve complicated problems in nature [11]. Accordingly, complex non–linear problems can be recognized by the human brain through the input–output mapping in a short time. For simplicity, the computation speed of ANN can be divided into three simplified layers. The first layer is on a mission to gather information, either through observation or other components. They are then transferred to the neurons in the network (the second layer), and herein, they are analyzed and computed. Activation functions may be applied to transfer the information, and finally, in the last layer, the output is computed/ forecasted.

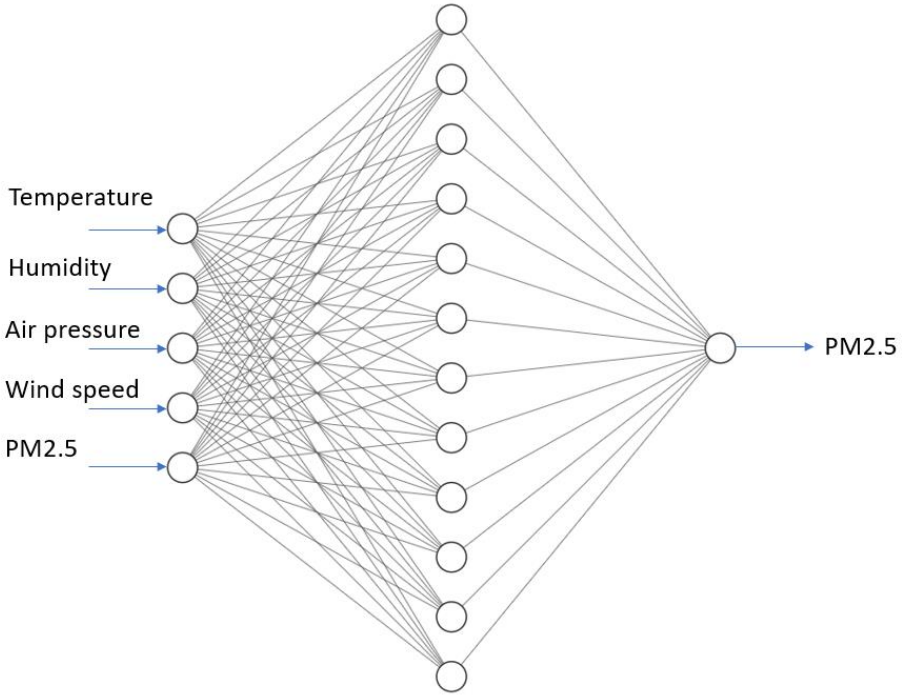


Figure 1. Architecture of the ANN model for forecasting PM2.5.

ANN applies the same logic to these layers. The first layer of the human brain is called the input layer in ANN with the same functions. They can be entered by human from the gathered information. Depending on each problem, the number of dimensions of the dataset may be different. It also depends on the difficulties of the data collection. Once the inputs are entered, they are transferred to the hidden layer (the second layer) to simulate the relationships between inputs and output. Similar to the human brain, ANN can consist of multiple hidden layers. In each hidden layer, a number of neurons (nodes) are designed for the computational functions. Finally, the desired variable based on the analyzed relationships will be predicted. Herein, ANN is used to forecast PM2.5 at an open–pit mine, and its structure is shown in Figure 1.

3. Application

As introduced above, this study intends to apply the ANN model to forecast PM2.5 and evaluate the air quality in open–pit mines. In the application field, a case study will be performed for this aim and the details are presented in the three sub–sections: data collection, design the topology network, and training the model.

3.1. Data collection

For forecasting PM2.5 in deep open–pit mines, the Coc Sau open–pit coal mine (Vietnam) was selected as a case study with a depth of –300 m (below sea level). It is worth noting that the air quality of the bottom is pretty bad at such depth. To collect the dataset, an air quality monitoring system based on the internet of things was designed and developed by the Innovations for Sustainable and Responsible Mining (ISRM) Research Group of the Hanoi University of Mining and Geology (HUMG). This system can monitor nineteen parameters using intelligent sensors and a wind measuring system, including air quality parameters and meteorological conditions. Many researchers indicated that meteorological conditions have significant effects on air quality index [12–14]. They are therefore used to forecast PM2.5 in this study. However, in this study, only five parameters were used for forecasting PM2.5, including temperature (T), humidity (H), air pressure (P), wind speed (WS) and PM2.5, and it is a multivariate dataset. This data represents a multivariate time series of PM2.5–related variables, that in turn could be used to model and even forecast future PM2.5 in open–pit mines. For this aim, the parameters were measured hourly from September 16 to October 28, 2020. Finally, a total of 1011 samples were recorded and compiled as the multi–variate dataset for forecasting hourly PM2.5. It is worth noting that the dataset measured using this monitoring system is the time series dataset and they are hourly observations. The dataset is visualized in Figure 2.

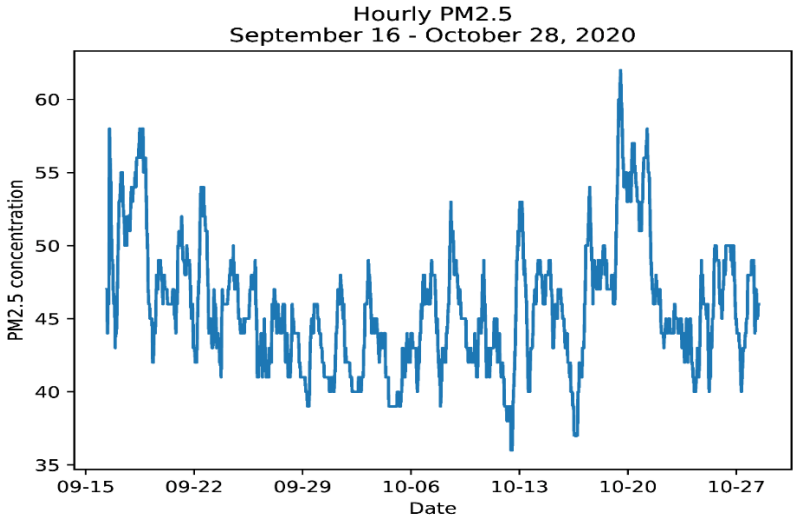


Figure 2. Timeseries dataset of the PM2.5 at the Coc Sau open–pit coal mine.

3.2. Design topology network

Prior to training the ANN model for forecasting PM2.5, a topology network was designed, including the number of hidden layers and the number of hidden nodes. Accordingly, a trial and error procedure was deployed with a different number of hidden layers and nodes through the MSE, which was used as the objective function. The dataset was normalized using the MinMax scaling method and the stochastic gradient descent (SGD) algorithm was used to train the ANN model with different topology networks, as shown in Figures 3 and 4. Please be noted that, due to the stochastic mechanism of the SGD algorithm, the results may be different with different runs.

Analyzing the performance of the ANN model with a different number of hidden layers, we can see that the best number of hidden layers is five hidden layers. However, the performances between different numbers of hidden layers are not too dissimilar. Furthermore, many researchers recommended that the ANN model with a single hidden layer can solve and model most problems [15–16]. Therefore, for simplicity, we selected the topology network with one hidden layer to improve the computational cost.

For the selection of the number of hidden nodes (Figure 4), it is clear that the ANN model with 15 hidden nodes provided the lowest MSE, and therefore, it was selected as the best unit of the ANN model. In other words, the ANN model with the structure of a single hidden layer and 15 nodes was selected for forecasting PM2.5 in this study.

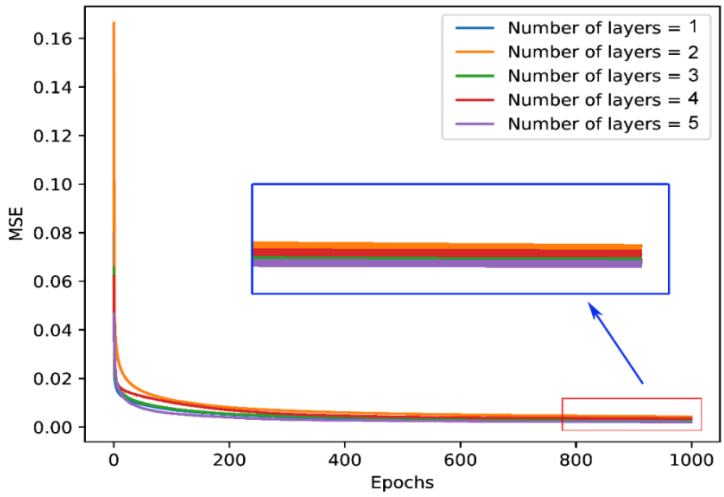


Figure 3. Selection of the optimal number of hidden layers for the ANN model.

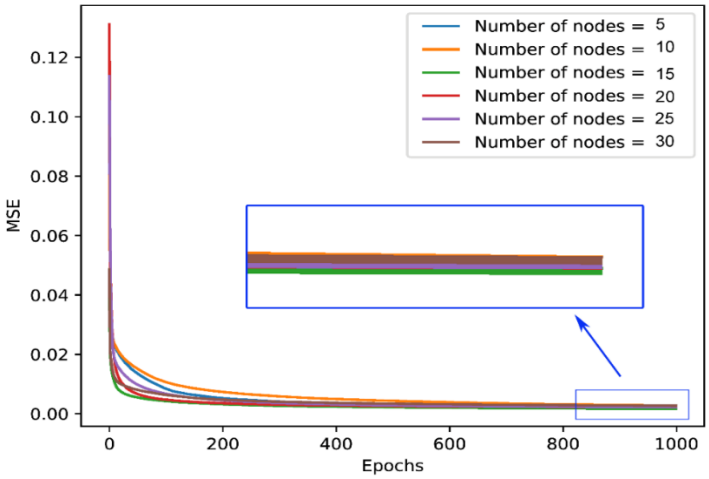


Figure 4. Selection of the optimal number of hidden nodes for the ANN model.

3.3. Training the model

Once the structure of the ANN model was defined, the SGD algorithm with the learning rate of 0.001 was applied to train the ANN model for forecasting PM2.5. The Minmax scaling method was applied to normalize the attributes in the range of 0 to 1. Besides, the selection of the learning rate of 0.001 is to avoid overfitting of the ANN model, and it was determined using the trial and error procedure, as well. It should be noted that 80% of the whole dataset was selected for this task, as shown in Figure 5.

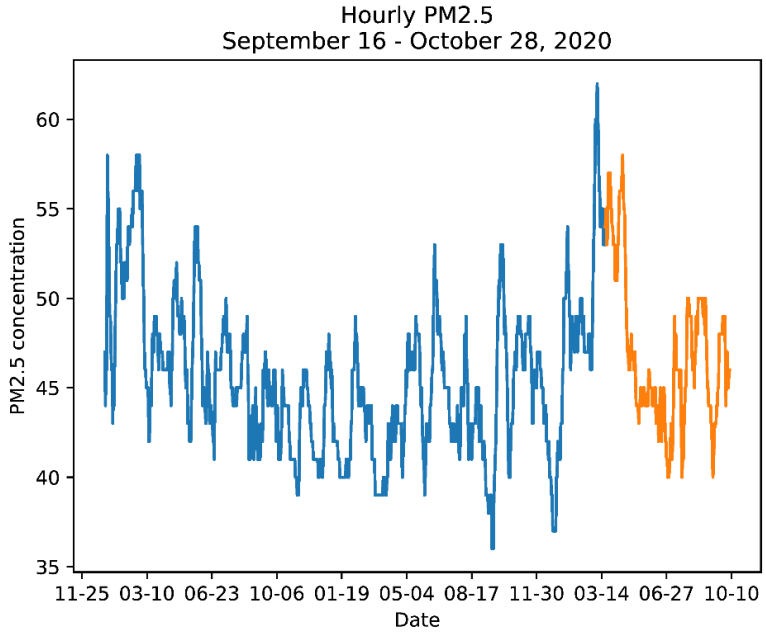


Fig. 5. Splitting dataset for training and testing the model (training dataset – blue line; testing dataset – orange line).

During training the model, to ensure the model's convergence, the training process was implemented with 1000 epochs through the loss function (i.e., MSE). In this study, multi–step time series forecasting model was used to forecast PM2.5 in the future. To do so, the number of time steps were selected as 10 to forcast the next time step. In other words, t–1, t–2, t–3, t–4, t–5, t–6, t–7, t–8, t–9, t–10 values were used to forecast the t value. The training results of the ANN model for forecasting PM2.5 are shown in Figure 6.

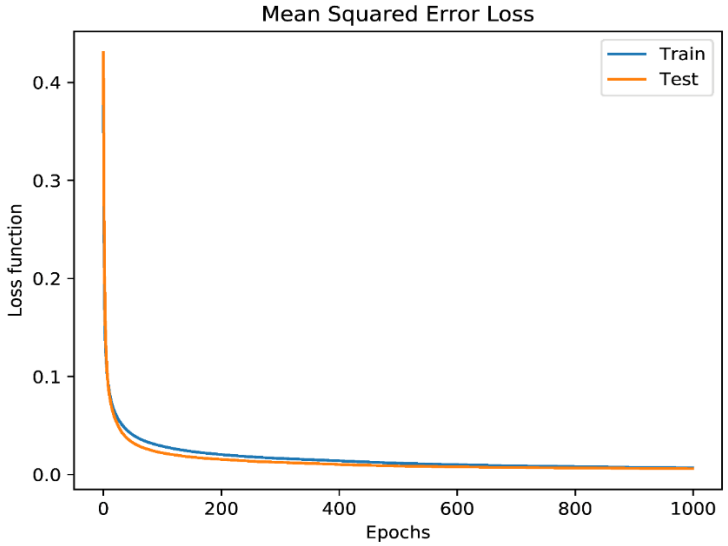


Figure 6. Training performance of the ANN model for forecasting PM2.5.

4. Results and discussion

As depicted in Figure 5, it is conspicuous that the performance curves are excellent, and the model has no overfitting. In other words, the training and testing performances are high convergence with very low lost function values, and the trained ANN model is feasible for forecasting PM2.5 at the Coc Sau open–pit mine.

Once the ANN was well–trained, it was applied to forecast PM2.5 in the testing dataset (out–of–samples) to check the model's goodness. It is worth mentioning that the out–of–samples have not been used before during training the model, as illustrated by the orange line in Figure 5. This step aims at providing the forecast results of PM2.5 in the future (e.g., next hours – hourly PM2.5) by the developed ANN model. The results are shown in Figure 7.

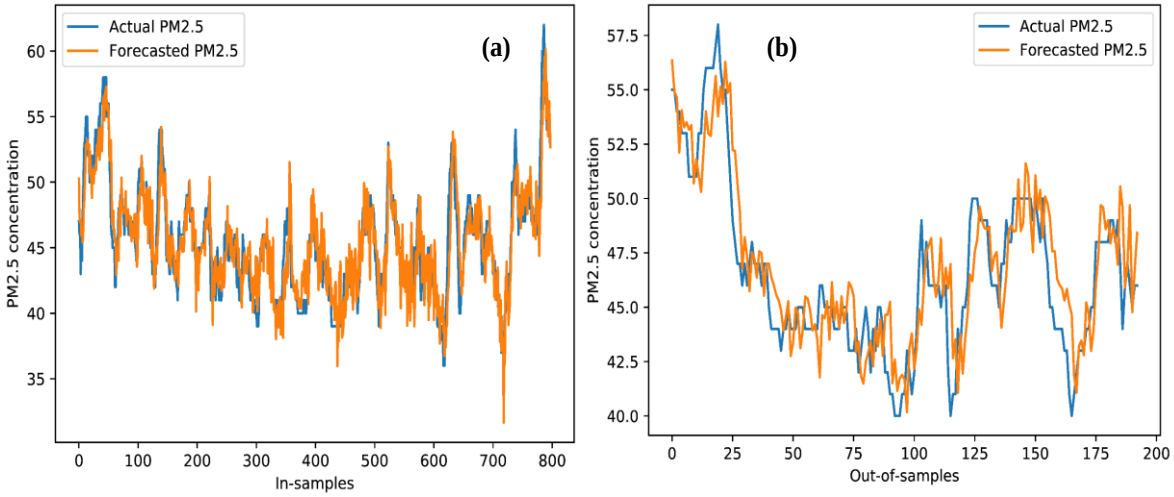


Figure 7. Comparison of the actual and forecasted PM2.5 by the trained ANN model (a) Training dataset; (b) Testing dataset.

As compared between the actual and forecasted PM2.5 values in Figure 7, we can see that the developed ANN model worked very well in both the in–samples and out–of–samples. The trends and forecasted values are pretty close to the actual values, even though at several peak points. On the other hand, it seems that the performance on the training dataset is slightly better than the performance on the testing dataset. Therefore, statistical metrics and regression analysis are necessary to evaluate whether the model is overfitted. Herein, the root mean squared error (RMSE) and mean absolute percentage error (MAPE) were used to consider the error of the model, as calculated using equations (1) and (2). In addition, the correlation analysis is performed in Figure 8 to show how the forecasted PM2.5 is far from the actual PM2.5.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)}{n}} \tag{1}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{2}$$

where y_i is the actual PM2.5; $\hat{y}_i$ is the forecasted PM2.5, and n denotes the number of samples.

Based on the calculations using equations (1) and (2), the accuracy of the ANN model obtained an RMSE of 2.152 and MAPE of 0.037 on the training dataset. Meanwhile, they are 2.000 and 0.035 for the RMSE and MAPE on the testing dataset, respectively. With these statistical metrics, we can claim that the developed ANN model has not been overfitted. In

contrast, its accuracy is very good, with a MAPE of 0.037 only on the training dataset and 0.035 on the testing dataset.

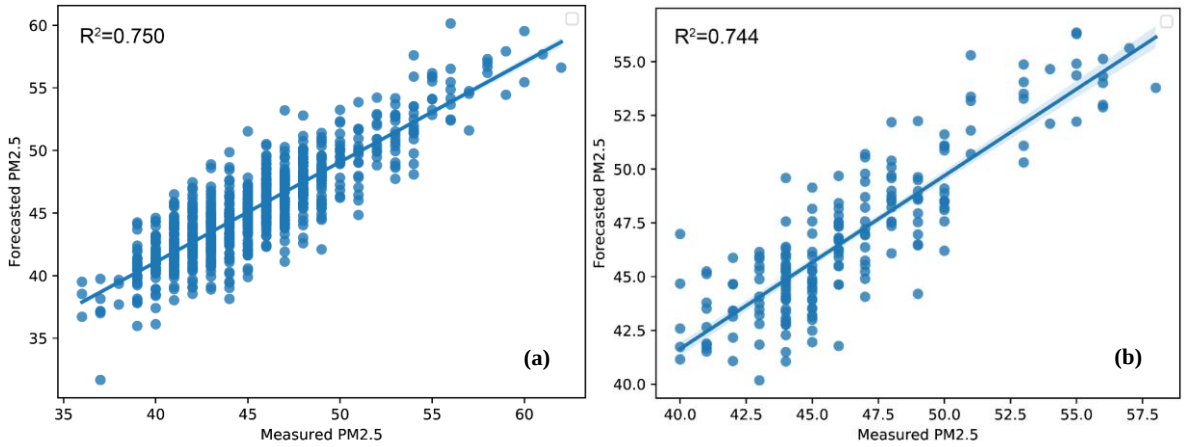


Figure 8. Regression analysis of the forecasted PM2.5 by the ANN model: (a) Training dataset; (b) Testing dataset.

Furthermore, the regression analysis in Figure 8 showed that the dataset used is equivalent suitable to the developed ANN model, although its determination coefficient (R^2) is only 0.75 and 0.744 on the training and testing datasets, respectively. Nonetheless, R^2 is rarely used to evaluate the accuracy of a time–series model [17]. Thus, RMSE, MAPE, and comparison in Figure 7, as well as the curves performance in Figure 6, are conspicuously enough to conclude that the developed ANN model is a good candidate for forecasting PM2.5 in open–pit mines, especially at the Coc Sau open–pit coal mine, as analyzed and discussed in this study.

5. Conclusion

This study discovered the feasibility of the ANN model for forecasting PM2.5 in deep open–pit mines under the consideration of meteorological conditions. The obtained results indicated that ANN is a good approach for forecasting PM2.5 in open–pit mines. By the use of the monitoring system combined with the ANN model, engineers can control the air quality in general, and especially PM2.5 in open–pit mines.

Despite this, further studies in the future are still needed to provide more detailed arguments for the application of AI models to forecast air quality in open–pit mines, as well as improve their accuracy.

Author contribution statement: X.N.B., H.N.: Conceptualization; Investigation; Resources; Writing - Original Draft, Writing - Review & Editing; Q.T.L., T.Q.H.: Formal analysis; Writing - Original Draft; Visualization.

Acknowledgments: This work was financially supported by the Ministry of Education and Training (MOET) in Viet Nam under the grant number B2018–MDA–03SP. The authors also thank the Center for Mining, Electro–Mechanical research of Hanoi University of Mining and Geology, Hanoi, Vietnam; Prof. Changwoo Lee and Dr. Nguyen Van Duc at the Dong–A University, Korea, for kind collaborations in research the grant.

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

References

1. Silvester, S.; Lowndes, I.; Hargreaves, D. A computational study of particulate emissions from an open pit quarry under neutral atmospheric conditions. *Atmos. Environ.* **2009**, *43(40)*, 6415–6424.
2. Alvarado, M.; Gonzalez, F.; Fletcher, A.; Doshi, A. Towards the development of a low cost airborne sensing system to monitor dust particles after blasting at open–pit mine sites. *Sensors* **2015**, *15(8)*, 19667–19687.
3. Ali, M.U.; Liu, G.; Yousaf, B.; Ullah, H.; Abbas, Q.; Munir, M.A.M. A systematic review on global pollution status of particulate matter–associated potential toxic elements and health perspectives in urban environment. *Environ. Geochem. Health* **2019**, *41(3)*, 1131–1162.
4. Qi, C.; Zhou, W.; Lu, X.; Luo, H.; Pham, B.T.; Yaseen, Z.M. Particulate matter concentration from open–cut coal mines: A hybrid machine learning estimation. *Environ. Pollut.* **2020**, *263*, 114517.
5. Shou, Y.; Huang, Y.; Zhu, X.; Liu, C.; Hu, Y.; Wang, H. A review of the possible associations between ambient PM2. 5 exposures and the development of Alzheimer's disease. *Ecotoxicol. Environ. Saf.* **2019**, *174*, 344–352.
6. Lu, X.; Zhou, W.; Qi, C.; Luo, H.; Zhang, D.; Pham, B.T. Prediction into the future: A novel intelligent approach for PM2. 5 forecasting in the ambient air of open–pit mining. *Atmos. Pollut. Res.* **2021**, *12(6)*, 101084.
7. Li, L.; Zhang, R.; Sun, J.; He, Q.; Kong, L.; Liu, X. Monitoring and prediction of dust concentration in an open–pit mine using a deep–learning algorithm. *J. Environ. Health Sci. Eng.* **2021**, *19(1)*, 401–414.
8. Hyder, Z.; Siau, K.; Nah, F. Artificial intelligence, machine learning, and autonomous technologies in mining industry. *J. Database Manage.* **2019**, *30(2)*, 67–79.
9. Ali, D.; Frimpong, S. Artificial intelligence, machine learning and process automation: existing knowledge frontier and way forward for mining sector. *Artif. Intell. Rev.* **2020**, *53(8)*, 6025–6042.
10. Soofastaei, A. The application of artificial intelligence to reduce greenhouse gas emissions in the mining industry. In Green Technologies to Improve the Environment on Earth: IntechOpen London, UK, 2018.
11. Prieto, A.; Prieto, B.; Ortigosa, E.M.; Ros, E.; Pelayo, F.; Ortega, J. et al. Neural networks: An overview of early research, current frameworks and new challenges. *Neurocomputing* **2016**, *214*, 242–268.
12. Zhu, Z.; Qiao, Y.; Liu, Q.; Lin, C.; Dang, E.; Fu, W. et al. The impact of meteorological conditions on Air Quality Index under different urbanization gradients: a case from Taipei. *Environ. Dev. Sustainability* 2021, *23*(3), 3994–4010. doi:10.1007/s10668-020-00753-7.
13. Zhang, Y. Dynamic effect analysis of meteorological conditions on air pollution: A case study from Beijing. *Sci. Total Environ.* **2019**, *684*, 178–185.
14. He, J.; Gong, S.; Yu, Y.; Yu, L.; Wu, L.; Mao, H.; et al. Air pollution characteristics and their relation to meteorological conditions during 2014–2015 in major Chinese cities. *Environ. Pollut.* **2017**, *223*, 484–496.

15. Madhiarasan, M.; Deepa, S. Comparative analysis on hidden neurons estimation in multi layer perceptron neural networks for wind speed forecasting. *Artif. Intell. Rev.* **2017**, *48(4)*, 449–471.
16. Niska, H.; Hiltunen, T.; Karppinen, A.; Ruuskanen, J.; Kolehmainen, M. Evolving the neural network model for forecasting air pollution time series. *Eng. Appl. Artif. Intell.* **2004**, *17(2)*, 159–167.
17. Armstrong, J.S. Evaluating forecasting methods. In Principles of forecasting, Springer, 2001, pp. 443–472.

Table of content