Elsevier required licence: C <2018>. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <u>http://creativecommons.org/licenses/by-nc-nd/4.0/</u> The definitive publisher version is available online at [insert DOI]

Integration of SWAT and QUAL2K for water quality modeling in a data scarce basin of Cau River basin in Vietnam

Abstract. Water quality modeling in a river basin often faces the problem of having a large number of parameters yet limited available data. The important inputs to the water quality model are pollution concentrations and discharge from river tributaries, lateral inflows and related pollution load from different sources along the river. In general, such an extensive data set is rarely available, especially for data scarce basins. This makes water quality modeling more challenging. However, integration of models may be able to fill this data gap. Selection of models should be made based on the data that is available for the river basin. For the case of Cau River basin, the SWAT and QUAL2K models were selected. The outputs of SWAT model for lateral inflows and discharges of ungauged tributaries, and the observed pollutant concentrations data and estimated pollution loads of sub-watersheds were used as inputs to the water quality model QUAL2K. The resulting QUAL2K model was calibrated and validated using recent water quality data for two periods in 2014. Four model performance ratings PBIAS, NSE, RSR and R² were used to evaluate the model results. PBIAS index was chosen for water quality model evaluation because it more adequately accounted for the large uncertainty inherent in water quality data. In term of PBIAS, the calibration and validation results for Cau River water quality model were in the "very good" performance range with |PBIAS|<15%. The obtained results could be used to support water quality management and control in the Cau River basin.

Keywords: Model integration; Water quality modeling; Data scarcity; SWAT; QUAL2K.

1. Introduction

Water quality models have been used extensively in water resources management to improve understanding of the system and support decision making. Computer models simulating water quality have undergone a long period of development since the Streeter and Phelps' water quality model was used to control river pollution in Ohio, US in 1925 (Streeter and Phelps, 1925). Starting with two water quality variables (DO and BOD), water quality models became more and more complex incorporating different transport and transformation processes to simulate eco-hydrological phenomena in rivers. Typical water quality variables simulated in water quality models are total suspended solids (TSS), nitrogen species (organic nitrogen, nitrate/nitrite, and ammonia/ammonium), phosphorus species (organic and mineral phosphorus), dissolved oxygen (DO), chlorophyll a, carbonaceous biochemical oxygen demand (CBOD), pesticides, and metals. In order to simulate water quality in river basin, two types of models were used: (1) watershed models with a focus on terrestrial processes and management options, but with simple hydraulic routing of pollutants in the river, or (2) river water quality models with detailed description of the riverine routing and pollutant transport processes (Debele et al., 2008).

Most watershed models, such as SWAT (Arnold et al., 1994), HSPF (Donigian et al., 1984) and MIKESHE (Refsgaard, 1997) are used to simulate the processes that take place in the upland watershed and streams with simplifying hydrodynamic and water quality assumptions. On other hand, river water quality models, such as QUAL2K, MIKE11, HEC-RAS, etc., limit their scope to solving the hydrodynamic routing and water quality processes in river network, utilizing flow and pollutants from the upland watershed inputs (EPA, 2000). In reality, most water resources management programs involve planning and implementation of a complex network of upland watershed and river systems. Thus, what is needed are models that simulate the processes that water undergoes in terms of quantity and quality in the upland watershed and in downstream river network.

Nowadays, there is a tendency to combine the two different model approaches to build a more realistic scheme of the ecohydrological processes in a watershed (Ambrose et al., 1993; Neitsch et al., 2001; Bicknell et al, 1996; Debele et al., 2008; Chihhao Fan et al., 2009; Hesse et al., 2012). Neitsch et al. (2001) and Bicknell et al. (1996) have linked hydrological and water quality models of an upland watershed and downstream river network in SWAT and HSPF model, respectively. In the SWAT model, the hydrologic and pollutant loads from the upland watershed are simulated in the SWAT model, whereas water quality in the stream-river system (organic and inorganic variables including TSS, TN, TP, DO, CBOD, metals, etc.) is simulated by the enhanced stream water quality model QUAL2E (Brown and Barnwell, 1987). Similarly, HSPF incorporates watershed-scale agricultural runoff and nonpoint source models into a basin-scale analysis framework that includes pollutant transport (organic and inorganic variables including TSS, TN, TP, DO, CBOD, metals, etc.) and transformation in streams. In general, it requires simulating the upland watershed hydrological and water quality variables using a watershed model, and utilizing the outputs from the watershed model as inputs into a hydrodynamic and water quality model appropriate for the river network.

However, water quality modeling in the upland watershed often face a challenge of requiring a large number of parameter values and limited available data. It requires many detailed spatially varied data such as spatial distribution of point and diffuse pollution sources, land-use and land cover maps, and soil property maps of the river basin. In general, such data sets are unlikely to be available and those that are available may have large uncertainties, especially for data scarce basins, typical of many developing countries. Without these data, a watershed model could not properly give good results for water quality in the basin. Nonetheless, most water resources management programs often focus on environmental problems of downstream river network around developing urban areas and economic centers. Thus, at a minimum, a well-performing model for the downstream river network is more often needed than for the whole river basin.

The important inputs for the water quality model for the downstream river network comprise pollution concentrations and discharge from river tributaries and lateral inflows from different sources along river bank. These data may be obtained from observed data and from model simulation results from an upland watershed model. In an ideal case, the use of observed data would be the best form of input for the downstream river network model because an upland watershed model can have large errors depending on the quality of input data. But in data sparse regions, there is unlikely to be enough observed data of discharge and pollution concentration for all river tributaries. Therefore, a better way for constructing a water quality model for a downstream river network is to combine two types of data: observed data and simulation data from upland watershed model. In the case where spatial data of pollution sources is scarce, the water quality outputs of upland watershed model may contain large errors, so the concentrations outputs of an upland watershed model cannot be used as input for the downstream river water quality model. In this case, for better water quality model results, observed water quality data at the outlets of tributaries should be used instead of using water quality outputs from upland watershed model. However, an upland watershed model could give reasonable results for hydrological outputs when discharge data is available. So flow results of the upland watershed model can be used for providing discharge of ungauged river tributaries as input data for a river water quality model in a data scarce river basin.

In this study of the Cau River basin, we used the SWAT model to provide the hydrological outputs from the upland sub-watersheds (river discharge at tributary's outlets). This together with observed water concentrations (water temperature, pH, BOD, DO, COD, SS, metals, ammonium

(NH₄-N), nitrate (NO₃-N), nitrite (NO₂-N), phosphate (PO₄-P), phenol, and coliform) in the upstream river tributaries as input data to the QUAL2K model (Chapra et al., 2012) to simulate the water quality (DO, BOD, COD, total nitrogen (TN) and total phosphorus (TP)) in the downstream river network. The simulation results of water quality of SWAT model cannot be used because of lack of many spatial data for pollution sources in Cau River basin. The models SWAT and QUAL2K were chosen for the study because they are well founded and verified public domain models. It should be noted that, although the latest version of SWAT integrates the QUAL2E water quality model, there is no options to input observed water quality data to river tributaries in SWAT during its simulation process to get proper water quality results in the downstream river network. However, the outputs of SWAT model may be processed and manually input to the QUAL2K model. The QUAL2K is an updated version of QUAL2E and there were some enhancements with respect to water quality simulation; however these are not the focus of this study.

The objective of this paper is to obtain a good water quality model in a data scarce river basin focusing on organic and nutrient pollution including water quality variables such as dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), total nitrogen (TN) and total phosphorus (TP). In the paper, we proposed a new procedure for integrating, in a unique and particular manner, a watershed model (SWAT) and stream-river water quality model (QUAL2K). Further, a new data analytical procedure for estimation of concentrations in lateral inflows (diffuse pollution sources) and missing N and P species in input data was developed based on observed data. Audet (2013) proposed a similar methodology to simulate water quality on the Cau River basin for coupling the hydrological model Hydrotel and in-stream water quality QUAL2E-GIBSI. However, in this work the missing water quality inputs were calculated based on consideration of regional characteristics of three regions (forest lands in mountains, agricultural lands in valleys and urban zones in plains); and the concentrations of diffuse sources were assumed to be equal the concentrations in the river segment in Audet (2103).

2. Cau River basin and data resources

The Cau River basin is situated in North-Eastern Vietnam covering an area of about 6,030 km². The river basin includes the whole Thai Nguyen province and parts of other six provinces. The Cau River is about 288 km long and flows though the provinces of BacKan, Thai Nguyen, Bac Giang and Bac Ninh. Its outlet is at Pha Lai which is the confluence of six different rivers. The main river tributaries of Cau River include Cho Chu River, Nghinh Tuong River, Du River, Cong River, Ca Lo River and Ngu Huyen Khe River (Fig. 1a). The Cau River basin has 68 rivers and streams with lengths longer than 9.0 km and the total length of rivers and streams is about 1,602 km.

Daily rainfall data were acquired from more than 40 stations within and nearby the watershed. Other meteorological data such as temperature, wind speed, relative humidity and solar incidence were available at 5 National Meteorological Stations. Hydrological variables (water level and discharge) in the mainstream were observed at 7 stations, but some of these stations ceased discharge measurement in 1980s. Recent river discharge is available at only in Gia Bay station in the center of river basin (Fig. 1b). The precipitation data indicates that the annual rainfall in the Cau River basin varies from 1400 to 2700 mm (with an average of 1680 mm). According to Nguyen (2013), the total average annual flow of Cau River to Thac Rieng (most upstream) hydrological station is about 564 million m³/year; the average annual flow at the Gia Bay station

is about 1,728 million m³/year and the average annual flow to Cau River's mouth at Pha Lai station is about 5,477 million m³/year. The water resources of Cau River basin are plentiful but varies seasonally. The total flow of five months in the rainy season (from June to October) accounts for 80% - 85% of the total annual flow. Some hydraulic works such as dams and reservoirs were constructed in river basin for flow regulation. Nui Coc reservoir is the biggest reservoir that was built in Cong River. It supplies water for Thai Nguyen City, Song Cong Town and for the agricultural activities in downstream areas. At Thai Nguyen City, the Thac Huong dam was built across the Cau River to raise the water level for irrigation to Bac Giang province. During some months of the dry season, the water level of Thac Huong Dam's upstream is lower than the level of spillway when most water of Cau River flows to irrigation canals and only a small amount of water flows to downstream areas of Thac Huong dam (Fig. 1a).



(a)

(b)

Fig. 1. (a) Rivernetwork and water quality moniroring stations; (b) Hydrological, meteorological and rainfall stations

The land-use/land-cover types of Cau River basin are described in Table 1 and Fig. 2a. The main land-uses in the watershed are forest (51.13%), followed by agriculture (38.11%) and residence (towns, villages, etc.) (7.36%) (IET, 2012).



Fig. 2. (a) Land-use map and (b) Sub-watersheds and river reaches of Cau River basin

Table 1. Percent of land-use/land-cover classification in Cau River basi
--

Land-use/Land-cover	Percent	Land-use/Land-cover	Percent
Residence	7.36	Shrubland	16.8
Water	3.4	Irrigated cropland	19.85
Rice	16.4	Dryland cropland	1.86
Forest	34.33		

Regular monitoring of the river water quality commenced in 2005 and occured four to six times a year for 42 stations by Center for Environmental Monitoring (CEM). The available data for water quality variables are daily average value of the following: water temperature (T), pH, biochemical oxygen demand (BOD), dissolved oxygen (DO), chemical oxygen demand (COD), suspended solids (SS), cadmium (Cd), arsenic (As), lead (Pb), zinc (Zn), manganese (Mn), iron (Fe), ammonium (NH₄-N), nitrate (NO₃-N), nitrite (NO₂-N), phosphate (PO₄-P), phenol, and coliform.

According to the Environment Report of Vietnam in 2006 (MONRE, 2006), the surface water in the Cau River basin is seriously polluted. The major sources of pollution are wastewater discharge from more intense activities during the last decade at craft villages and the newly developed industrial zones. However, due to the implementation of the Vietnam government's master plan for water quality control in Cau River beginning in 2008 based on the Prime Minister's decision (Ha Ngoc Hien et al., 2016), the water quality of Cau River appears to be improving in recent years. Fig. 3 shows the variation of DO, BOD, COD and NH4-N concentrations (seasonally averaged for period 2010-2015) along Cau River mainstream at 19 stations from station 1 to station 19 (Fig. 1a). It can be seen that compared with surface water quality standards QCVN 08-MT:2015 (MONRE, 2015), the water quality of the Cau River is generally below the Grade A2 and the noncomplying parameters are BOD, COD and NH₄-N. The monitoring results indicate that the water quality of Cau River decreases a little from upstream to downstream, then increase clearly after station 10 and especially after station 13 due to discharge of the local wastewater of craft villages along Ngu Huyen Khe tributary in Bac Ninh province. There is still need to improve the water quality in Cau River basin for good environmental health of the river. Therefore, the objective of this study is to implement the QUAL2K model to calibrate and validate the water quality parameters of the Cau River. The model would provide a tool for environmental management and pollution control of the river with the goal of assisting decision-makers to better use water resources and forecast the impending damages caused by socio-economic developments in the Cau River basin.



Fig. 3. Variation of DO, BOD, COD, and NH₄-N along Cau River mainstream (seasonally averaged for period 2010-2015)

3. Methodology

One of the most difficult yet important issue in water quality modeling is the availability of detailed temporal and spatial recorded data based on which a model could be established. In Vietnam, water quality data are mostly collected through grab samples, and rarely on a continuous basis owing to the expense. For the purpose of environmental planning, we chose low flow periods for constructions of water quality model, because at these hydrological conditions river's environmental health is at its worst (US EPA 2010). For rivers where the tidal effects are negligible, as in this study, the river flow in low flow periods may be considered as steady and water quality variables may be considered as constant during the periods. Therefore under these circumstances, dynamic simulation of water quality is not necessary, and steady state models for

water quality simulation can be used. This is one reason that the QUAL2K model was chosen for the study.

Selection of simulation periods (for model calibration and validation) is important and must be based on the availability of observed data of river discharge and water quality. It is assumed that the hydrological and water quality variables are constant for the simulation periods (but they can vary along the river). Based on available data of the Cau River basin, we have defined two low flow periods to calibrate and validate the water quality model:

- Period for model calibration from 11/03/2014 to 19/03/2014: based on observed water quality data conducted by CEM from 11/03/2014 to 19/03/2014 and six-year monthly (March) average data between 2010-2015 at 19 water quality stations along the river (Fig. 1a).
- Period for model validation from 03/11/2014 to 09/11/2014: based on observed water quality data conducted by CEM from 03/11/2014 to 09/11/2014 and six-year monthly (November) average data between 2010-2015 at the same 19 water quality stations sampled in calibration period.

The year of 2014 was chosen for the simulation because it had an annual flow close to annual average flow for the period 2000-2015 observed at hydrological station Gia Bay (see Fig. 1b). For water quality data, one composite sample was taken for each station in accordance with the Vietnamese Standard TCVN 6663-6:2008 (ISO 5667-6:2005) (MONRE, 2008). That means samples are taken at 3 locations in a river cross-section: ¹/₄ distance from one river side, middle of the stream, and ³/₄ distance from the river side. The samples were mixed to form a composite sample for the station. The average discharges for the simulation periods obtained from a calibrated SWAT model for the river basin were used as flow inputs data for QUAL2K model. In the next sections, details of the calibration and validation results and integrating process of SWAT and QUAL2K models for Cau River basin are presented. The assumption for steady state for these periods was justified because the river daily discharge did not vary as much (i.e. no rain during the monitoring periods). At Gia Bay station for the first period (in March 2014), discharges changed from 21.7 to 22.9 m³/s and for second period (in November) from 33.0 to 37.8 m³/s.

3.1. SWAT model

Input data: The input data to construct the SWAT model consists of:

- DEM based on 30mx30m resolution DEM (IET, 2012).
- Cover map based on 2007 land-use map (IET, 2012) and updated data from Provincial Statistical Books 2014.
- Soil map from FAO Asia soil map (FAO, 2011).
- Daily hydro-meteorological data acquired from more than 40 stations from 2010-2015.

Sub-watersheds and river reaches in the SWAT model is presented in the Fig. 2b.

Calibration and validation results: The SWAT model was calibrated for river discharge using observed discharge data at the Gia Bay station for the period 2001-2012, and validated for the period 2013-2015. Four model performance ratings NSE (Nash and Sutcliffe efficiency), RSR (observations standard deviation ratio), PBIAS (percent bias) (Moriasi et al., 2007) and R^2 (correlation coefficient) (Parajuli et al., 2009) were used for assessment of the model. The performance ratings for calibration and validation processes are presented in the Table 2.

Evaluation	Calibration period	Donformon oo notin o	Validation period	Performance
statistics	2001-2012	Performance rating	2013-2015	rating
NSE	0.70	Good	0.60	Satisfactory
PBIAS(%)	25.0	Satisfactory	14.8	Good
RSR	0.55	Good	0.63	Satisfactory
\mathbb{R}^2	0.78	Very good	0.71	Good

Table 2. Performance ratings for SWAT model

According to Moriasi et al. (2007), the SWAT model performance is "good" or "satisfactory" for all evaluation statistics and for both calibration and validation periods. Therefore, the model can be accepted for use. As an example, the validation results for the period 2014–2015 are displayed in Fig. 4.



Fig. 4. Validation results of SWAT for 2014-2015 (daily simulation)

3.2. QUAL2K model

Study area and river network: The water quality model was constructed for the mainstream of Cau River from Cho Moi (at border of province Bac Kan and Thai Nguyen) to Pha Lai. The upstream of the model was set at Cho Moi because there is limited information on pollution load and water quality data in Bac Kan, which is the upstream province of the basin.

Figure 1a shows the river network that consists of the main river from Cho Moi to Pha Lai and 7 major tributaries: Nghinh Tuong, Du, Ben Oanh, Bach Duong, Cong, Ca Lo and Ngu Huyen Khe. Table 3 summaries the locations of upstream inlet headwater (Cho Moi) and 7 tributary outlets and their upstream sub-watersheds (see also Fig. 1a).

	Table 3.	Location	of headwater	and tributaries'	outlet and their	· upstream	sub-watersheds
--	----------	----------	--------------	------------------	------------------	------------	----------------

Headwater/tributaries' name	Distance from watershed outlet (km)	Upstream sub- watersheds	Drained area (km ²)
Cho Moi	206.9	1-20	1635.8
Nghinh Tuong	191.7	22,23	429.8
Du	165.6	24,25,29	367.7

Ben Oanh	154.9	26,27,31	198.8
Bach Duong	146.2	34	140.3
Cong	82.9	37-40,43	412.6
Ca Lo	65.1	49,53-54	826.2
Ngu Huyen Khe	42.7	55-57	308.9

In the water quality model, the river is divided into 46 segments based on data of 50 measured river cross-sections between Cho Moi and Pha Lai (IET, 2017). Schematic segmentation of the Cau River and locations of 7 tributaries and direct wastewater sources from 12 facilities into mainstream of Cau River is shown in Fig. 5. The surveyed river cross-sections are approximated to trapezoidal shapes (as required by QUAL2K) described by three parameters: bottom width and two side slopes.



Fig. 5. Schematic segmentation of Cau River for QUAL2K model: segments, locations of tributaries outlets and direct point sources from the outlet boundary

The Manning roughness coefficients for the Cau River have been chosen according to Chow et al. (1988). Based on observed river surfaces and topology (windingness or how windy), the Cau River from Cho Moi to Pha Lai was divided into two segments with different roughness ranges. The first segment from Cho Moi to the confluent with Cong River is characterized by windy reaches and some weeds on the surface. The Manning roughness coefficient for this segment could range from 0.03 to 0.05. In contrast, the second segment from the confluence with Cong River to Pha Lai is characterized by straighter less windy reaches and smooth surface (with mud), so the

Manning roughness coefficient for this segment could range from 0.02 to 0.03. In reality, the Manning roughness coefficient can vary within segments and it could be determined by variation of water level and discharge along the segment. However in the Cau River basin where there is not enough observed data for such an analysis, a constant Manning roughness coefficient was assumed for each segment with the value 0.035 for river segments from Cho Moi to Cong confluence and 0.025 for river segments from Cong confluence to Pha Lai.

The Thac Huong dam has a height of 6.5m (from riverbed) and width of 100m. The irrigation channel from the Thac Huong dam is regulated by Da Gan gates. The important water intake for irrigation is during January and February, so for simulation periods it can be assumed that the gates are closed and no water was taken from the dam.

Flow inputs from the SWAT model into the QUAL2K model: The flow boundary conditions inputs for QUAL2K model were obtained from flow outputs of the calibrated SWAT model (section 3.1). Discharge inputs were required for upstream inlet of Cho Moi and the outlets to the mainstream river of 7 river tributaries (Fig. 2b and Fig. 5). Further the lateral inflow inputs to QUAL2K from sub-watersheds bordering the Cau River were needed along the mainstream reaches.

The average of daily discharges from the SWAT model at Cho Moi and 7 tributaries outlets for simulation periods was obtained and used as discharge inputs in QUAL2K.

The water yields along the mainstream reaches from sub-watersheds bordering to the river obtained from SWAT were used for lateral inflow inputs to QUAL2K. They can be calculated from the equation:

$$Q_{WVSWAT} = SURQ + LATQ + GWQ - TLOSS - Abstractions$$
(1)

Where Q_{wySWAT} is the water yield along a reach, *SURQ* is the surface runoff contribution to streamflow, *LATQ* is the lateral flow contribution to streamflow (from sub-surface layer), *GWQ* is the groundwater contribution to streamflow, *TLOSS* is the water loss from reach by transmission through the streambed and *Abstractions* is water amount abstracted from the reach.

The pollution load input: Pollution load analysis shows that wastewater from domestic, animal husbandry and craft villages for food production are the major source of pollution. Therefore, organic and nutrient pollution were modeled in the study. The following water quality variables were simulated: water temperature (T), DO, BOD, COD, total nitrogen (TN) and total phosphorus (TP). Other contaminant parameters in the QUAL2K model were assumed to be negligible.

The pollution load input to the model consists of:

1. Pollution load from river upstream of Cho Moi and from 7 tributaries: These pollution loads were calculated using the input discharges (from SWAT) and measured concentrations of water quality variables (DO, BOD, COD, etc.) at outlets of the river tributaries during the simulation periods. These pollution loads represent the contribution of all pollution sources (point and diffuse sources) from upstream sub-watersheds of the tributaries into Cau River mainstream (see Fig. 1b).

2. Pollution from point sources: The term of "point source" here refers to all pollution sources that discharge directly to the mainstream of Cau River. In the study area, 12 point pollution sources were considered as inputs data to the QUAL2K model (see Table 5).

3. Pollution loads from diffuse sources of sub-watershed bordering the main river were calculated with appropriate removal rates from interior locations to Cau River. The term "diffuse pollution" here refers to all pollution sources except point sources that discharge directly to the

mainstream river or are connected to the sewage networks. The method for estimation of removal rates for different regions from observed data is given next.

One problem of data scarce basin is that observed water quality data often does not contain all N and P species, while QUAL2K model requires the input for all species of nitrogen (NH₄-N, NO₃-N, NO₂-N and Organic N) and phosphorus (Dissolved P and Inorganic P) for calculation of Total Nitrogen (TN) and Total Phosphorus (TP) respectively. To fill this gap in data, we calculated the average species ratios for N and P from field data for pollution sources and from CEM observed data (of period 2010-2015) for river water quality. Then, the concentration of missing species (of N or P) in the data set was calculated from TN or TP as follows:

$$C_s = \frac{C_{Tx} \times r_s}{100} \tag{2}$$

Where C_s is the concentration of the species, C_{Tx} is the concentration of TN or TP and r_s is the species ratio in percent.

The average species ratios for N and P in Cau River basin are presented in Table 4.

Species	Cau River	Pollution point	Species	Cau River water	Pollution point
	water	source	species	Cuu Hiver water	source
NO ₃ +NO ₂ -N	36.4	23.0	InOrg P	36.4	62.0
NH ₄ -N	20.0	34.8	P Dis	63.6	38.0
N Org	43.6	42.1			
TN	100	100	TP	100	100

Table 4. Species ratios of N and P pollution (%)

Headwater and direct point source inputs: As mentioned above, the concentration inputs at Cho Moi and 7 tributary outlets were taken from observed data for the simulation periods or obtained by using Eq. (2) for missing species of N and P as required.

Locations of 12 direct point pollution sources, their discharge and pollution concentrations are presented in Table 5 (see also Fig. 5 for their locations).

No.	Facility's Name	Distance from outlet (km)	Discharge (m ³ /day)	BOD (mg/L)	COD (mg/L)	TN (mg/L)	TP (mg/L)
DS 1	Hoang Van Thu Paper JSC	161.7	1650	20.6	45.0	4.4	0.19
DS 2	Cao Ngan Thermo- Power Plant	160.3	1680	56.5	84.3	2.2	0.72
DS 3	Hop Thinh Industrial Cluster	83.1	4050	33.2	50.5	14.7	1.32
DS 4	Vong Nguyet Craft Village	62.0	1728	33.2	50.5	14.7	1.32
DS 5	Dai Lam Villages for Wine Brewery	55.5	70	36.0	53.0	79.1	4.30
DS 6	Tho Ha Craft Village	53.6	384	1186.5	2184.0	22.7	3.58
DS 7	Van Craft Village	49.8	350	1930.0	8652.0	79.1	4.30

Table 5. Data for direct point pollution sources

DS 8	Viglacera Dap Cau Sheet Glass JSC	43.8	112	12.1	34.1	27.1	0.03
DS 9	Hung Phat Urban Environment Ltd.	41.1	200	35.0	78.5	0.9	0.37
DS 10	Que Vo IP1+Que Vo IP2	31.0	143424	36.0	60.0	18.3	0.51
DS 11	VK Environment Company Limited	20.0	60	32.6	58.0	2.8	0.04
DS 12	Que Vo IP3	14.0	60480	5.4	18.0	1.1	0.04

Removal rates: The concentration of diffuse pollution sources were estimated based on the average removal rates of pollution species calculated from observed data for simulation periods. The removal rate expresses the loss rate of pollutant as it is transported over the land surface to the water body and depends on surface conditions and physical, chemical and biological processes. Pollutant removal rates were studied experimentally by many investigators. The results have shown that for constructed wetlands or riparian zones, the removal rates are about of 70%, 65%, and 90% for BOD, COD, TSS respectively (Çakir et al., 2015). For nutrients, it can be larger than 95% (NOAA, 1999). For the Elbe drainage area, it has shown that on average 0.4–3% of the applied manure and fertilizer flows directly into surface water and drains away and a runoff factor of 2% was used for heavy metals (Vink et al., 1999; Raad et al., 1993). In our study, the removal rates were calculated based on observed data for a sub-watershed or a set of sub-watersheds that have a water quality station at their outlet. Based on available data of Cau River basin, the removal rates can be calculated at four water quality stations: Cau Da Phuc, Giang Tien, Yen Ninh and Chua Hang (see Fig. 2b and Table 3). The removal rate of a station can be obtained by the following equation:

$$R_{Rp} = \left(1 - \frac{L_{p-st}}{L_{p-sub}}\right) \times 100 \tag{3}$$

where R_{Rp} is the removal rate of pollutant p, L_{p-st} (kg/day) is the pollution load of pollutant p at the station, L_{p-sub} (kg/day) is pollution load of pollutant p of a sub-watershed or set of sub-watersheds (including point and diffuse sources except point sources discharging directly to the river). The removal rates at four stations were calculated for BOD, COD, TN and TP based on observed data in March and November for six years (2010-2015). The average values of these removal rates are presented in Table 6. These values were used for calculating pollution loads of sub-watershed bordering the river mainstream that were used as lateral pollution load inputs for QUAL2K model for the two simulation periods (in March and November), respectively.

Station	Period	BOD	COD	TN	ТР	Upstream Sub- watersheds
Cau Da Phuc	March	89	91	96	94	27 28 20 40 42
	November	80	75	90	85	57,58,59,40,45
Giang Tien	March	91	93	94	98	24.25.20
	November	85	82	88	94	24,25,29
Yen Ninh	March	97	98	98	98	13,16,18,19,20

 Table 6. Removal rates at four water quality stations (%)

	November	83	78	90	92	
Chua Hang	March	95	94	96	96	26 27 21
	November	79	77	91	93	20,27,31

Lateral inputs: The lateral discharge from sub-watershed adjacent to the river was calculated as follows,

$$Q_{lat,i} = Q_{wySWAT,i} + Q_{waste,i} \times k_{runoff} \tag{4}$$

where *i* is the index of sub-watershed, $Q_{lat,i}$ (m³/s) is the lateral inflow input for QUAL2K river reach, $Q_{wySWAT,i}$ (m³/s) is the water yield of sub-watershed *i* obtained from SWAT model (containing surface and groundwater flow), $Q_{waste,i}$ (m³/s) is wastewater discharge from different pollution sources within sub-watershed, k_{runoff} is annual average of runoff coefficient of the watershed.

The pollution load of a sub-watershed L_{p-sub} in Eq. (3) was estimated based on the observed data and rapid load estimation method (WHO, 1993). The results for Cau River basin were presented in the report of IET (2017).

The concentration $C_{p,i}$ of lateral discharge $Q_{lat,i}$ of pollutant p in sub-watershed i is calculated as follows,

$$C_{p,i} = \frac{L_{p-sub,i}}{Q_{lat,i}} \tag{5}$$

Where $L_{p-sub,i}$ is the pollution load of pollutant p in sub-watershed i.

Finally, as required for input in QUAL2K, the concentrations of nitrogen and phosphorous species of $Q_{lat,i}$ were calculated using Eq. (2) and species ratios presented in Table 4.

QUAL2K parameters: The exponential model was selected for oxygen inhibition of BOD hydrolysis, COD oxidation, denitrification and nitrification. The single pool option (fast CBOD) was chosen to simulate BOD, and COD was simulated as "Constituent Γ " in QUAL2K model (Chapra et al., 2012). The influence of algae and phytoplankton were ignored. There are ten degradation parameters including Fast CBOD: Oxidation rate (k_{dc}); Organic N: Hydrolysis rate (k_{hn}), Settling velocity (v_{on}); Ammonium: Nitrification rate (k_{na}); Nitrate: Denitrification rate (k_{dc}), Settling velocity (v_{op}); Inorganic P: Settling velocity (v_{ip}); and COD: Oxidation rate (k_{dcod}). These were obtained by trial and error iterations to get a best fit with observed data. The re-aeration rates were assumed to be constant for upstream and downstream parts of the river and were also obtained by trial and error iterations. The remaining parameters were kept at the default values in the QUAL2K model (Rode et al., 2007).

4. Results and discussions

According to Moriasi et al. (2007), in general, model simulation can be judged as satisfactory if NSE is greater than 0.50 and RSR is less than 0.70, and if absolute value of PBIAS (|PBIAS|) is less than 25% for streamflow, less than 55% for sediment, and less than 70% for both N and P. In water quality modeling, simultaneously satisfying three evaluation criteria is very difficult. There are large uncertainties of observed data because of the water quality monitoring budget is often limited and laboratory conditions are not always as good as required, especially in developing countries. Moreover, many water quality studies (Moriasi et al., 2007; Debele et al., 2008; Hesse

et al., 2012) showed that their models did not satisfied this requirement, and some performance ratings were, in general, far outside the evaluation criteria values proposed by Moriasi et al. (2007).

PBIAS measures the average tendency of simulated data versus observed data. It does not require the data to fit strictly on the 1:1 line as NSE and RSR do (Moriasi et al., 2007). So PBIAS is appropriate for evaluation of water quality modeling, because in reality, the water quality data do vary depending on many other conditions at the time of sampling and samples are often obtained with large uncertainty. To our knowledge, there was no works concerning the PBIAS threshold values for other water quality variables such as T, DO, BOD or COD. However, these parameters (except water temperature T) are measured with more uncertainty than sediment or streamflow, because they depend on many in-stream processes like components of N and P. So we can assume that they have at least the same threshold as N and P as proposed by Moriasi et al. (2007). Therefore, we propose to take the PBIAS as main evaluation criterion for water quality modeling in Cau River basin. So, according to Moriasi et al. (2007), the model simulation can be judged as "very good" if |PBIAS|<25%, "good" if 25%≤|PBIAS|<40%, "satisfactory" if 40% <->
PBIAS <components of N and P. Note that PBIAS may be low if the overestimation and underestimations cancel each other and PBIAS should not be used alone for performance rating. The correlation coefficient R^2 does not require the data to fit strictly on the 1:1 and avoids the cancelation by overestimation and underestimation values. Parajuli et al. (2009) require an R²>0.25 as "fair performance" for the simulation for flow, sediment and TP. Therefore, the R² was used together with the PBIAS to assess the model performance.

Table 7 presents the performance indices for the calibration period for different water quality variables. The PBIAS is in the range of 2% < |PBIAS| < 14% and according to the criterion presented above, the model performance is "very good". Moreover, we have NSE ≥ 0.13 and R² ≥ 0.32 for all variables, these generally viewed as an acceptable level of performance (Moriasi et al., 2007; Parajuli et al., 2009). The RSR values vary between 0.63 and 0.93 and these magnitudes are as good as or better than those regularly reported in water quality studies (Moriasi et al., 2007; Debele et al., 2008; Hesse et al., 2012; Chihhao Fan et al., 2009). Hesse et al. (2012) reported RSR values in the range between 0.17 and 1.71 for water quality variables.

Fig. 6 presents simulation results in comparison with the observed data at 19 stations along mainstream of the Cau River. The simulation results show good agreement with observed data for the variables T, DO, BOD, COD, TN and TP. Fig. 6 also shows the gap between observed data for the calibration period (March 11-19, 2014) and the average values of the observed data collected in March between 2010 and 2015. The variations of observed water quality data along the river may have resulted from the grab collecting method of water quality sampling. This lends supports for using PBIAS as a more suitable evaluation criteria of a water quality model.

Evaluation statistics	Т	DO	BOD	COD	TN	TP
PBIAS (%)	-1.68	1.00	12.56	6.49	13.73	9.92
RSR	0.69	0.93	0.86	0.92	0.62	0.87
NSE	0.53	0.13	0.26	0.16	0.62	0.24
\mathbb{R}^2	0.72	0.63	0.45	0.38	0.77	0.32

Table 7. Performance ra	atings for	QUAL2K	model	calibration	period
-------------------------	------------	--------	-------	-------------	--------

The calibration results (Fig. 6) showed very good agreement between simulated and observed values for T. It is clear that the good agreement is often achieved for water temperature (T) because

this variable depends mainly on mean air temperature in the basin which is often the most reliable of all the variables in the observed database. DO seems to be overestimated in the upstream part of the river (to station 10) and underestimated for downstream part of the river. This may be caused by the parameters related to DO that were omitted from the modeling, e.g. the growth of algae or DO sediment exchange. The peak of DO at station 7 is explained by aeration occurring at the Thac Huong dam.

Other variables such as BOD, COD, TN and TP are subject to many complex processes and derived from different sources along the river. In general, the concentrations of these variables increase downstream due to accumulation of pollution load discharged along the river; the large increase after the outlet of the Ngu Huyen Khe tributary is due to untreated wastewater discharge from many craft villages in Bac Ninh province. The concentrations decrease before the outlet of the watershed at Pha Lai station due to the self-purification process of the river. We can see that for these variables, the simulation results also produce good agreement with observed data. The largest PBIAS are for TN and BOD: 13.73% and 12.56%, respectively.



Fig. 6. Calibration results of water quality model in Cau River basin for data on March 11-19, 2014

Table 8 lists all calibration parameters that were implemented in the model with their possible ranges recommended by Neitsch et al., 2001; Zhang et al., 2012; Hesse et al., 2012; EPA, 1978 and the values used for the Cau River basin in this study. Note that all the values used in the study

were in their recommended ranges and obtained by trial and error fitting to observed data. Nonetheless, the calibration parameters were not optimized and indeed, it is possible for a different set of calibration parameters values to yield a better model performance.

Parameter	Value	Units	Symbol	Range
Oxygen:				
+ Reaeration model	User specified			
Fast CBOD:				
+ Oxidation rate	0.1	/d	k _{dc}	0.02-3.4
Organic N:				
+ Hydrolysis	0.05	/d	k _{hn}	0.02-0.4
+ Settling velocity	0.001	m/d	Von	0-2
Ammonium:				
+ Nitrification	0.05	/d	k _{na}	0-10
Nitrate:				
+ Denitrification	0.075	/d	k _{dn}	0-2
+ Sed denitrification transfer coeff.	0.001	m/d	V _{di}	0-1
Organic P:				
+ Hydrolysis	1.2	/d	\mathbf{k}_{hp}	0-5
+ Settling velocity	0.8	m/d	Vop	0-2
Inorganic P:				
+ Settling velocity	0.85	m/d	V _{ip}	0-2
+ Sed P oxygen attenuation half sat constant	0.5	mgO ₂ /L	k _{spi}	0-2
Constituent I (COD)				
+ First-order reaction rate	0.08	/d	k _{dcod}	0.02-4.2

Table 8. Calibrated model parameters for Cau River water quality modeling in 2014

The model was validated with observed water quality data from November 3-9, 2014 using parameters that were derived from calibrating the model with the observed data from the March 11-19, 2014 period. The validation results (Table 9 and Fig. 7) showed that the calibrated parameters used in the model were able to reproduce the observed data in the validation period. Table 9 presents the performance ratings for the validation period for different water quality variables. It shows that 0.41% <|PBIAS| <10.35%, and according to the PBIAS performance criteria, the model performance is "very good". The worse performance were for TN with PBIAS=-10.35, NSE=-0.36, RSR=1.16 and R²=0.31. Fig. 7e shows the variation of TN along the river. It shows that, for upstream part of the river (upstream of station 15), the fit between observed and simulated TN values is good but at the downstream part (downstream of station 15) the observed values decrease more rapidly than the simulated ones. This results in large difference between observed and simulated values. The difference may be explained in several ways. For example, it can be explained by the different nitrogen degradation parameter for upstream and downstream river segments while the model uses only one parameter for the river or by perhaps a poor nitrogen load calculation for downstream sub-watershed. This could be improved if additional data were available.

Table 9. Performance ratings for QUAL2K model validation period

Evaluation statistics	Т	DO	BOD	COD	TN	TP	
PBIAS (%)	0.41	-6.36	5.91	2.17	-8.33	-10.35	
RSR	0.66	0.78	0.88	0.76	1.16	0.83	
NSE	0.57	0.39	0.23	0.42	-0.36	0.30	
\mathbb{R}^2	0.63	0.57	0.51	0.45	0.31	0.36	



Fig. 7. Validation results of water quality model in Cau River basin for data on November 3-9, 2014

Further, some errors in this modeling are inevitable as field work consisted of collecting a single grab sample at each station due to the limited budget. In addition, the model predictions are daily averaged while the observed water quality might have been sampled at times when conditions were different. For example, the observed DO can change considerably during the day due to the change of water temperature and the rate of photosynthesis of river plants. Observed BOD may change due to irregular/uncontrolled discharge at some pollution sources. In spite of these errors, the modeling results were quite acceptable given the limited data conditions that exist in developing countries, where the financial resources are constrained. However, a greater accuracy may be achieved by more reliable water quality data. In this respect, the model performance can also guide future sampling campaigns. The model can be also used to guide for modest management plans.

5. Conclusions

Water quality modeling has been shown to be a useful tool for water resources management. However, the use of a basic model for water quality simulation of a river may not be satisfactory when the available data on hydraulic and water quality characteristics are limited.

This paper presents an attempt to integrate the outputs from a hydrological model (SWAT) with a water quality model (QUAL2K) to simulate water quality focusing on organic and nutrient pollution (DO, BOD, COD, TN, TP) in a data scarce river basin. The SWAT model was used to capture detailed hydrological processes in the upland watershed and small river tributaries. The QUAL2K model was used to simulate water quality processes in the downstream river network. The SWAT model outputs provided the hydrological boundary inputs for QUAL2K model. The water quality outputs of SWAT model was not used as inputs for QUAL2K model because the lack of spatial data may have caused large errors in the SWAT results for water quality variables. To fill this gap in data, available observed water quality data were used and an intermediate data analysis processes was developed to provide the required data.

PBIAS performance rating was chosen the model evaluation because of the large uncertainty inherent in water quality data. In term of PBIAS, the calibration and validation results for Cau River water quality model were in the "very good" performance range with |PBIAS|<15% for all water quality variables. However, PBIAS may be low if the overestimation and underestimations cancel each other, so other indicators such as NSE and R² were calculated together with PBIAS to avoid unwanted cancelation.

Despite the absence of extensive measured water quality data and spatial data for pollutant sources in the river basin, the model was able to reliably simulate most water quality variables in the Cau River basin by using these two models. This study showed encouraging results in water quality modeling despite the paucity of water quality data. This should not mean that there is no need for such data for establishing a good water quality models. Rather it shows how a good model could still be developed in data scarce basins based on procedures presented above. The results of the model framework could be used as a management tool to prioritize water quality data that should be collected to improve the rigour of the model simulations and reduce uncertainty, to prioritize implementation of watershed management, and to alleviate the significant water quality problems in the river basin.

References

- Ambrose, R.B., Wool, T.A., & Martin, J.L., 1993. The Water Quality Analysis Simulation Program WASP5, Part A: Model Documentation, Version 5.10. US Environmental Protection Agency, Athens, Georgia.
- Arnold, J. G., Williams, J. R., Srinivasan, R., King, K. W., & Griggs, R. H., 1994. SWAT: Soil and Water Assessment Tool. Temple: USDA, Agriculture Research Service, Grassland, Soil and Water Research Laboratory.
- Audet L., 2013. Modélisation de la qualité de l'eau de la rivière Cau au Vietnam. M.Sc. thesis, ETE, Université du Québec, INRS.
- Bicknell, B. R., Imhoff, J. C., Kittle, J. L., Jr., Donigian, A. S., Jr., and Johanson, R.C., 1996. Hydrological Simulation Program - FORTRAN User's Manual for Release 11. Environmental Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Athens, GA.
- Brown, L. C., & Barnwell, T. O, Jr., 1987. The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS: Documentation and user manual. US Environmental Protection Agency, Athens, Georgia.
- Chapra, S.C., Pelletier, G.J., Tao, H., 2012. QUAL2K: A Modeling Framework for Simulating River and Stream Water Quality, Version 2.12: Documentation and Users Manual. Civil and Environmental Engineering Dept., Tufts University, Medford, MA.
- Chow, V.T., Maidment, D.R., Mays, L.W., 1988. Applied hydrology. 1st Edn., McGraw-Hill Education, New York, pp: 572.

- Chihhao Fan, Chun-Han Ko, Wei-Shen Wang, 2009. An innovative modeling approach using Qual2K and HEC-RAS integration to assess the impact of tidal effect on River Water quality simulation. Journal of Environmental Management 90, 1824–1832.
- Çakir R., Gidirislioglu A., Çebi U., 2015. A study on the effects of different hydraulic loading rates (HLR) on pollutant removal efficiency of subsurface horizontal-flow constructed wetlands used for treatment of domestic wastewaters. Journal of Environmental Management.164,121-128.
- Debele, B., Srinivasan, R., Parlange, J.Y., 2008. Coupling upland watershed and downstream waterbody hydrodynamic and water quality models (SWAT and CE-QUAL-W2) for better water resources management in complex river basins. Environmental Modelling Assessment. 13, 135–153.
- Donigian, A.S.Jr., Imhoff, J.C., Bicknell, B.R, Kittle, J.L.Jr., 1984. Application guide for Hydrological Simulation Program Fortran (HSPF), US Environmental Protection Agency, Athens, Georgia.
- EPA (Environmental Protection Agency), 1978. Rates, constants, and kinetics formulations in surface water quality modeling, Athens, Georgia.
- EPA (Environmental Protection Agency), 2000. National water quality inventory report. Washington, DC.
- FAO/UNESCO Soil map of the world

http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/faounesco-soil-map-of-the-world/en/ [Accessed 22 June 2018].

- Ha Ngoc Hien, Bui Huy Hoang, Tran Thi Huong, Tran Thanh Than, Pham Thi Thu Ha, Ta Dang Toan, Nguyen Minh Son, 2016. Study of the climate change impacts on water quality in upstream portion of the Cau River Basin, Vietnam. Environmental Model Assessment. 21, 261-277.
- Hesse, C., Krysanova, V., Voß, A., 2012. Implementing instream nutrient processes in large-scale landscape modeling for the impact assessment on water quality. Environmental Modeling and Assessment.17,589–611.
- IET (Institute of Environmental Technology), 2012. Integrated water resources management in Cau River basin. Institute of Environmental Technology: Vietnam Academy of Science and Technology Project final report.
- IET (Institute of Environmental Technology), 2017. The Work for Simulation Modeling in Cau River Basin, Vietnam Environment Administration: Project final report.
- MONRE (Ministry of Natural Resources and Environment), 2006. Environment Report of Vietnam: The current state of water environment in 3 river basins of Cau, Nhue Day and Dong Nai river system.
- MONRE (Ministry of Natural Resources and Environment), 2008. Water quality. Sampling. Part 6: Guidance on sampling of rivers and streams: TCVN 6663-6:2008 (ISO 5667-6:2005).
- MONRE (Ministry of Natural Resources and Environment), 2015. National technical regulation on surface water quality: QCVN 08-MT:2015/BTNMT.
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., Veith, T. L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE. 50(3), 885-900.
- Neitsch, S. L., Arnold, J. G., Kiniry, J. R., Williams, J. R., 2001. Soil and water assessment tool (SWAT) user's manual version 2000. Grassland Soil and Water Research Laboratory. Temple, Texas.
- Nguyen Phuong Nhung, 2013. Annual flow characteristics of Cau river basin. VNU Journal of Science: Natural Sciences and Technology. 29(2), 152-158.
- NOAA, 1999. Reducing nutrient load, especially nitrate-nitrogen, to surface water, ground water, and the Gulf of Mexico, NOAA Coastal ocean program, decision analysis series No. 19. Maryland.
- ParajuliP.B., Nelson N.O., Frees L.D., and MankinK.R., 2009. Comparison of AnnAGNPS and SWAT model simulation results in USDA-CEAP agricultural watersheds in south-central Kansas. Hydrol. Process. 23, 748–763.
- Raad, J.S., PeerboomR.A.L., Dorland C., Hanegraaf M.C., FeenstraJ.F., 1993. Diffuse bronnen van waterverontreinigingen in het oppervlaktewater in Friesland. Inventarisatieenkwantificering. Institute for Environmental Studies, Amsterdam.
- Refsgaard, J. C., 1997. Parameterization, calibration and validation of distributed hydrological models. Journal of Hydrology, 198,69–97.
- Rode, M., Suhr, U., Wriedt, G., 2007. Multi-objective calibration of a river water quality model-information content of calibration data. Ecological Modelling. 204, 129–142.
- Streeter, HW, Phelps, EB, 1925. A Study of the pollution and natural purification of the Ohio river. III. Factors concerned in the phenomena of oxidation and reaeration, Public Health Bulletin no. 146.
- USEPA 2010. U.S. EPA NPDES Permit Writers' Manual. EPA-833-K-10-001. United States Environmental Protection Agency, Office of Water, Washington, DC. https://www.epa.gov/sites/production/files/2015-09/documents/pwm_2010.pdf.

- Vink, R.J., Behrendt, H., Salomons, W., 1999. Point and diffuse source analysis of heavy metals in the Elbe drainage area: Comparing heavy metal emissions with transported river loads. Hydrobiologia. 410, 307–314.
- WHO, 1993. Assessment of sources of air, water, and land pollution A guide to rapid source inventory techniques and their use in formulating environmental control strategies. Part one: rapid inventory techniques in environmental pollution. Geneva.
- Zhang R., Qian X., Yuan X., Ye R., Xia B., Wang Y., 2012. Simulation of water environmental capacity and pollution load reduction using QUAL2K for water environmental management, International Journal of Environmental Research Public Health. 9, 4504-4521.