Evaluating the Effect of Multi-Objective Calibration on the Modeling of Catchment-Scale Hydrology and Nitrogen Load

Von der Fakultät für Umwelt und Naturwissenschaften
der Brandenburgischen Technischen Universität Cottbus-Senftenberg
zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs

genehmigte Dissertation

vorgelegt von

Master of Science

Qianwen He

aus

Jiangxi Province, China

Gutachter: apl. Prof. Dr.-Ing. Frank Molkenthin
Gutachter: Honorarprof. Dr. rer. nat. Frank Wendland
Tag der mündlichen Prüfung: 03. Juli 2020

Acknowledgment

First I would like to express my great appreciation to my supervisors, Prof. Frank Molkenthin and Prof. Frank Wendland. Prof. Molkenthin has been offering massive support that helped me to conduct my research and improve the dissertation. His encouragement inspired me to overcome hard times. Prof. Wendland has always provided me valuable advice for my research and publications.

I would also like to thank all my colleagues and friends. I am very grateful that Mrs. Nora Reichert cared for me all the time and helped me to integrate into the local life. Mr. Bernard Fosu Frimpong and Mr. Manh Trinh Xuan always encouraged me and helped to correct my English writing. I am also very grateful for the support from Mrs. Uta Warstat and Mr. Mirko Filetti. Special thanks to Mr. Daneish Despot for his continuous support for my research and life. Many thanks to my friends, Ms. Zhengzheng Hao, Mr. Lutz Philip Hecker, Mr. Wenchao Xu, Mr. Wenchao Zou, and Mr. Yangwen Zhang, who colored my life in Cottbus.

Besides, I would like to thank the Chinese Scholarship Council (CSC) for financial support. I would also like to thank the Chinese Academy of Science, Jiangkou Hydropower Station, Yichun Hydrologic Bureau, and Xinyu Environmental Protection Bureau for providing me valuable data for my research.

Last but not the least, I would like to thank my parents, Mr. Dongsheng He and Mrs. Lu Xiao, for their endless love. Finally, I would also like to thank myself, for the passion I put into my life.

Table of contents

List of Figures

List of Tables

Acronyms and abbreviations

AET	Actual evapotranspiration
AW	Average Weight
AET_{cover}	The coverage of the simulated AET by the uncertainty band of the observed AET
CSAs	Critical source areas
DEM	Digital elevation model
ED	Euclidean Distance
ET	Evapotranspiration
FDC	Flow duration curve
HRU	Hydrologic response unit
JK	Jiangkou hydrometric station
LHS	Latin Hypercube Sampling
LX	Luxi hydrometric station
LX_g (MZ_g, JK_g)	The PBIAS of the simulated to the observed baseflow ratio of LX MZ, or JK
LX_s (MZ_s, JK_s)	The PBIAS of the simulated to the observed surface runoff ratio of LX MZ, or JK
MOD16 ET	MODIS global monthly evapotranspiration data
MZ	Maozhou hydrometric station
NPS	Non-point source
NSE	Nash-Sutcliffe Efficiency
NSE_LX (NSE_MZ, NSE_JK)	NSE of LX, MZ or JK
NSE_L (NSE_F, NSE_L)	NSE of 1_Y, 4_F, or 6_L
NSE_in	Inverse Nash-Sutcliffe Efficiency
NSE_in_LX (NSE_in_jz, NSE_in_JK)	Inverse NSE of LX, MZ or JK

NSE_in_L (NSE_in_F, NSE_in_L)	Inverse NSE of 1_Y, 4_F, or 6_L
NSGA-II	Non-Dominated Sorting Genetic Algorithm
PBIAS	Percent BIAS
PS	Point source
RMSE	Standardized mean square error
RSR	RMSE-observations standard deviation ratio
SUFI-2	Sequential Uncertainty Fitting, ver.2
SWAT	Soil and Water Assessment Tool
TN	Total nitrogen
YRC	Yuan River Catchment
1_Y	Yangjiang water quality station
4_F	Fuqiao water quality station
6_L	Luofang water quality station
95PPU	95% prediction uncertainty

Abstract

Process-based hydrological models, which simulate nitrogen load from its sources to the receiving waterbody, play an important role in supporting catchment management. The reliability of those models, such as the representative model SWAT used in this study, is determined by a sound calibration and the analysis of the prediction uncertainty. The multi-objective calibration approach prevails the classic single-objective calibration on the spatial parameterization of specific processes. However, the requirement of additional observations and practical procedures limits its application. Moreover, the prediction uncertainty of nitrogen load is inevitable and should also be quantified. This study is a scientific contribution to catchment management by overcoming the challenge with a systematic, well-founded concept of multi-objective calibration and uncertainty analysis for nitrogen load simulation in data-scarce catchments. The concept is tested and proofed by its practical applicability on the Yuan River Catchment (YRC) in China and leads to a generalized, recommended procedure for catchment management application as valuable progress in this field.

The study proposed to apply three groups of objectives, multi-site, multi-objective-function, and multi-metric. The applicability and the advantages of two multi-objective calibration approaches, Euclidean Distance and Non-Dominated Sorting Genetic Algorithm-II were analyzed. To quantify the prediction uncertainty, the study proposed to use the simulations with the highest or the lowest percent bias to represent the uncertainty band of the nitrogen load from the critical source areas (CSAs) and to the stream. The data-scarcity of the YRC was overcome by metrics obtained from open-access satellite-based datasets and metrics extracted from the existing discharge observations. Results show that multi-objective calibration has ensured the model's better performance in terms of the spatial parameterization, the magnitude of the output time-series and the water balance components, in comparison to single-objective calibration. The predicted CSAs showed that 50% of the total nitrogen (TN) loading to the stream was from 26.3% to 37.1% of the area in the YRC. Meanwhile, over 50% of those TN were from the paddy field. Recommendations for application go to the multi-objective calibration considering all three groups of objectives. Approaches to obtain multi-metric objectives in the YRC are also applicable for catchments with data-scarcity. Recommendations for nitrogen management in the YRC is to emphasize the CSAs identified, especially the paddy field.

Kurzfassung

Prozessbasierte hydrologische Modelle zur Simulation des Stickstofftransports und der resultierenden Belastung in Gewässern spielen eine wichtige Rolle bei der Unterstützung des Managements von Wassereinzugsgebieten. Die Zuverlässigkeit dieser Modelle, wie das in dieser Studie verwendete repräsentative Modell SWAT, wird durch eine sorgfältige Kalibrierung und die Analyse der Vorhersageunsicherheit bestimmt. Eine Kalibrierung auf der Basis von mehrkriteriellen Optimierung (multi-objective calibration) hat sich gegenüber der Kalibrierung mit einer Zielfunktion (single-objective calibration) insbesondere zur räumlichen Parametrisierung relevanter Prozesse bewährt. Sie erfordert jedoch umfangreiche zusätzliche Beobachtungs- und Messwerte sowie zielführende Vorgehensweisen/Verfahren, die nicht immer vorliegen bzw. entwickelt wurden. Die Anwendung der mehrkriteriellen Kalibrierung ist daher insbesondere bei datenarmen Einzugsgebieten nur eingeschränkt möglich. Die Schwierigkeit der Quantifizierung der zugehörigen Vorhersageunsicherheit der Stickstoffbelastung erhöht sich ebenfalls deutlich. Diese Dissertation ist ein wissenschaftlicher Beitrag zum Management von datenarmen Wassereinzugsgebieten bzgl. Stickstoffbelastung der Gewässer, indem sie die gegebene Herausforderung mit einem systematischen, fundierten Konzept der mehrkriteriellen Kalibrierung und zugehörige Unsicherheitsanalyse für die Simulation der Stickstoffbelastung angeht. Das Konzept wird hinsichtlich der praktischen Anwendbarkeit am Beispiel des Yuan River Catchment (YRC) in China entwickelt und getestet. Daraus wird eine allgemeine Vorgehensweise für die Anwendung der mehrkriteriellen Kalibrierung von Modellen zur Simulation der Stickstoffbelastung in datenarmen Wassereinzugsgebieten als wissenschaftlicher Beitrag in diesem Bereich abgeleitet.

In der Dissertation werden die verschiedenen Zielkriterien bei der Modellkalibrierung in drei Gruppen klassifiziert: multi-site, multi-objective-function und multi-metric. Die Anwendbarkeit und die Vor- und Nachteile von zwei mehrkriteriellen Kalibrierungsverfahren, Euclidean Distance und Non-Dominated Sorting Genetic Algorithm II werden eingesetzt und für verschiedene Szenarien ausgewertet. Um die Vorhersageunsicherheit zu quantifizieren, werden die Simulationen mit der höchsten oder niedrigsten prozentualen Abweichung für die Spezifikation der Unsicherheitsbandbreite verwendet. Die Datenknappheit des Wassereinzugsgebietes YRC wurde durch die Verwendung von Messungen aus mehrerer, verschiedenen Quellen (multi-metric)

überwunden: satellitenbasierte Open-Access-Datensätzen und die vorhandenen Abflussmessungen im Gewässer. Die Ergebnisse zeigen, dass die mehrkriteriellen Kalibrierung die Qualität des Simulationsmodells in Bezug auf die räumliche Parametrisierung, die Kenngrößen der Ausgabezeitreihen und die Komponenten der Wasserhaushaltsbilanz im Vergleich zur Kalibrierung mit einer Zielfunktion signifikant erhöht hat. Die vorhergesagten kritischen Stickstoffeintragsbereiche (critical source areas, CSAs) zeigen, dass 50% der gesamten Stickstoffbelastung des Yuan River 26,3% bis 37,1% der Fläche im YRC ausmachen. Dabei stammen über 50% dieser Stickstoffbelastung von Reisfeldern. Für die praktische Umsetzung beim Management von Wassereinzugsgebieten wird empfohlen, für die mehrkriteriellen Kalibrierung in datenarmen Einzugsgebieten wie dem YRC alle drei Gruppen von Zielkriterien zu verwenden. Spezifisch für das YRC wird empfohlen, bzgl. der Stickstoffbelastung ein besonderes Augenmerk auf die Reisfelder im Bereich der CSAs und deren Bewirtschaftung zu legen.

The Kurzfassung was written with the support of Ms. Nora Reichert.

摘要

水文模型对流域范围的水文和氮负载的过程模拟有助于流域规划，并在近年来有越加广泛的应用。在利用以 SWAT 为代表的基于过程的模型进行模拟时，合理的校准和模拟结果的不确定性分析能够提升模型的可靠性。与传统单一目标校准相比，多目标校准能在空间上对特定水文过程获得更准确的参数数值。然而多目标校准方法需要更多的实测数和实用的应用步骤，这也在一定程度上限制了此方法的推广。氮负载的模拟结果的不确定性不可避免，因此对于不确定性的定量分析也同样重要。本研究的核心及创新性在于提出了并且验证了在数据有限的流域上，多目标校准方法以及一种定量模拟结果不确定性的方法的适用性。

本研究首先提出将校准的目标分为三组，包括“多站点”，“多目标函数”和“多指标”，并且以此为基础验证了两种多目标校准方法，Euclidean Distance 和 Non-Dominated Sorting Genetic Algorithm-II，的适用性和优越性。对于预测河流中的氮含量和重点氮排放区域，提出利用校准过程中有最高和最低百分偏差的模拟结果来代表预测排放重点区域的不确定性的上下限。研究的区域为面临河流氮超标并且建模数据有限的袁河流域。在校准目标的获取中，“多指标”利用了 MODIS 的全球土地蒸腾量数据和本有的实侧流量数据，弥补了流域本身的数据缺乏。研究结果表明多目标校准能明显优化参数的空间取值特异性，并且提高时间序列和水平衡各组成部分的模拟结果。氮负荷的预测表明，排入河流中 50% 的总氮来源于 26.3% 至 37.1% 的流域面积，而这些总氮有超过 50% 来自于水稻种植区域。基于研究结果，应用多目标校准时可同时考虑上述提出的三组目标，并且利用开放获取的卫星数据产品和本有的实侧数据来弥补流域数据有限的情况。对于袁河流域的氮负载情况，本研究建议关注预测结果中的氮排放重点区域，尤其是水稻种植区域。

1 Introduction

Excess nitrogen in the water body has been a threat to sustaining a healthy aquatic ecosystem and thus a potential risk to human health. It is not unique to specific catchments but a global issue. The sources of nitrogen are relatively complex and are mainly catchment individually that can be grouped into point sources (PS) and non-point sources (NPS) (WHO, 2011). Nitrogen from PS is loading directly to the stream and is mainly regulated, e.g. effluent from industrial sewers. However, nitrogen from NPS, e.g. fertilizer operation, is released from scattered sources and spatially distributed over the entire land phase (USEPA, 2017). Among all the sources, agricultural activities have been emphasized as the main NPS (Ansari and Gill, 2014). According to the report from FAO (Mateo-Sagasta et al., 2017), the pressure from agriculture worldwide on the water quality is increasing for decades due to the increasing food demand, which caused a doubled area of irrigation since 1961, a tripled amount of livestock since the 1970s, an over 20-fold amplification of aquaculture since 1980s, and 10 times more mineral fertilizer since 1960s. Also, the nitrogen used by agriculture is estimated to have an annual growth rate of 1.5% globally.

In the high-income countries or many emerging economies, the pollution from agriculture has been the major factor in the degradation of the water bodies, e.g. by 2015, 38% of the water bodies in the EU have been significantly affected by the agricultural pollution (WWAP, 2015). In low-income countries, besides agriculture, major concerns also go to the nitrogen from untreated municipal and industrial wastewater. In the developed countries, the regulations to manage the nitrogen from agriculture have been implemented since the 1990s, e.g. the Council Directive 91/676/EEC (European Commission, 1991) published in 1991 to regulate nitrogen from agricultural sources. In China, nitrogen discharge from municipal and industrial sources was regulated since 1988 (Ministry of Ecology and Environment of the PRC, 1996). However, only from the establishment of the Water Act in 2015 (Ministry of Ecology and Environment of the PRC, 2015), the nitrogen pollution from the agricultural non-point sources has been emphasized by the legislators, which indicated the increasing necessity to have a better understanding of the behavior of the nitrogen pollution.

1.1 Problem statement

The simulation of the nitrogen transport and distribution spatially and temporally from its sources to and in the river system is supportive to derive cost-effective pollution control strategies for catchment management (Carpenter, 2008; Lindgren et al., 2007). Process-based hydrological models, which develop the system with physically-based equations to describe the hydrological cycle in consideration of the case-specific topography, geology, climate conditions, and management operation, are useful tools to analyze the NPS from its sources to its receiver. Soil and Water Assessment Tool (SWAT) (Neitsch et al., 2011) is one of the representative models. It was selected in the study that it integrates the simulation of chemical yields and management practices to the hydrological model and it has been widely applied on catchments with varied conditions (Merwade et al., 2017).

Process-based hydrological models such as the SWAT model rely on the parameter calibration to obtain a reliable simulation. The classic calibration approach applied the regression-based summary statistics (Gupta et al., 2008), e.g. Nash-Sutcliffe Efficiency (NSE), to evaluate the goodness-of-fit of the simulations to the observations at the outlet of the catchment. The limitations of the approach have been emphasized to be insufficient to indicate the individual hydrological processes and to derive the parameters to represent the spatial heterogeneity of the catchment (Gupta et al., 1998; Gupta et al., 2008; Arnold et al., 2015). However, to quantify the nitrogen load, e.g. to identify the critical source areas (CSAs), the accuracy of the model to simulate the spatial and temporal variation is also critical. The classic calibration approach is referred to as single-objective calibration that the calibration is only implemented to optimize one output variable, e.g. discharge at the outlet. In contrast, the multi-objective calibration that implements the optimization of multiple objectives is a more balanced approach for process-based hydrological models applied at the catchment scale (Gupta et al., 1998; Pfannerstill et al., 2017). However, multi-objective calibration has not been widely applied in hydrological modeling at the catchment-scale.

The limitations of implementing the multi-objective calibration include its requirement of additional observations and a practical application procedure. The additional observations are critical to catchments with only limited observations, e.g. only measured discharge and water quality for calibration. Therefore, the first problem to be solved is the acquisition of the observations as the additional objectives, especially for catchments with data-scarcity. Meanwhile,

the process-based hydrological models like SWAT have varied processes to be calibrated, and therefore the selection of the objectives is also of great importance in a calibration procedure. The calibrated hydrological model in the study intends to simulation the nitrogen load for the study catchment. However, the prediction uncertainty is inevitable of the process-based model like SWAT and thus should be quantified in terms of interpreting the nitrogen analysis for catchment studies.

Yuan River Catchment (YRC) in Jiangxi Province in China is a representative catchment that has the concern of excessive nitrogen load in the river, especially during the dry season (Fang, 2011). Meanwhile, when setting up a hydrologic and nitrogen load model for the YRC, it also confronts the data-scarce problem that only the measured discharge and total nitrogen load at limited stations are available as observations for calibration. Previous studies to analyze the total pollutant control were limited in this region with a focus on the discharge from PS (Zhang et al., 2011a; Li and Liu, 2015). However, the study (Wu, 2012) also showed that the nitrogen fertilizer application, untreated waste from the rural households, and the feedlots, as dominant NPS in the Jiangxi Province, should not be neglected when managing nitrogen pollution.

1.2 Research aim

The study aims to improve the reliability of the process-based hydrological model by applying the multi-objective calibration approach, quantifying the prediction uncertainty in terms of simulating the hydrology and nitrogen load on the data-scarce catchment.

The research was conducted to answer the following questions:

1) How to derive and apply the multiple objectives in the calibration for catchments with data scarcity like the YRC?
2) How is multi-objective calibration superior to single-objective calibration?
3) How to interpret the nitrogen load and quantify its prediction uncertainty at the catchment scale?

To handle the data-scarce condition for the study catchment, the study first proposes generally available sources to obtain additional observations. To enable the practical implementation of the multi-objective calibration, the study proposes to group the objectives for calibration as multi-site, multi-objective-function, and multi-metric. Two calibration approaches, Euclidean Distance and

NSGA-II are to be applied as representative multi-objective calibration approaches. The features of the calibration approaches will be analyzed by focusing on the comparison of the spatial parameterization, the uncertainty range of the results, and the performance of the calibrated output variables. In terms of the model calibration, the quantification of the prediction uncertainty of the nitrogen load will also be proposed and its benefits to interpret the nitrogen condition spatially and temporally in the YRC are to be discussed.

The content of the dissertation was arranged into the following chapters: Chapter 2 summarized the current state of the research related to the study by a literature review; Chapter 3 stated the condition of the study catchment and the processing of the required datasets; Chapter 4 narrated the development of the research concept for the study; Chapter 5 listed the methods selected to conduct the research; Chapter 6 displayed the results and Chapter 7 stated the discussion and the outlook. The conclusion can be found in Chapter 8, followed by a summary. The appendix described the tools implemented in the research, the R scripts to implement the NSGA-II, and the data of the Yuan River Catchment. Besides, a digital appendix was offered that contains the R scripts and the data used in the study*.

**Due to the agreement with the data providers, the data are not open to the public. If interested, the reader can contact fg-hydrologie@b-tu.de for more information.*

2 Literature review

Section 2.1 summarized the characteristics of the process-based hydrological models. Section 2.2 explained the features and the application of the single-objective calibration. Section 2.3 introduced the multi-objective calibration in terms of the objectives and the algorithms of the representative approaches; Section 2.4 described the uncertainties introduced by the hydrological models.

2.1 Hydrological and nitrogen load models

The nitrogen load to the stream from the land phase goes through complex processes from its varied sources to the final receivers. The processes at the catchment-scale are triggered by the hydrological cycle. The surface runoff, lateral flow, infiltration, and groundwater flow, along with the anthropogenic facilities, e.g. tile drainage system, transport the soluble nitrogen species, e.g. nitrate (NO_3^--N). The soil water content affects the soil oxygen and thus determines the nitrogen transformation processes, e.g. the denitrification. The estimation of the nitrogen load requires the analysis of nitrogen in a spatially distributed manner from its input sources, transformation processes, and the transport paths to the receivers, i.e. waterbody, soil, or plant, which is also determined by the local geographical and anthropogenic factors. Therefore, models to estimate the nitrogen load should be able to consider the heterogeneity of the hydrology and the nitrogen varied in space and time.

Models that simulate the hydrology and nitrogen load, according to its complexity to compute the natural system, can be classified as empirical or conceptual models and the process-based models (Shen et al., 2012; Bouraoui and Grizzetti, 2014). Empirical models implicitly represent the processes of hydrology and nitrogen load with statistical relations or simplified physically-based representations. Export coefficient (Johnes, 1996; Alvarez-Cobelas et al., 2008) is one of the empirical models and it attributes an empirical nitrogen export coefficient to each source. GREEN (Grizzetti et al., 2008) is one of the conceptual models and it considers both point and non-point sources and assigns specific coefficient to describe the consumption of the individual paths of each source. The relation of water quality and nitrogen quantity in the runoff (Shen et al., 2012), which assumes nitrogen from non-point sources only loads when surface runoff occurs, is also applied to predict the nitrogen load from the land. These methods require a smaller amount of data and a

more simplified algorithm compared to the process-based models. However, due to the simplification, the hydrological processes are often not explicitly simulated which leads to a lack of proper temporal and spatial representation of the nitrogen load.

Process-based models, which aim at explicitly representing the individual physical processes of nature, develop the algorithms to simulate the hydrological cycle with physically-based equations and derives the computational unit integrating information of the topography, geology, and climate condition, or even the management operations (Borah and Bera, 2003). Process-based models are capable of computing the complex processes of nitrogen transformation and transport in the soil and the stream. Moreover, the models generate a simulation with a higher spatial and temporal resolution by its discretization approach, to maintain the heterogeneity of the catchment (Arabi, 2010). The commonly applied process-based models are often coupled systems and integrated models. The coupled systems combine the hydrological model with an independent nitrogen cycling model or other additional processes, e.g. HBV-N/HYPE that couples the watershed scale water balance model HBV (Bergström, 1995) with the one-dimensional nitrogen transformation model SOILNDB (Johnsson et al., 2002). The integrated process-based hydrological models compute multiple hydrology relevant processes in one model, e.g. HSPF (Bicknell et al., 1997) and SWAT (Arnold et al., 1998), which integrate the hydrological, chemistry, management of hydraulic construction, etc., within the model system.

The focus of this research is on the process-based hydrological models and the following classifications are within the range of process-based hydrological models.

2.1.1 Characteristics of the process-based hydrological model

2.1.1.1 Spatial scale

Based on the spatial scale, the process-based models can be categorized as point-scale, field-scale, and watershed-scale (Moriasi et al., 2012; Arnold et al., 2015). Point-scale models often simulate primarily the flow, heat and solute transport processes at vertical direction through soil column in a finer time scale (e.g. minutes) by solving equations, e.g. Darcy's law (Darcy, 1956) to simulate saturated flow, and the advection-reaction-dispersion equation (Bear, 1988) to simulate the reactive transport process. Point-scale models are often applied at locations where parameters are well captured and examples include HYDRUS (Simunek et al., 2012) and STANMOD (van Genuchten et al., 2012). Field-scale models are often developed to simulate the soil-water-plant

system to predict the effect of management practices on plantation activities. The models commonly simulate the system in a spatial unit with homogeneous weather, soil, and cropping system. Field-scale models calculate more detailed processes of plant growth and agricultural management processes when compared to watershed-scale models. Examples of field-scale models include EPIC (Sharpley and Williams, 1990) and CREAM (Knisel, 1980). Watershed-scale models are developed to simulate the hydrological processes within catchments of heterogeneous topography, weather, landuse, and soil. The models also combine terrestrial processes and channel processes. Compared to the field-scale and point-scale models, the watershed-scale models tend to simplify some of the processes, e.g. a homogeneous horizontal water content distribution at a soil layer within a computational unit to enable the practical computation of watersheds with diverse conditions geographically.

The discretization of the domain by the watershed-scale model also determines the model's application to individual usage. The discretization is generally categorized as lumped, semi-distributed, or gridded. A watershed is firstly delineated into sub-watersheds based on its topography. Lumped models uniform the characteristics of individual sub-watershed and gridded models discretize the watershed into structured grids. The semi-distributed model tends to retain the hydrological characteristics and preserve the heterogeneity in one sub-watershed. Table 2-1 listed the spatial/computational unit adopted by the four widely applied watershed-scale hydrological models.

2.1.1.2 Temporal scale

Watershed-scale models, which were developed to support water and landuse management, are often designed to adopt a temporal scale from daily or sub-daily to monthly and annually (Borah and Bera, 2003). The temporal scale enables the model to capture the response of the catchment within a specific period and it is categorized into event-based (the temporal scale is allowed to be hourly to seconds) or long-term based (the temporal scale is from daily to annual). Event-based models are designed to simulate the water and pollutant dynamics in a short period, e.g. a flood event. Long-term based models aim to predict the impact of management practices for years. The variation of the time scale determines the model's algorithms to simulate the identical hydrological process (Moriasi et al., 2012; Arnold et al., 2015). The Green-Ampt approach applied by SWAT assumes a wetting front between the saturated and the unsaturated zone to represent the water

content change and it is widely used by the long-term based models. Richard equation fully captures the dynamics of water moving vertically in the unsaturated zone and it is adopted by event-based models, e.g. MIKE-SHE. Due to the different algorithms applied by models with different temporal scales, the respective spatial heterogeneity of the input data also varies (see Table 2-1 and Table 2-2).

Table 2-1 Spatial unit and temporal scale of the selected watershed-scale hydrological models

Category	Model	Spatial/computational unit	Temporal scale
Lumped	AnnAGNPS (Bingner et al., 2011)	AnnAGNPS cell: subcatchment determined by DEM and user-defined size of threshold and length of the channel, and assigned with dominant landuse and soil type	Constant at sub-daily, daily and long-term
Semi-distributed	SWAT (Neitsch et al., 2011)	HRU (hydrologic response unit): regions with identical landuse, soil and slope class combination	
	HSPF (Singh et al., 2005)	Subcatchment divided into the impervious and the pervious region	
Fully-distributed	MIKE-SHE(Refsgaard and Storm, 1995)	2-D Fixed rectangular grids with the resolution corresponding to the input spatial data	Flexible, from minutes to days

2.1.1.3 Computing approaches

The processes computed by the hydrological models are governed by equations with varying complexity in the same model and between different models. The fundamental processes computed are hydrology with subordinate modules linked to sediment and chemical processes. Additional modules include plant growth, management strategies, hydraulic constructions, etc.

2.1.1.3.1 Hydrology

The flow governing equations, which simulate the excess rainfall and surface runoff or overland flow, can be categorized into the empirical or conceptual based approach and mechanical based approach. The former approach conserves the mass balance by the water budget within the time step, e.g. SCS curve number method applied by the SWAT and the AnnAGNPS. The mechanical approach applies the continuity equation of mass, momentum, and energy, e.g. diffusive wave equation by MIKE-SHE and storage based equation by HSPF. The SCS curve number method and the storage based equations are robust approaches but are sometimes oversimplified, whereas the

diffusive wave equation simulates flow dynamics but is more computationally intensive. The methods of the individual hydrological processes adopted by the selected process-based models are listed in Table 2-2. Besides the basic hydrological components, some models also capture specific hydrological processes, e.g. pumping rate and seepage flow simulated by the SWAT model.

Table 2-2 The approaches computing individual hydrological processes

Examples	Surface runoff/overland flow	Infiltration and lateral flow	Channel process	Groundwater flow
SWAT	SCS curve number method	Green-Ampt infiltration approach and kinematic storage for lateral flow	Variable storage or for routing and Manning's equation for flow	The exponential decay weighting function
AnnAGNPS	SCS curve number method	Darcy's equation for lateral flow	Manning's equation	Hooghoudt's equation for subsurface drainage flow
HSPF	Outflow depth and detention storage relation	Empirically based areal distribution of infiltration	Manning and continuity equation	Darcy's equation for gravity flow
MIKE-SHE	2-D diffusive wave equation	Richard equation	The 1-D diffusive wave equation	The 3-D groundwater flow equation

2.1.1.3.2 Nitrogen transport and transformation

Table 2-3 below lists three of the representative process-based models that simulate complete nitrogen transport and transformation processes from the source to the receiving waterbody for long-term analysis at catchment-scale. The similarities of the models are the consideration of the dynamics of the individual nitrogen species from its sources to the receivers. The differences lie in the approach used for individual processes and whether specific processes are computed by the model, e.g. the septic tank system simulated by SWAT, and the nitrogen dynamics in the reservoir or in-the stream dynamics by SWAT and HSPF.

Table 2-3 Nitrogen processes computed by SWAT, HSPF, and AnnAGNPS

Processes	SWAT	HSPF	AnnAGNPS
Nitrogen input	The input nitrogen (N) can be set spatially as direct N input from the sources, e.g. manure or fertilizer, atmospheric deposition	The input N can be set spatially and however, only computed as a coefficient of specific land uses instead of as a net increase to the spatial unit; atmospheric deposition is also computed	The input N can be set spatially as direct N input from the sources, e.g. manure, fertilizer, atmospheric deposition
Nitrogen transport	Inorganic N by plant-uptaken; nitrate in runoff, infiltration, or leaching to the subsurface; organic N in sediment	Nitrate and ammonia N by plant-uptaken; soluble N in the runoff or infiltration; particulate organic N in the sediment	Inorganic N by plant-uptaken, in runoff or leaching to subsurface flow; organic N in sediment
Nitrogen transformation in the soil	Mineralization / immobilization, denitrification, nitrification and volatilization	Mineralization / immobilization, ammonium N adsorption/desorption, denitrification, nitrification and volatilization	Mineralization, denitrification
In-stream processes	Algal consumption, riverbed organic N exchanging and other biochemical processes	Biochemical transformations of nitrate, ammonia N, and refractory organic N	No direct transformation is calculated but an equilibration model to transfer between absorbed and solute N

2.1.2 The application to nitrogen load analysis

In terms of nitrogen load analysis, a comprehensive understanding and a sound estimation of the nitrogen from its sources to its load to the final receiving water body is critical to deriving a practical nitrogen management strategy, especially for the nitrogen from the non-point sources (NPS). Critical source areas (CSAs) identification is one of the approaches to interpret the simulated nitrogen load and is regarded as the prerequisite of developing a cost-effective strategy for catchment management (Lindgren et al., 2007).

The spatial distribution of the nitrogen sources and their contribution as nitrogen load to the stream are not homogeneous across the catchment, due to the differences of topography, weather, agricultural operations, and management strategies. However, only a few regions of catchments are producing the highest nitrogen load to the stream (Ouyang et al., 2009; Liu et al., 2016). Case studies have shown that area with less than 10% loads more than 50% of the total nitrogen (TN)

in the stream (Folle et al., 2007; Niraula et al., 2011). The region contributing to a high load of nitrogen is defined as critical source areas (CSAs). Moreover, the varied nitrogen transformation processes in the soil and the stream make it challenging but also necessary to analyze the in-stream nitrogen contribution from point sources or non-point sources and their dynamics spatially and temporally. (Yang et al., 2013; Yang et al., 2016)

2.2 Single-objective calibration

The catchment-scale process-based models, especially those applying empirical equations, require quantities of parameters to compute individual processes. The parameters can be categorized as physical parameters and process parameters (Sorooshian and Gupta, 1995). The former is determined by the physical characteristics of the catchment, e.g. slope, elevation, or land cover, which is obtained from measurements or approximation based on the reality; the latter, however, cannot be measured or estimated by prior information (Beven, 2012), e.g. Manning's n value for the overland flow applied by the SWAT model. Therefore, calibration, which is to adjust the sensitive parameters to derive the simulation with the least residue to the observations, is essential to evaluate the model's capability. The difficulty of calibrating process-based models lies in the validation of the spatially varied catchment characteristics and the interacted complex processes. Mainly due to the lack of information on parameterization or a better representation of specific natural processes, the model is not able to calibrate each process and at each location. The commonly applied procedure to calibrate the process-based hydrological model is by screening the parameter values within their reasonable value ranges and thus to minimize the residual between the simulated and the observed variables. The parameter sets, which lead the model to reach the residue minimum of the simulated variables, are the suitable and calibrated parameter sets. Due to equifinality (Beven, 1993), more than one parameter set could be obtained. The model updated with any of those parameter sets is referred to as calibrated.

The classic calibration procedure is to calibrate at least the streamflow flux (e.g. discharge or water quality component) at the catchment outlet by evaluating the residue. The magnitude of the residue is evaluated by one or two residual-based summary statistics (Gupta et al., 2008). Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970) and Percent Bias (PBIAS) (Gupta et al., 1999) are among the most commonly applied indexes to quantify how good the simulated variables are to fit the observed ones. If only one variable is evaluated, the calibration procedure is referred to as the

single-objective/criteria calibration (Efstratiadis and Koutsoyiannis, 2010). A commonly applied procedure is described as follows (see Figure 2-1) that 1) select the sensitivity parameters and obtain their reasonable value ranges; 2) select the random parameter sets recursively by a sampling technique, e.g. the Latin Hypercube Sampling (LHS) technique (McKay et al., 1979), and run the simulation sequentially; 3) evaluate the performance of each simulation on the targeted variable (the objective) by evaluation statistics; 4) obtain the calibrated parameter sets that meet the criteria, e.g. NSE larger than 0.5 for daily streamflow (Moriasi et al., 2015b).

Auto-calibration is discussed in this section, which is less subjective and capable of searching individuals from an extensively large number of parameter sets with an acceptable effort. There are varied single-objective auto-calibration approaches developed to handle the relatively large number of parameters applied by the process-based models. SUFI-2 (Sequential Uncertainty Fitting, ver.2) (Abbaspour, 2005) assumes the uniform distribution of the parameter values to perform the LHS and it evaluates the parameter set based on the Cremer-Rao theorem to optimize the parameter value range. SCE-UA (Shuffled Complex Evolution) (Duan et al., 1992) also applies the random sampling to obtain the objective function for each simulation, which is evaluated by the competitive complex evolution algorithms. GLUE (Generalized Likelihood Uncertainty Estimation) (Beven and Binley, 1992) estimates the weights of probabilities for the parameter sets and obtains the prediction probability based on the posterior distribution of the model derived from a subjective likelihood measure. The differences between the approaches are the algorithms applied to generate the parameter set clusters for calibration and the algorithms to derive the parameter uncertainty (Yang et al., 2008).

To derive the calibrated parameter set for a hydrological model, the single-objective calibration should at least be performed. The model, after calibration and validation, can obtain a more reliable simulation of the targeted variable or objective. However, the limitations of the single-objective approach are also obvious. One drawback is to neglect the tradeoff among interactive processes and the responses at the interior location of the watershed, e.g. only to calibrate the nitrate load in the stream will lead to the fuzzy estimation of the nitrate contributed by surface and subsurface flow, thus to mislead the nutrient abatement strategies (Yen et al., 2014). Secondly, if the model is only calibrated at one location, e.g. the outlet, the heterogeneity of the catchment cannot be represented by the model, which will lead to the parameters selected inconsistent with their

physical meanings (Zhang et al., 2008). Thirdly, applying only one objective function tends to derive a biased assessment, e.g. the NSE emphasizes the high values more than the low values when evaluating the fitness (Pushpalatha et al., 2012) and leads to a less reliable simulation of the low flow scenario, which is however critical for hydrology and water quality analysis. Therefore, the prediction uncertainty of single-objective calibration could be relatively high especially when the uncalibrated processes are also critical for assessment. The single-objective calibration is also regarded as a biased approach (Arnold et al., 2015).

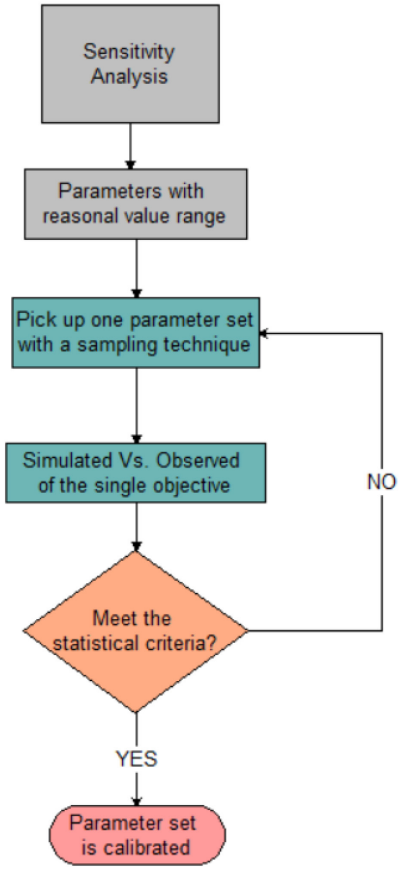

Figure 2-1 The workflow of the commonly applied single-objective calibration procedure

2.3 Multi-objective calibration

To overcome the drawbacks of single-objective calibration, multi-objective calibration for hydrological models as a diagnostic approach (Gupta et al., 1998) was proposed. Instead of only constraining parameter uncertainty via the residue-based objective function for one variable, it proposed also to use signature measures that are theoretically relevant to the system process, e.g. quantitative description of a process characteristics, as the additional variables/objectives (Gupta

et al., 2008; Yilmaz et al., 2008). When compared to single-objective calibration, multi-objective calibration is regarded as a more balanced calibration approach (Efstratiadis and Koutsoyiannis, 2010).

2.3.1 The objectives applied in the calibration

Definition: the term "objective" in this study is a variable to be calibrated in the model.

Studies have been conducted to include varied objectives for calibration. Van Werkhoven et al. (2009) proposed two types of objectives/metrics: statistics metrics and hydrologic metrics. Statistics metrics were the objective functions tackling the different aspects of the simulated time-series, e.g. the NSE for peak flow and the transformed root mean square error (TRMSE) to account for the low flow. Hydrologic metrics are the auxiliary information, e.g. the slope of the normal flow (van Werkhoven et al., 2009; Zhang et al., 2016) or the baseflow index (Arnold et al., 2000). Pushpalatha et al. (2012) also proposed to apply different objective functions as efficiency criteria, e.g. the transformed NSE to account for both peak and low values of time-series. Seibert and McDonnell (2002) proposed to categorize the data for calibration to be soft and hard data. Whereas hard data refer to the measured time-series, e.g. daily discharge or water quality data, as used in the classic calibration approach and soft data is the information of individual processes, which can hardly be directly measured at a frequent time step, e.g. the regional estimation of the water balance components. (Arnold et al., 2015)

The above studies on multi-objective calibration can be summarized to improve the model performances by one of the following aspects: 1) to better represent the parameters in a spatially distributed manner; 2) to match different phases of the measured time-series, e.g. the hydrograph; 3) to evaluate the performance of multiple hydrological components. An effective approach to achieving the first aspect is to apply the measured time-series at the internal stations instead of only at the outlet, which is mostly applied when multiple measurements are available and is referred to as multi-site calibration (Zhang et al., 2008; Lerat et al., 2012). The second aspect requires a completed analysis of the hydrograph itself, to extract additional information to represent individual periods of the hydrograph, e.g. low flow period (Pfannerstill et al., 2014; Zhang et al., 2016; Haas et al., 2016) or baseflow index (Arnold et al., 2000; Zhang et al., 2011b). Moreover, the proper objective functions should also be selected, e.g. different objective functions to emphasize peak values and low values (Pushpalatha et al., 2012; Molina-Navarro et al., 2017).

The third aspect requires the involvement of external data which can quantify additional processes in the model (Nair et al., 2011; Yen et al., 2014; Pfannerstill et al., 2017), e.g. the water balance condition of the catchment as soft data. Those external data are often regarded as the input of expert knowledge to ensure a more reasonable and consistent model structure and simulation (Arnold et al., 2015).

As a conclusion of the above state of the art, a conceptual approach in this study is developed that the objectives are classified into three groups, which are 1) multi-site, including the internal sites of observed discharge or water quality; 2) multi-objective-function, evaluating different phases of the multi-site objectives; 3) multi-metric, including metrics extracted from hydrograph and soft data, each assessed by proper evaluation statistics.

2.3.1.1 Multi-site

Multi-site aims to handle the spatial variation of the parameterization especially for process parameters, which are difficult to quantify by field measurements. Single-objective calibration tends to calibrate the parameters in a lumped manner, e.g. assigning one value for a parameter over the entire catchment. If multi-site is applied, each group of subcatchments controlled by each hydrometric or water quality station can be assigned with an individual parameter value. Therefore, the parameterization varied spatially according to the internal catchment measurements.

Studies have analyzed the effect of multi-site calibration on both discharge and the water quality components and they have shown an obvious improvement in the gauged location where the measurements were taken (Zhang et al., 2008; Lerat et al., 2012; Shrestha et al., 2016). However, a trade-off at the outlet is also inevitable. Two approaches were commonly applied. One is to calibrate in the sequential order from the upstream site to the downstream site that the parameters at the upstream are first calibrated and fixed before calibrating the downstream subcatchments. The second approach calibrates the parameters with a distinct value to each subcatchment cluster simultaneously.

2.3.1.2 Multi-objective-function

The objective function is used as the evaluation statistic to quantify the goodness-of-fit of the simulation to the observation.

NSE, as a dimensionless measure, is one of the most widely applied objective function to evaluate the performance of time-series. It has the advantages of the good applicability on continuous long-term simulations of different variables, e.g. flow and nutrients, and it can robustly represent how well the performance is to fit the 1:1 line plotted from the simulation and the observation. (Harmel et al., 2010; Molina-Navarro et al., 2017). However, NSE, as a distance-based or residue-based objective function (Legates and McCabe, 1999), has its drawbacks. It is more sensitive to the peak values that it squares the residue between the simulation and the observation and thus the low values are less emphasized (Krause et al., 2005).

For hydrological studies, the low flow period is critical to nitrogen load analysis that the deteriorating water quality often occurs during this period due to a weaker self-purification ability of the waterbody. To evaluate low flow, the transformed NSE was also applied by many studies (Krause et al., 2005; Le Moine, 2008; Pushpalatha et al., 2012). The transformation of the original NSE aims to decrease or inverse the magnitude of the peak flow. Table 2-4 lists the transformations of logarithms-based, root-square-based, and inverse-based form. The commonly applied transformed NSE is logarithms-based, however, it is more sensitive to low values than NSE but also sensitive to peak values. The inverse NSE, due to the inverse transformation, shows no sensitivity to peak flows (Pushpalatha et al., 2012).

The transformation of the original objective function was implemented in many studies (Gupta et al., 1999; Bekele and Nicklow, 2007; Confesor and Whittaker, 2007; Shafii and De Smedt, 2009; Herman et al., 2018). The conclusion has been apparent that multi-objective-function could derive a more reliable simulation at the low values. In the above studies, two different approaches were applied. One approach was to implement the original formula for the peak values, and the transformed one for the low values. When the log NSE was implemented, the separation of the time-series would avoid the log NSE's sensitivity to peak values. The other approach was to apply both the original formula and the transformed formula at the entire time-series, which avoids the subjective separation of peak and low flows.

Table 2-4 NSE and the transformed NSE

Objective functions	Formulation	Reference	Characteristics
NSE	$1 - \frac{\sum_{i=1}^{n}(Q_i^{obs} - Q_i^{sim})^2}{\sum_{i=1}^{n}(Q_i^{obs} - Q_{mean}^{obs})^2}$	Nash and Sutcliffe (1970)	Sensitive to peak flows
Root square NSE	$1 - \frac{\sum_{i=1}^{n}(\sqrt{Q_i^{obs}} - \sqrt{Q_i^{sim}})^2}{\sum_{i=1}^{n}(\sqrt{Q_i^{obs}} - \sqrt{Q_{mean}^{obs}})^2}$	Oudin et al. (2006)	All-purpose efficiency measure, not specifically focusing on low flows
Log NSE	$1 - \frac{\sum_{i=1}^{n}(lnQ_i^{obs} - lnQ_i^{sim})^2}{\sum_{i=1}^{n}(lnQ_i^{obs} - lnQ_{mean}^{obs})^2}$	Krause et al. (2005)	Higher sensitivity to low flows but also sensitive to peak flows
Inverse NSE	$1 - \frac{\sum_{i=1}^{n}(\frac{1}{Q_i^{obs}} - \frac{1}{Q_i^{sim}})^2}{\sum_{i=1}^{n}(\frac{1}{Q_i^{obs}} - \frac{1}{Q_{mean}^{obs}})^2}$	Le Moine (2008)	Sensitive to low flows

2.3.1.3 Multi-metric

Multi-metric discussed in this section is the additional data other than the objectives of multi-site evaluated by the multi-objective-function, e.g. observations of additional hydrological or nitrogen relevant processes as soft data, and the data extracted from existing observed time-series. The attempt to consider the multi-metric is to set criteria for individual watershed processes (Yen et al., 2014) to reduce the uncertainty of the processes. Though the extra metrics are often limited by data availability and with less accuracy in terms of spatial and temporal scale, they help to ensure the relevant processes within a reasonable boundary (Seibert and McDonnell, 2002).

2.3.1.3.1 Water balance components

Water balance components, e.g. actual evapotranspiration (AET), runoff coefficient, infiltration rate, lateral flow rate to the discharge, etc., are rarely measured at a competitive scale to discharge or water quality components but can be used as descriptive data (Arnold et al., 2015). AET, groundwater return flow, and surface runoff have been applied in many studies as the additional information to calibrate the water balance processes. Pfannerstill et al. (2017) obtained the mean annual water balances by estimating the ratio of the evapotranspiration (ET), baseflow, and surface runoff to precipitation. Studies (Zhang et al., 2011b; Arnold et al., 2000) also calibrated the baseflow index and surface runoff ratio to better represent processes determining the discharge.

Immerzeel et al. (2008) and Herman et al. (2018) used satellite-based ET to calibrated the simulation process of AET.

One limitation to apply the water balance components is the lack of reliable measured or estimated data. However, the open-access global satellite-based data sets, e.g. the MODIS global terrestrial ET product from NASA (Running et al., 2018) that provides the long-term AET globally could be an option (van Griensven et al., 2012; Abiodun et al., 2017). Besides the direct water balance components, other relevant processes, e.g. soil water residence time (Vaché and McDonnell, 2006) and the shape of the recession curve (Winsemius et al., 2009) were also applied in the previous studies as metrics for calibration. Moreover, metrics could be obtained from the existing hydrograph, e.g. the baseflow index filtered from hydrograph (Arnold et al., 1995; Ladson et al., 2013), which works as the supplementary information to assess the model's performance at the low flow period and is also recommended to be considered in the calibration. (Zhang et al., 2011b; Arnold et al., 2012a)

2.3.1.3.2 Metrics extracted from the hydrograph

The single-objective calibration applying one objective function as the performance measure can hardly account for different phases of the hydrograph. Studies indicated that different periods of the hydrograph are dominantly controlled by different hydrological processes and can be captured by hydrograph analysis (Carillo, 2011). Pfannerstill et al (2014) analyzed the magnitude of different flow periods by evaluating the goodness-of-fit of the flow duration curve sections. Zhang (2016) proposed to generate metrics that represent the timing of a hydrograph, e.g. the day of annual maximum discharge, to optimize the simulation of flow characteristics and the parameter uncertainty.

2.3.1.3.3 Nitrogen metrics

The nitrogen load prediction is with higher uncertainties than streamflow simulation in terms of input data, model algorithms, and the input parameters. Nitrogen variables, especially in the data-scarce catchment, have a longer sampling time step, e.g. weekly, monthly, or seasonally when compared to flow measurements. Therefore, the information extracted from the measured time-series of the nitrogen load is rather limited. However, the complex transformation processes among the different species in both land phase and the in-stream phase make the calibration only based on the measured nitrogen load in the stream less reliable, especially if the measured nitrogen

species is limited, e.g. only with total nitrogen (TN) measured. Studies tried to calibrate the nitrogen transformation processes indirectly. The expert knowledge of the annual denitrification rate was applied to quantify denitrification (Yen et al., 2014), which was the main process to determine the nitrate loss in the soil. Calibration of the crop yield was also implemented to quantify the reduction of water and nitrogen in the soil, as plant uptaken is the main loss of the water and nitrogen (Nair et al., 2011; Malagó et al., 2017). Haas (2016) has proposed to obtain the magnitude index for calibration from the nitrate duration curve, which was plotted by the time-series of daily nitrate to describe the low to high nitrate level in the stream.

2.3.2 The calibration algorithms

2.3.2.1 The multi-objective optimization concept

Multi-objective calibration optimizes the parameters by employing vector search instead of scalar search to meet the multiple criteria that evaluate each objective. In multi-objective calibration, one simulation $\overrightarrow{S_\iota} = [b_1^i, b_2^i, \dots, b_m^i]$ is a vector containing all the objectives and b_m^i is the value of the evaluation statistic of one objective. Theoretically, if any b_m^i of $\overrightarrow{S_\iota}$ is the optimum compared to other simulations, $\overrightarrow{S_\iota}$ is the optimal solution. However, due to the existing conflicts between objectives in the hydrological modeling system, it is often not feasible to optimize all the objectives simultaneously, thus trade-offs in the solutions should be considered.

In multi-objective optimization, the dominance, e.g. $\overrightarrow{S_1}$ dominates $\overrightarrow{S_2}$, is defined as 1) $\overrightarrow{S_1}$ is no worse than $\overrightarrow{S_2}$ in any objectives; 2) $\overrightarrow{S_1}$ has at least one objective better than $\overrightarrow{S_2}$ (Deb et al., 2002). Otherwise $\overrightarrow{S_1}$ is not dominated by $\overrightarrow{S_2}$ and $\overrightarrow{S_1}$ is the non-dominated set of $\overrightarrow{S_2}$. A scenario is displayed in Figure 2-2. A and B are two objectives. Magnitudes at the axis stand for the value of the evaluation statistics for A and B; the closer the objective function values to zero are, the better the solution is. Each point stands for one simulation. Based on the dominance definition, points 1, 2, and 3 are the dominant solution to the rest and the three points form the Pareto Front. Pareto Front is defined as the optimal solutions, consisting of a set of non-dominated solutions, and no objective of those solutions can be improved without sacrificing at least one other objective. Points that lie upon the Pareto Front are the optimal solutions. Those solutions are not dominated by the other solutions at the Pareto Front but have equivalent performance when compared with each other. If there are more than two objectives, the Pareto Front is multi-dimensional.

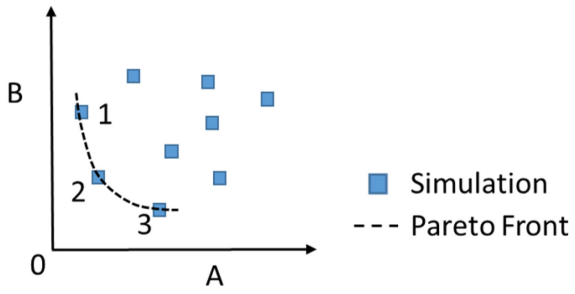

Figure 2-2 Pareto Front obtained by the optimization of two objectives A and B

2.3.2.2 Algorithms of multi-objective calibration

2.3.2.2.1 The classic aggregation approach

The aggregation approach is to aggregate all evaluation statistics into a scalar function, which converts the multi-objective optimization into the evaluation of one scalar function. The typical aggregation method is the weighted average that assigns an individual weight to each of the objectives to obtain their arithmetic mean value (Zhang et al., 2016). Another aggregation method is the Euclidean Distance (Gupta et al., 2009; Haas et al., 2016; Pfannerstill et al., 2017), which calculates the Euclidean Distance of the objectives.

Previous multi-objective calibration studies applied the aggregation approach to evaluate the performance of a relatively larger number of variables, e.g. 16 flow magnitude and timing objectives by Zhang et al. (2016); 8 flow magnitude and water balance component objectives by Pfannerstill et al. (2017). One advantage of the classic aggregation approach is its easy implementation. Another advantage is its theoretically unlimited number of objectives. However, its disadvantages are also obvious that the scalar function may hide the competition among conflicting objectives (Efstratiadis and Koutsoyiannis, 2010). Therefore, the aggregated scalar cannot represent the optimization of individual objectives. As a scalar indicator, it also bears the disadvantage of single-objective calibration as to have only one best simulation obtained without the consideration of the equifinality.

2.3.2.2.2 The Multi-objective evolutionary algorithms (MOEAs)

The MOEAs was first proposed in the 1980s (Schaffer, 1984). The algorithm applies the parallel search of feasible space, through a set of randomly generated points that evolves based on stochastic transition schemes and then spreads the points along the Pareto Front (Efstratiadis and Koutsoyiannis, 2010). Several MOEAs have been developed and applied in the hydrologic sector,

including the Non-Dominated Sorting Genetic Algorithm (NSGA) and its improved version (NSGA-II, ε-NSGA-II) (Deb et al., 2002; Reed et al., 2003), the Strength Pareto Evolutionary Algorithm-II (SPEA-II) (Zitzler et al., 2001), Multi-objective shuffled complex evolution Metropolis algorithm (MOSCEM) (Vrugt et al., 2003), etc.. The approaches were developed to increase the efficiency of MOEAs of the dominance evaluation and the evolution algorithms (Tang et al., 2006; Zhang et al., 2008; Efstratiadis and Koutsoyiannis, 2010; Kan et al., 2017). MOEAs handle the nonlinear optimization problems and apply the heuristic search algorithm to update the parameter sets. It maintains the equifinality of the parameters by generating the Pareto Front instead of a single solution. However, MOEAs require a rather high computational effort especially when a large number of objectives, e.g. larger than five, are optimized (Tang et al., 2006).

2.3.2.2.2.1 Non-Dominated Sorting Genetic Algorithm-II (NSGA-II)

NSGA-II (Deb et al., 2002) is one of the representative MOEAs to have been widely applied. It is a fast and efficient population-based optimization technique (Ercan and Goodall, 2016). The genetic algorithm (GA) imitates the natural evolution process to derive and optimize the parameter sets. GA has also been proved to perform well on complex optimization problems involving nonlinear and noisy functions (Schwefel, 1995). Figure 2-3 displays the procedure in the GA and the respective natural evolution process. In GA, the parameters are decoded as binary values and simulations with better performance are to be selected for the next generation. The crossover and mutation in GA are also performed on the binary value level, which is then decoded back to the real value. Afterward, the evolution of one generation is completed and this process repeats until the user-defined criteria are met, e.g. the performance of all the objectives is acceptable. Key features of NSGA-II should be pointed out. Firstly, it applies the crowding distance algorithm, which gives higher weight on simulations with the most varied performance if they are not dominating each other, to keep the diversity in a population. Secondly, it keeps the elitism of the population by selection and crossover on the combination of the parent population, obtained from the previous generation and the child population, obtained from the current generation.

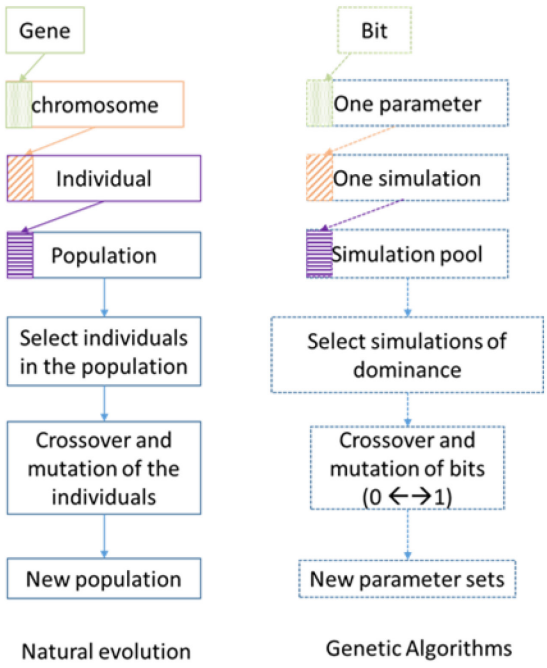

Figure 2-3 Natural evolution steps and the corresponding genetic algorithm steps

Previous hydrological studies have implemented NSGA-II to evaluate 2 to 3 objectives, e.g. the AET and discharge (Herman et al., 2018), the transformed NSE and root mean square error (RMSE) to account for low flow (Bekele and Nicklow, 2007; Shafii and De Smedt, 2009) or multi-site (Confesor and Whittaker, 2007; Ercan and Goodall, 2016). The advantages of the NSGA-II lie in the following aspects that it ensures a less subjective optimization of the objectives by non-dominated sorting, and equifinality by obtaining multiple optimized simulations at the Pareto Front, where the users are to determine the best one according to the existing priorities. Meanwhile, NSGA-II also has a comparably easier algorithm compared to other MOEAs and maintains both the elitism and the variety of the population. However, the applicability and efficiency of the non-dominated sorting are limited by the number of objectives that it can fail to obtain a Pareto Front for a large number of objectives (Tian et al., 2017). The larger the number of the objectives is, the less likely one simulation can dominate the other, also the longer computational time is required to obtain the dominance. Meanwhile, the population size and the number of generations will theoretically introduce uncertainty in the calibration but have not been summarized by previous hydrological studies. Therefore, compared to the single-objective or aggregation approach, MOEAs requires a comparatively high effort for implementation.

2.4 Uncertainty analysis

The uncertainty of process-based hydrological models, in summary, is resulted from the inability to fully representing the natural system. The uncertainty is inevitable due to the intrinsic complexity of the hydrological system (Efstratiadis and Koutsoyiannis, 2010). The sources of the uncertainty include 1) input data uncertainty, e.g. the error of the measuring equipment and the inappropriate spatial or temporal representation of the study region; 2) model structure uncertainty, e.g. assumptions created by the algorithms for simplification and the inconsistency of the model complexity to the available information to describe the system; 3) parameter uncertainty, e.g. the estimated parameter values applied in the model. The uncertainty sources aggregate and result in the prediction uncertainty (Daren Harmel and Smith, 2007; Abbaspour et al., 2018). Figure 2-4 displays the impact of parameter uncertainty, which is represented by a value range instead of a specific value, on the predication uncertainty, which is also led to a value range at every time step. The uncertainty of the model system also results in the fact that an identical output of a variable could be reproduced by multiple different parameter sets. This phenomenon, which is inherent to the process-based hydrological model (Abbaspour et al., 2018), is referred to as non-uniqueness or equifinality (Beven, 1993). Therefore, ignoring the model uncertainty and only deriving one suitable parameter set is a biased calibration result.

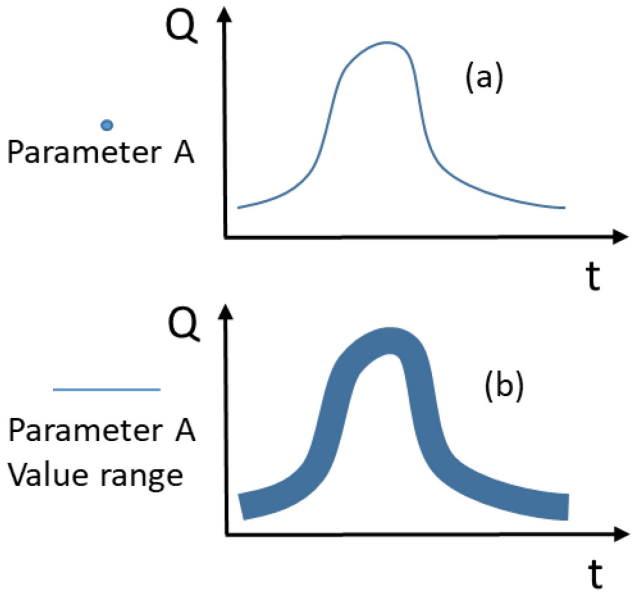

Figure 2-4 The prediction uncertainty due to the uncertainty of parameter A. (a) if parameter A has a certain value; (b) if parameter A is uncertain

Uncertainty analysis is defined as the propagation of uncertainties from all sources to the model outputs (Abbaspour et al., 2018). Yang (2008) concluded the previous application of uncertainty analysis. Bayesian approaches only considered the parameter uncertainty and thus derived the incomplete uncertainty bounds of prediction. GLUE and SUFI-2 are among the approaches that account for all sources of uncertainty by enhanced parameter uncertainty. Markov Chain Monte Carlo (MCMC) is the representative approach that applies the additive error model (Yang et al., 2007) to express the input and model structural uncertainty. Integrated Bayesian uncertainty estimation (Ajami et al., 2007) applies an improved likelihood function to explicitly represent input and model structure errors. Yang also pointed out that SUFI-2 is less computationally demanding to generate prediction uncertainty, whereas GLUE derives the widest uncertainty range and MCMC with a proper likelihood function achieves a good approximation of the maximum of the posterior.

Though the uncertainty is inevitable, the reduction of the uncertainty, besides improving the observed data accuracy and the model algorithms, can also be achieved by calibration. Calibration filters parameter sets from their uncertainty range to satisfy the assumptions and the prior information (Abbaspour et al., 2007), i.e. via narrowing the parameter ranges to meet the performance assessment criteria. However, single-objective calibration generates a good fit for one output variable but no information of the performance of the interior processes, and thus conceals the issues resulting from the uncertainty sources (Efstratiadis and Koutsoyiannis, 2010). In contrast, the multi-objective calibration set criteria to ensure the performance of multiple aspects of the model (Abbaspour et al., 2018), e.g. multi-site objectives for the spatial parameterization, multi-objective-function for individual periods of time-series, and multi-metric objectives for more processes.

3 Study catchment

This section summarized the characteristics of the study case, the Yuan River Catchment. The basic condition, e.g. geography, was described in Section 3.1–3.5. The nitrogen condition was explained in detail in Section 3.6. In Section 3.7, Table 3-2 lists the available data applied in the research with their sources and the remaining part describes the processing of the nitrogen input data from each source.

3.1 Geography

The Yuan River Catchment (YRC) is located in the central west of Jiangxi Province, China (113°50′~115°32′E; 27°22′~28°08′N). Yuan River is one branch of the Gan River, and the Gan River is one tributary of the Yangtze River and flows into the Poyang Lake. The Yuan River network flows through four administrative districts, including Pingxiang City, Yichun City, Xinyu City, and Fenyi Town. The YRC covers an area of 6223.69km^2. Figure 3-1 below displays the digital elevation map (DEM) of the YRC and the administrative districts it covers.

The Yuan River originates from the south of the YRC, at an elevation of approx. 1900m, and joins the Gan River in the north, at an elevation of approx. 20m. The mountainous region (elevation > 500m) is located at the source of the Yuan River and occupies less than 5.7% of the entire catchment. The plain region (elevation < 100m) is located at the flood plain and the downstream of the YRC, occupying approx. 39.5% of the catchment area. The Yuan River has a length of approx. 273km with a deeper slope at the upstream (approx. 1.815%) and a shallower slope at the downstream (approx. 0.026%). In the mid-downstream of the Yuan River, Jiangkou (JK) Reservoir has been operated since the 1960s to generate hydropower and to manage the flood downstream.

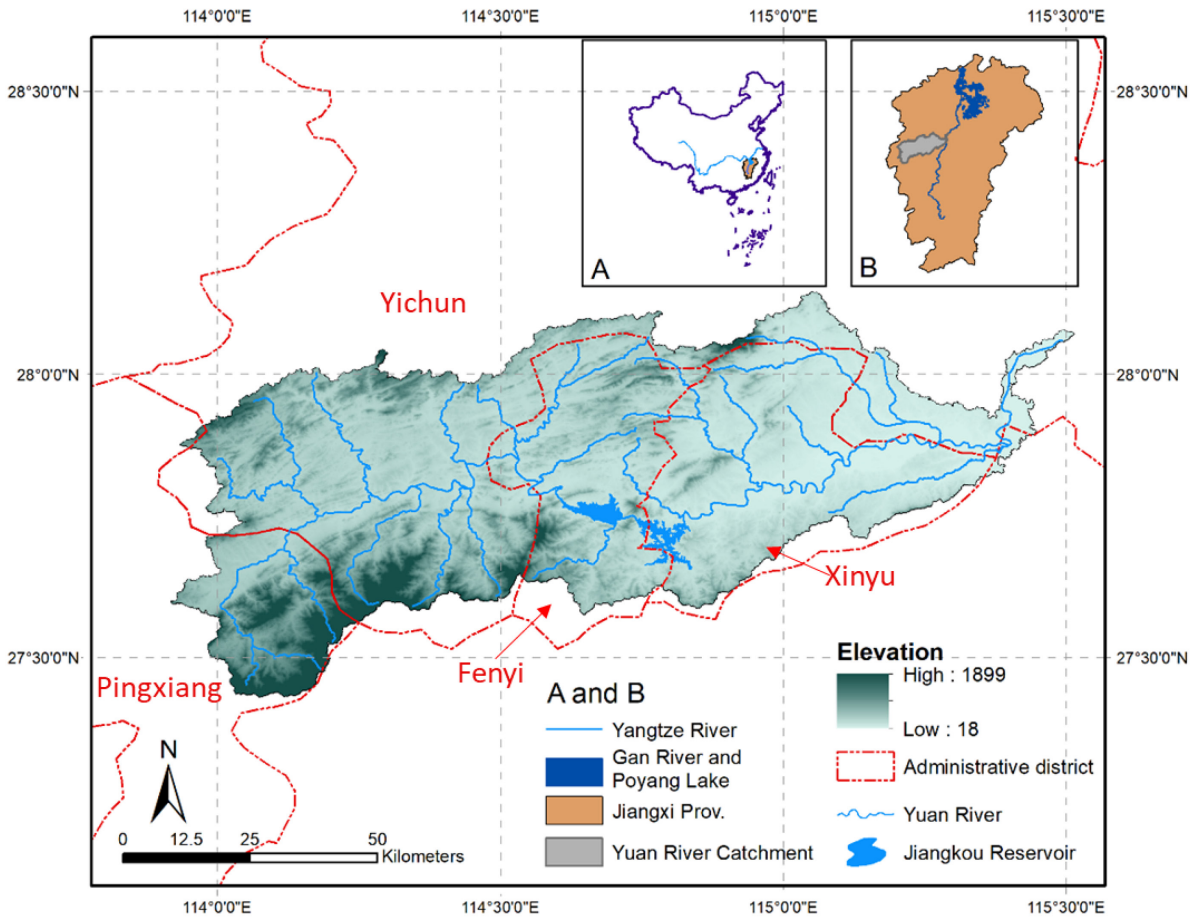

Figure 3-1 The digital elevation model (DEM) of the Yuan River Catchment and the Yuan River network

3.2 Climate

The YRC is located in the subtropical monsoon climate zone. The annual rain season of the YRC is from April to June and it is dominated by frontal rainfall. The long-term statistics from 1980 to 2010, obtained from the National Meteorological Information Center of 16 stations, demonstrated the average annual precipitation to be 1668mm, with the rain season of the three months occupying 43.6% of the total precipitation. The average annual lowest temperature is 2.8°C in January and the highest temperature is 34.2°C in July. The annual pan evaporation rate is approx. 1200mm (Fang, 2011). The potential evapotranspiration (PET) is calculated based on the Penman-Monteith equation (Allen, 1998) and displayed below in Figure 3-2, the annual PET is approx. 1123mm. The monthly average precipitation (PCP) and temperature (T) are also demonstrated in Figure 3-2.

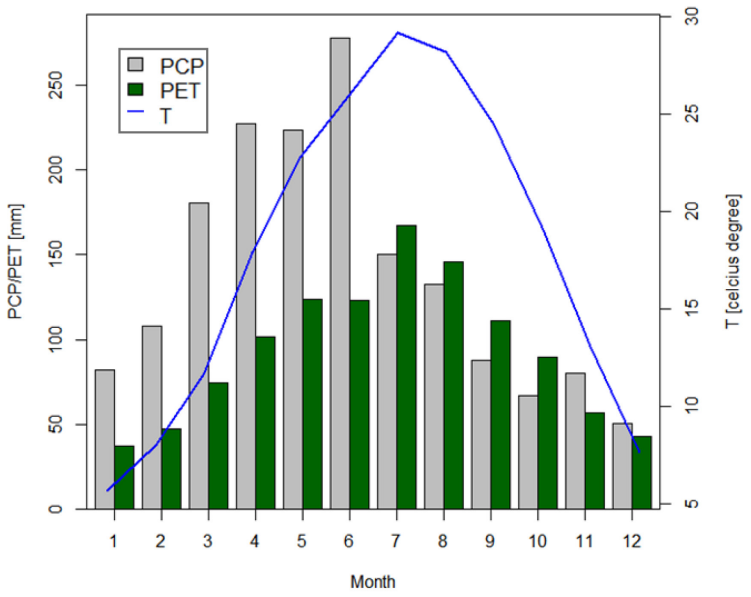

Figure 3-2 Long term monthly climate statistics in YRC

3.3 Landuse and land cover

The landuse distribution in 2010 of the YRC is displayed in Figure 3-3. The main agricultural activity in the YRC is the rice plantation in the paddy field, which occupies 29.8% of the entire catchment and consists of plain, hilly, and mountainous regions. The rice is planted in two periods, from April to July and from July to October. The paddy field in the YRC is mainly distributed along the flood plain and the downstream plain region. Besides paddy field, dry arable field and orchard occupy approx. 5.2% of the land with vegetables and orange as the main crop. Forest, including evergreen forest and shrubland, occupies 54.8%. The total residential area, including industrial land, urban and rural residential, occupies 5.6%. Cities and towns could be observed in Figure 3-3 and small villages are distributed scattered over the YRC, especially in the plain region. The YRC has a population in 2010 of 2.4 million with an average annual growth rate of 7.5‰ (Jiangxi Statistical Bureau, 2008-2014). 54.6% of the population is rural residents, living in the towns and small villages that occupying 4.1% of the YRC.

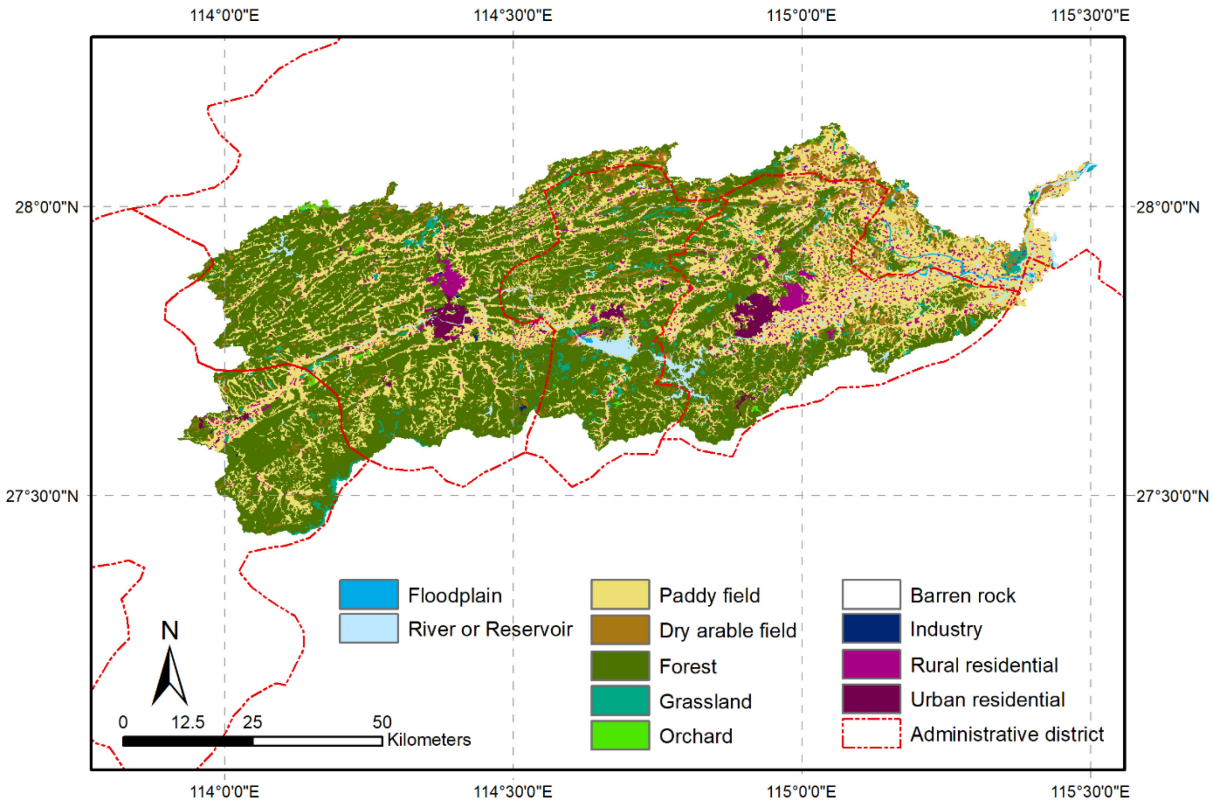

Figure 3-3 Landuse classes of the YRC in 2010

3.4 Soil and geology

Red soil (see Figure 3-4) occupies 50.3% of YRC and it is a typical soil in the southern side of the Yangtze River, which is distributed mainly in the lower hills. Due to the relatively high iron content, its pH ranges from neutral to acidic. Red soil belongs to soil hydrologic group D (USDA-NRCS, 1993), with a high potential to generate runoff. However, the third layer of the red soil has a sharp increase in the hydraulic conductivity (approx. 30mm/hr) compared to the conductivity of the first and second layer (approx. 4mm/hr) that makes the water easier to penetrate, where the content of sand in the soil increases from 30% to 60%. Paddy soil, including the hydromorphic paddy soil (30.2%) and submergic paddy soil (0.3%), is the soil modified by rice plantation activities, which benefits the accumulation of organic matter (approx. 3.3%). Hydromorphic soil belongs to the hydrological soil group C, with a low infiltration rate but higher than group D. It has a lower clay content (approx. 9.2%) and a higher sand content (approx. 54.2%) at the first layer. The soil conductivity from the top layer to the bottom layer ranges around 20mm/hr to 30mm/hr. Paddy soil is also with high nitrification ability which increases the potential of denitrification processes when soil oxygen content is low (e.g. during flood events). Terra Rossa, which is mainly located

in the north of YRC, occupies 9.8%. It is a category of lime soil and is distributed in the relatively dry region when compared to the region where red soil is located.

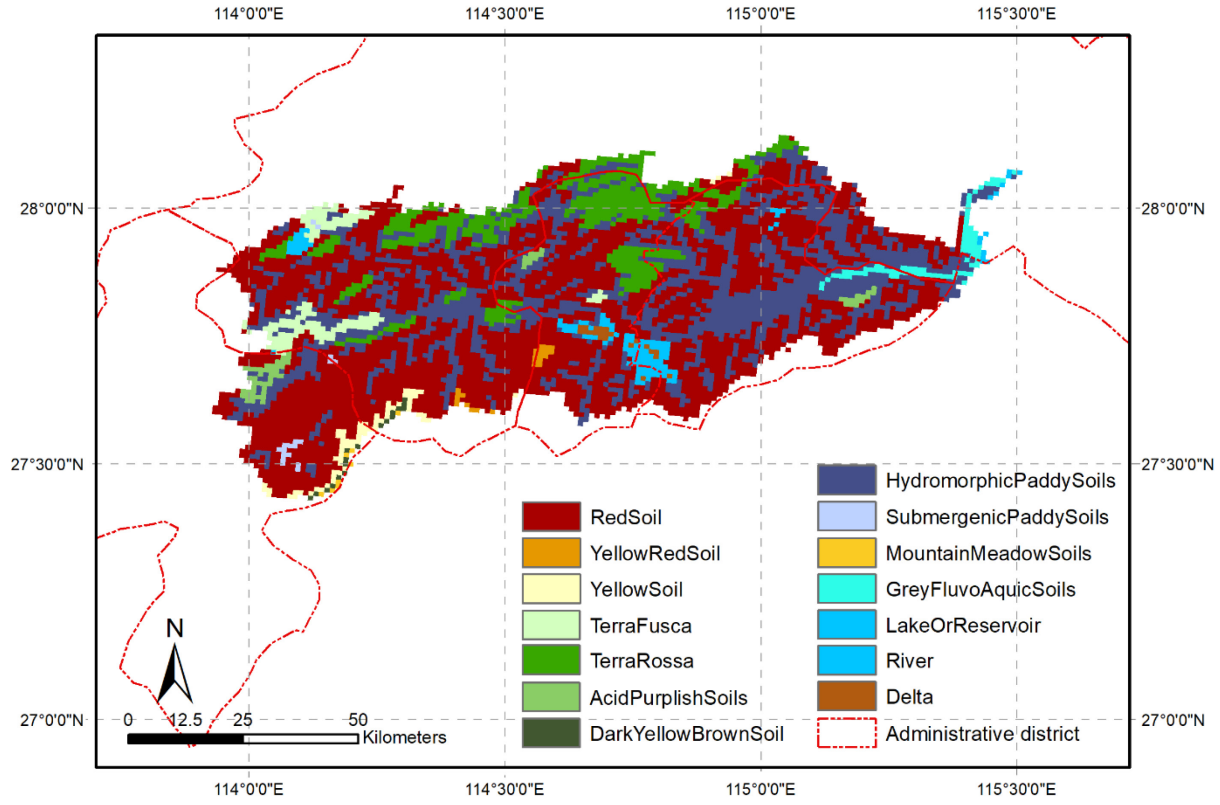

Figure 3-4 The soil type of the YRC

3.5 Hydrology and geology

The discharge of the Yuan River varies seasonally. The flood season is from April to July and occupies approx. 56.7% of the total annual discharge. The average discharge at the Jiangkou (JK) Reservoir (see Figure 3-5) outlet is 111 m^3/s (Fang, 2011). Three hydrometric stations are configured on the Yuan River, LX (Luxi) at the upstream, MZ (Maozhou) at the up-to-midstream, and JK (Jiangkou) at the mid-to-downstream. No hydrometric stations are available in the study at the downstream of the JK reservoir. The available period of the measured daily discharge is listed in Table 3-1. MZ stations had no discharge records since 1989. Therefore, the data from 2008 to 2010 were extrapolated by the linear regression relation between LX and MZ from 1985 to 1987 in this study. The JK reservoir was constructed in the 1960s with the aim of flood management and hydropower generation. The discharge at JK station was calculated based on the measurement of the total volume of the reservoir and the reservoir volume-inflow relation. The management of

the JK reservoir is not considered in the study, instead, a measured daily discharge released by the JK reservoir is used as an input of reservoir discharge in the model set-up procedure. The mean daily discharge at the hydrometric station is also listed in Table 3-1.

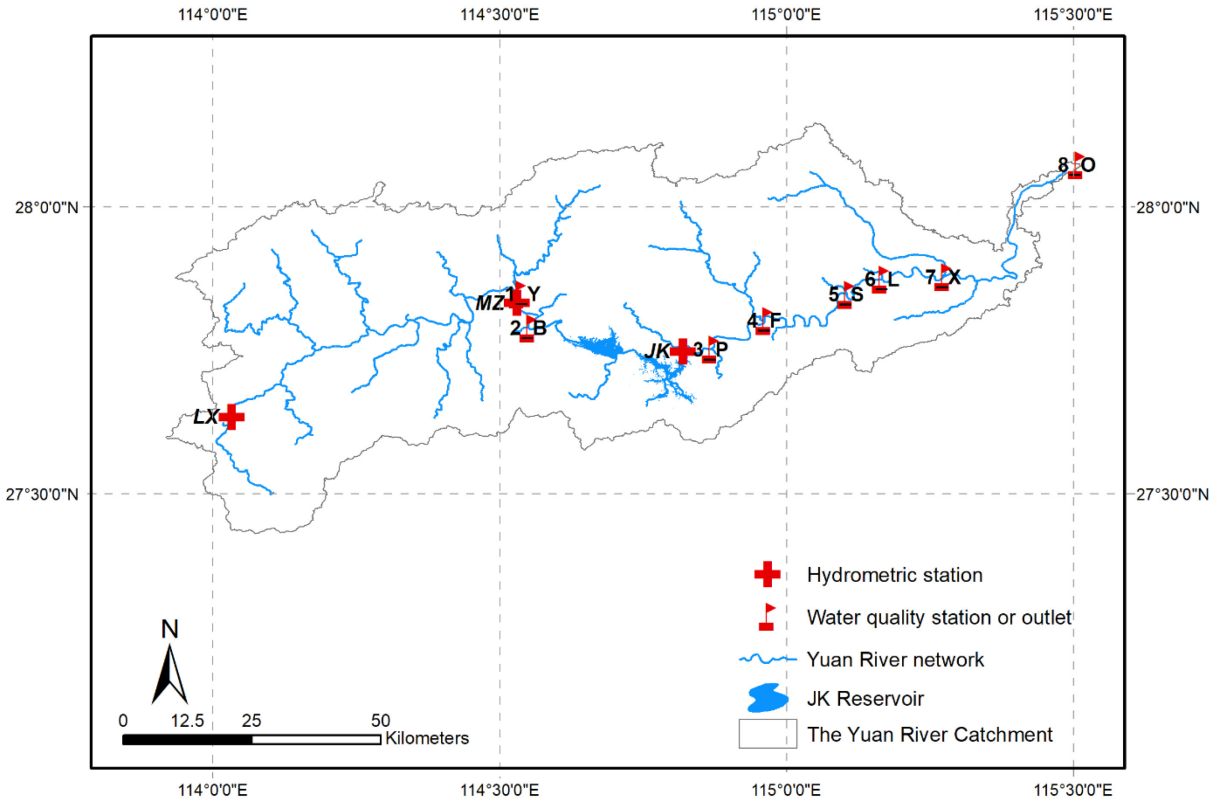

Figure 3-5 The hydrometric stations and the water quality stations

Table 3-1 The available measured daily discharge

Hydrometric stations	Available period	Average daily discharge (m3/s)
LX (Luxi)	1983 to 1986; 2008 to 2010	11.5
MZ (Maozhou)	1983 to 1986; 2008 to 2010 (extrapolated)	100.1
JK (Jiangkou)	2008 to 2014	115.1

The groundwater system in the YRC mainly consists of fracture water in the bedrock, karst water, and pore phreatic water in the loose rock mass. The shallow aquifer and the weak phreatic aquifer are the main groundwater sources for exploitation in YRC and groundwater consists of only 2% of the total water supply (Fang, 2011). Groundwater is mainly recharged in the YRC by

precipitation with a seasonal variation that the baseflow rate is increasing in the rain season and groundwater is recharged by surface flow in the flood season (see Figure 3-6).

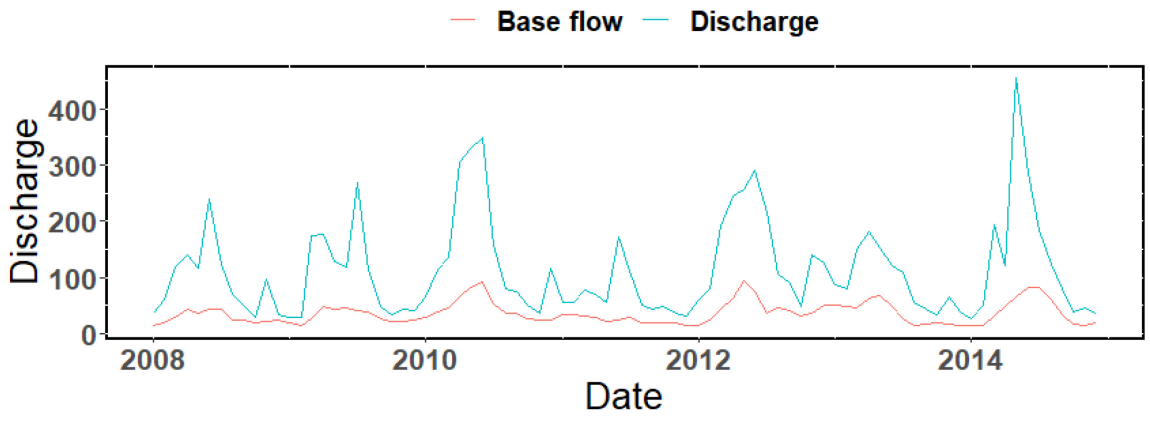

Figure 3-6 The monthly discharge and the base flow rate at JK from 2008 to 2014

3.6 Nitrogen condition

According to the investigation from the local Hydrologic Bureau (Fang, 2011), excess ammonia nitrogen was monitored in and before 2015 in the dry season in the Yuan River. Based on the environmental quality standards for surface water (Ministry of Ecology and Environment of the PRC, 2002), the Yuan River should not exceed the constituents' concentration threshold of category III, with an upper limit of ammonia nitrogen and total nitrogen (TN) no higher than 1.0mg/L. Published water quality data by the local Environmental Bureau and investigations (Huang and Yuan, 2003; Hu et al., 2011; Li et al., 2011) have displayed a deteriorating water quality in terms of nitrogen from the upstream to the downstream. The TN concentration at 17 cross-sections was sampled by Hu (2011) and they are displayed below in Figure 3-7. The cross-sections were located from upstream to downstream and the samples were taken at both low flow and high flow period in March and June in 2010. In this study, the measured concentration of TN was available at three water quality sampling stations, which are 1_Y (Yangjiang), 4_F (Fuqiao), and 6_L (Luofang), as displayed in Figure 3-5. 2_B, 3_P, 5_S, 7_X, and 8_O are outlets demonstrated without observations.

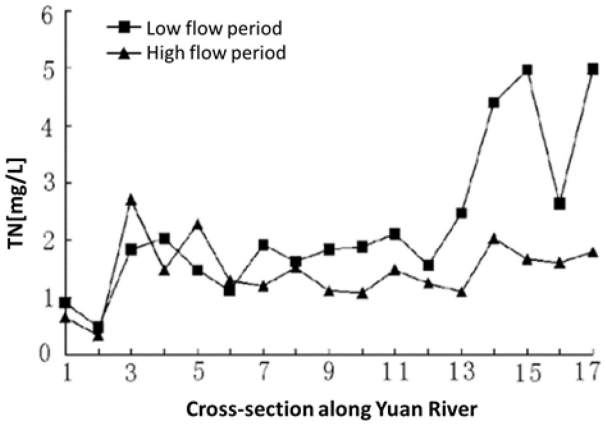

Figure 3-7 Total nitrogen concentration along 17 cross-sections from upstream to downstream of the Yuan River

3.6.1 Nitrogen from point sources

The main nitrogen point sources (PS) in the YRC include sewage discharge from industry and the municipal wastewater treatment plant. Industries are mostly located in Yichun and Xinyu district in the urban region, displayed in Figure 3-1. The discharge of 17 factories and 3 wastewater treatment plants was monitored and published by the local Environmental Bureau every 1 to 3 months, including the concentration of TN and NH_4-N. More details are narrated in section 3.7.1.2.

3.6.2 Nitrogen from non-point sources

The three dominant non-point sources (NPS) in the YRC are the application of nitrogen fertilizer, waste from the unregulated feedlot, and waste from the unregulated rural household. The excess usage of nitrogen fertilizer is a typical issue in the rice plantation region and the YRC is not an exception. The rice plantation requires a large amount of nitrogen fertilizer, e.g. urea or compound fertilizer (Yan, 2008). Waste from scattered feedlots and rural households are often disposed of with little or no treatment (Ongley et al., 2010; Liu et al., 2013). Besides the above NPS that were directly loaded by human activities, nitrogen from atmospheric deposition due to the omission of the waste gas is another source of nitrogen in the YRC. The detailed description can be found in section 3.7.1.2.

3.6.3 Previous studies of nitrogen pollution

Zhang (2011) generated the Loading Duration Curve, i.e. the total loading versus the frequency, to quantify the nitrogen load at different locations of the YRC. The study assumed the load from diffusive sources occurring only in the high flow days due to the higher frequency of rainfall. It

concluded that industrial discharge in Xinyu city was the main contribution of the ammonia nitrogen and the total nitrogen. However, the limitation of the assumption was also obvious that it underestimated the load from non-point sources, especially on days without precipitation. Other studies (Li and Liu, 2015; Li, 2008) were implemented to only sample the sewage discharge along the Yuan River without the analysis of quantifying nitrogen from NPS.

3.7 The available data for the research

Table 3-2 below lists data applied in the study. Based on the review of previous studies, the input datasets, e.g. DEM, landuse, and soil property were of suitable resolution. DEM of 30m×30m resolution is commonly used by previous studies to delineate the watershed. The 30m×30m landuse is able to display the scattered villages distributed in the YRC that determines the location of the nitrogen from the rural households. However, the quantity of the precipitation data, including 12 stations from ground measurements at the upstream of JK station and 6 stations from the assimilation dataset at the downstream of JK station, indicated that one rain gauge in the YRC covered approx. 300km^2.Though the recommended standard by the National Ministry of Water Resources is at least 12 gauges per 3000km^2 (Standard SL 34-2013). In terms of discharge calibration, the observed discharge as stated above was only available at three hydrometric stations (LX, MZ, and JK), including the data from MZ derived from extrapolation. Moreover, LX and MZ only covered three years.

The data listed in Table 3-2 are provided in the digital appendix and can be viewed by requesting fg-hydrologie@b-tu.de.

Table 3-2 The summary of the available data

Data	Resolution or scale	Sources
Hydrology		
DEM	30m×30m	Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Science
Yuan River network	1:200,000	
Landuse	30m×30m	
Soil type and properties	1000m×1000m	Institute of Soil Science, Chinese Academy of Science
Precipitation (2008–2014)	12 stations, daily	JK Hydropower Station
	1/3 degree, daily	China Meteorological Assimilation Driving Datasets for the SWAT Model (2008–2014), Chinese Academy of Science
Climate data (2008–2014)	1/3 degree, daily	
Reservoir discharge	Daily	JK Hydropower Station
Industry and domestic water use (2008-2010)	Annual average	Statistic Year Book of Jiangxi Prov., 2008–2014
Discharge (2008–2010 of LX and MZ; 2008–2014 of JK)	3 stations, daily	Yichun Hydrologic Bureau, JK reservoir power station
Nitrogen		
Crop and fertilization	Fertilizer application per hectare	Local survey and literature
Animal production (2008–2014)	Annual average	Statistic Year Book of Jiangxi Prov., 2008-2014
Rural population (2008–2014)	Annual average	
The effluent of sewage from industry and wastewater treatment plant (2008–2014)	The urban region, seasonally	Yichun and Xinyu Environmental Protection Bureau, public available
Monitored total nitrogen at 1_Y, 4_F, and 6_L (2008–2014)	3 stations, monthly or seasonally	Xinyu Environmental Protection Bureau

3.7.1 Nitrogen data description

3.7.1.1 Nitrogen sources identification

The dominant nitrogen sources identified in the YRC were:

1) Nitrogen fertilizer;
2) Untreated waste from feedlots;
3) Untreated water and waste from rural households;

4) Effluent from industries and municipal water treatment plant as point sources;
5) Atmospheric deposition.

Simplifications were assumed for each source:

1) The fertilizer was applied to paddy field (rice) only;
2) The path of the nitrogen from the feedlot and rural household was apportioned as manure applied to the crops or discharged directly to the river;
3) Point source nitrogen was aggregated at the outlet of each subcatchment, based on the measured data;
4) The atmospheric deposition was uniformly distributed over the YRC.

3.7.1.2 Sources description

3.7.1.2.1 Nitrogen fertilizer

The operation schedule for the rice as the model input was derived from both literature (LÜ et al., 2016) and local agricultural department interview by the author, as listed below in Table 3-3.

Table 3-3 The rice operation in the YRC

Season	Operation	Time
Spring	Rice seeding	Mid-Mar. (approx. 20th, Mar.)
	Base fertilizer application	Mid-Apr. (approx. 15th, Apr.)
	Rice transplanting	Mid-Apr. (approx. 20st, Apr.)
	Topdressing application	Mid-May (approx. 10th, May)
	Harvest	Mid-Jul. (approx. 15th, Jul.)
Autumn	Base fertilizer application	Mid-Jul. (approx. 20th, Jul.)
	Rice transplanting	Late-Jul. (approx. 25st, Jul.)
	Topdressing application	Mid-Aug. (approx. 10th, Aug.)
	Harvest	Mid-Oct. (approx. 25th, Oct.)

The fertilizer application was derived from a survey conducted in 2012 with samples in multiple administrative districts in Jiangxi Prov., including Xinyu and Yichun (Qi, 2013). The amount of fertilizer was converted to net nitrogen (N) in Urea and compound fertilizer, as listed in Table 3-4. The management operation behavior of individual farmers was not the consideration of this research, thus, the paddy field was operated with the uniform plan listed in Table 3-3.

Table 3-4 Net N usage of fertilizer in 2012

Season		Net N in Urea (kg/ha)	Net N in compound fertilizer (kg/ha)
Spring	Base fertilizer	66.21	76.09
	Topdressing	43.00	50.56
Autumn	Base fertilizer	59.97	68.92
	Topdressing	32.13	36.92

3.7.1.2.2 Waste from feedlots

Table 3-5 The average cattle production of the four districts in the YRC

City	Pig	Cow	Goat	Fowl
Yichun	1083750	77504	14618	2135425
Xinyu	504077	63229	2885	1795486
Fenyi	301911	24660	6994	1002386
Luxi	378052	24577	46163	1719223

The average cattle production (Jiangxi Statistical Bureau, 2008-2014) in the four administrative districts in the YRC is listed in Table 3-5 and was assumed to be all from the scattered feedlots in the rural region. 86.8% was disposed directly as manure in Jiangxi Prov. (Xie et al., 2015) and the rest 13.2% was discharged to the river network. The waste from each species is listed below in Table 3-6 and they were converted to net nitrogen in Table 3-7 (Yi, 2013).

Table 3-6 Daily waste discharge from feedlot per animal

Waste discharge (kg/d)	Pig	Cow	Goat	Fowl
Feces	2.2	30	2.6	0.15
Urea	3.2	18	1.3	0

In this study, the spatial distribution of the feedlot was assumed to be determined by the size of the village, which was defined as the size of the rural patch in the landuse map (Figure 3-3). The number of the cattle in each of the four administrative districts were then apportioned to an individual rural patch by the size of the patch and aggregated to each subcatchment. The manure at each subcatchment from the feedlot was assumed to be applied to the cropland in the individual subcatchment. Cropland here included paddy field, dry field, and the orchard, as shown in Figure

3-8. The manure discharged to the river network was seen as a point source at the outlet of each subcatchment.

Table 3-7 The nitrogen content (%) in the feces

% in Feces	Pig	Cow	Goat	Fowl
mineral N	0.304	0.095	0.308	0.32
organic N	0.246	0.285	0.522	1.28
NH_3-N in mineral N	99	99	99	99

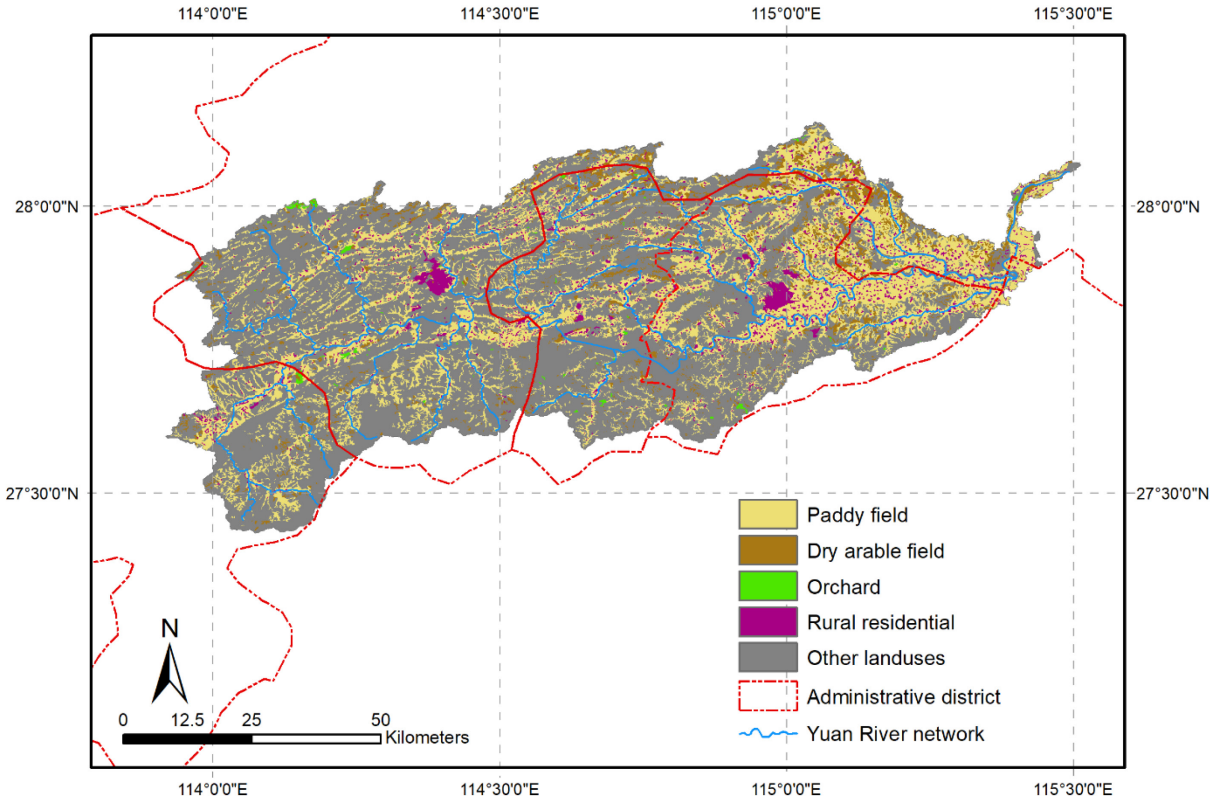

Figure 3-8 The cropland and rural distribution in the YRC

3.7.1.2.3 Waste from rural households

The path of the waste from rural households is listed below in Table 3-8. The data were derived from a survey by Li (2013) in Jiangxi Prov.. 55.4% of the total wastewater in the rural households was discharged untreated to village gullies and then routed to the surface water body (71%) or the cropland (29%). Except for 25.7% of the waste from rural households (urea and feces) discharged to the sewage system in the rural region, the rest was also disposed directly to the water body (17.2%) or applied as manure (57.1%).

Table 3-8 The path of waste and wastewater from rural households

Untreated waste	Directly to surface water body	To agricultural land as manure
Wastewater	39.3%	16.1%
Waste	17.2%	57.1%

Based on the survey from Wu (2012), the total nitrogen (TN) released in the household wastewater in the rural patch per person was 0.584kg/a with the ratio of ammonia nitrogen (NH_4-N) in the TN estimated at 44%. The TN in human waste was 3.060 kg/a, with the content of organic nitrogen and NH_4-N estimated at 80.8% and 20.2% (Bai and Wang, 2011; Rose et al., 2015). The rural population (Jiangxi Statistical Bureau, 2008-2014) of each four administrative districts are listed below in Table 3-9. The population was then distributed to each rural patch weighted by the size of the patch.

Table 3-9 The population at the 4 administrative districts in the YRC from 2008 to 2014

Administrative region	Mean population
Yichun	104200
Xinyu	582900
Fenyi	324000
Luxi	290000

The spatial distribution of rural households' waste was processed with the same approach as the waste from feedlots. The waste and wastewater disposed as manure was applied to all the cropland whereas the amount discharged to the river was assumed as a point source at the outlet of each subcatchment.

Therefore, the path of nitrogen from the feedlot and rural households was started either as manure distributed at each subcatchment or as a point source discharged at the outlet of each subcatchment.

3.7.1.2.4 Effluent from point sources

The effluent from point sources was identified from the industries and the municipal wastewater treatment plants. The outlets were mainly distributed in the urban region in the YRC (see Figure 3-3). 20 point-source (PS) outlets, including three outlets from the wastewater treatment plants, were monitored by the local environment protection agency every one to three months and established to the public, including the concentration of TN and NH_4-N. The discharge of the TN

ranged from approx. 103 to 756kg per day and NH_4-N ranged from approx. 25 to 255kg per day. These 20 outlets were considered in the study as all the nitrogen PS in the YRC. However, they were only the representatives and the major PS regulated in the YRC. The effluent of the individual PS was aggregated at the outlet of each subcatchment.

3.7.1.2.5 Atmospheric deposition

The atmospheric deposition in the YRC was assumed to be similar as in the Qianyanzhou experiment field, which is approx. 50 km in the south of the YRC with a monitored monthly nitrogen concentration in the precipitation in 2013 and 2014 (Xu et al., 2016b). The annual average concentration of TN (1.1mg/L), nitrate (NO_3-N) (0.3mg/L) and NH_4-N (0.3mg/L) in the precipitation was taken to describe the condition in YRC.

3.7.2 Water balance estimation

3.7.2.1 The MODIS global terrestrial evapotranspiration product

The MODIS global terrestrial evapotranspiration product (Running et al., 2018) is part of NASA's Earth Observing System Project (EOS) to estimate global terrestrial ET from earth land surface by using satellite remote sensing data. The MODIS global monthly ET data (MOD16 ET) was one of the products with a 0.05-degree resolution available from the year 2000. The data were computed based on the Penman-Monteith equation with daily meteorological reanalysis data from NASA's Global Modeling and Assimilation Office and the 8-day remotely sensed vegetation property dynamics from MODIS. The average daily ET was validated globally at 46 field-based eddy covariance flux towers and the average mean absolute errors (MAE) were 24.1% (Running et al., 2018). Figure 3-9 below demonstrates the long term spatially distributed ET and the catchment average ET is 839mm.

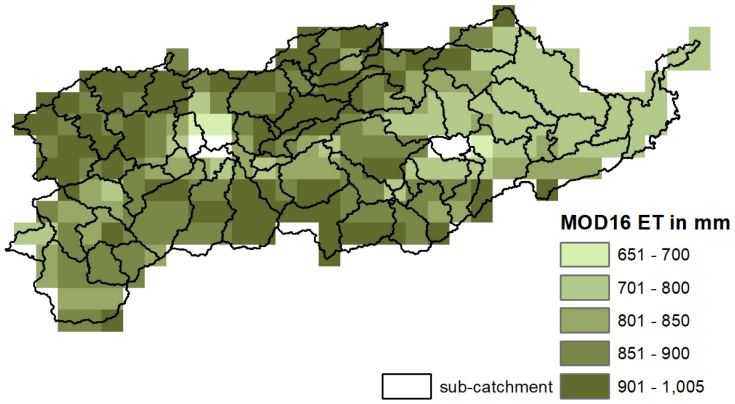

Figure 3-9 The long term average (2008–2014) distributed ET at YRC

3.7.2.2 Baseflow index of YRC

The baseflow index that estimates the ratio of the groundwater return flow contributing to the streamflow discharge was computed by separating the measured hydrograph at the three hydrometric stations with the Lynne-Hollick baseflow filer (see Section 5.3.1) and are listed below in Table 3-10.

Table 3-10 The baseflow index at LX, MZ and JK station

	LX	**MZ**	**JK**
Baseflow index	0.53	0.61	0.48

3.7.2.3 Flow duration curve of each hydrometric station

Figure 3-10 displays the flow duration curve (FDC) at LX, MZ and JK station and Table 3-11 listed the threshold value at each FDC segment (see Chapter 5 Section 5.3.1). MZ and JK are located in the mid-stream and mid-to-downstream at the Yuan River. The reservoir at JK station regulated the discharge and thus the low flow at JK was lower than the discharge at MZ.

Table 3-11 The threshold flow value at each flow duration curve segment

Flow value (m^3/s)	**Q5**	**Q20**	**Q70**	**Q95**
LX	37	14	5	3
MZ	293	122	49	36
JK	344	146	70	26

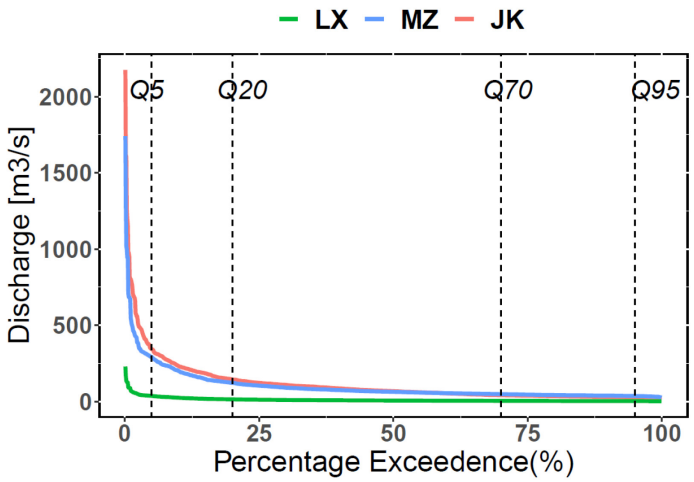

Figure 3-10 The FDC of measured daily discharge at LX, MZ and JK station

4 Research concept development

This chapter explained the development of the research concept based on the research aim and the current data condition (e.g. availability of data as described in Chapter 3) of the Yuan River Catchment (YRC). Section 4.1 explained the criteria to select the proper model; Section 4.2 stated the selection of the objectives and the approaches adopted to perform the multi-objective calibration; Section 4.3 stated the implementation of the uncertainty analysis.

4.1 Model selection requirements

Harmel et al. (2014) have defined the intended use of the hydrologic and water quality models into exploratory, planning and regulatory that helps to determine the tolerance of the model uncertainty, where an exploratory study with the highest tolerance and the regulatory application with the lowest tolerance. Moriasi (2012) summarized 25 models based on the scale and the respective processes computed and Tuo et al. (2015) summarized the amount of input data required of a hydrological model for nitrogen simulation. They indicated that the more detailed a hydrological model can simulate, the higher the amount of input data and parameters are required. Therefore, a compromise should also be considered between the aims and the current data condition.

According to the aims of this study, the model selection procedures are developed to consider the following aspects: 1) the temporal and spatial scale; 2) the critical hydrological processes to be simulated; 3) the expected accuracy considering the data availability. Table 4-1 lists the modeling aims decomposed according to the above procedures.

Table 4-1 The requirements based on the YRC condition for the model selection

Aspects	Requirements
Scale	Temporal scale: be able to capture the impact of the hydrological cycle on the nitrogen dynamics; Spatial scale: spatial characteristics should represent individual land uses to indicate the spatial distribution of the nitrogen sources
Targeted processes	A robust and accurate surface runoff; a detailed nitrogen cycle simulation to capture the nitrogen sources and the respective transport and transformation processes
Expected accuracy	Be able to identify the nitrogen critical sources to the specific land uses and to quantify the in-stream nitrogen load; the intended use is exploratory

1) The scale

When analyzing the nitrogen load at catchment-scale in consideration of the anthropogenic sources, including agricultural activities, the time scale of the input data is often seasonal and annual, instead of daily, or sub-daily and thus the impact of the nitrogen load should also be evaluated at the equivalent temporal scale. Therefore, models that are able to simulate the long-term effect should be considered instead of the event-based models.

The spatial scale determines the model's algorithms to represent the heterogeneity of the study catchment. The proper spatial heterogeneity representation of the nitrogen sources is critical to identify the relative nitrogen contribution from individual regions (Xu et al., 2016a). Moreover, the variation of the catchment conditions in terms of topography, crops and agricultural activities is critical to define the nitrogen input as well as the transport and transformation processes. However, the model is not required to accept only accurate input of every activity in the catchment but representative information within a computational unit, e.g. dominant agricultural activities or dominant landuses to adequately represent the spatial feature and to optimize the computational effort (Yalew et al., 2010). Therefore, models that have a certain extent of simplification to maintain the heterogeneity but also computationally practical are preferred. Semi-distributed models (see Table 2-1) are more suitable for the study than the lumped or fully-distributed models.

2) The targeted processes

In the YRC, the excessive nitrogen load to the surface water body is the concern of the catchment management (see Section 3.6), and the available data for calibration is the measured discharge and total nitrogen concentration. Therefore, the algorithms of computing the surface runoff are critical. However, a simplification, e.g. the groundwater recharge and the interactions between the aquifers, could be accepted in the groundwater system due to the very low water supply from groundwater (2%) in comparison to surface water (98%) (Fang, 2011).

In terms of nitrogen load simulation, CSAs require the model to quantify the in-stream nitrogen by tracing back to its load from the land, therefore, a detailed spatial representation of the nitrogen dynamics of the individual nitrogen species is required.

3) The expected uncertainty considering the data availability

The detailed representation of the nitrogen relevant processes requires a high amount of input data and parameters. The approximation of the input data for nitrogen input, the estimation of the unknown parameters and the limited data for calibration would result in a higher model uncertainty. However, due to the intended use of the study is exploratory, higher tolerance of the uncertainty could also be acceptable.

4.1.1 The advantages of the SWAT model

The aims and the catchment data condition determined the temporal scale of the model to be long-term and the spatial scale to be catchment-scale with a semi-distributed discretization approach. The accuracy needed for the potential model application required the model to provide an integrated simulation to account for all the nitrogen dynamics. Meanwhile, as an exploratory study, a higher parameter and prediction uncertainty due to limited available data could also be accepted. Therefore, in this study, SWAT was selected, though similar models, e.g. HSPF (Hydrological Simulation Program – Fortran) and AnnAGNPS (Annual Agricultural Non-point Source model) (see Table 2-1 and Table 2-2) could also meet the requirements in terms of proper scale and the application. Other than the requirements listed above met by the model, SWAT has additional advantages based on its cumulative applications that enable the users to lower the learning costs, which, in the author's view, is also a critical factor to model selection.

In terms of the developed purposes, governing flow equations and the applicability, Borah and Bera (2003, 2004) summarized 11 watershed-scale models and concluded that though models similar to SWAT have a comparably high amount of input data and parameter, it is still a promising tool in terms of continuous simulations in predominantly agricultural watersheds. This conclusion was also supported by more recent model comparison studies (Moriasi et al., 2012; Bouraoui and Grizzetti, 2014; Tuo et al., 2015). Massive applications in the past decade have indicated the models' ability to simulate the hydrology, sediment, and chemicals (Merwade et al., 2017). More specific applications in terms of nitrogen load analysis have also be implemented. Malago (2017) applied the model to simulate the water and nitrogen fluxes in the Danube Basin. Folle (2007) has derived the CSAs of TN load in southern Minnesota and evaluated the management practices in terms of reducing the load. Tran (2017) applied the SWAT model to the Cau River Basin in Vietnam with 36% of the area occupied by crops including rice, to assess the nitrogen load

condition in the complex pollution system. SWAT has also been applied to catchments in China. Huang (2009) simulated the streamflow and nutrient load in the lower reach of the Yangtze River basin. Xu (2016a) analyzed different DEM resolutions in the simulation of nitrogen load in one of the tributaries of the Three Gorges Reservoir. The worldwide application of the SWAT model has indicated its capability of representing the varied processes of hydrology and thus a more detailed analysis of the nitrogen transport and transformation mechanisms can be derived.

Supportive information provided by the SWAT user community is a pragmatic benefit. In terms of modeling procedure, Arnold et al. (2012a) and Abbaspour et al. (2015) generated guidelines for the modeling framework, i.e. the sequential calibration procedure for discharge, sediment and nutrients; the statistics applicable for calibration evaluation; approaches for calibration and recommended parameters; and the interpretation of the results considering the uncertainty. Moriasi et al. (2015b) and Abbaspour et al. (2018) summarized the main issues of the previous studies and proposed recommendations or the protocol to the respective issues, which provided an insight into the current focus of the model development and applications. SWAT, as an open-source tool, also provides a supportive guide to certain topics, i.e. theoretical report (Neitsch et al., 2011), documentation on input data preparation (Arnold et al., 2012b), the GIS interface developed, automatic calibration tools (Abbaspour, 2015), and the model user community.

4.2 Multi-objective calibration

The first step of applying the multi-objective calibration is to determine the objectives to be calibrated. In Section 2.3.1, the author summarized the objectives into three groups as an important conceptual approach in this study: 1) multi-site; 2) multi-objective-function; 3) multi-metric. Previous studies have demonstrated improved performance by applying the individual objective group. Therefore, a potential improvement might also be expected on multiple aspects of the model if all three groups of objectives are simultaneously calibrated.

Additional observations, recommended by previous studies, e.g. qualitative water balance estimation, are often limited at data-scarce catchment and lead to the multi-objective calibration inapplicable. Therefore, in this study, it is proposed to generate additional objectives by making the best use of the currently available datasets, which are previously served as the fundamental requirements for model calibration, e.g. measured discharge, or the data which are available globally, e.g. the satellite-based open-access dataset.

In terms of the calibration algorithms, both the classic aggregation approach and NSGA-II, as a representative MOEAs approach, are applied and compared to analyze their effectiveness. The classic aggregation approach AW or ED is used as a post-processing tool to screen the best simulation by sorting the value of the scalar function. Therefore, it requires a comparably low effort to combine with the single-objective calibration SUFI-2. NSGA-II enables a less subjective optimization of the objectives and, comparing to other evolution algorithms, it keeps both the variety and the elitism of the species. Therefore, these two approaches are selected.

4.2.1 The selection of objectives

The selection of the multiple objectives should be determined by 1) the availability of all the possible observations for calibration; 2) the aim of the model application. In this research, the objectives selected are listed below in Table 4-2.

Table 4-2 The list of objectives applied in the multi-objective calibration

Category	Objectives
Multi-site	Hydrology: three hydrometric stations; Nitrogen: total nitrogen (TN) load at three water quality monitoring stations
Multi-objective-function	Apply both NSE and inverse NSE for evaluating the discharge and the total nitrogen
Multi-metric	MODIS global monthly ET data (MOD16 ET); The baseflow ratio obtained from filtering the hydrograph (Section 5.3.1)

The YRC has only measured discharge and total nitrogen. However, the limitation of additional observations can be overcome by extracting useful information from existing data. The intention is to capture the features that were previously ignored but critical to be involved in the calibration, e.g. the baseflow ratio obtained by filtering the hydrograph with the Lynne-Hollick baseflow filter (Ladson et al., 2013). Another metric applied in the study is the evapotranspiration data obtained from the MOD16 ET dataset (Running et al., 2018). Besides the metrics, the multi-objective-function applied in the calibration is to target the different value magnitudes of the hydrograph, i.e. both peak values and low values.

The YRC is of varied geography (e.g. mountains and plains) and anthropogenic activities (e.g. fertilizer applied in the rural region and point source nitrogen discharged at urban region) that differ from the upstream to the downstream. Therefore, the consideration of the data within the

catchment domain other than only at the outlet is of great importance in parameterization over the catchment. When analyzing the nitrogen load, the impact is more critical on the low flow period that faces the higher concentration of the pollutants, thus the consideration of multi-objective-function and the baseflow ratio could benefit a more reliable simulation of both the peak flow rate and the low flow rate. Moreover, the evapotranspiration rate is another key water balance component and impacts the crop growth condition in the YRC, which is the main consumption of the fertilizer applied.

4.2.2 Calibration algorithms

The characteristics of the three approaches have been described in Section 2.3.2. Though the advantages of the NSGA-II are emphasized, however, when compared to AW and ED, the limited number of objectives considered and the effort to couple the approach with the SWAT model made NSGA-II less competitive. However, those limitations are not the issue with AW and ED. When comparing AW with ED, ED is a less subjective selection approach to optimize the selection of the best simulation (Pfannerstill et al., 2017).

Due to the limited number of objectives to be practical in the NSGA-II, it is proposed in this study to aggregate the individual objectives into three categories. These three groups were referred to as categorized objectives in the context, which are 1) NSE for the three hydrometric stations; 2) inverse NSE for the three stations; 3) the water balance components. Meanwhile, to prove the applicability of the substitution, the consistency of the categorized objectives should be first analyzed (see Section 6.3.1.1). The more consistent approach ED is then selected instead of AW to further compare with single-objective calibration and NSGA-II.

4.3 Uncertainty analysis

The uncertainty analysis implemented previously, especially the prediction uncertainty, was often derived for only one objective. The multi-objective calibration with boundaries set on additional processes is theoretically to better narrow the parameter uncertainty than the single-objective calibration. In this study, a comprehensive assessment is performed to analyze how the multi-objective calibration impacted the prediction and the parameter uncertainty, in terms of applying the SWAT model on the simulation of hydrology and the nitrogen load analysis for the Yuan River Catchment.

4.3.1 The prediction uncertainty of nitrogen load

Though the reliability of the SWAT model has been proved in many studies, and the prediction uncertainty has been emphasized in terms of model application. The approaches applied by the SWAT model to simulate the nitrogen transformation and transport processes are dominated by empirical relations, which are determined by the coefficients (Neitsch et al., 2011) to be calibrated, e.g. nitrate percolation coefficient, partitioning the nitrate between surface flow and percolation. The calibration is often based on the measured in-stream nitrogen concentration, which is also with higher measurement uncertainty than discharge (Harmel et al., 2006). Those uncertainties along with the parameter uncertainty will amplify the uncertainty in the TN prediction. Therefore, the prediction uncertainty of the nitrogen load enables a better understanding of the mode's reliability of the individual applications.

The prediction uncertainty can be derived by the aggregation of all the simulations in a calibration procedure, e.g. the 95PPU that denotes the coverage of 2.5^{th} and 97.5^{th} percentiles of the simulations. However, it is not practical to also generate the uncertainty band for CSAs by this approach, due to the costly effort for storing and post-processing the TN load at HRU level for each simulation. Instead, the simulations that have the highest or the lowest PBIAS to the calibrated simulation of the TN load, which are the simulated highest or lowest TN load to the stream, can also be representatives of the upper and lower bound of the uncertainty. Its applicability was proved at the YRC by the author's previous study (He et al., 2019). Therefore, the parameter sets, resulting in the highest/lowest PBIAS, are used to generate the uncertainty band of CSAs.

4.4 Contribution of the research

The research put a scientific contribution to the calibration and the uncertainty analysis of modeling hydrology and nitrogen load on a data-scarce catchment. A multi-objective calibration framework is proposed and described to overcome the known deficiency in conventional calibration procedures. The study proves the applicability and the advantages of this framework on the Yuan River Catchment and proposed a related approach to quantify the uncertainty band of the nitrogen load prediction. The applicability of the multi-objective calibration is first achieved by utilizing the open-access satellite-based dataset and extracting additional metrics from existing data to overcome the limitation of lacking observations. The effect of calibrating three groups of

objectives simultaneously, multi-site, multi-objective-function, and multi-metric, is demonstrated by the reduced uncertainty range of the simulation and the parameters comparing to classic single-objective calibration. The prediction uncertainty is quantified by applying the two simulations, the simulations with the lowest or the highest PBIAS to the best-calibrated simulation, to represent the range of the uncertainty. This approach makes it practical to analyze the uncertainty of the predicted critical source areas and thus enables a more reliable interpretation of the nitrogen load analysis.

The calibration procedure implemented in the research, independent from the study catchment, also demonstrates its potential to be applied by similar catchment studies with the SWAT model. The preparation of the objectives can be achieved by considering the proposed three groups of the objectives. The calibration approaches, ED and NSGA-II, are analyzed to identify their suitability and thus offers a useful reference for implementation in future studies. Besides, an executable R-scripts assembled in the Appendix gives an example of coupling the NSGA-II with the SWAT model as a technical contribution to the research community.

5 Methodology

In this chapter, the methods applied to implement the research were stated. Section 5.1 described the features and the algorithms of the SWAT model in brief; Section 5.2 explained the algorithms of the single-objective calibration approach SUFI-2; Section 5.3 explained the multi-objective calibration, from the acquisition of the multiple objectives to the algorithms of the calibration approaches; Section 5.4 demonstrated the flowchart to explain the procedure of implementing calibration; Section 5.5 stated the methods applied to identify the critical source areas; Section 5.6 stated how the uncertainty analysis was performed to interpret the nitrogen load simulation.

5.1 SWAT model overview

5.1.1 The hydrology module

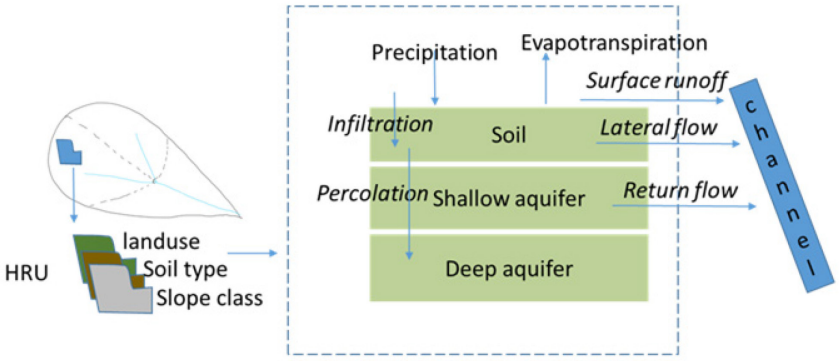

Figure 5-1 HRU and hydrological processes computed in SWAT

The Soil and Water Assessment Tool (SWAT) is an integrated process-based hydrological model to predict the long-term impact of land management practices on the water, sediment, and agricultural chemical yields in catchments (Neitsch et al., 2011). The computational unit of SWAT is the hydrologic response unit (HRU), which is a combination of a certain landuse, soil, and slope at a location within a subcatchment (see Figure 5-1). The hydrological cycle is based on the water balance listed below in Equation 1. The hydrological processes computed by SWAT and the respective methods are listed in Table 5-1. SWAT adopts homogeneous precipitation within a subcatchment. Each soil layer is assumed to be homogeneous horizontally. Green-Ampt infiltration method and the storage routing method are implemented to calculate the infiltration and percolation respectively. The groundwater system is simplified by the SWAT model to be a

shallow aquifer and a deep aquifer. The water yield of each HRU, which is routed to the outlet of the channel in the subcatchment, consists of surface runoff, lateral flow, and groundwater return flow.

Equation 1 The water balance equation applied by SWAT

$$SW_t = SW_0 + \sum_{i=1}^{t}(R_{day} - Q_{surf} - E_a - w_{perc} - Q_{gw})$$

When the time step i is the daily base, SW_t and SW_0 are the soil water content at the end and the beginning of a day; R_{day} is the daily precipitation; Q_{surf} is the surface runoff; E_a is the actual evapotranspiration; w_{perc} is the percolation exiting the soil profile bottom; Q_{gw} is the groundwater return flow or baseflow; The time step can also be monthly or annually.

5.1.2 The nitrogen module

The nitrogen module of SWAT calculates both transport and transformation processes in the land phase and the stream. Figure 5-2 below demonstrates the processes considered in the nitrogen module of the land phase. The nitrogen species are categorized into five pools: the ammonia nitrogen (NH_4^+-N), the nitrate (NO_3^--N), and the active, stable, and fresh organic nitrogen (OrgN). Each pool follows the individual transformation and transport paths (see Figure 5-2). The transformation processes of the land phase are mainly computed by empirical equations of soil water content and soil temperature and the function is determined by a rate coefficient (see Table 5-2). Therefore, the accuracy highly relies on calibration.

The OrgN is transported by the erosion process at the topsoil layer. NO_3^--N is transported as a solute either in the surface runoff and lateral flow or infiltrated to the groundwater system and transported as return flow (see Figure 5-3). However, NH_4^+-N is not transported directly to the stream but retained or transformed in soil. The mass of the nitrogen being transported in the model depends on the rate of sedimentation or runoff and also determined by a transport rate to be calibrated, e.g. organic nitrogen enrichment ratio and nitrate percolation coefficient.

Table 5-1 The main hydrological processes, the respective method and parameters applied in SWAT

Processes	Method	Main parameters
Evapotranspiration	Penman-Monteith equation	Meteorological variables
Surface runoff	SCS Curve Number (USDA, 1972)	CN2: Moisture condition II curve number SOL_AWC: available water capacity ESCO: soil evaporation compensation coefficient SOL_K: saturated hydraulic conductivity HRU_SLP: the average slope of the subcatchment
Infiltration	Green & Ampt infiltration method	
Soil water	Evaporation rate by compensation factor; Percolation to occur when water content higher than field capacity	
Lateral flow	Kinematic storage model	
Percolation	Storage routing method in the soil layer	
Groundwater recharge	Exponential decay weighting function; User-defined coefficient to partition recharge between the shallow aquifer and deep aquifer (Sloan and Moore, 1984)	GW_DELAY: delay time for groundwater recharge RCHRG_DP: aquifer percolation coefficient from shallow to the deep aquifer
Groundwater return flow or base flow	Steady-state within the time step, a function of hydraulic conductivity, travel distance and recharge rate	GWQMN: the threshold water level in the shallow aquifer for baseflow to occur
In-channel process	Manning's equation for flow rate and velocity; Variable storage coefficient for discharge	CH_N: manning's n value CH_S: the slope of the channel

SWAT applies the QUALE2 (Brown and Barnwell, JR., 1987) to compute the in-stream processes of nitrogen transformation. The in-stream processes are shown in Figure 5-3 and they are determined by functions of stream condition (temperature, oxygen content) and the algal growth. The SWAT model applies the Heat Unit approach (Barnard, 1948) to simulate plant growth. Plant residue is computed by the model as the nitrogen input and the nitrogen is consumed by the growth of the plant. The nitrogen input from anthropogenic sources is added to the respective nitrogen pools. The non-point source nitrogen, e.g. fertilizer was inserted to each HRU and the point source nitrogen was added to the respective outlet of the subcatchment.

Table 5-2 The main nitrogen processes, the respective method and parameters applied in SWAT

Transformation processes	Method	Main parameters
Initial nitrate content	NO_3-N: user-defined; or a function of soil depth OrgN: a function of soil organic carbon content and portioned (0.98 to 0.02) between active and stable pools; fresh organic nitrogen in the topsoil layer is set to be 0.15% of the residue content	Soil depth; soil organic carbon content; soil bulk density; residue content of the topsoil layer
Immobilization /decomposition /mineralization	Adapted PAPRAN equation, sources for immobilization are fresh OrgN and the active OrgN	Soil humus and residue content; soil temperature and water content; rate coefficient of mineralization according to crop type or defined by the user
Nitrification and ammonia volatilization	1st-order kinetic rate equation	Temperature, soil water content, soil depth
Denitrification	The soil water content is exceeding a user-defined threshold for the process to occur; the denitrification rate is determined by the soil carbon content	Rate coefficient for denitrification, the threshold value of nutrient cycling water factor for denitrification to occur; soil organic carbon content
Atmospheric deposition	User-defined nitrogen content in the precipitation and added to the land as an input	-
Fixation	Determined by the legumes in the catchment; independent from the soil layer	-
Nitrate in the aquifer	An exponential decay weighting function (Venetis, 1969) to calculate the nitrogen in the recharge	User-defined delay time
Transport process		
Nitrate in the flow	Nitrogen content is a function of soil porosity, saturated water content	Nitrogen percolation coefficient to partition nitrogen to the runoff and the leaching process
In-stream	QUAL2E	Algae growth rate, hydrolysis rate and the settling rate of either NH_4-N or NO_3-N
Erosion	Loading function to calculate the amount of organic nitrogen in the sediment and is proportional to the sediment yield	Nitrogen enrichment ratio

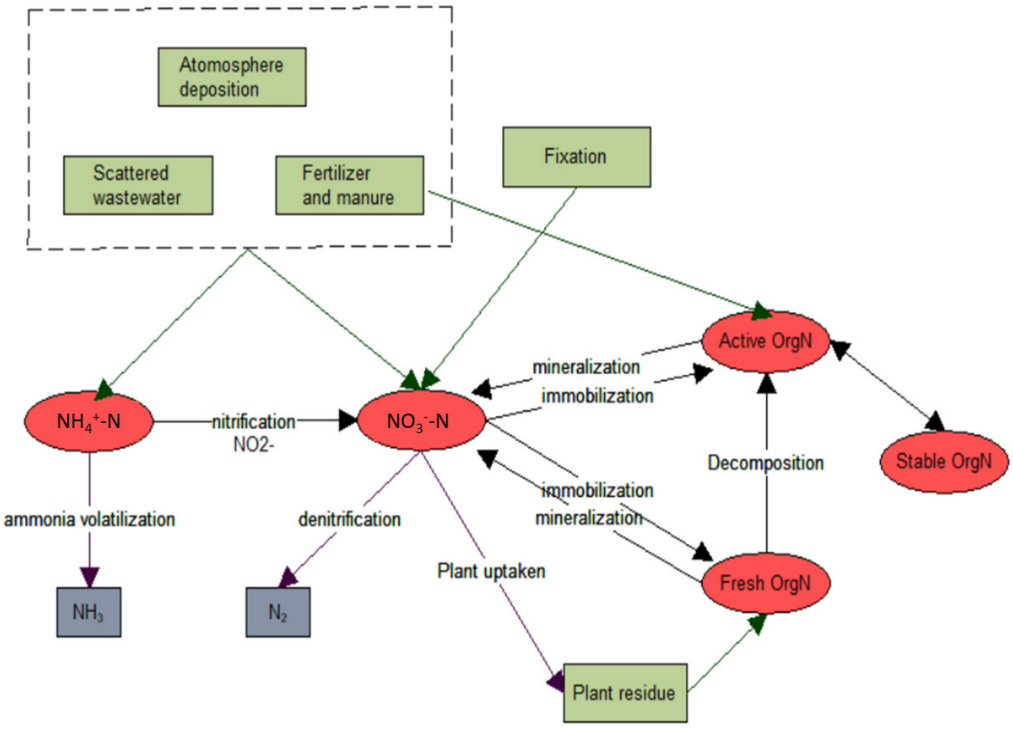

Figure 5-2 The processes at the land phase computed in the SWAT nitrogen module

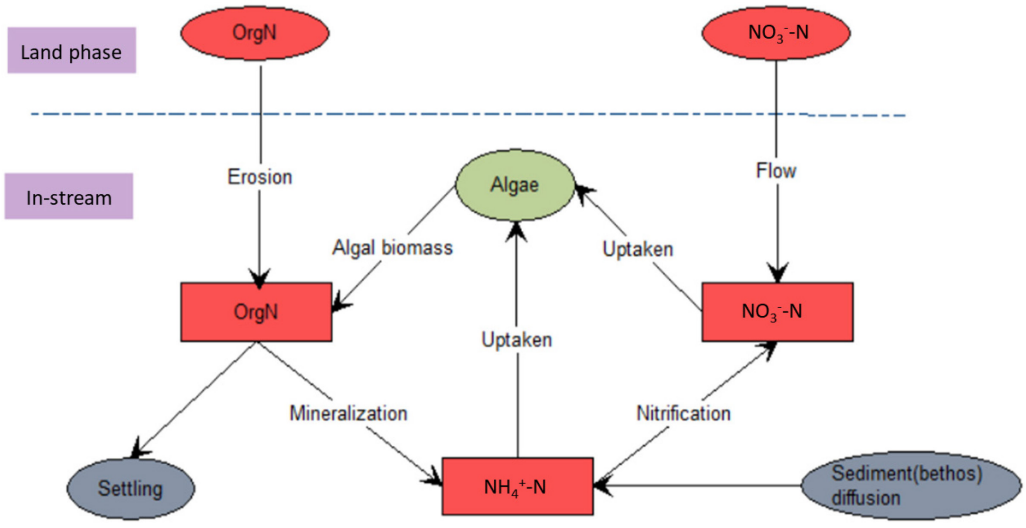

Figure 5-3 The in-stream transformation processes of the SWAT model

5.2 Single-objective calibration with SUFI-2

5.2.1 The algorithm

SUFI-2 (Sequential Uncertainty Fitting, ver.2) (Abbaspour, 2005) is a widely applied auto-calibration and uncertainty analysis approach. It is an efficient approach that requires fewer iterations compared to other approaches (Yang et al., 2008). The working procedure is displayed in Figure 5-4. The parameter sets are generated by the Latin Hypercube Sampling (LHS) technique (McKay et al., 1979) for global sensitivity analysis, and m parameters are then selected for the calibration. After the first iteration with n simulations are executed, the objective function of each simulation is calculated and evaluated. If the user-defined criteria are not met, an updated parameter range is derived to start the next iteration until the criteria are met. The algorithms can be found in detail in Abbaspour (2004) and explained below.

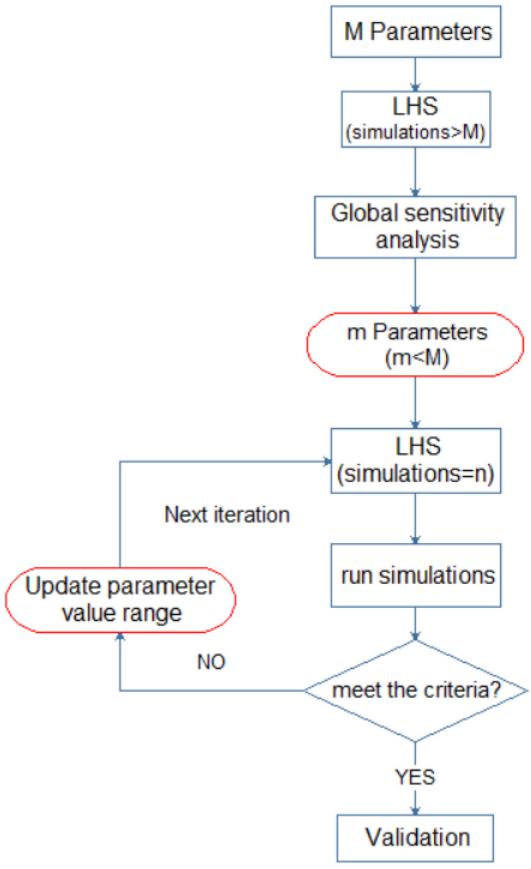

Figure 5-4 The working procedure of SUFI-2

a. Select potential parameters and LHS

The potential parameters with the initial value ranges can be obtained both from the literature review (Shrestha et al., 2016; Liu et al., 2016; Abbaspour et al., 2018) and the observation of the offset between the measured and simulated data (Abbaspour et al., 2015), which should be also physically meaningful (if applicable). In this study, the parameters were selected according to the processes related and were separated into hydrology relevant and only nitrogen module relevant. The parameter values are assumed to be uniformly distributed within its value range and LHS generates a group of random parameter sets from them. The number of parameter sets is determined by the size of simulations of an iteration.

b. Global sensitivity analysis

The model is first executed with the initial parameter sets to analyze the sensitivity of the individual parameters. The sensitivity is evaluated by global sensitivity analysis, which sorts the relative sensitivity of a parameter while the other parameters are also dynamic.

The sensitivity of parameter j is calculated as S_j:

Equation 2 The sensitivity of parameter j

$$S_j = \overline{b_j} \frac{1}{C_n^2} \sum_{i=1}^{C_n^2} \left| \frac{\Delta g_i}{\Delta b_j} \right|; \mathrm{i} = 1, \dots, C_n^2; \ \mathrm{j} = 1, \dots, \mathrm{m}$$

$\overline{b_j}$ is the average value of the parameter *j*, n is the number of the parameter sets (the number of simulations to run in the calibration) ; C_n^2 is the number of possible combinations of two simulations of n simulations, and $C_n^2 = \frac{n!}{2! \cdot (n-2)!}$; m standards for the number of parameters in the calibration; $\left| \frac{\Delta g_i}{\Delta b_j} \right|$ is the determinant of the Jacobian matrix $\frac{\Delta g_i}{\Delta b_j}$, where Δg_i stands for the difference of the objective function values picked from two random simulations and Δb_j is the difference of the parameter j value applied in these two simulations.

c. Update the parameter range after each iteration

The Jacobian matrix $\boldsymbol{J}$ is derived after finishing the first iteration; the Hessian matrix $\boldsymbol{H}$ is then derived based on the Gauss-Newton method that only considers the first-order derivatives in the

Jacobian matrix. Then the Cremer-Rao theorem is applied to obtain the lower bound of the parameter covariance matrix $\boldsymbol{C}$, the equations are: $\begin{cases} J_{ij} = \frac{\Delta g_i}{\Delta b_j} \\ \boldsymbol{H} = \boldsymbol{J}^T\boldsymbol{J} \\ \boldsymbol{C} = s_g^2(\boldsymbol{J}^T\boldsymbol{J})^{-1} \end{cases}$ s_g^2 is the variance of the objective function values and $s_g^2 = \frac{1}{n}\sum_i^n(g_i - \bar{g})$. The 95% confidence interval of a parameter j was then computed, as displayed below.

Equation 3 The 95% confidence interval of parameter j

$$\begin{cases} b_{j,lower} = b_{j,best} - t_{v,0.025}s_j \\ b_{j,upper} = b_{j,best} + t_{v,0.025}s_j \\ s_j = \sqrt{C_{jj}} \end{cases}$$

C_{jj} is the diagonal elements of $\boldsymbol{C}$; the $b_{j,best}$ is the value of parameter j that has the best objective function value; $t_{v,0.025}$ is the t distribution with the degrees of freedom to be $v = (n - m)$.

The range of b_j was then to be updated, as displayed below.

Equation 4 Updating the parameter range

$$\begin{cases} b_{j,min}^{new} = b_{j,lower} - \max[\frac{b_{j,lower} - b_{j,min}}{2}, \frac{b_{j,max} - b_{j,upper}}{2}] \\ b_{j,max}^{new} = b_{j,upper} + \max[\frac{b_{j,lower} - b_{j,min}}{2}, \frac{b_{j,max} - b_{j,upper}}{2}] \end{cases}$$

d. Generate 95% prediction uncertainty

After each iteration, the 95% prediction uncertainty (95PPU) is also generated to represent the coverage of 2.5th and 97.5th percentiles of all the simulations in the iteration. Two indicators, p-factor, and r-factor are to assess the prediction uncertainty. p-factor defines the percentage of observations covered by the 95PPU and it is recommended to be higher than 0.7. r-factor is the average thickness of 95PPU, which is the ratio of the average 95PPU thickness to the standard deviation of the observed data. The lower the r-factor is, the smaller the prediction uncertainty is.

5.2.2 The objective function applied in SUFI-2

Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970) was used in this study as the objective function of SUFI-2 to evaluate the performance of the discharge and the total nitrogen content.

The criteria to assess the model performance was to achieve a monthly flow simulation of NSE > 0.5, and a monthly nitrogen simulation of NSE > 0.35 (Moriasi et al., 2015a). The equation of NSE is listed below.

Equation 5 The formula of NSE

$$NSE = 1 - \frac{\sum_{i=1}^{n}(Y_i^{obs} - Y_i^{sim})^2}{\sum_{i=1}^{n}(Y_i^{obs} - Y_{mean}^{obs})^2}$$

Y_i^{obs} is the observed data at time step *i*, Y_i^{sim} is the simulated data at time step *i* and, Y_{mean}^{obs} is the mean value of the observed data over the simulation period *n*.

5.3 Multi-objective calibration

5.3.1 Objectives

5.3.1.1 Multi-site

At the sequential order from upstream to downstream, the subcatchments were grouped into clusters based on the hydrometric or water quality stations. Therefore, the value of each parameter was set to be specific for each subcatchment cluster. Figure 5-5 below explains how one parameter was set. Subcatchment cluster C1 consists of subcatchment S1 and S2 and its outlet is M1, where the hydrometric or water quality station is located; C2 consists of S3 and its outlet is M2. Parameter P is then assigned as P1 for S1 and S2 and P2 for S3. Therefore, the parameterization is regionalized and calibrated simultaneously at each cluster.

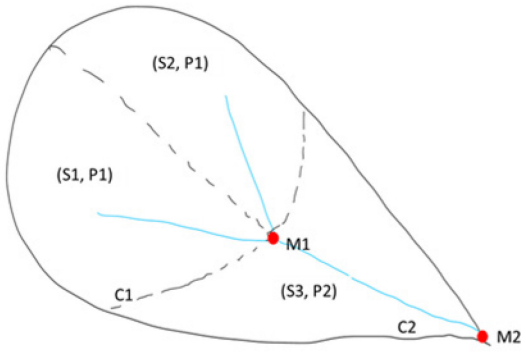

Figure 5-5 The schematic plot to demonstrate the multi-site parameterization

5.3.1.2 Multi-objective-function

Two objective functions were applied as the evaluation statistics to assess the simulated discharge or nitrogen load simulation at each hydrometric or water quality station, to evaluate the simulation

of both peak value and low value. The two objectives were NSE (Equation 5) and NSE_in (Pushpalatha et al., 2012; Le Moine, 2008), the inversed form of NSE. When applying the NSE_in, a 50 (unit: m^3/s) was added to each value to avoid the high values obtained by the inversed form of very low flow values.

Equation 6 The modified formula of inverse NSE

$$NSE\_in = 1 - \frac{\sum_{i=1}^{n}(1/(Y_i^{obs} + 50) - 1/(Y_i^{sim} + 50))^2}{\sum_{i=1}^{n}(1/(Y_i^{obs} + 50) - 1/(Y_{mean}^{obs} + 50))^2}$$

The standardization of NSE and NSE_in are listed in Table 5-4.

5.3.1.3 Multi-metric

a. Actual evapotranspiration (AET)

The observed AET data were obtained from the MOIDS global monthly ET data (MOD16 ET) from 2008 to 2014. A global average mean absolute errors (MAE) of 24.1% was obtained from validating at 46 field-based eddy covariance flux towers (Running et al., 2018). The spatial resolution of the datasets is 0.5 km. The data were extracted within the border of the YRC and the catchment average monthly AET was computed. In the study, the uncertainty band of the MOD16 ET was set to be covered by the MAE at ±24.1%.

The simulated AET was evaluated by the percentage of the monthly simulated AET covered by the MOD16 ET uncertainty band. Figure 5-6 displays the MOD16 ET and the simulated AET with 47.6% covered by the uncertainty band, the equation is displayed below.

Equation 7 The formula of multi-metric AET_{cover}

$$AET_{cover} = \frac{No\_of\_covered}{N}$$

AET_{cover} stands for the percentage of the simulated AET that is covered by the MOD16 ET uncertainty band. $No\_of\_covered$ stands for the number of simulated AET in the simulation period covered by the MOD16 ET uncertainty range and N stands for the total number of simulated AET in the simulation period. AET_{cover} is also standardized as listed in Table 5-4.

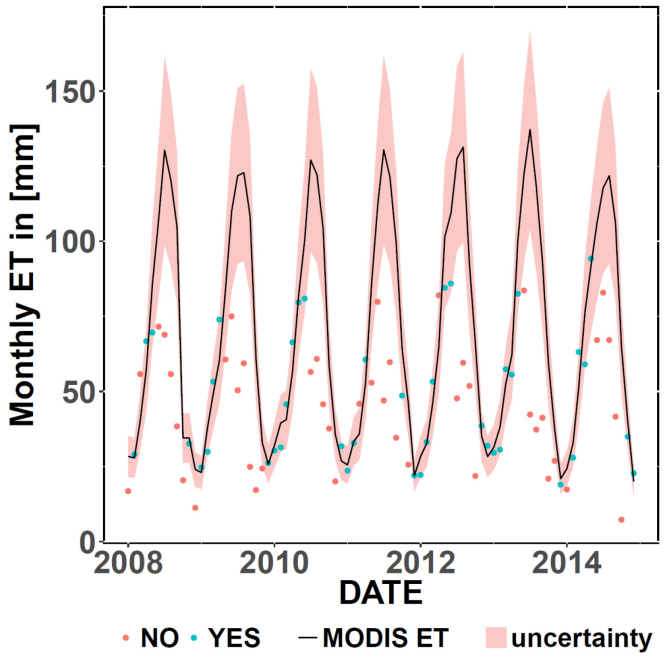

Figure 5-6 The uncertainty band of Monthly MOD16 ET and a demonstration of the simulated AET covered by the band (YES) or not (NO).

b. Baseflow and surface runoff

The mean ratio of baseflow and surface runoff to the discharge was derived from the separation of the hydrograph by applying the Lynne-Hollick baseflow filter, which was standardized by Ladson (2013) and published as an R package (hydrostats). The basic equation is listed below.

Equation 8 The Lynne-Hollick baseflow filter

$$q_f(i) = \begin{cases} \alpha q_f(i-1) + \dfrac{(1+\alpha)}{2}[q(i) - q(i-1)]; if\ q_f(i) > 0 \\ 0; \qquad\qquad\qquad\qquad\qquad\qquad if\ q_f(i) = 0 \end{cases}$$

$$q_b(i) = q(i) - q_f(i)$$

$q_f(i)$ is the quick flow response time in day i; $q(i)$ is the measured streamflow and $q_b(i)$ is the baseflow rate; α is the filter parameter to represent the attenuation of the hydrograph, with a suggested value to be 0.9 to 0.98. The filter will be passed by the time step of the dataset recursively multiple times both forward and backward. Ladson proposed to consider a warmup period at both the begin and the end of the dataset, e.g. 30 days with a flexible initial value of q_f for each pass

by updating it with the last q_b obtained in the previous pass; three passes were applied for the daily streamflow dataset.

In this study, the α cannot be estimated by prior knowledge, therefore, the mean annual ratio of baseflow to streamflow was derived by computing the filter with α value ranges from 0.90 to 0.98 to obtain the mean value to represent the baseflow ratio, as presented in Figure 5-7 for the three hydrometric stations. The higher the α value was, the lower the baseflow index was derived.

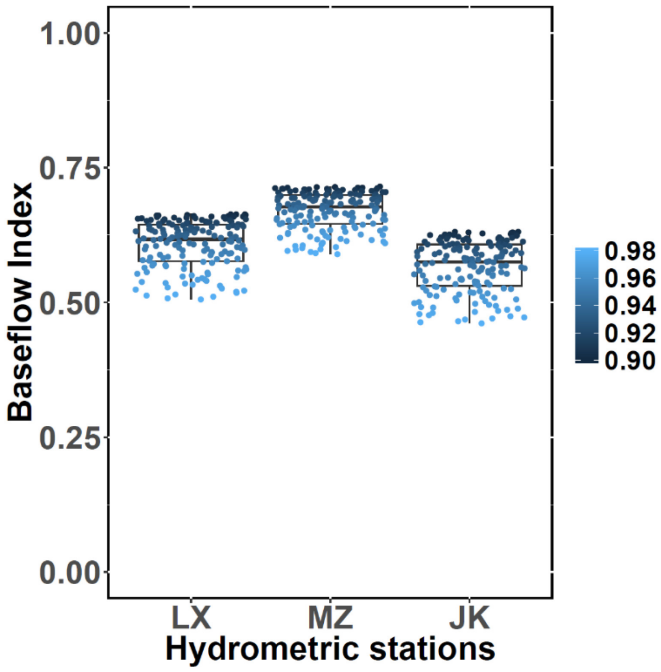

Figure 5-7 The baseflow index calculated by Lynee-Hollick approach with the alpha value from 0.90 to 0.98

In the study, the surface runoff was computed as the summary of the surface runoff and the lateral flow; the rate of baseflow was the rate of the groundwater return flow as the model output variable. The simulated monthly ratio of surface runoff to water yield and the ratio of groundwater return flow to water yield were computed for the subcatchments aggregated at the upstream of each hydrometric station. The Percent Bias (PBIAS) was calculated to evaluate the deviation of the simulated ratio to the ratio obtained from the baseflow filter. The standardization of PBIAS is displayed in Table 5-4.

Equation 9 The formula of PBIAS

$$PBIAS = \frac{\sum_{i=1}^{n}(Y_i^{sim} - Y_i^{obs}) \times 100}{\sum_{i=1}^{n}(Y_i^{obs})}$$

Y_i^{obs} is the observed data at time step i, Y_i^{sim} is the simulated data and, Y_{mean}^{obs} is the mean value of the observed data over the simulation period n.

5.3.1.4 The summary of the objectives

Table 5-3 summarized the objectives applied in the multi-objective calibration. As mentioned in Section 4.2.2, the three groups of the objectives were divided into three categorized objectives to ensure the applicability of the non-dominated sorting approach.

Table 5-3 The summary of all the objectives for multi-objective calibration

The categorized objectives	Objectives	Quantity	Evaluation statistics	Observations
Hydrology				
Multi-site with NSE	Daily discharge at LX, MZ, and JK	3	NSE	Measured daily discharge at LX, MZ, and JK
Multi-site with NSE_in		3	NSE_in	
Multi-metric	Actual evapotranspiration	1	AET_{cover}	Monthly MOD16 ET
	(Surface runoff)/(water yield) at LX, MZ, and JK	3	PBIAS	Hydrograph separation at LX, MZ, and JK
	(Groundwater return flow)/(water yield) at LX, MZ, and JK	3		
Nitrogen				
Multi-site with NSE	Monthly TN load at 1_Y,4_F, and 6_L	3	NSE	Measured monthly or seasonally TN load at 1_Y,4_F, and 6_L
Multi-site with NSE_in		3	NSE_in	

As stated above, the original value range of the evaluation statistics differed, and they were standardized to have the optimal value of 0, as listed below in Table 5-4.

Table 5-4 The standardization of the evaluation statistics

Evaluation statistics	Original value range	Standardization	Standardized value range
NSE	$(-\infty, 1)$	$1-NSE$	$(0,+\infty)$
NSE_in	$(-\infty, 1)$	$1-NSE\_in$	$(0,+\infty)$
AET_{cover}	(0,1)	$1/AET_{cover}$	(0,1)
PBIAS	$(-\infty, +\infty)$	$\|PBIAS\|/100$	$(0,+\infty)$

5.3.1.5 Other evaluation approaches

The simulated discharge was also evaluated by RSR (RMSE-observations standard deviation ratio) to assess the fitness of each flow duration curve (FDC) segment.

a. Flow duration curve

FDC demonstrates the percentage exceedance of a flow value over the entire measured flow period. The daily discharge values were first sorted in descending order; the lower the order was, the lower the discharge and the higher the percentage of exceedance would be. Figure 5-8 demonstrates the FDC at JK station and was separated into five segments: <Q5, Q5Q20, Q20Q70, Q70Q95, and >Q95. In this study, discharge values that occurred higher than 70% (Q70) within the measured period were regarded as low flow and the discharges higher than 95% (Q95) were very low flow. The normal flow was within the segments Q20Q70. Q5Q20 was the transient period from normal to peak flow. <Q5 represented the flow values that occurred lower than 5% within the period and they were the peak flow values. The FDC is an indicator to represent the magnitude of a hydrograph at each flow segment.

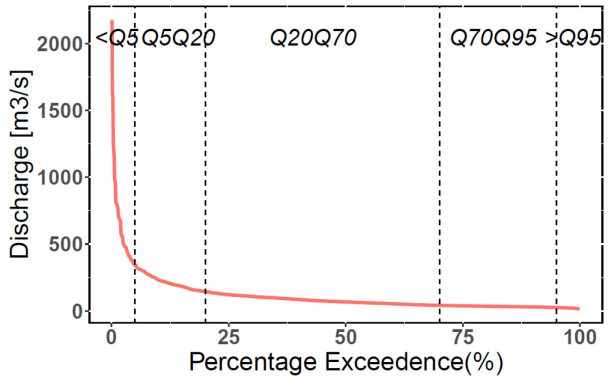

Figure 5-8 The FDC of measured daily discharge at JK station

b. RSR

RSR is the standardized form of the root mean square error (RMSE) to measure the residue of the simulation and the observation. RSR combines the error-index and a scaling normalization factor. It was first proposed by Singh (2005) and evaluated as dimensionless model evaluation statistics by Moriasi (2007). Pfannerstill (2017) proposed to apply the RSR to evaluate the FDC segments. The equation is listed below.

Equation 10 The formula of RSR

$$RSR = \frac{RMSE}{SD_{obs}} = \frac{\sqrt{\sum_{i=1}^{n}(1/Y_i^{obs} - 1/Y_i^{sim})^2}}{\sqrt{\sum_{i=1}^{n}(1/Y_i^{obs} - 1/Y_{mean}^{obs})^2}}$$

Where SD_{obs} is the standard deviation of the observed data; RSR equaling to zero is a perfect model simulation.

5.3.2 Calibration algorithms

5.3.2.1 Average Weight and Euclidean Distance

Average Weight (AW) and Euclidean Distance (ED) were applied and evaluated in the study as the classic aggregation approach to compare with NSGA-II. Two scenarios were created, the AW or ED for individual objectives (13 objectives in hydrologic calibration, see Table 6-5) and the AW or ED for categorized objectives (3 categories, see Table 6-5).

The equation to calculate the average weight:

Equation 11 The formula of Average Weight

$$\mathrm{AW} = \frac{\sum_{i=1}^{N} Obj_i}{N}$$

$$\begin{cases} \mathrm{AW\_cate} = \dfrac{\sum_{i=1}^{M} cate\_Obj_i}{M} \\ cate\_Obj_i = \dfrac{\sum_{i=1}^{J} Obj_{i\_J}}{J} \end{cases}$$

The equation to calculate the Euclidean Distance:

Equation 12 The formula of Euclidean Distance

$$\mathrm{ED} = \sqrt{[\sum_{i=1}^{N}(Obj_i)^2]/N}$$

$$\begin{cases} \mathrm{ED\_cate} = \sqrt{[\sum_{i=1}^{M}(cate\_Obj_i)^2]/M} \\ cate\_Obj_i = \sqrt{[\sum_{i=1}^{J}(Obj_{i\_J})^2]/J} \end{cases}$$

Obj_i stands for the individual objective, $cate\_Obj_i$ stands for the categorized objective. *N* stands for the total number of the individual objectives, *M* stands for the number of the objective categories, and *J* stands for the number of objectives in each category. The formulas to evaluate Obj_i, $cate\_Obj_i$, and Obj_{i_J} were standardized as stated above. The optimal ED or AW is 0. The lower the ED or AW values of the simulation were, the better the performance of the simulation was in the study.

5.3.2.2 Non-Dominated Sorting Genetic Algorithms II (NSGA-II)

The concept of NSGA-II (Deb et al., 2002) is to optimize the parameter set based on the evolution algorithm. The individuals of a population are evaluated by the non-dominated sorting approach and evolved via crossover and mutation (see Figure 5-11) to form the next generation. The evolution process will be accomplished if the number of generations or the Pareto Front of the last generation is to meet the criteria defined by the user. The key components of NSGA-II are listed below.

a. Non-dominated sorting and Pareto Front

Simulation $\overrightarrow{S_1}$ dominates $\overrightarrow{S_2}$ is defined as 1) $\overrightarrow{S_1}$ is no worse than $\overrightarrow{S_2}$ in any objectives; 2) $\overrightarrow{S_1}$ has at least one objective better than $\overrightarrow{S_2}$. If $\overrightarrow{S_1}$ is not dominated by $\overrightarrow{S_2}$ in any objectives, $\overrightarrow{S_1}$ is the non-dominated set to $\overrightarrow{S_2}$, e.g. set $\overrightarrow{S_1} = (O1, O2)$ and $\overrightarrow{S_2} = (O3, O4)$, $Oi \in [0, +\infty)$ is the objective function and the lower value is optimal. if $O1 < O3$ and $O2 < O4$, then $\overrightarrow{S_1}$ dominates $\overrightarrow{S_2}$; if $O1 > O3$ and $O2 < O4$, the $\overrightarrow{S_1}$ is non-dominated by $\overrightarrow{S_2}$.

The non-dominated sorting is applied in all random two simulations in a population by tournament selection. A sequence of Pareto Fronts is then derived according to the non-dominated sorting. The first Pareto Front consists of the simulations that are not dominated by any other simulations; the second Pareto Front consists of the simulations that are only dominated by the first Pareto Front; third Pareto Front consists of the simulations that are only dominated by the first and the second Pareto Front, and thus all the simulations are assigned with a Pareto Front rank accordingly. Figure 5-9 displays the Pareto Front derived by non-dominated sorting. Objective 1 and 2 were objectives evaluated by the non-dominated sorting and three (first, second, and third) Pareto Fronts were highlighted in the figure, with the dots representing individual simulations.

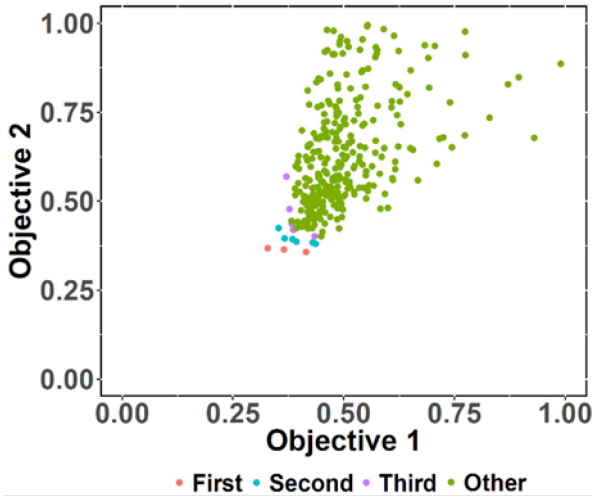

Figure 5-9 The Pareto Front (first, second, and third) displayed

b. Crowding distance

If the simulations are at the same Pareto Front, then their priority to be selected for the evolution is evaluated by the crowding distance. This crowding distance of one simulation is the summed distance (the difference of the objective function values) of this simulation to its adjacent two simulations. Figure 5-10 shows the crowding distance of simulation i, which is the sum of d1 and d2 and standardized by dividing the longest distance at this Pareto front. The larger the crowding distance is, the higher the priority the simulation is at its Pareto Front. Applying the crowding distance guarantees the diversity to be maintained in the evolution.

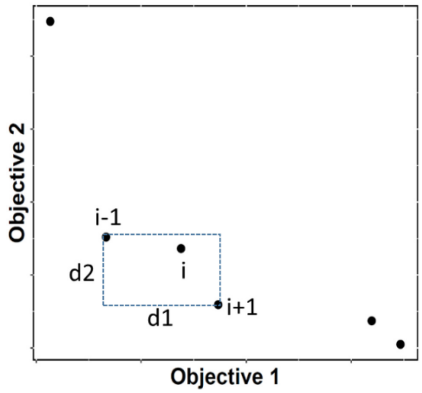

Figure 5-10 The crowding distance at one Pareto front with two objectives evaluated

c. Crossover and mutation

The individual parameter value of a simulation is a chromosome (see Figure 2-3). To perform the crossover and the mutation, each parameter value is assigned with a certain number of binary values (e.g. 6 bits, 010011), by a decoding approach and each bit is referred to as a gene. For parameter P1, its binary value is a vector $\overrightarrow{gene}$. The conceptual R code of the decoding approach is listed below. And the binary values of a parameter will follow the scheme reversely to be decoded back to the real value. The length of $\overrightarrow{gene}$ is bits, e.g. $\overrightarrow{gene}$=[0,1,0,0,1,1], then bits=6. The value range of P1 is [P1_min, P1_max].

```
sum = (P1-P1_min)×(2·bits-1)/(P1_max-P1_min)
For (i in 1:bits){
     If (sum%%2==1) {gene[i] = 1; sum = sum-1; sum = sum/2}
     else {gene[i] = 0; sum = sum/2}}
```

The crossover and mutation are then performed at the gene level, as displayed in Figure 5-11. The mutation occurs on the chromosome itself and crossover was performed between individuals of two populations on the chromosome at the same location. After crossover and mutation, via the reversed decoding approach, the binary values were transformed back into a real parameter value. Coefficients were also assigned to define the possibility of crossover and mutation occurring in one population.

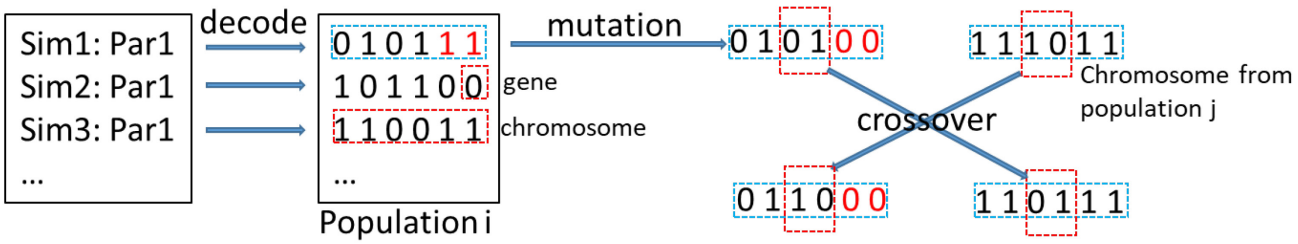

Figure 5-11 The demonstration of population, chromosome, gene; mutation and crossover process

In the study, the application of NSGA-II coupled with the SWAT model was performed in R and the script was adapted from the R code established by Whittaker (2016) and the Python code established by Ercan and Goodall (2016). The structure of the R scripts and execution code developed by the author as part of this study are listed in Appendix A.

5.3.2.3 The working procedure of SWAT with NSGA-II

The working procedure of applying NSGA-II to calibrate the SWAT model in this research is shown below in Figure 5-12.

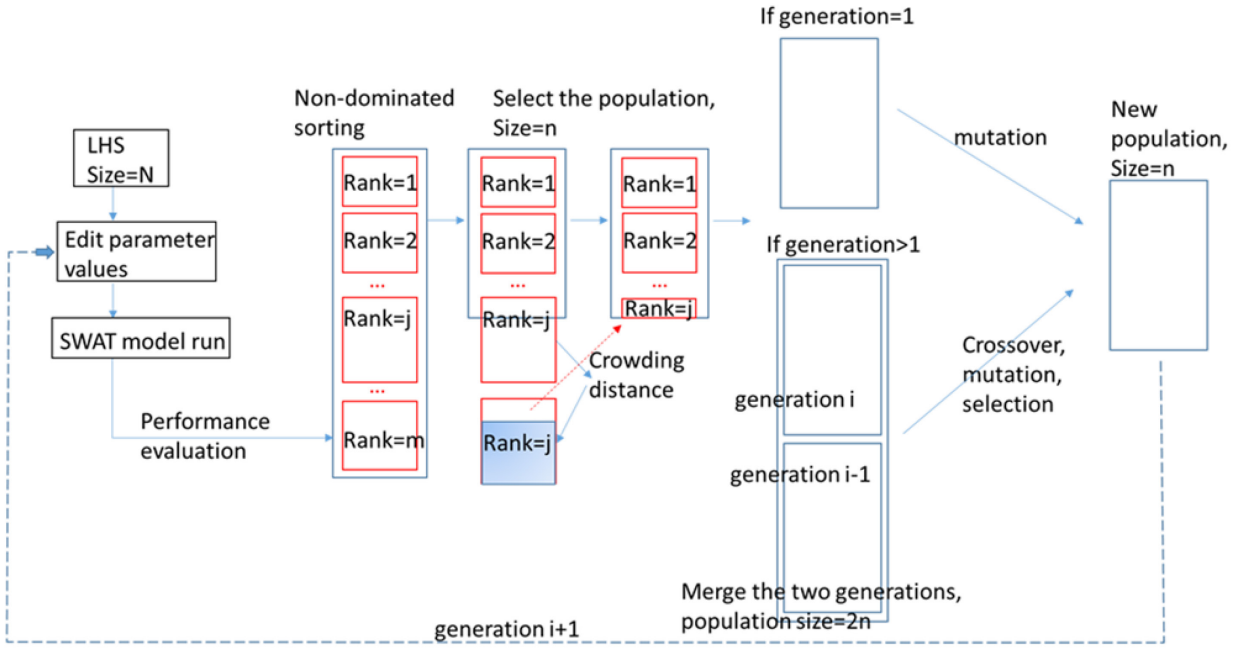

Figure 5-12 The calibration procedure of NSGA-II for SWAT mode, modified from Ercan and Goodall (2016)

a. The initial model execution

The SWAT model was first run with parameter sets obtained from LHS to derive the initial group of the simulations as the initial population (size=N).

b. Non-dominated sorting and crowding distance to derive the population

The non-dominated sorting was then performed on the initial population with N simulations to generate the sequential Pareto Fronts ('Rank' in Figure 5-12). The population was selected from the first Pareto Front until the size m was fulfilled. If the last Pareto Front can only be partially selected, e.g. Rank=j in Figure 5-12, then the crowding distance was calculated to rank the simulations in Rank j to complete the population.

c. Crossover and mutation

Every generation of NSGA-II can be seen as an iteration of simulations. In the first generation, only the mutation was applied in the old population to generate the new population. After the first generation, the new generation to be derived would firstly crossover with the previous generation, then merged with this previous generation and a combined population with size= 2n was derived. Then the mutation was performed on this combined population. Non-dominated sorting and crowding distance were performed again to derive the new generation with size=n from the combined population. The selection from the combined population aimed to keep the elitism of the population that the first Pareto Front of both previous and current populations can be selected to form the next population.

d. To continue the iteration

This newly generated population was then decoded reversely to derive the new parameter sets for the SWAT model to run. In this study, an upper limit number of generations was given to uniform the simulations executed by the three calibration approaches. When the upper limit was met, the iteration would terminate and the Pareto Front of the last generation was the result of the calibration.

5.3.2.4 C-function for convergence evaluation

The convergence of the Pareto Front indicates how well the former generation can surpass the latter generation, as a sign of whether more generations are needed for the iteration. The convergences are quantified by the C-function proposed by Zitzler and Thele (1999) and applied by Shafii and De Smedt (2009). The 'surpass' is indicated by how many of the simulations in the

current generation are dominated by the simulations from the previous generation. The definition of C-function is displayed below:

Equation 13 The formula of C-function

$$C(G_{n-1}, G_n) := \frac{|\overrightarrow{S'} \in G_n; \exists\, \vec{S} \in G_{n-1}: \overrightarrow{S'} \preceq \vec{S}|}{|G_n|}$$

G_n and G_{n-1} are the n and (n-1) generation; where $\vec{S}$ and $\overrightarrow{S'}$ are one random simulation in the generation, which is a vector of all objectives; $\overrightarrow{S'} \preceq \vec{S}$ means that $\overrightarrow{S'}$ is dominated or equal to $\vec{S}$. $C(G_{n-1}, G_n)$ is a fraction, of which the denominator is the number of simulation in G_n and the nominator is the number of simulation in G_n that is dominated by or equal to simulations in G_{n-1}. When $C(G_{n-1}, G_n)$ equals 1, all simulations in G_n are dominated by or equal to G_{n-1}; When $C(G_{n-1}, G_n) = 0$, all simulations in G_n dominate G_{n-1}.

Shafii and De Smedt (2009) proposed one criterion in their study to evaluate the convergence of NSGA-II that if C-function values remain equal to 1 for a certain number of consecutive generations (10 was used in their study), the iterations have converged and can be terminated.

5.4 The calibration procedure

Figure 5-13 displayed the calibration procedure applied in the study. 10 sensitive parameters were selected for hydrological calibration and each of them was calibrated separately at the three catchment clusters, therefore, 30 parameters were considered in total for the hydrological calibration in the study. The initial iteration of simulations was set to have the size of 2000 based on the author's experience. The initial parameter sets generated from LHS were applied by both single-objective and multi-objective calibration.

SUFI-2 was applied in the single-objective calibration by evaluating the NSE at the JK station. The updated parameter sets for the next iteration of another 2000 simulations were also obtained by SUFI-2. When applying the AW or ED, parameter sets for the next iteration were still generated by the LHS of the initial value ranges and should not overlap the ones from the initial iteration. When applying NSGA-II, the simulations obtained from the initial iteration were evaluated by the three categorized objectives, which were 1) the NSE of discharge at the three stations; 2) the inverse NSE of discharge at the three stations; 3) water balance components. NSGA-II was set to

have a population size of 50 and to be iterated with 40 generations (see Section 6.3.1.2). Therefore, each calibration procedure implemented 4000 simulations.

For the comparison of the single- and multi-objective calibration approaches, the top simulations were selected for each approach. The simulations at the Pareto Front of the last generation of NSGA-II were selected. The size of the simulations was applied to select the number of simulations obtained by ED or AW after sorted according to the performance.

The best simulation of ED was then selected to be the base model for nitrogen calibration. The objectives applied in the nitrogen calibration were the NSE and the inverse NSE at the three water quality stations. An initial iteration of 1000 simulations was first implemented. Then SUFI-2, ED, and NSGA-II were applied for another 2000 simulations.

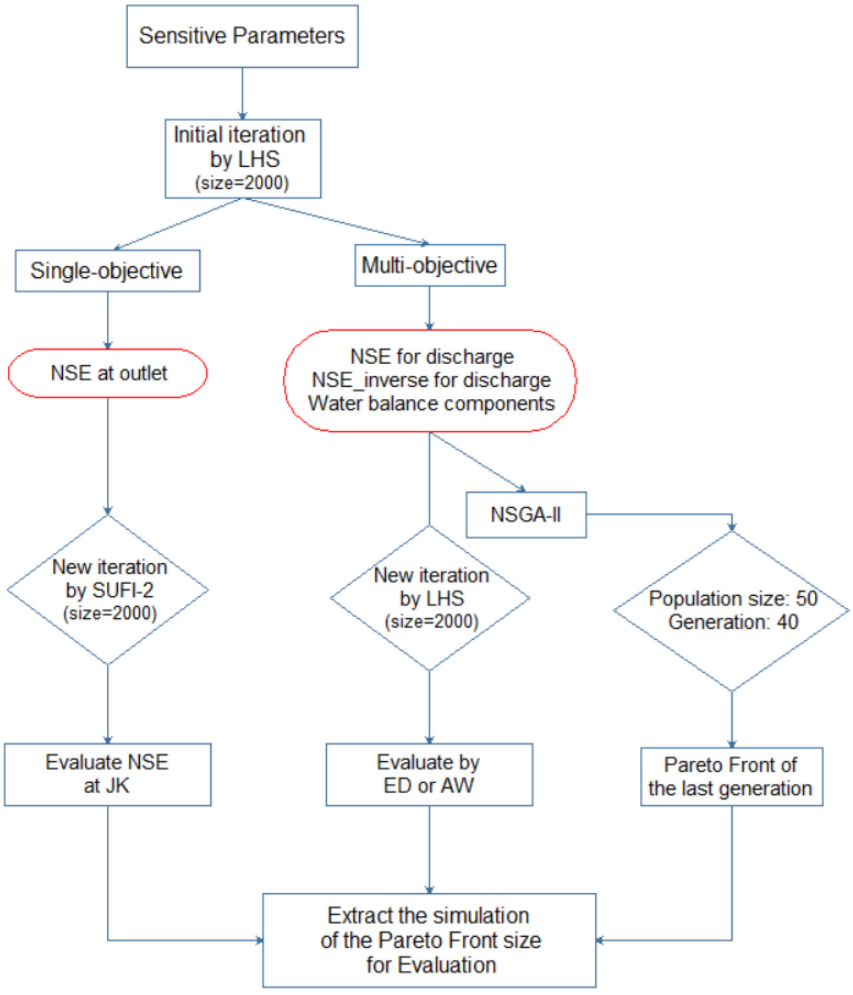

Figure 5-13 The hydrological calibration procedure

5.5 Nitrogen load analysis

In order to identify the critical source areas (CSAs), the total nitrogen (TN) contributed to the stream from each HRU was derived. The TN load at each HRU was represented by the total annual load (kg) to the stream, which included the nitrate transported with the surface runoff, lateral flow, and groundwater return flow as a solute and the organic nitrogen absorbed to the sediment and transported. The TN to identify the CSAs was only from NPS and the TN from industrial and municipal sewers were not included. Then, the TN load of each HRU were ranked at a descending order and the cumulative TN was calculated as to classify the classes: Class I included the HRUSs contributing the first 25% of the TN; Class II included the HRUs of the second 25% (class I and II contributes 50% of the TN in total); Class III the third 25% and Class VI the remaining 25%. The lower the area of the region that covers the individual class, especially Class I, II, and III, the more critical the region is defined.

5.6 Uncertainty analysis

5.6.1 Nitrogen load simulation

To quantify the uncertainty of the nitrogen load simulation, the concept of 95PPU and r-factor were applied in the study. In SUFI-2, the 95PPU was applied for all simulations in an iteration. In this study, only the algorithm of 95PPU that denotes the coverage of 2.5^{th} and 97.5^{th} percentiles of the simulations, and r-factor were implemented. The simulations to obtain the 95PPU in nitrogen load simulation were the first 1000 simulations of each calibration approach.

5.6.2 The uncertainty band of the CSAs

As stated in Section 4.3.1, the prediction uncertainty for the CSAs was generated in a simplified process. Only two of the parameter sets and the respective simulation results represented the range of the uncertainty band of the in-stream nitrogen load. For single-objective-calibration, these two parameter sets were the ones that derive the highest and the lowest PBIAS to the best nitrogen load simulation at LF station. For multi-objective calibration, the PBIAS were computed for all the stations, and the lowest/highest PBIAS were obtained by sorting the Euclidean Distance of the two multi-objective calibrations.

6 Results

In this chapter, the analysis of the results was narrated from model set-up, single-objective and multi-objective calibration to nitrogen load analysis, concerning the research procedure. Section 6.1 describes the HRUs and nitrogen input generated for the model; Section 6.2 explained the parameters selected to calibrate the hydrological and the nitrogen processes sequentially. Section 6.3 focused on the comparison of the three calibration approaches in terms of their performances on the individual objectives; Section 6.4 displayed the results of applying the calibrated model to predict the nitrogen load.

6.1 The description of model construction

6.1.1 Watershed delineation and the HRU classification

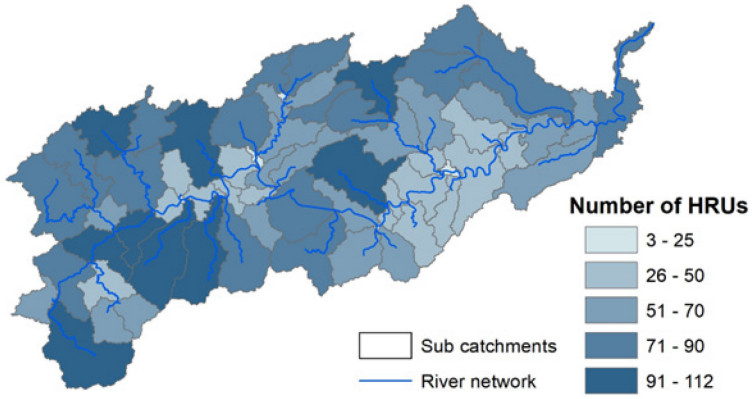

Figure 6-1 The subcatchment delineated by the model and the river network

The Yuan River Catchment (YRC) was delineated into 67 subcatchments based on the DEM (30m×30m) and no subcatchment was smaller than 5000ha as displayed in Figure 6-1. Xu (2016a) has concluded that the TN load and the critical source areas were sensitive to the watershed delineation and the HRUs' generation. Therefore, to maintain the characteristics of the YRC in terms of landuse and soil distribution, no landuse or soil type was eliminated when generating the HRUs. The slope was classified into three groups as lower than 10%, 10% to 30%, and higher than 30%, to identify the plains and mountainous regions. 4136 HRUs were identified in total. The number of HRUs in each subcatchment was displayed in Figure 6-1. The number of HRUs in the downstream, though with a relatively large area, is low. This is due to the reason that the landuse at the downstream is dominated by large areas of paddy fields. The cumulative plot of the HRUs

and the respective areas are displayed in Figure 6-2, as 1000 HRUs occupies approx. 90% of the entire YRC. The large number of HRUs with small areas was partially contributed by villages spreading over the catchment.

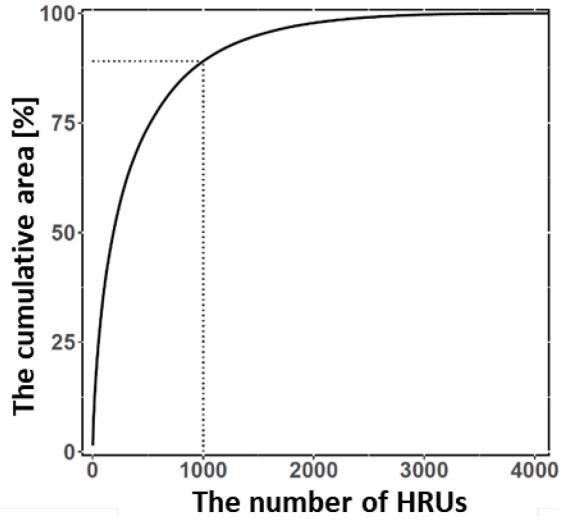

Figure 6-2 The cumulative plot of HRUs

6.1.2 The input total nitrogen from individual nitrogen sources

The annual total nitrogen (TN) from each source is displayed below at each of the subcatchment, as shown in Figure 6-3 a-e, and the portion of each source is displayed in Figure 6-3 f.

a. Fertilizer

86.4% of the total nitrogen input was contributed by the fertilizer. At the subcatchment scale, the amount of fertilizer as input was determined by the area of the paddy field in the subcatchment. A higher rate of fertilizer can be observed at the downstream of the YRC with a denser distribution of the paddy field.

b. Feedlots and rural households

Nitrogen from feedlots and rural households only contributed 2.3% of the total nitrogen input. Their distribution was determined by the rural landuse size at each of the subcatchment. Therefore, the higher nitrogen input in some subcatchments denoted a relatively higher rural population.

c. Point sources

Nitrogen from point sources in the YRC occupied 0.6% of the total nitrogen input. The sources were mainly distributed at the subcatchments where the cities and big towns were located. Only published sources were considered in the research and other regions were unrecorded (missing).

d. Atmospheric deposition

The atmospheric nitrogen content was assumed to be homogeneous over the YRC, therefore, the distribution of atmospheric deposition was determined by the annual precipitation of each subcatchment. The atmospheric deposition of nitrogen occupies 10.7% of the total nitrogen input.

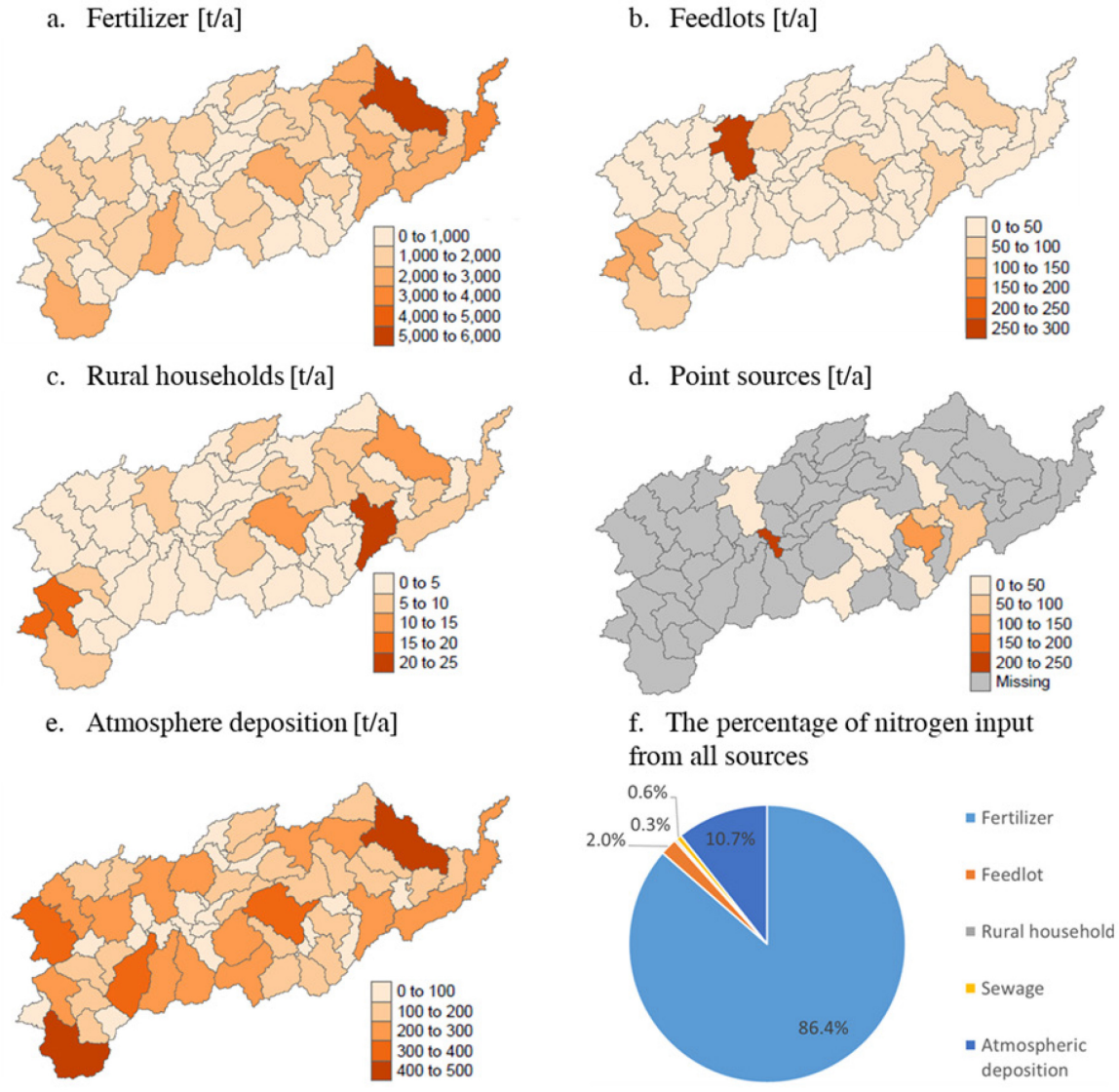

Figure 6-3 The annual total nitrogen input (t) from all the main sources in YRC and their portions

6.2 The evaluation of the sensitive parameters

The initial selection of the parameters in the study was based on previous catchment studies using the SWAT model (see Section 4.1) and the analysis of the relevant. The global sensitivity analysis was then performed to obtain the first 10 sensitive parameters. The processes targeted by the discharge calibration were 1) the surface runoff; 2) the groundwater recharge; 3) the channel process. The targeted nitrogen relevant processes were 1) the nitrogen transformation in the soil layers, including the plant-uptaken; 2) the transport of each nitrogen species; 3) the in-stream process.

The 10 parameters were assigned with a specific value at each subcatchment cluster (see Figure 6-4). The subcatchment clusters controlled by the individual hydrometric (LX, MZ, and JK) or water quality stations (1_Y, 4_F, and 6_L) are displayed in Figure 6-4. Therefore, overall, 30 parameters were selected for both hydrology and nitrogen calibration. The hydrometric or the water quality station is the outlet of the related subcatchment cluster. However, there were no measurements available at the outlet of the YRC, the subcatchments downstream of JK or 6_L were assumed to share the same parameter value with the subcatchments between MZ and JK or between 4_F and 6_L.

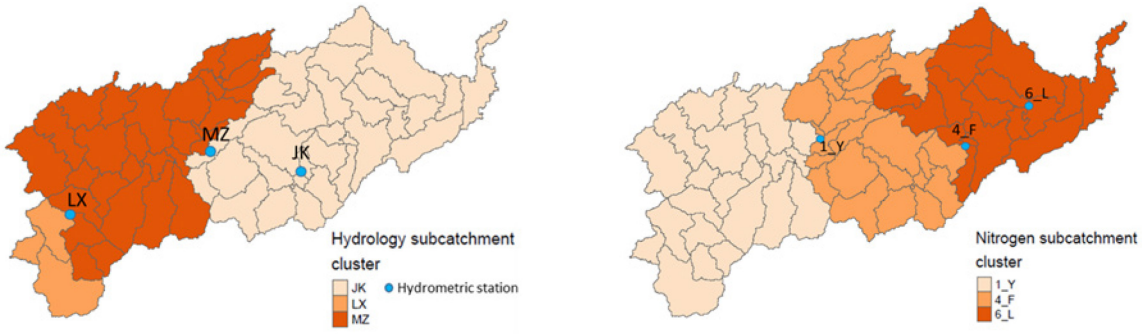

Figure 6-4 The subcatchment clusters applied in the calibration of hydrology and nitrogen.

6.2.1 Hydrology

Table 6-1 lists the 10 sensitive parameters and their calibrated values at the three subcatchment clusters (see Figure 6-4) by the three calibration approaches. The calibrated values were from the best simulation (Section 6.3.2) of each approach.

Table 6-1 The selected parameters for the hydrological model calibration

Parameter	Description	Default/ value range	Cluster	Calibration approach		
				SUFI-2	ED	NSGA-II
OV_N	Manning's n value for overland flow	0.01-0.15/ (0.01-0.5)	LX	0.35	0.11	0.27
			MZ	0.14	0.07	0.14
			JK	0.02	0.36	0.21
ESCO	Soil evaporation compensation factor	0.90/(0-1)	LX	0.31	0.93	0.317
			MZ	0.21	0.17	0.984
			JK	0.13	0.63	0.063
POT_FR	The fraction of HRU area that drains into the pothole	0/(0-1)	LX	0.01	0.69	0.51
			MZ	0.09	0.17	0.32
			JK	0.98	0.59	0.22
POT_VOL	The initial volume of water stored in the pothole [mm]	0/(0-100)	LX	41	47	0
			MZ	34	43	56
			JK	53	39	75
DEP_IMP	Depth to the impervious layer for modeling perched water tables [mm]	6000/(0-6000)	LX	5822	3161	3143
			MZ	3592	3886	5429
			JK	2855	3459	2952
DIS_STREAM	Average distance to the stream [m]	30/(0-1000)	LX	32	317	556
			MZ	300	419	63
			JK	426	680	460
CN2	SCS curve number for moisture condition II	-[a]/(-0.2-0.2)	LX	0.09	0.06	-0.02
			MZ	0.16	0.03	0.08
			JK	0	-0.05	0.14
SOL_AWC	Available water capacity of the soil layer [mm/mm]	-[a]/(-0.2-0.2)	LX	0	-0.10	0.18
			MZ	0.01	-0.14	-0.19
			JK	0.18	-0.01	0.07
GWQMN	Threshold depth of water in the shallow aquifer required for return flow to occur [mm]	1000/(0-5000)	LX	3016	2262	1508
			MZ	3921	785	79
			JK	2362	2169	317
CH_N2	Manning's n value for the main channel	-[a]/(-0.2-0.2)	LX	-0.11	-0.04	0.05
			MZ	-0.06	0.01	-0.15
			JK	-0.13	-0.03	0.12

a. *The parameter values were spatially distributed at the HRU level, and a percentage (relative) change was applied.*

The parameters to determine the soil properties are SOL_AWC, ESCO, and DEP_IMP. SOL_AWC is the available water capacity, which is to compute the field capacity water content in the soil; only if the current soil water content is higher than the field capacity, the layer is regarded as being saturated and the vertical (percolation and evaporation) or horizontal (lateral flow) water movement occurs. Therefore, the higher the SOL_AWC is, the higher the water content needs to be in one soil layer to trigger the movement of the water. ESCO is the coefficient to compensate for the evaporation rate in the soil when the surface layer of the soil cannot support enough amount of evaporation. The lower the value of ESCO is, the higher the evaporation rate is contributed by the soil. DEP_IMP is the depth of the impervious layer. The actual value of DEP_IMP, if deeper than the soil profile, does not impact the amount of percolation water. However, if the soil is fully saturated, it can increase the height of the water table, which would then decrease the baseflow. The calibrated ESCO of SUFI-2 (single-objective calibration) led to an increase of the soil evaporation rate in all clusters, whereas ED had an increase in MZ and NSGA-II had an increase in LX and JK. The calibrated SOL_AWC of SUFI-2 tended to set a higher threshold for water to move in the JK cluster whereas ED had the opposite effect in all clusters and NSGA-II in MZ.

The surface runoff is impacted by CN2, OV_N, POT_FR, and POT_VOL. CN2 is directly proportional to the surface runoff rate. The parameters to impact the overland flow from the paddy field were also sensitive in the model, as the paddy field occupies approx. 30% of the YRC. POT_FR controls the fraction of the water yield, including surface runoff, baseflow, and lateral flow in an HRU that flows into the paddy field. When the maximum water storage in one paddy field is constant, the higher the POT_FR and POT_VOL are, the higher the amount of water will flow out of the paddy field if the water storage is exceeded or if the operation to release the water is conducted. The traveling time of the surface runoff is affected by OV_N and DIS_STREAM, as the higher the OV_N and DIS_STREAM, the longer the travel time the overland flow will have. The individual HRUs in a subcatchment have no spatial information of the distance to the stream, therefore DIS_STREAM needs to be calibrated. The three calibration approaches all tended to increase the surface runoff at the MZ cluster. The value of POT_FR of SUFI-2 differed from ED and NSGA-II to only have a higher rate of water draining to the paddy field at JK clusters.

The groundwater flows to the river were impacted by GWQMN, only if the groundwater table exceeds the GWQMN, base flow would occur. CH_N2 is the manning's n value to control the velocity of the flow in the channel. GWQMN with SUFI-2, increased the threshold of the baseflow to occur at all stations, whereas with ED in MZ and NSGA-II in MZ and JK, the threshold values decrease as to increase the potential of baseflow to occur.

6.2.2 Nitrogen

The sensitivity analysis of the nitrogen module calibration was performed after fixing the hydrological parameters (see Table 6-1) with the best simulation obtained by ED. Three parameters CDN, SURLAG, and NPERCO were global parameters and one value was assigned to the entire catchment. The remaining 9 parameters were assigned to each subcatchment clusters (see Figure 6-4) 1_Y, 4_F, and 6_L individually.

The nitrogen transformation and transport processes include the in-soil and in-stream processes. The 12 sensitive parameters (3 global parameters) with their default values and the reasonable value ranges are listed in Table 6-2 and Table 6-3. Nitrate transport in the surface runoff was impacted by SURLAG, NPERCO, and LAT_TTIME. SURLAG determines surface runoff reduction in the region with the concentration-time larger than one day and the transport of nitrogen in sediment with the runoff is also reduced due to the lag time. NPERCO determines the portion of the mobile nitrate in the soil to surface runoff, lateral flow, and percolation. LAT_TTIME determines the travel time of the lateral flow. In terms of the parameter values, SUFI-2 resulted in a longer lag time with a higher portion of nitrate in the runoff; whereas ED or NSGA-II resulted in a shorter lag time with lower nitrate portion. Meanwhile, the LAT_TTIME of ED and NSGA-II obtained a longer travel time at 6_L and shorter at 1_Y or 4_F, whereas SUFI-2 resulted in the opposite effect.

Table 6-2 The selected parameters for the nitrogen module calibration – the global parameter

Parameter	**Description**	**Default/ value range**		**Calibration approach**		
			Cluster	SUFI-2	ED	NSGA-II
CDN	Denitrification rate coefficient	1.40/ (0-3)	-	0.47	0.57	0.19
SURLAG	Surface runoff lag time [days]	4/(1-24)	-	8.76	21	11
NPERCO	Nitrogen percolation coefficient	0.2/(0-1)	-	0.2	0.96	1

Table 6-3 The selected parameters for the nitrogen module calibration – the cluster-specific parameter

Parameter	Description	Default/ value range		Calibration approach		
			Cluster	SUFI-2	ED	NSGA-II
MUMAX	Maximum specific algal growth rate at 20℃	2/(1-3)	1_Y	1.90	1.90	2.59
			4_F	1.58	1.08	1.38
			6_L	1.02	1.37	1.03
HLIFE_NGW	The half-life of nitrate reduced by a biological and chemical process in the shallow aquifer [days]	0/(0-200)	1_Y	96	148	197
			4_F	10	149	19
			6_L	28	76	140
P_N	Algal preference factor for ammonia	0.5/(0-1)	1_Y	0.13	0.96	0.06
			4_F	0.22	0.48	0.25
			6_L	0.43	0.03	0.89
BC3	The local rate constant for the hydrolysis of organic nitrogen to NH_4 at 20℃ [/day]	0.21/ (0.2-0.4)	1_Y	0.21	0.20	0.30
			4_F	0.22	0.30	0.30
			6_L	0.37	0.31	0.20
RK4	Sediment oxygen demand rate in the reach at 20°C [mg $O_2/(m^2 \cdot day)$]	2/(0-100)	1_Y	26	32	25
			4_F	34	65	52
			6_L	95	72	22
SOL_CBN	Organic carbon content	-[a]	1_Y	-0.11	0.20	0.03
			4_F	0.07	-0.14	-0.12
			6_L	-0.12	0.17	0.17
RS4	Rate coefficient for organic N settling in the reach at 20°C	0.05/ (0.001-0.100)	1_Y	0.07	0.08	0.08
			4_F	0.53	0.41	0.27
			6_L	0.03	0.10	0.06
LAT_TTIME	Lateral flow travel time [days]	0/(0-180)	1_Y	13	8	9
			4_F	29	4	8
			6_L	12	20	25
ERORGN	The organic N enrichment ratio	(0-5)	1_Y	3.3	1.64	0.48
			4_F	2.4	0.39	0.63
			6_L	4.9	1.34	0.32

a. *The parameter values were spatially distributed at the HRU level, and a percentage (relative) change was applied.*

The in-soil transformation processes were impacted by CDN, HLIFE_NGW, and SOL_CBN. CDN determines the denitrification reaction rate in the soil. The higher the CDN is, the higher the consumption of nitrate in the soil is. HLIFE_NGW determines the amount of the nitrate consumed

the shallow aquifer after the biological and chemical processes; it is inversely proportional to the amount of nitrate remained and the nitrate concentration in the groundwater return flow. The SOL_CBN was to calibrate the organic carbon content in the soil and the mineralization would occur in the soil layer when the ratio of C: N was lower than 20:1. The mineralization would result in an increase of ammonia nitrogen and nitrate in the soil. ED tended to have a higher reduction of the in-stream nitrate from baseflow at all clusters, NSGA-II at 1_Y and 6_L, and SUFI-2 only at 1_Y. Opposite impact of SOL_CBN can be observed that SUFI-2 tended to increase the carbon content in 4_F and decrease the other two clusters, whereas ED and NSGA-II tended to decrease the value in 4_F and increase the rest. Therefore, SUFI-2 could result in a lower content of ammonia nitrogen and nitrate in the soil in 1_Y and 6_L.

The in-stream processes were impacted by RK4, BC3, P_N, and MUMAX. MUMAX determines the growth rate of the algae at 20℃, and it is proportional to the algae concentration and thus more biomass was transformed to the organic nitrogen; RK4 determines the amount of oxygen required to decompose the organic nitrogen, and therefore impacts the oxygen net increase; BC3 determines the conversion rate of organic nitrogen to ammonia nitrogen via mineralization and it was inversely proportional to the organic nitrogen concentration in the stream. P_N determines the preference portion of algae to uptake the nitrogen from the ammonia nitrogen pool. The BC3 value and RK4 showed the similarity of decreasing organic nitrogen in the stream by all the calibration approaches.

6.3 Single-objective vs. multi-objective calibration

2000 simulations with the parameter sets created by Latin Hypercube Sampling (LHS) were first executed by each approach and followed by another 2000 simulations of each approach, as summarized in Table 6-4. NSGA-II used the population size 50 and 40 generations. The computational time of each approach is also listed, for the PC of core i7, 4.20GHz, and 16G RAM.

Table 6-4 The summary of the calibration approaches

Calibration category	**Approach**	**Number of simulations**	**Computational time**
Single-objective	SUFI-2	2000 (LHS)+2000 (one iteration)	Approx. 1.5 min/simulation and 2 days/2000 simulations
Multi-objective	Euclidean Distance (ED)	2000 (LHS)+2000 (one iteration)	
	NSGA-II	2000 (LHS)+40 Generation(/50)	Approx. 2 min/simulation, and 2.5 days/2000 simulations

Table 6-5 lists the 13 objectives considered in the multi-objective calibration for hydrology. The individual 13 objectives and the three categorized objectives: C_NSE, C_NSE_in, and C_WB. NSE at JK (NSE_JK) station is the only variable to be calibrated by single-objective calibration. The abbreviation would be used further to represent each objective. The standardization of the evaluation statistics of each objective is listed in Table 5-4.

Table 6-5 The individual objectives and the categorized objectives

Categorized objectives	Abbreviation	Individual objectives	Description	The standardized value range
Multi-site with NSE	C_NSE	NSE_LX; NSE_MZ; NSE_JK	**NSE_**: the NSE value at LX, MZ or JK station; **C_NSE**: the ED of the three objectives	(0,+∞)
Multi-site with NSE_in	C_NSE_in	NSE_in_LX; NSE_in_MZ; NSE_in_JK	**NSE_in_**: the inverse NSE value at LX, MZ or JK station; **C_NSE_in**: the ED of the three objectives	(0,+∞)
Water balance components	C_WB	LX_s; MZ_s; JK_s; LX_g; MZ_g; JK_g	**_s** and **_g**: the bias of the simulated to the observed surface runoff ratio and the baseflow ratio of the discharge in LX, MZ, and JK subcatchment cluster	(0,+∞)
		AET_{cover}	AET_{cover}: the coverage of the AET by the MOD16 ET uncertainty range	(0,1)
Single objective		NSE_JK	NSE at JK station	(0,+∞)

6.3.1 Multi-objective calibration algorithms

6.3.1.1 AW vs. ED

To ensure the efficiency of non-dominated sorting, the number of objectives needed to be reduced. The 13 objectives were thus categorized into C_NSE, C_NSE_in, and C_WB. The categorized objective was an aggregated value that represented the performance of all the objectives under this category, e.g. C_NSE is the aggregated ED of the three objectives, which are NSE_LX, NSE_MZ, and NSE_JK.

However, to compare the ED or AW with NSGA-II, the reliability of substituting the 13 individual objectives by the 3 categorized objectives needed to be proved. This reliability was evaluated by the consistency of the performance of the categorized objectives, in comparison to the individual objectives. The top 10 simulations, selected by categorized objectives and by individual objectives are displayed in Figure 6-5 a-c, the abbreviations of the scenarios are listed in Table 6-6.

Table 6-6 The abbreviations of the scenarios to compare ED and AW

Abbreviation	Scenarios
Cate_ED	Categorized objectives selected by ED
Cate_AW	Categorized objective selected by AW
Indi_ED	Individual objectives selected by ED
Indi_AW	Individual objectives selected by AW

Figure 6-5 a-c demonstrates the performance of all the 13 individual objectives from Indi_AW, Cate_AW, Cate_ED, and Indi_ED,. The violin plot displays the distribution of the objective function values and the shape of objects indicated the density of the value distribution of the NSE, NSE_in, and the water balance objectives. Among all the 13 objectives, when comparing the consistency of the individual and the categorized objectives, it can be observed that majority of the objectives selected based on the Cate_ED and Indi_ED derived the value distribution with no significant differences, except the objectives LX_g (Figure 6-5 c) and NSE_in_MZ (Figure 6-5 b), where Cate_ED produced a more concentrated distribution of the top 10 simulations than Indi_ED. However, by AW, only the objective NSE_LX can be observed with no significant difference between Cate_AW and Indi_AW. Therefore, ED was selected instead of AW for the study, due to its higher consistency of the performance of the categorized objectives. On the other hand, the results also indicated the applicability to use categorized objectives to represent individual objectives.

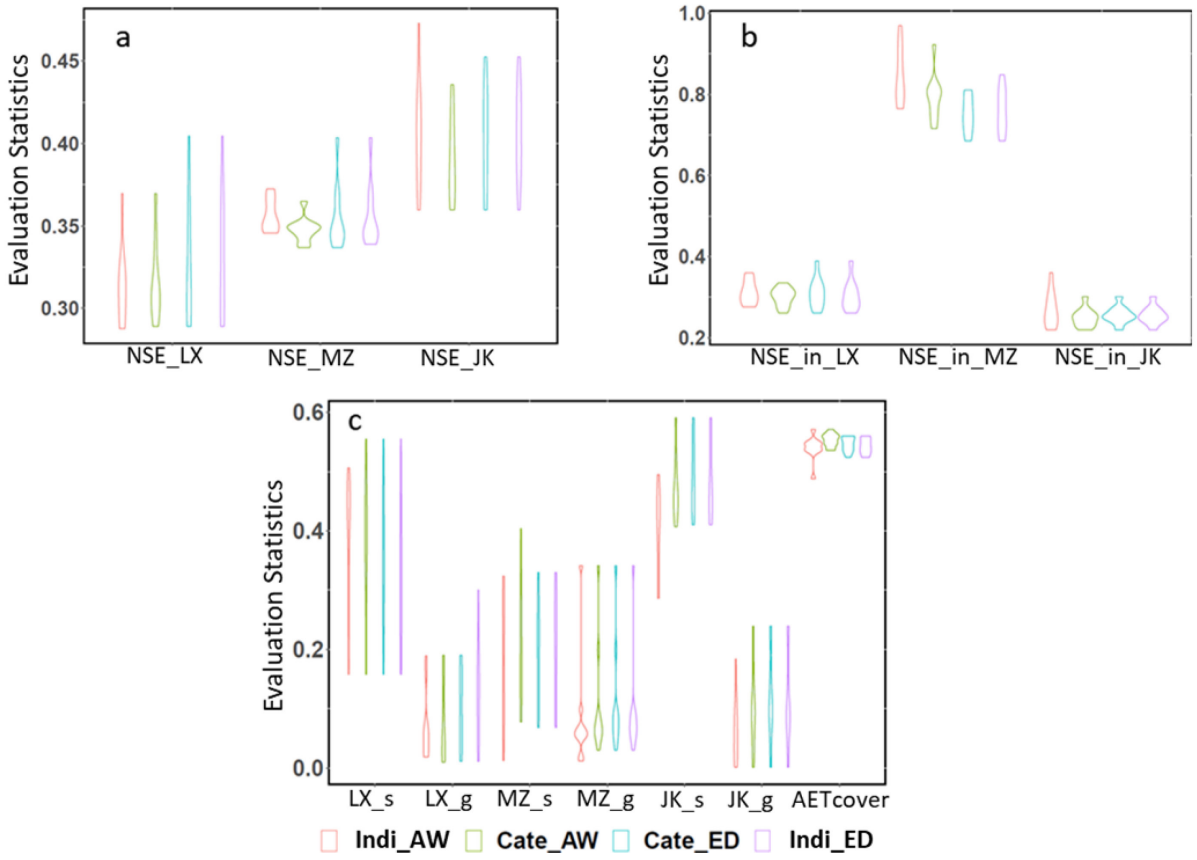

Figure 6-5 a-c. The comparison of individual and categorized objectives with ED or AW displayed by the violin plot

6.3.1.2 The population size of NSGA-II

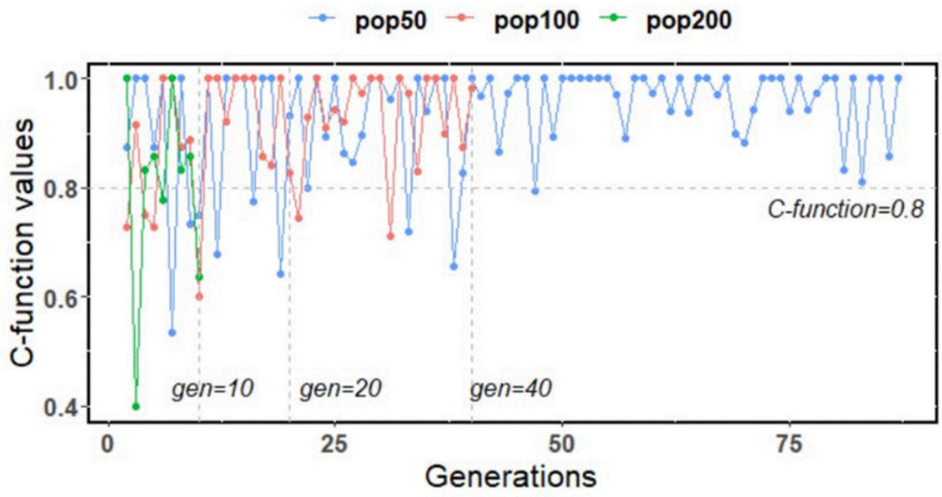

Figure 6-6 The convergence evaluated by C-function values of population size 50, 100 and 200

No clear guide to the population size has been summarized by previous studies. Therefore, the proper population size was first analyzed in this study. The population size 50, 100, and 200 were

selected and the number of generations was listed below in Table 6-7. The convergence of the Pareto Front was evaluated by the C-function. The value of the C-function indicates the ratio of the results in the former generation that dominates the current generation. Figure 6-6 displays the relation between the C-function value and the number of generations.

Table 6-7 The selected population size and number of generations

Population size	Generations	Total number of simulations
50	80	4000
100	40	4000
200	10	2000

Due to the comparably long time required to execute the calibration, e.g. 4000 simulations for each approach and approx. 2 mins/simulation (approx. 2.5 days/2000 simulations), the number of simulations is a critical factor to determine the efficiency of the calibration approaches in this research. To be consistent with the simulations of the second iteration by SUFI-2 and ED, the population size that can reach the convergence within 2000 simulations was preferred. Shafii and De Smedt (2009) proposed to determine the convergence reached if the C-function is 1 in a consecutive 10 simulations. In this study, due to the preferred lower number of simulations, the convergences were defined to be reached that C-function can be always above 0.8 after a certain number of generations.

In Figure 6-6, the C-function values are fluctuating. As the generations increase, the lowest value of the C-function is also increasing, which is shown by the population size 50 and 100. With population size 50, when the generations exceeded 40, its C-function value all reached 0.8 with only one exception. With population size 100, when the generations exceeded 20, 2 exceptions under 0.8 can be observed. However, with population size 200, though only 10 simulations were executed, it can be estimated that more generations were required to reach the convergence. When to compare the convergence in this case, independent from the number of simulations, the larger the number the generation had, the higher the possibility the convergence can be reached. Therefore, the population size 50 and 100 were more suitable than population size 200.

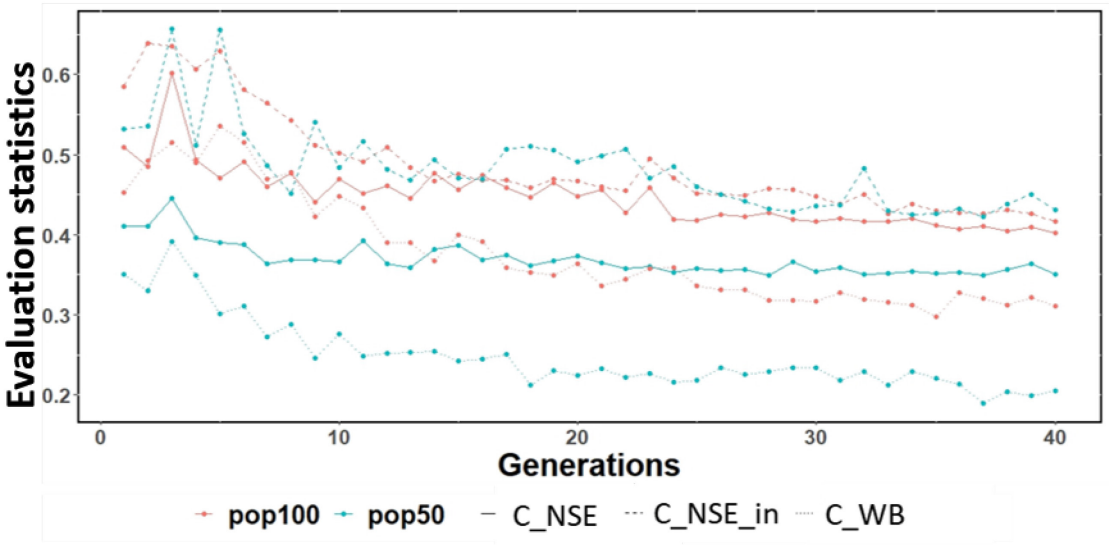

Figure 6-7 The performance of each generation with population size 50 and 100

In Figure 6-7, the performance of the population size 50 (in blue) and 100 (in red) are displayed for the three categorized objectives in different line types. Independent from the population size, before 10 generations, both population size 50 and 100 derived fluctuating evaluation statistics on all objectives. A more stable performance can be reached as more generations were implemented. However, no significant improvement of the performance can be observed after 20 generations. Moreover, the population size 50 derived a significantly better performance of C_NSE and C_WB than the population size 100.

The fluctuation of the convergence is determined by the evolution algorithm used by NSGA-II that the mutation and crossover would increase the variety in a population and meanwhile, elitism can also be maintained. If the number of simulations is limited, the smaller the population size is, the more generations can be implemented. Therefore, in this study, the population size 50 was selected for both hydrological calibration and nitrogen calibration.

6.3.1.3 Pareto Front of NSGA-II

Figure 6-8 displays the performance of the categorized objectives (C_NSE, C_NSE_in, and C_WB) of the 40 generations (2000 simulations) derived by NSGA-II. Two Pareto Fronts were highlighted, the Pareto Front (3_Pareto_Front) of the 40th generation (size=31) and the Pareto Front (2_Pareto_Front_aggr) of the 2000 simulations (size=96, excluding the overlapped ones with 3_Pareto_Front). The performance of the simulations at the Pareto Front dominated the other

simulations outside the Pareto Front, but they were not dominated by each other and thus had an equivalent performance. Due to the possibility of mutation and crossover (the possibility were set to be 40% in this study) in every generation, Pareto Fronts of each generation were not identical. Thus, more simulations could be observed at 2_Pareto_Front_aggr, and simulations with better performance also existed. As displayed in Figure 6-8, the 2_Pareto_Front_aggr consists of the tangent plane as a curved surface above the 2000 simulations, whereas the 3_Pareto_Front forms the tangent plant of the 40th generation.

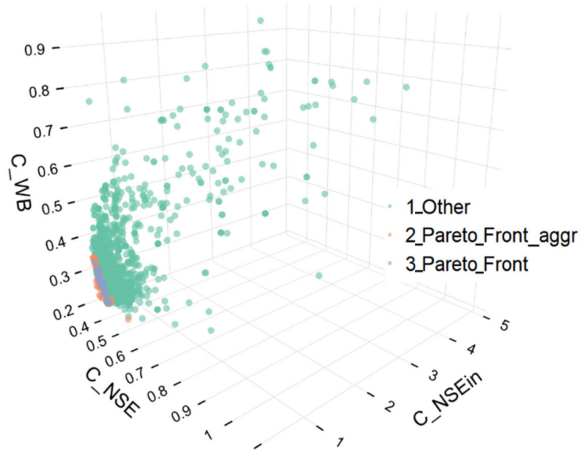

Figure 6-8 The Pareto Front after 40 generations

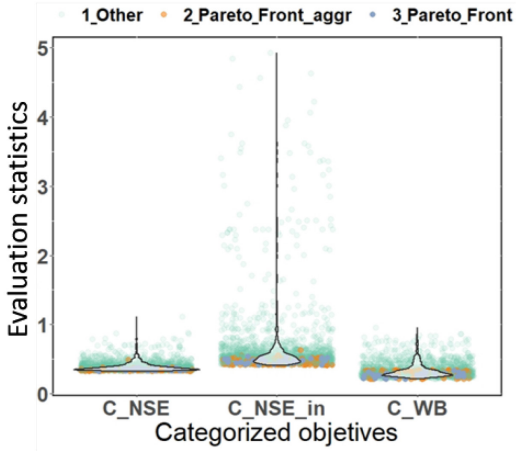

Figure 6-9 The performance of the 2000 simulations derived by NSGA-II, displayed by a violin plot

The three categorized objectives display different performances, as indicated in Figure 6-9. C_NSE has a lower uncertainty as the majority of the values are concentrated below 0.5 and the highest value of C_NSE is slightly over 1. C_NSE_in has the largest uncertainty, where values are concentrated at around 0.5 and the highest range could reach 5. Figure 6-10 demonstrates the performance of the 13 individual objectives of the 31 simulations at 3_Pareto_Front with a radar chart, where each axis demonstrates the value of one objective. Simulations selected by the varied ratio of categorized objectives, i.e. 100% C_NSE, 100% C_NSE_in, or 100% C_WB, are also displayed in Figure 6-10. The uncertainty demonstrated in Figure 6-9 is consistent with the value ranges of the individual objectives in Figure 6-10, as a rather high uncertainty of NSE_in_MZ (discharge evaluated by NSE_in at MZ) can be observed. This higher uncertainty might due to the lower reliability of the observations obtained from extrapolation and the uncertainty was especially high at the low to normal values.

C_WB is the categorized objective of the water balance components and its objective values should all within 0 to 1. AET_{cover} ranges from approx. 0.45 to 0.6 (in Figure 6-10) and it demonstrates the insensitivity of the evapotranspiration (ET) processes of the Pareto Front simulations to the calibration that only a 40% to 55% of the ET covered by the MOD16 ET uncertainty band. When compared to the PBIAS of surface runoff and baseflow ratio, LX derived the highest uncertainty and JK derived a larger bias of the surface runoff ratio. The simulations highlighted in Figure 6-10 indicates the trade-off among objectives when one objective was emphasized over the others. The simulations of 100% C_NSE derived the simulation with the lowest values of C_NSE, however, neglecting the performance of C_NSE_in (e.g. NSE_in_MZ) and C_WB (e.g. LX_g). The trade-off among other objectives can also be observed with simulations of 100% C_NSE_in or 100% C_WB. When C_NSE, C_NSE_in, and C_WB were assigned the same weight, the performance of the objectives was also in-between the extreme cases.

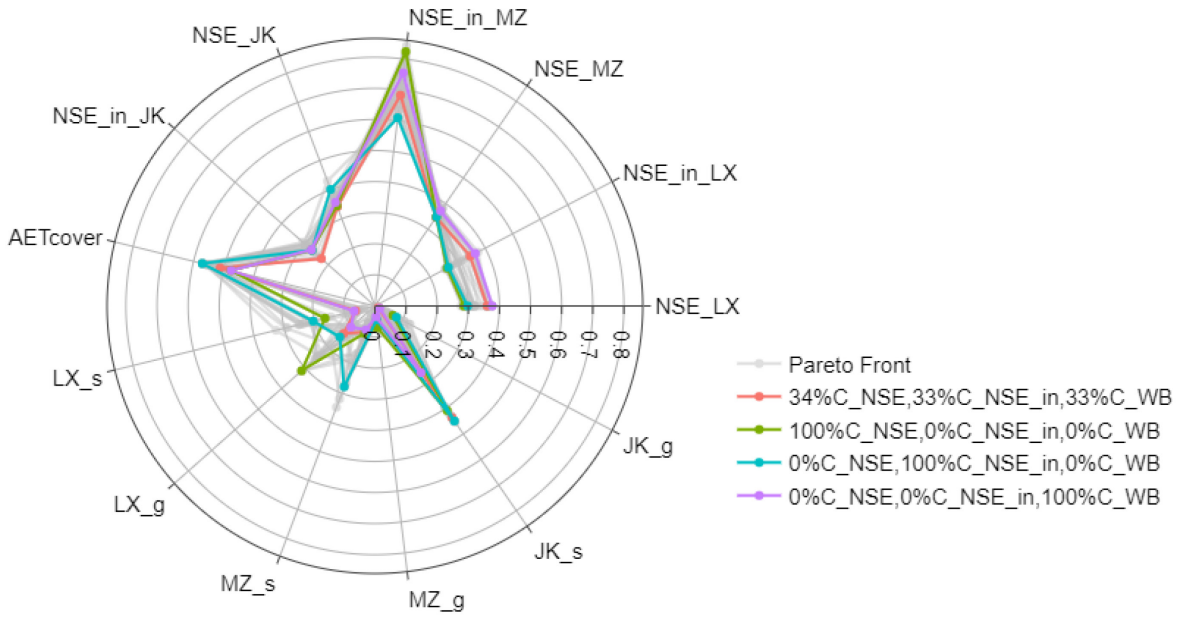

Figure 6-10 The radar chart of the individual objectives at the Pareto Front

6.3.2 Hydrological calibration

31 simulations were obtained at the Pareto Front (3_Pareto_Front) by applying NSGA-II. Therefore, the top 31 simulations were also sorted from single-objective calibration (SUFI-2) and multi-objective calibration with Euclidean Distance (ED). The best simulation of SUFI-2 is the one with the lowest NSE_JK and the best simulation of ED is the one with the lowest Euclidean

Distance. As for NSGA-II, the Euclidean Distances were then computed for simulations at the Pareto Front and thus the best simulation of NSGA-II is the one also with the lowest Euclidean Distance. The comparison was displayed to analyze the impact of multi-objective calibration on the three objective groups: multi-site, multi-objective-function, and multi-metric.

6.3.2.1 The parameter values

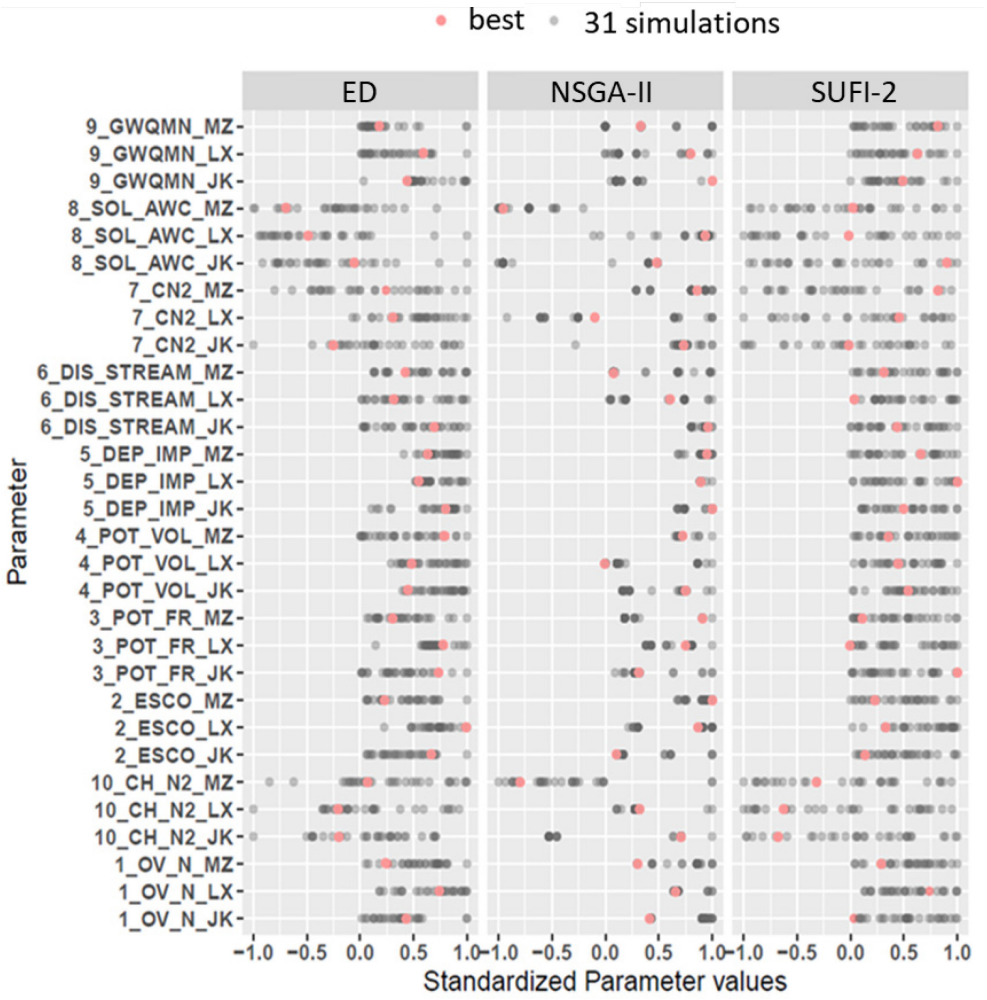

Figure 6-11 The distribution of the standardized parameter values of the 31 simulations selected from single- and multi-objective calibration

Figure 6-11 displays the parameter values applied by the 31 simulations of SUFI-2, NSGA-II, or ED. The parameter values were standardized to the range of [0, 1] (parameters with an absolute value change, e.g. GWQMN) or [-1,1] (parameters with a relative value change, e.g. CN2). The best simulation among the 31 obtained from each calibration approach is highlighted in red. 30 parameters (10 parameters for each subcatchment cluster: _MZ, _LX, or _JK) were included. The

analysis of the standardized parameter values was focused on interpreting the pattern of the value distribution in Figure 6-11, whereas the parameters derived by Latin Hypercube Sampling (LHS) of the initial iteration (2000 simulations) were uniformly distributed within their value ranges.

Among the selected 31 simulations, none of the parameters was constrained by any calibration approach into one value, instead, they were all scattered distributed, indicating the equifinality resulted from the complex processes computed by SWAT as an integrated hydrological model. However, certain unique patterns are still observed among the approaches. When analyzing the density of the distribution along with the value range, SUFI-2 had the most obvious uniform distribution and the majority of the parameters still possessed the original value ranges. On the contrary, the unique pattern was derived by NSGA-II, that most parameters were constrained to certain values, e.g. GWQMN_MZ had only four parameter values; OV_N_JK and OV_N_LX had only two narrowed value ranges. Other parameters could also be observed with the narrowest value range, e.g. DIS_STREAM_JK, DEP_IMP, and POL_VOL_LX. The calibrated parameter range of ED is between SUFI-2 and NSGA-II, e.g. GWQMN at the three subcatchment clusters, DEP_IMP, POT_VOL_LX, POT_VOL_JK, and OV_N_JK. Some parameters of ED also had a more similar pattern to SUFI-2, e.g. DIS_STREAM_LX, DIS_STREAM_JK, and POT_VOL_LX.

The respective impact of the constrained parameter values on the processes is also illustrated in Figure 6-11. DEM_IMP of ED and NSGA-II were from 0.5 to 1, whereas, of SUFI-2, the value was from 0 to 1. The constrained value of DEP_IMP from ED and NSGA-II indicated a tendency of a longer depth to the impervious layer. GWQMN had also a shorter value range of ED and NSGA-II. These two variables indicated that the processes of baseflow and lateral flow were affected directly by ED and NSGA-II. Another example can be observed at ESCO and SOL_AWC. ED and NSGA-II had an obvious preference at all the three subcatchment clusters that the processes of the evapotranspiration and soil water storage were directly impacted. The characteristics of parameters at the individual cluster can also be observed. CN2_LX of ED and CN2_MZ of NSGA-II were distributed from 0 to1, indicating an increase of the surface runoff preferred, whereas SUFI-2 did not show the identical trend. Higher values of POT_FR_LX and POT_VOL_LX from ED were demonstrated, indicating a higher surface runoff preferred.

6.3.2.2 Multi-site

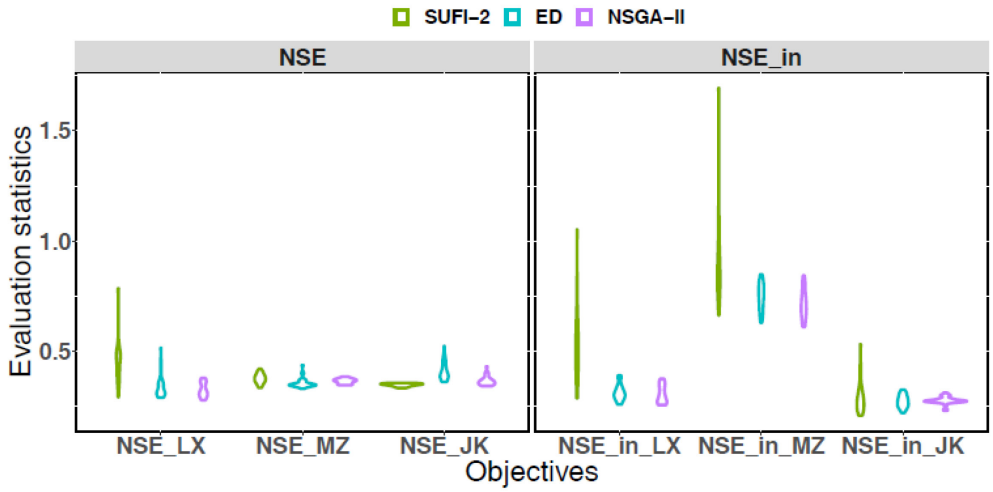

Figure 6-12 Comparison of the SUFI-2, ED, and NSGA-II on discharge evaluated by NSE and inverse NSE

Figure 6-12 displays the performance of the 31 simulations obtained from the three calibration approaches. The objectives displayed are the NSE and inverse NSE (NSE_in) at the three hydrometric stations: LX (upstream), MZ (midstream), and JK (mid-downstream). The interpretation of the results is focused on the uncertainty range of each objective. In Figure 6-12, the violin plot demonstrates the value distribution of the objectives and the width of the shape is the density of the distribution at a certain value.

The 31 simulations of SUFI-2 have the lowest NSE_JK among the 4000 simulations. Therefore, when compared to other calibration approaches, NSE_JK of SUFI-2 had the lowest prediction uncertainty and its values were distributed between approx. 0.33 and 0.35. Meanwhile, NSE_MZ of SUFI-2 also showed a relatively low uncertainty, from approx. 0.33 to 0.41. However, NSE_LX of SUFI-2 showed an obvious larger uncertainty range, from approx. 0.29 to 0.78, whereas ED and NSGA-II had the value from approx. 0.29 to 0.51 and a denser distribution can be found at low values. If simulations with the NSE larger than 0.5 are satisfactory (Moriasi, 2015), discharge at LX of SUFI-2 would be with unacceptable performance. Meanwhile with ED, both in LX and JK, unacceptable simulations also existed, however with smaller numbers than SUFI-2, whereas NSGA-II derived an acceptable NSE at all the three stations.

When to observe the performance of NSE_in at the three stations, SUFI-2 derived a significantly larger uncertainty than NSE, especially at MZ and LX. Though NSE_in_JK had a smaller

uncertainty than NSE_in_LX or NSE_in_MZ, the performance was still not equivalent to NSE_JK. On the contrary, ED and NSGA-II obtained a similar and much narrower uncertainty range of NSE_in at all stations. The NSE_in_LX and NSE_in_JK of ED and NSGA-II ranged from 0.22 to 0.39, indicating an acceptable simulation of low flow values. However, the NSE_in_MZ ranged from 0.61 to 0.85, though with lower uncertainty, the performance is not comparable to LX and JK that even the lowest value cannot be lower than 0.5. A biased calibration was obtained with single-objective calibration where the performance at the inner stations cannot all be guaranteed. ED and NSGA-II are able to ensure an improved performance at LX and MZ, especially at the low flow period. When comparing ED and NSGA-II, NSGA-II derived an overall better performance in terms of the NSE values and the uncertainty ranges of the top 31 simulations.

6.3.2.3 Multi-objective-function

Table 6-8 Mean RSR of each FDC segment

Station		**<Q5**	**Q5Q20**	**Q20Q70**	**Q70Q95**	**>Q95**
		Peak flow	Normal to peak flow	Normal flow	Low flow	Very low flow
JK	SUFI-2	0.68	0.30	0.41	1.86	3.01
	ED	0.58	0.31	0.54	1.91	1.66
	NSGA-II	0.62	0.19	0.58	1.44	1.21
MZ	SUFI-2	0.60	0.40	0.50	5.18	9.83
	ED	0.46	0.29	0.59	4.09	8.93
	NSGA-II	0.52	0.23	0.52	4.46	9.24
LX	SUFI-2	0.50	0.84	1.42	2.94	2.66
	ED	0.33	0.50	0.79	1.56	2.35
	NSGA-II	0.37	0.49	1.27	2.29	0.64

The multi-objective-function was evaluated by the goodness-of-fit of the flow duration curve (FDC) segments, which was assessed by the RSR to indicate the performance of simulating the discharge magnitude. Figure 6-12 displays the NSE_in values at individual hydrometric stations and Table 6-8 displays the mean RSR of each FDC segment. The segments represent the flow magnitude from peak flow (<Q5) to very low flow (>Q95), where <Q5 or >Q95 stands for the occurrence of a flow value lower than 5% of the entire period or higher than 95%. The higher the occurrence, the lower the flow value is. The lower the RSR, the higher the fitness of the magnitude is.

When analyzing the RSR of SUFI-2, the significantly lower RSR can be observed at the flow segment <Q5, Q5Q20, and Q20Q70 than Q70Q95 and >Q95. In JK, lower RSR values than in MZ or LX can be found at segments Q5Q20, Q20Q70, and Q70Q95. The highest values can be found in MZ at segment Q70Q95 and >Q95 and in LX at segment Q20Q70. This higher discrepancy is consistent with the results in Figure 6-12 that NSE_in_MZ, NSE_LX, and NSE_in_LX also display a large uncertainty range and a higher NSE or NSE_in value. When the NSE_in was considered in the calibration, i.e. by ED or NSGA-II, though the RSR of Q70Q95 and >Q95 was still high, improvements can be observed in comparison to SUFI-2. In terms of the periods with low RSR values, the three approaches have similar behavior that they derived the lowest RSR at LX in peak flow period and at JK or MZ at normal to peak flow period.

The lowest RSR of each FDC segment is also highlighted in red. NSGA-II obtained the lowest RSR at Q5Q20, Q70Q95 in JK, and >Q95 in LX. In comparison to the other segments, NSGA-II had a more significant decrease of RSR at Q70Q95 and >Q95. In terms of the flow magnitude, single-objective calibration only performed better at Q5Q20 in MZ and Q20Q70 in JK, whereas multi-objective calibration not only improved the simulation of the flow magnitude at low flow period but also the normal to peak and peak flow period.

6.3.2.4 Multi-metric

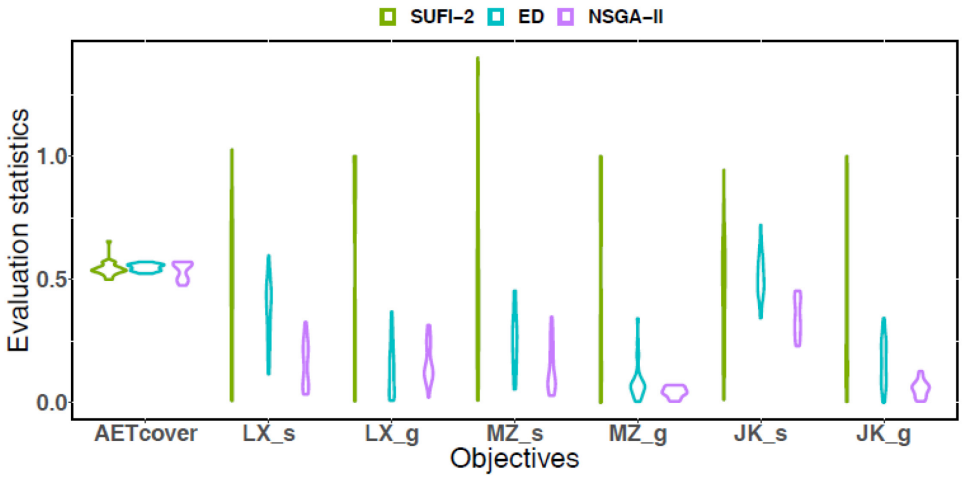

Figure 6-13 Comparison of SUFI-2, ED, and NSGA-II on water balance components

Figure 6-13 demonstrates the performance of the multi-metric, including the AET (AET_{cover}) and the ratio of surface runoff (LX_s, MZ_s, JK_s) and baseflow (LX_g, MZ_g, JK_g) to the water yield (see Table 6-5). AET_{cover} represents the coverage of simulation by the MOD16 ET uncertainty band; _s and _g represent the PBIAS of the simulated to the observed ratio. Figure 6-14 displays the best simulated daily discharge and monthly baseflow filtered from the discharge at JK in 2010 from January to June.

In accordance with the prediction uncertainty of the discharge, NSGA-II also derived a larger uncertainty range of all metrics, especially the surface runoff and the base flow ratio, where the 31 simulations had a rather scattered distribution of the PBIAS values, i.e. the value ranged from approx. 0 to larger than 1. NSGA-II derived an overall lower uncertainty at most metrics, except AET_{cover}, and it also generated the best performance at most metrics, except LX_g. AET_{cover} demonstrated a more concentrated distribution of all the three approaches that the value ranged from 0.5 to 0.6, indicating the insensitivity of the evapotranspiration procedure to the selected parameters.

When analyzing the monthly baseflow filtered from discharge at the JK station in Figure 6-14, SUFI-2 underestimated baseflow at the entire period, whereas ED and NSGA-II overestimated before June and underestimated after. In Table 6-9, the average baseflow index and its percent bias to the observed (residue/absolute) are listed that NSGA-II had the lowest bias of baseflow index. The respective daily discharge has shown that with SUFI-2, at no peak flow period, lower simulated values than ED and NSGA-II can be observed. When analyzing the discrepancy of the simulated discharge to the observed discharge, none of the calibration approaches were able to capture the magnitude of the high peak flow from April to June. However, NSGA-II still derived the highest values during the most peak flow periods.

Table 6-9 The mean baseflow index at JK station

Mean baseflow index	JK	Residue/Absolute
SUFI-2	0.36	-0.33
ED	0.65	0.26
NSGA-II	0.57	0.16
Observed	0.48	-

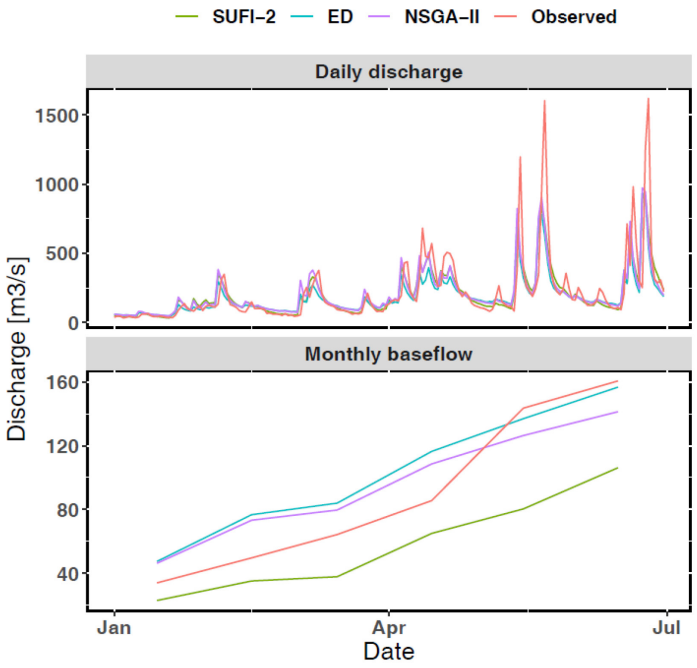

Figure 6-14 Daily discharge and the monthly baseflow filtered from hydrograph at JK station in 2010 from Jan. to Jun.

6.3.2.5 Best performance comparison

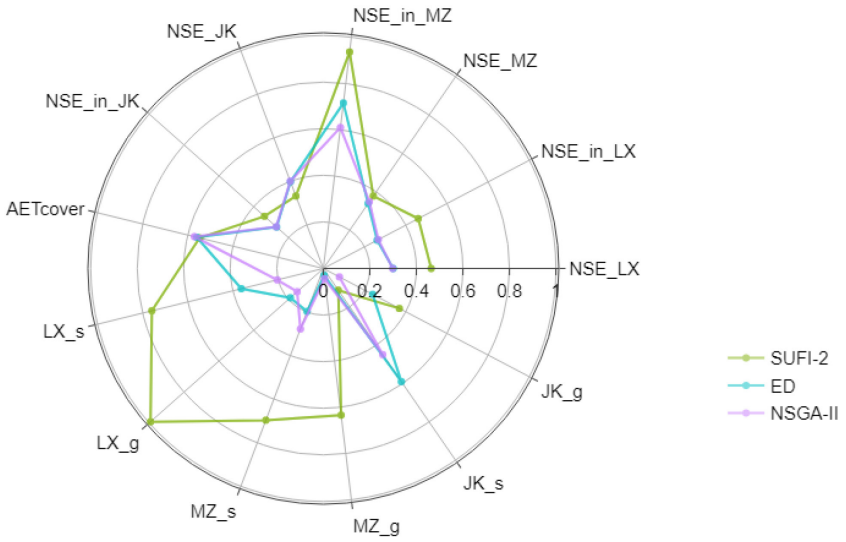

Figure 6-15 The radar chart of the best simulation of SUFI-2, ED, and NSGA-II

The radar chart Figure 6-15 displayed the best simulation of SUFI-2, ED, and NSGA-II. It is obvious to visualize the biased results obtained from SUFI-2 that except NSE_JK and AET_{cover}, significantly larger values of the remaining objectives were obtained. In contrast, with ED and NSGA-II, though NSE_JK was slightly higher, the AET_{cover} was equivalent to SUFI-2, and JK_s

had a higher deviation; whereas the rest objectives showed a significantly better performance than SUFI-2. If evaluating ED and NSGA-II by the 13 objectives, these two approaches generated an equivalent simulation of the discharge at all stations, and NSGA-II had a lower bias of discharge components at LX and JK.

Figure 6-16 displayed the daily discharge at the LX, MZ, and JK station. In terms of daily and seasonal variations, simulations from the three calibration approaches had similar performance. No significant differences can be observed at the JK station among the three approaches. However, at LX and MZ, ED generated the highest peak flow values at the flood season; at MZ, SUFI-2 had the highest discrepancy at the low flow period. Though NSE_JK had shown a satisfactory value of daily discharge, all the approaches failed to surpass the highest observed discharge. Moreover, the higher the magnitude of the peak flow values was, the more notable the difference was between the simulation and the observation.

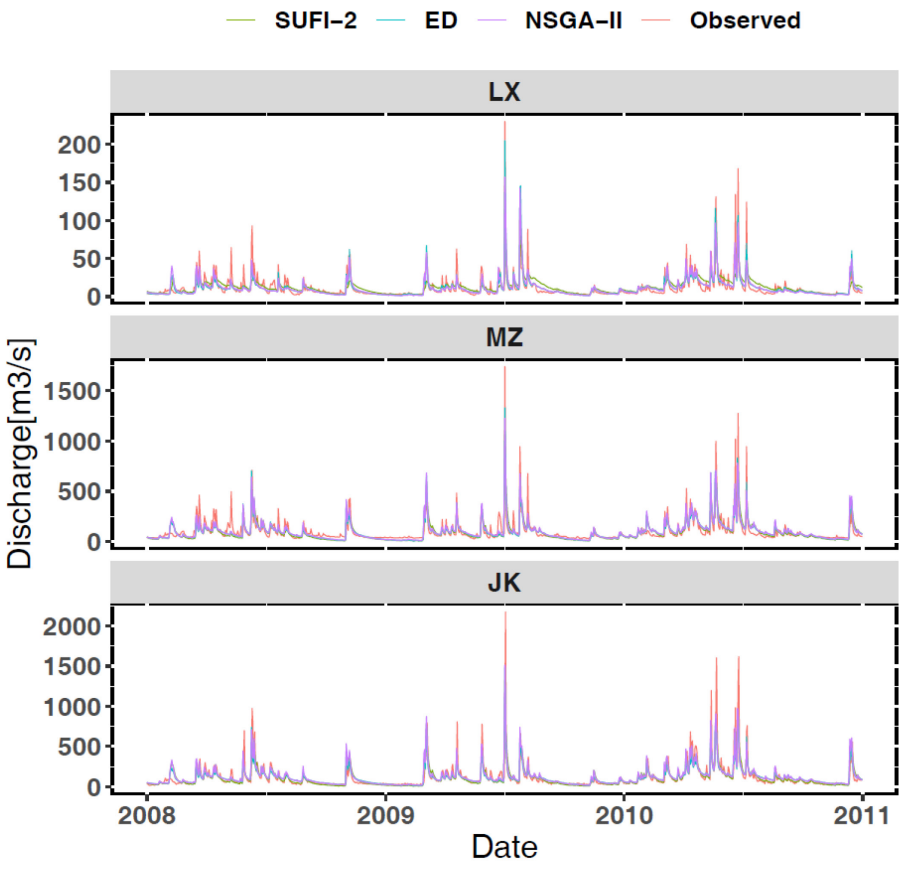

Figure 6-16 The simulated discharge at the LX, MZ, and JK station

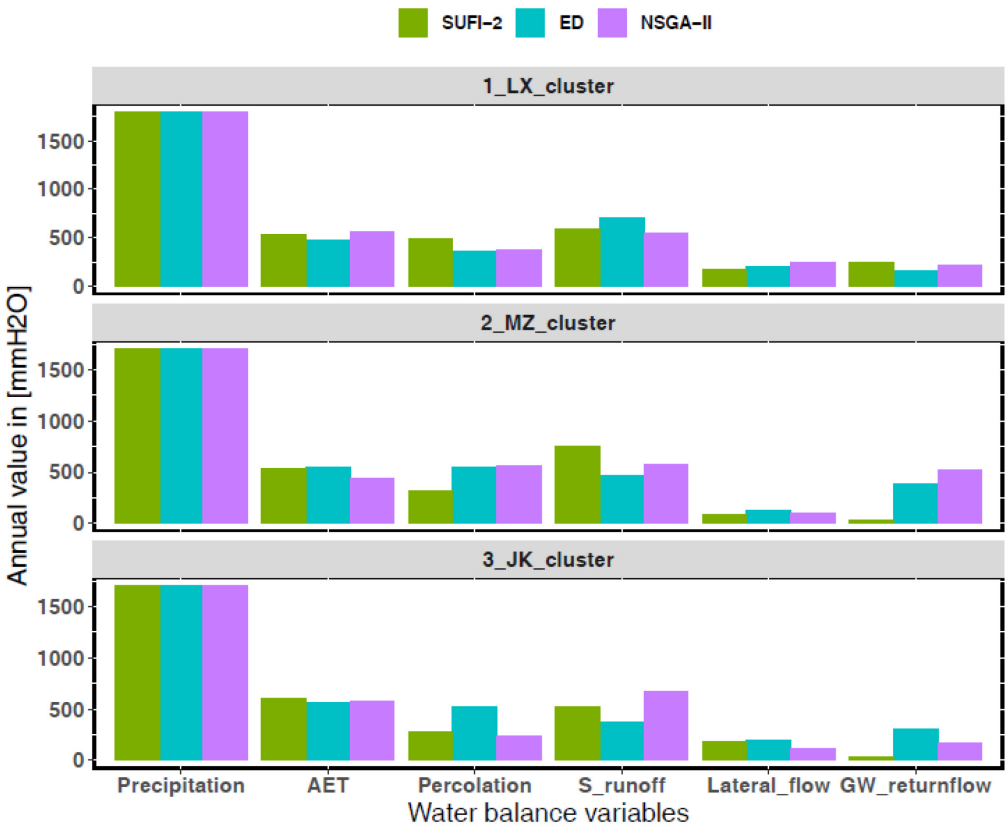

Figure 6-17 The catchment average annual water balance components

The water balance components were generated at each subcatchment cluster and the components were the annual average value in mmH_2O of all HRUs in the calibration period. The simulated actual evapotranspiration (AET) was approx. 500mm without a significant difference among clusters or calibration approaches. However, the annual average AET of the MOD16 ET was 839mm (see Section 3.7.2) and the uncertainty ranged from 637mm to 1041mm, which indicated the model's underestimation of AET in the YRC. Besides AET, lateral flow also obtained a low variation among calibration approaches. In MZ and JK cluster, the value ranged from 78mm to 190mm, where the highest lateral flow was derived from ED and the lowest from NSGA-II; at LX, the lateral flow was from 177mm to 242mm and NSGA-II derived the highest values.

At JK and MZ, a larger difference of percolation can be observed among calibration approaches, S_runoff (surface runoff), and GW_returnflow (baseflow) obtained a much higher S_runoff at all the stations than baseflow. On the contrary, at MZ, both NSGA-II and ED indicated a comparable

surface runoff and baseflow rate. However, at JK, NSGA-II obtained a much higher surface runoff and lower baseflow rate than ED. When tracing back to section 6.3.2.1, the variation of the water balance components was consistent with the variation of the parameter values among the stations and the calibration approaches. Parameter GWQMN of SUFI-2 was comparably higher at all clusters. The direct impact was thus the higher threshold for baseflow to occur and consequently a lower simulated baseflow than the other two approaches. A higher POT_VOL and CN2 at JK of NSGA-II would also contribute to the higher surface runoff rate.

6.3.2.6 Model validation

The hydrological components were validated from 2011 to 2014 and the objectives evaluated were displayed in Figure 6-18. The discharge and the ratio of surface runoff and baseflow, due to data availability, were only validated at the JK station. Figure 6-18 displays the value of the best simulation of ED and SUFI-2, and NSGA-II at all individual objectives. Figure 6-19 displays the prediction uncertainty of the 31 simulations.

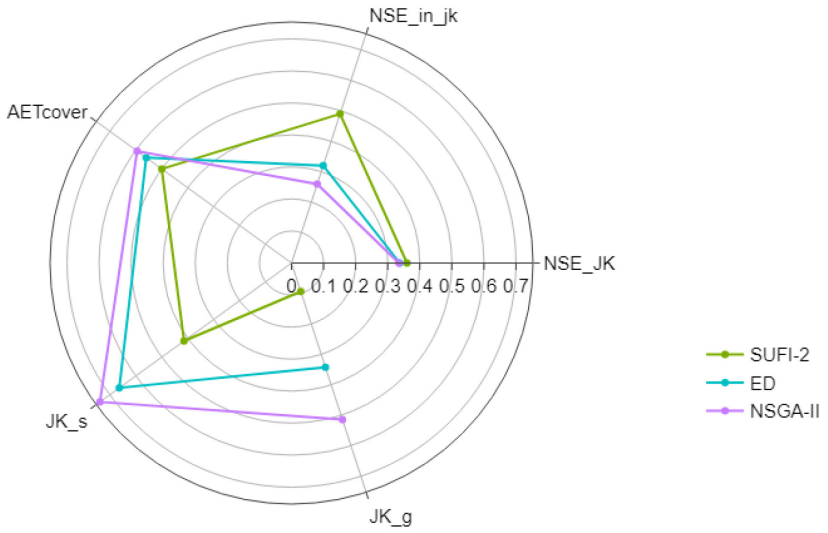

Figure 6-18 The radar chart of the validation period (2011-2014)

In Figure 6-19, for objective NSE and NSE_in of ED and NSGA-II, the performance was consistent with the calibration period that NSGA-II derived the least uncertainty range (approx. 0.4 to 0.5) and followed by ED (approx. 0.4 to 0.7). For objective JK_g and JK_s, ED displayed an equivalent uncertainty range as in the calibration period, whereas NSGA-II indicated a larger range of the prediction uncertainty. At the opposite, SUFI-2 obtained a much lower uncertainty

range in the validation than in the calibration period of JK_g and JK_s. Meanwhile, the best simulation of NSGA-II also indicated a higher bias of JK_s and JK_g in Figure 6-18. AET_{cover} can be observed with a smaller uncertainty of all calibration approaches at the validation period.

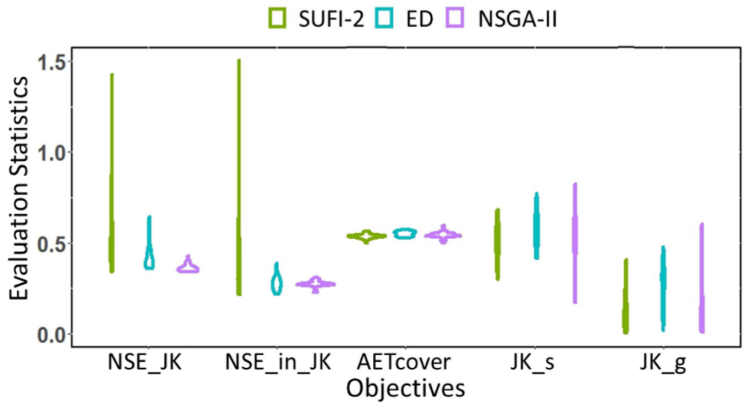

Figure 6-19 The prediction uncertainty of the individual objectives of the validation

6.4 Nitrogen load analysis

The simulation of the in-stream nitrogen load was highly sensitive to the flow calibration. Therefore, to reduce the uncertainty introduced by the differences in the hydrological model, the identical hydrological model was used as the base model for the nitrogen module calibration. The base hydrological model selected was the best simulation of ED since it had a similarly good performance to NSGA-II in the calibration period and a better performance of JK_s and JK_g in the validation period. The objectives applied for nitrogen module calibration were NSE and NSE_in at the three outlets 1_Y, 4_F, and 6_L. The NSE at 6_L (NSE_L) was the objective of SUFI-2. Therefore, the three approaches were applied to the same hydrological model to calibrate the nitrogen related parameters. The nitrogen load was interpreted to the critical source areas (CSAs) and the annual total load at the 8 outlets (see Figure 6-28).

6.4.1 The calibration of the total nitrogen load in the stream

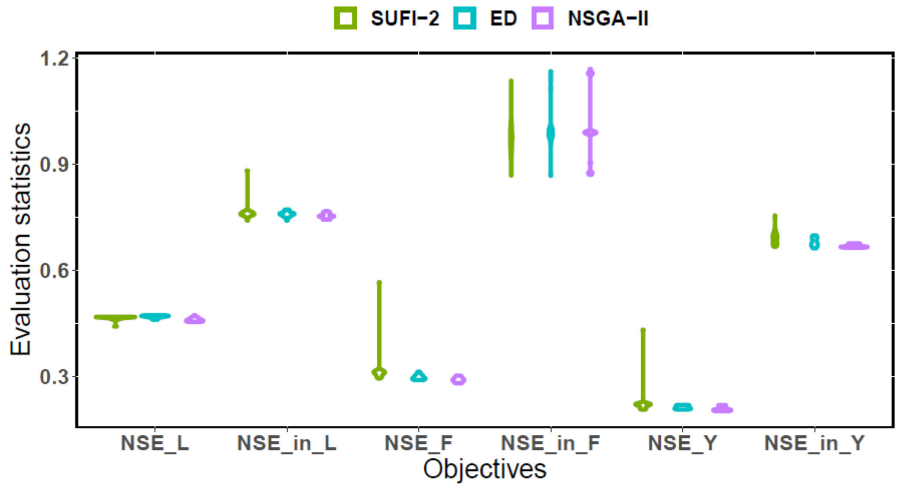

Figure 6-20 Comparison of SUFI-2, ED, and NSGA-II on TN at station 1_Y, 4_F, and 6_L

The Pareto Front obtained from the 40th generation contained 25 simulations. Figure 6-20 demonstrates the uncertainty range of the first 25 simulations from each calibration approach on the 6 objectives. The consistent performance of the calibration approaches on the objectives can also be observed that SUFI-2 performed the best at NSE_L whereas NSGA-II and ED had a lower uncertainty at the extra objectives. The only exception was NSE_in_F that the three approaches all derived relatively higher values and larger uncertainty range.

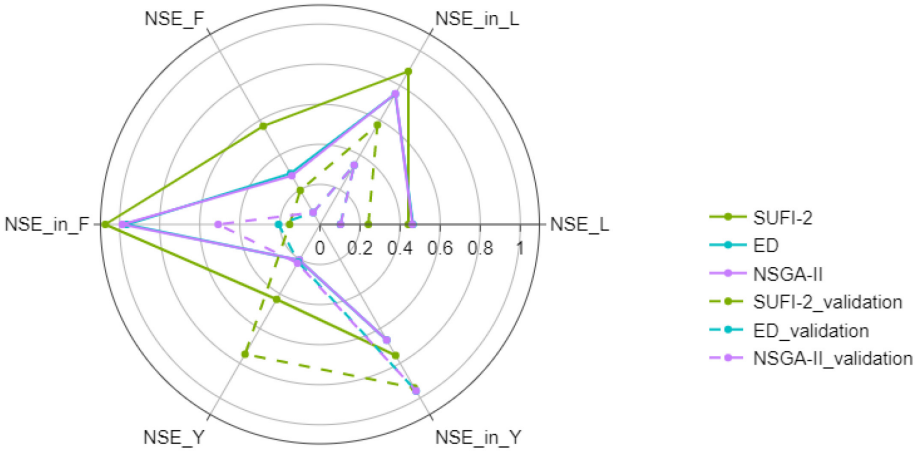

Figure 6-21 The radar chart of the best simulation in the calibration and validation period

Figure 6-21 displayed the NSE and NSE_in at the station 1_Y, 4_F, and 6_L in both the calibration period (2008 to 2010) and the validation period (2011 to 2014). In the calibration period, SUFI-2, ED, and NSGA-II were able to obtain a satisfactory NSE (standardized NSE < 0.65) for all stations, whereas in the validation period, SUFI-2 was not able to derive a reliable simulation at 1_Y (NSE_Y > 0.65). In terms of NSE_in, ED and NSGA-II performed better than SUFI-2 at all stations in the calibration period. ED and NSGA-II had a similar performance of the best simulation in both calibration and validation period, except NSE_in_F in the validation period. However, the overall better performance of validation than the calibration period could be due to the large discrepancy in the year 2008 as part of the calibration period (see Figure 6-22).

6.4.2 The prediction uncertainty of the in-stream nitrogen load

Table 6-10 The r-factor of the 95PPU uncertainty band

Calibration approaches	Water quality stations	r-factor	
		Calibration	Validation
SUFI-2	1_Y	0.49	0.85
	4_F	0.52	0.57
	6_L	0.60	0.70
ED	1_Y	0.15	0.21
	4_F	0.15	0.18
	6_L	0.17	0.18
NSGA-II	1_Y	0.20	0.27
	4_F	0.22	0.24
	6_L	0.25	0.24

The prediction uncertainty was generated from sorting the first 1000 out of the 3000 simulations according to their performances for each calibration approach. The first 1000 simulations were able to derive a comparably good performance that simulations with low reliability can be excluded. Therefore, the first 1000 simulations were selected. The 95PPU was the coverage of the 2.5th and 97.5th percentile of these 1000 simulations. Figure 6-22 displayed the observed and simulated discharge with the 95 PPU. The results of applying the simulations with the highest and lowest PBIAS to the best simulation are also displayed. Applying the highest and lowest PBIAS to represent the 95PPU was proposed that they are the largest deviation of the simulations to the best one. Figure 6-22 indicated the applicability of the assumption by visualization that the Pb_highest and Pb_lowest are able to substitute the range covered by the 95PPU. The statistical

comparison of the width of the uncertainty band (r-factor) after calibration and validation are listed in Table 6-10 for each station.

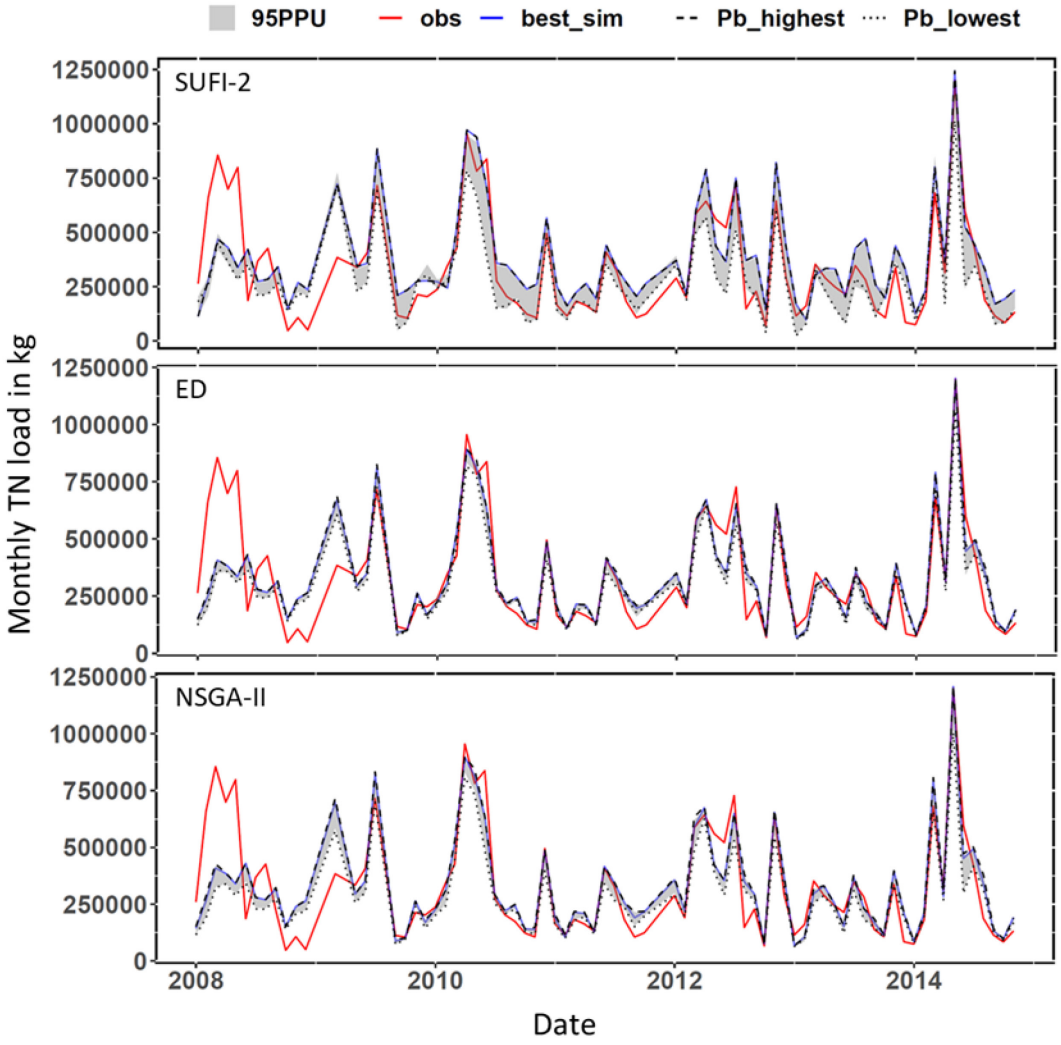

Figure 6-22 The total nitrogen load prediction uncertainty at 6_L station, including the 95PPU and the highest and lowest PBIAS

The larger prediction uncertainty can also be observed at the TN load time-series of SUFI-2 than of ED or NSGA-II that the 95PPU of SUFI-2 was significantly wider than ED or NSGA-II, as displayed in Figure 6-22. Meanwhile, the Pb_highest of SUFI-2 is the same one as the best simulation. This is also reflected by the r-factor value that SUFI-2 derived a value of 0.49 to 0.60, whereas ED was 0.15 to 0.17 and NSGA-II was 0.20 to 0.25. ED and NSGA-II generated similar TN load of the best simulation, however, the prediction uncertainty of NSGA-II was slightly higher than ED. When analyzing the performance of the simulation, NSE at 6_L was 0.44 by SUFI-2,

0.46 by ED, and 0.46 by NSGA-II, as displayed in Figure 6-21. Though all the calibration approaches derived an acceptable NSE at 6_L, all did not capture the magnitude of the TN load in 2008, neither can the uncertainty band of the calibration.

6.4.3 The nitrogen load prediction

6.4.3.1 The identification of the CSAs

The 4137 HRUs were sorted in descending order based on the annual TN load (kg) and were identified as Class I, II, III, and IV to demonstrate the critical regions. Class I consists of the area contributing to the first 25% of the TN load, Class II the second 25%, Class III the third 25%, and Class IV the remaining 25%. The smaller the area occupied by the Class I, II, or III, which included 75% of the TN load, the more concentrated the CSAs were. Figure 6-23 demonstrates the CSAs at the HRU level obtained from the best simulation of each calibration approach. Figure 6-24 displays the cumulative area of the HRUs corresponding to the cumulative TN load. Figure 6-25 displayed the percentage of landuse RICE (paddy field), Forest, and Dry_arable (dry arable land for crops) in Class I, II, and III.

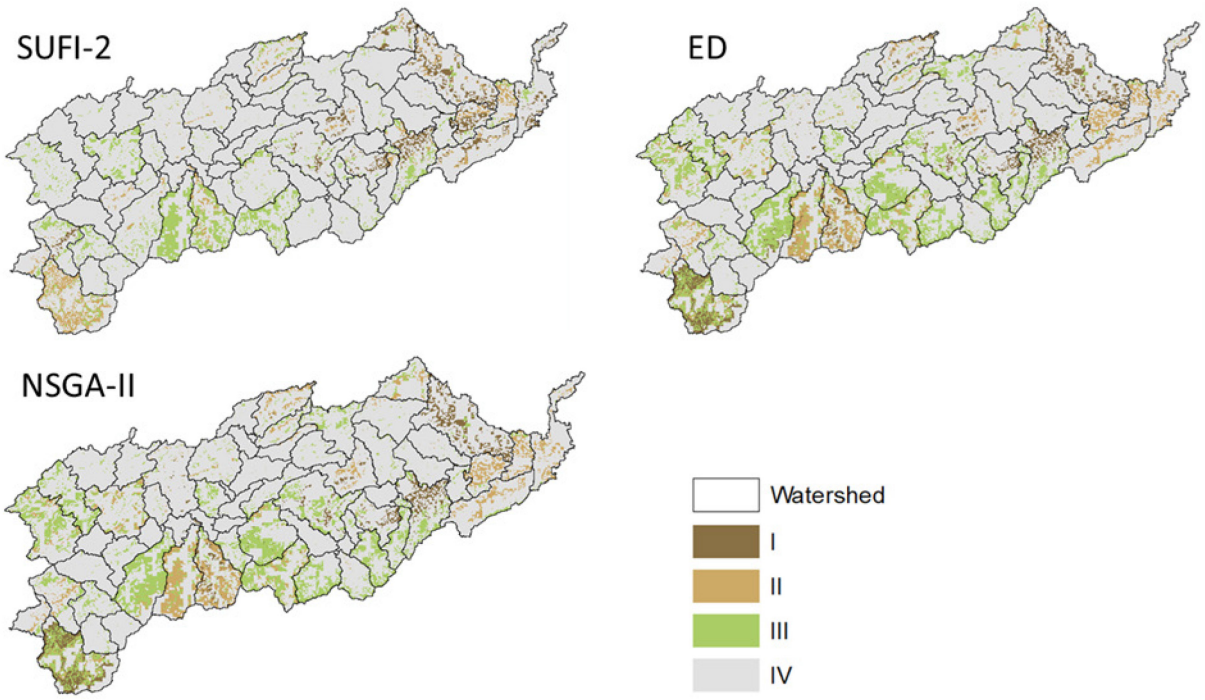

Figure 6-23 The critical source areas (CSAs) identified in YRC

The distribution of the CSAs showed similarities among calibration approaches. The downstream plain region along the Yuan River and its tributaries were mainly distributed within Class I and

Class II. The plain region was occupied by paddy field and denser rural population, which were the main source of the TN from fertilizer, feedlots, and rural households, distributed within the 6_L subcatchment cluster (see Figure 6-4). When compared to the other regions with paddy fields, its relatively high TN load could result from a higher leaching rate and higher runoff rate. The distribution of TN in subcatchment cluster 1_Y and 4_F between SUFI-2 and ED or NSGA-II was more distinguished due to the multi-site calibration applied by ED and NSGA-II. In the 4_F cluster, a larger area was identified in Class III by ED or NSGA-II, including forest and paddy field; in the 1_Y cluster, CSAs were classified at the mountainous region, where SUFI-2 derived the Class II and III in a more concentrated area. The up to midstream mountainous region is occupied with evergreen forest and at the valley with paddy field. The deeper slope and the forest residue might account for the high TN load in the forest.

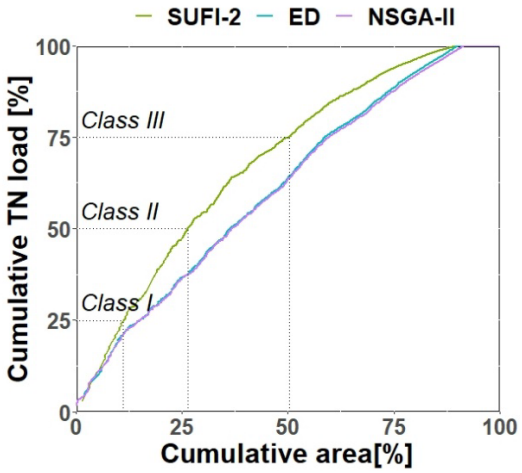

Figure 6-24 The cumulative plot of area and the TN annual load

The landuses covered by Class I and Class II were analyzed. The percentage of rice, forest, and dry arable land and their TN load to the stream are displayed in Figure 6-25. In correspondence to the CSAs distribution, paddy field with rice occupied above 50% of Class I and II, which also contributed to above 60% of the TN load in Class I and above 50% in Class II. The dry arable land was dominated by vegetables and contributed to lower than 10% of the TN load. Moreover, SUFI-2 had a lower TN load from the forest that none was included in Class I and a lower contribution in Class II. The cumulative plot (Figure 6-24) demonstrates the concentration of the area to the TN load, where the curves of ED and NSGA-II are almost overlapped. Class I occupied 10% to 15% of the YRC, and Class I and II occupied an area of 26% to 37%, which indicated that 50% of the

total TN load to the river was from lower than 40% of the area. Dry arable land in Class III, though still occupied a small area, load the highest TN per area. ED and NSGA-II derived the 35% of the TN load from paddy field that occupying 33% area in Class III, whereas SUFI-2 derived the 55% of the TN load from paddy field occupying 28% are in Class III.

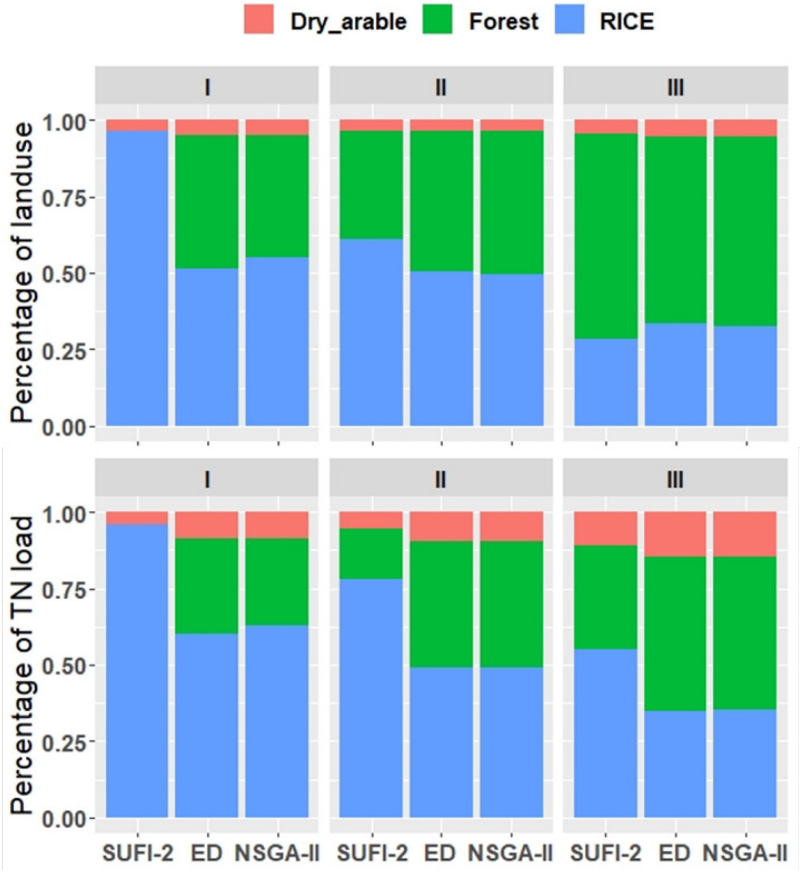

Figure 6-25 The percentage of landuse and TN load in CSAs Class I and Class II

In terms of the calibration approaches, the ED and NSGA-II derived a similar prediction of CSAs. However, the significant differences can be seen between single-objective calibration and multi-objective calibration in terms of the classification of the CSAs and the TN contribution from rice fields and forest. The cumulative plot showed the differences of Class I with 5% and of Class II with 11% that larger area was identified by ED or NSGA-II. Moreover, TN from the forest was identified with a relatively higher load by multi-objective calibration. These differences can be visualized in the subcatchment cluster 1_Y and 4_F, indicating the higher uncertainty caused by no consideration of cluster-specific parameterization when applying the single-objective calibration.

6.4.3.2 The prediction uncertainty of the CSAs

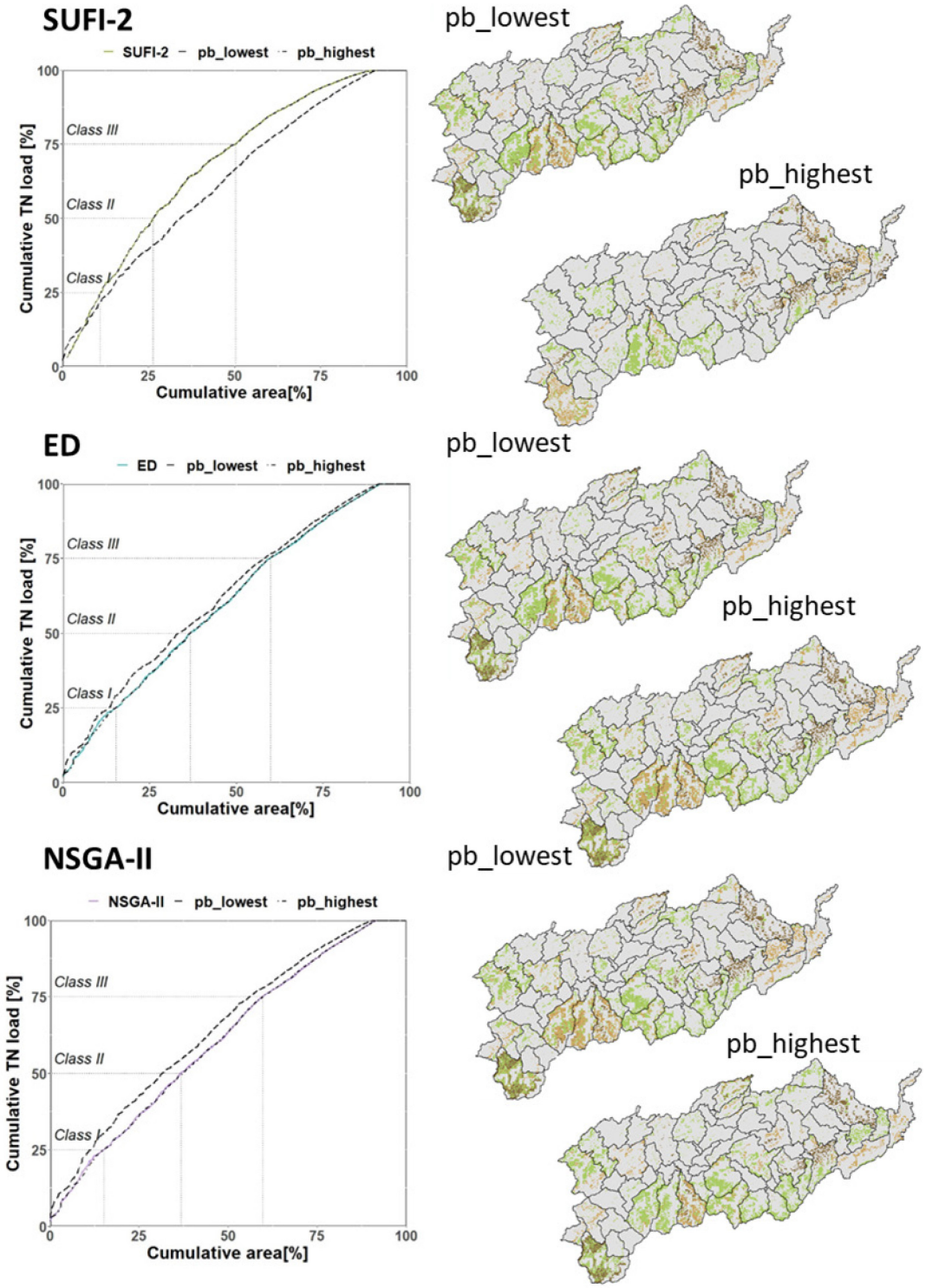

Figure 6-26 The prediction uncertainty of CSAs from SUFI-2, ED, and NSGA-II

The prediction uncertainty of CSAs was the result of the simulations with the lowest or the highest PBIAS. These two simulations, as stated above, with the largest deviation to the best simulation of the selected 1000 simulations, could represent the range of the prediction uncertainty.

Figure 6-26 shows the largest difference between the pb_highest and pb_lowest of the cumulative TN load obtained by SUFI-2, which is consistent with Figure 6-22, demonstrating the largest uncertainty range of TN load at 6_L station. Each class was affected by the parameter uncertainty to a different scale. From pb_lowest to pb_highest, Class I ranged from 11% to 13.6%, Class I and II ranged from 26.3% to 34.3%, and Class I, II and III ranged from 50.3% to 58.5%, as displayed by the cumulative plot. As stated above, Pb_highest was obtained to be the same simulation as the best simulation. When analyzing the distribution of CSAs, the obvious impact can be visualized at the up and midstream of the catchment, where the high load from the mountainous region was the main contribution. The downstream paddy field was also affected that certain subcatchments were not included by Class I with pb_lowest simulation.

The lower residue of classes II, III, and IV between pb_highest and pb_lowest was obtained by ED or NSGA-II. Pb_highest was less deviated from the best simulation than pb_lowest. Class I and II of ED ranged from 35.2% to 37.1% and NSGA-II ranged from 33.9% to 37.2%; Class I, II, and III of ED ranged from 58.2% to 59.7% and NSGA-II ranged from 56% to 60%. When analyzing the total area occupied by Class I, II, and III, no significant difference can be observed between the best simulation and pb_highest. Similar to SUFI-2, differences can also be observed in the mid-stream mountainous region and downstream paddy field in certain subcatchments that with NSGA-II or ED, the region was included in Class II by pb_highest and in Class III by pb_lowest.

6.4.3.3 The annual TN load at each subcatchment

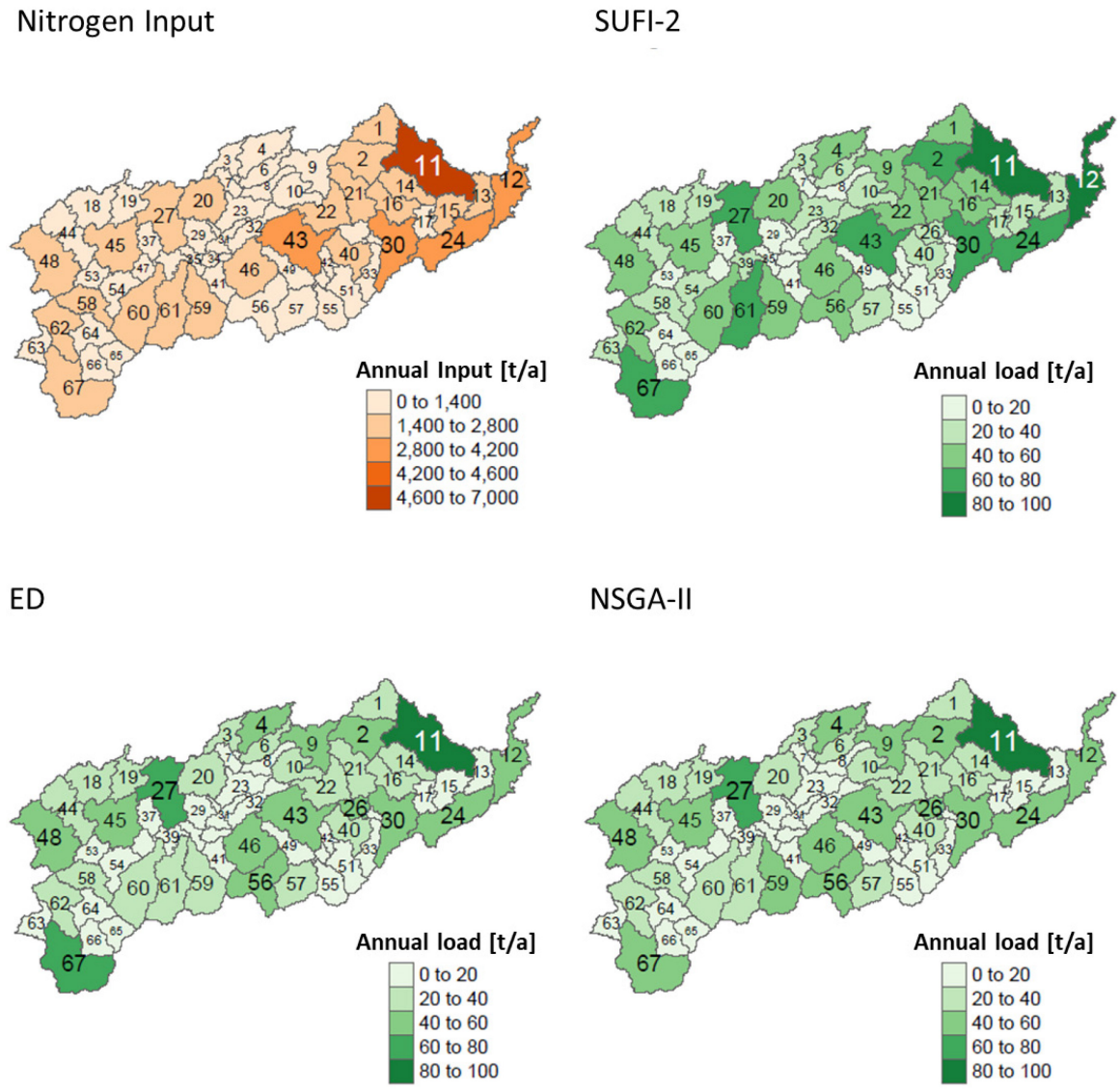

Figure 6-27 The annual nitrogen input and the load to the stream obtained from SUFI-2, ED and NSGA-II at subcatchment scale

The annual TN load of each subcatchment was aggregated from the HRU level and displayed in Figure 6-27. The nitrogen input from the non-point sources (NPS) in the YRC is also displayed. The NPS, as mentioned in Section 3.6, includes fertilizer applied to the paddy field, waste from feedlots, and rural households disposed as manure, and net nitrogen released from atmospheric deposition along with precipitation. The TN load was the TN generated and drained to the outlet at each subcatchment, which would contribute to the TN concentration in the stream. The species of the TN load include the organic nitrogen transported by sediment, which can be traced back to the manure and plant residue, and the nitrate transported by surface runoff, lateral flow and

baseflow, which can be traced back to the fertilizer and the initial nitrogen content in the soil. The assumption of the SWAT model defines the ammonia-nitrogen immobile, and thus the ammonia-nitrogen in the stream was derived from the in-stream transformation, e.g. mineralization, or point sources.

The differences of the CSAs identified by SUFI-2, ED, or NSGA-II were reduced at the subcatchment level. According to the legend displayed in Figure 6-27, besides subcatchment 67 with a higher TN load level (60 to 80 t/a) of ED, the spatial distribution of the nitrogen load level annually was identical to the three calibration approaches. The highest TN load at the downstream can be found from subcatchment 43, 30, 24, 12, and 11, located in the subcatchment cluster 4_F and 6_L. These five catchments were the ones with a relatively high input of all NPS (see Figure 6-3), where the denser rural population was located and dominated by paddy field and thus led to a relatively higher amount of nitrogen input from feedlots, rural households and fertilizer. The CSAs identification indicated the paddy field in subcatchment 11 and 30 to be dominated by Class I, and in subcatchment 12 and 24 to be dominated by Class II of ED or NSGA-II, and by Class I of SUFI-2. Subcatchment 43 were dominated by Class III of ED or NSGA-II and fewer CSAs of Class III were identified by SUFI-2. Subcatchment 27 was also with higher TN input that it was also dominated by paddy field and a large concentrated rural residential region, where scattered CSAs were identified, belonging to Class I and Class III. The CSAs identified the mountainous region at subcatchment 59, 60, 61, and 67 to have a higher load from the forest, however, the nitrogen input and output of these subcatchments also indicated the high contribution from the anthropogenic nitrogen sources.

6.4.3.4 The in-stream TN load

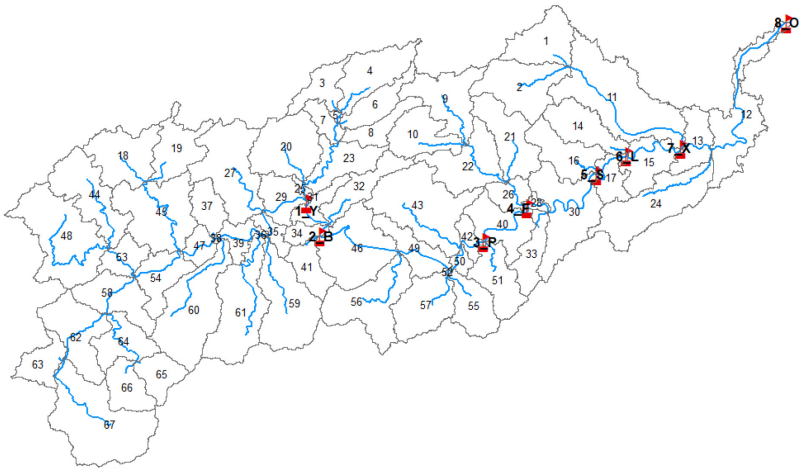

Figure 6-28 The location of the 8 outlets along the Yuan River

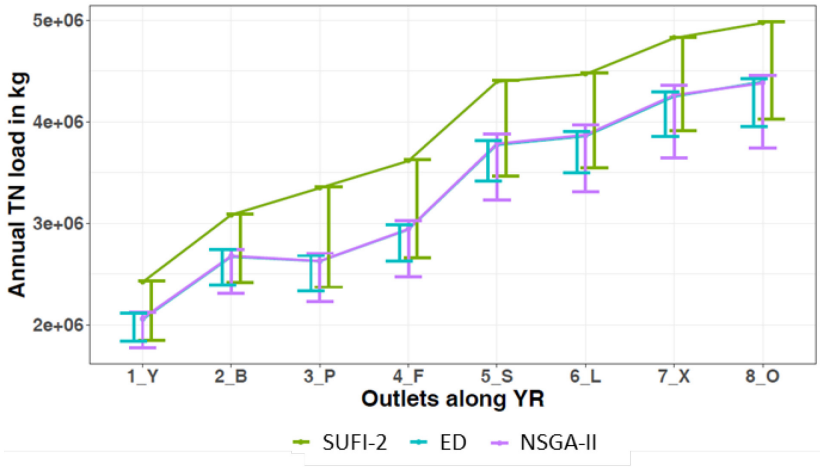

Figure 6-29 The TN load in the stream from the 8 Outlets along the Yuan River and their prediction uncertainty

The annual in-stream nitrogen content at the 8 outlets (see Figure 6-28) along the Yuan River was displayed in Figure 6-29. The prediction uncertainty of the annual content obtained by SUFI-2 kept being the largest at all the outlets. The ratio of the average uncertainty bandwidth to the average TN load is 22%, 11%, and 17% of SUFI-2, ED, and NSGA-II respectively.

Point sources (PS) were located at the upstream of 1_Y, between 2_B and 5_S (see Figure 6-3 and Figure 6-28). The urban regions with industrial discharge were located at the upstream of 1_Y and between 3_P to 5_S. Meanwhile, between 2_B to 5_S, high TN load from the NPS can also be found. In Figure 6-29, a significant increase among outlets can be observed between 4_F and 5_S,

and between 6_L and 7_X. In-between the relatively high TN input in subcatchment 30 (Figure 6-27) also led to a high load and at the downstream of 6_L, where the higher TN load for rice plantation resulted in a high annual load, as mentioned above. A similar change of gradients between the outlets can be observed among the three calibration approaches, except between 2_B to 3_P that ED and NSGA-II showed a slight decrease of TN load in-stream, though the subcatchments in-between were not with lower TN load (e.g. subcatchment 46 in Figure 6-27), which could due to the higher transformation rate of the in-stream nitrogen with ED and NSGA-II.

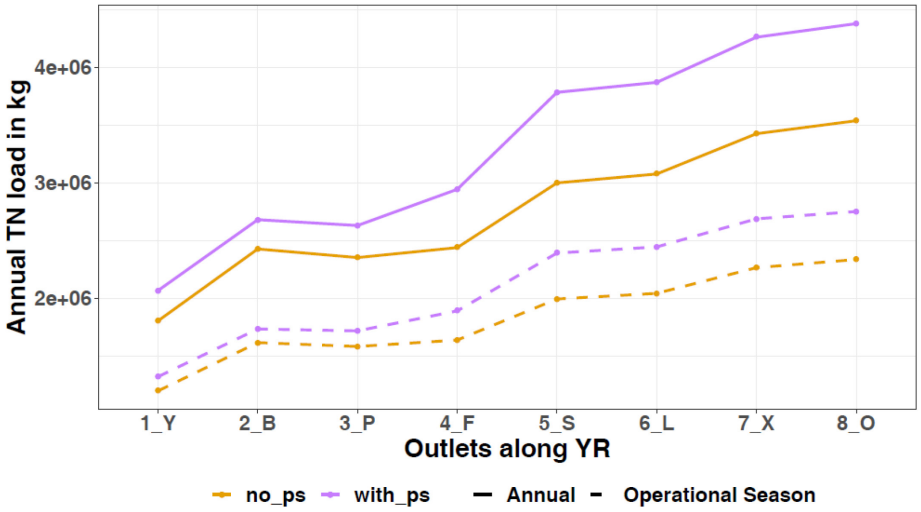

Figure 6-30 The annual TN load in the stream of NSGA-II with point sources (with_ps) or without (no_ps) and the TN load only in the operational season (From April to October)

The operational season consumes the majority of the anthropogenic nitrogen from non-point sources. The waste from feedlots and households were distributed to the paddy field as part of the top dressing in the operational season (see Section 3.6.2) and applied to the dry arable land and orchard all year round. TN from the point sources was discharged all year round in the YRC.

Figure 6-30 displays the TN load from the 8 outlets along the Yuan River with point source discharge (with_ps) and without the point source discharge (no_ps). The scenario no_ps was obtained by running the model without the point source input to the SWAT model. Results indicated that averagely, TN derived from the no_ps is approx. 75% of the TN derived from with_ps. However, it should be mentioned that this scenario did not consider the impact of the magnitude of nitrogen on the in-stream nitrogen transformation processes and assumed that this impact could be ignored. Based on this assumption, in the YRC, the nitrogen load from PS

contributed to 25% of the in-stream nitrogen and 75% is from the NPS. The operational season, from April to October, was also displayed. The operational season covers two rice planting seasons when fertilizer and the majority of the waste as manure from feedlots and rural households were applied. The TN load from the operational season occupied 63% or 66% of the annual TN if the point sources are counted (the with_ps scenario) or not (the no_ps scenario).

7 Discussion

The results were discussed in this chapter with the focus on two aspects: 1) the analysis of the multi-objective calibration, especially in comparison to the single-objective calibration; and 2) the interpretation of the nitrogen load simulation in the Yuan River Catchment (YRC). Section 7.1 summarized the new insights by this research in terms of the characteristics of the multi-objective calibration framework, including the impact of individual objectives on the calibration and the evaluation of ED and NSGA-II. Section 7.2 explains the simulation results of CSAs and in-stream nitrogen load, including the calibration result, prediction uncertainty, and the input data uncertainty. Section 7.3 stated the outlook in the aspect of the multi-objective calibration procedure, future research, and the YRC model application.

7.1 The applicability and advantage of multi-objective calibration

In section 6.3.2, the comparison between the single-objective calibration approach SUFI-2 and the multi-objective calibration approach ED and NSGA-II was displayed on the 13 hydrological objectives (3 categorized objectives: C_NSE, C_NSE_in, and C_WB) and the 6 nitrogen objectives. The results displayed the prediction uncertainty of the best (see Figure 6-15) and the top 31 simulations for hydrology (see Figure 6-12 and Figure 6-13) and of the top 25 simulations for nitrogen (see Figure 6-20 and Figure 6-21).

Biased calibration results could be observed from single-objective calibration that the optimized statistics can be obtained for the targeted objective (NSE_in_JK or NSE_L). However, the unsatisfied statistics of the remaining objectives were derived that the significant wider range of the prediction uncertainty and the larger deviation of the best simulation to the observations can be observed. On the contrary, ED and NSGA-II were able to derive a significantly better prediction of the additional objectives without deteriorating the performance of NSE_in_JK or NSE_L. Meanwhile, the parameter uncertainty also indicated the direct impact of the multi-objective calibration on the relevant hydrological or nitrogen processes. In terms of the objectives, multi-site significantly improved the model's performance at the inner stations and multi-objective-function was proved to better capture the magnitude of each flow period. Multi-metric allowed the model to also focus on the processes which were previously ignored by calibration but were also sensitive to the parameter uncertainty.

In terms of the comparison between single-objective calibration and multi-objective calibration, the key findings of this research are summarized as follows:

1) an obvious biased calibration result can be observed with single-objective calibration;
2) parameter uncertainty and prediction uncertainty were significantly reduced by the ED and NSGA-II.

7.1.1 The impact of the objectives

It has been emphasized that single-objective calibration results in a trade-off between the targeted objective and the other model output variables, whereas multi-objective calibration provides additional boundaries to ensure the reliability of the intra-processes. In this study, the objectives have been grouped into three categories, multi-site, multi-objective-function, and multi-metric. The objectives for the YRC were then selected accordingly. Considering the lack of additional observations in the YRC, the objectives were developed also from the fundamental observations, i.e. measured discharge or nitrogen load.

The impact of the multiple objectives has been indicated in Section 6.3. The consideration of the internal sites, LX at the upstream and MZ at the up to mid-stream, has ensured the model's performance at LX with an improved value of NSE_LX, whereas an acceptable performance of the targeted objective NSE_JK was also guaranteed (see Figure 6-12). The same impact on improving the inner stations can also be found at the in-stream nitrogen load simulation (see Figure 6-21). The application of multi-objective-function, both NSE and NSE_in, has a significant impact on the magnitude of the simulated discharge that not only the flow magnitude of low to very low flow period has been improved at all the three sites, but also the flow magnitude from normal to peak flow period at MZ and LX (see Table 6-8). The bias of multi-metric in terms of water balance components, which were of a large deviation by single-objective calibration, has been significantly reduced at LX_s, LX_g, MZ_s, MZ_g, and jx_g (see Figure 6-13). Meanwhile, a lower discrepancy of the baseflow can also be observed (see Figure 6-14 and Table 6-9). Also, the consideration of the objectives had a direct impact on narrowing the parameter uncertainty (see Figure 6-11). A similar impact can also be observed at the nitrogen load simulation with the multi-site and multi-objective-function implemented (see Figure 6-20 and Figure 6-21).

The multi-metric objectives, excluding the AET, were obtained from the existing measured discharge, which is the fundamental data for model calibration. Multi-objective-function is the statistics applied to ensure both low and peak values of the time series. These two groups were applied in the calibration without the effort of collecting additional observations. Therefore, only with the fundamental observations for calibration, multiple objectives can still be derived and worked as effective additional criteria for calibration.

7.1.1.1 The evaluation of the objectives

The AET obtained from the MOD16 ET dataset was applied previously as a reference to assess the evapotranspiration simulation of the hydrological model and van Griensven (2012) has concluded an underestimation of the AET by SWAT. However, the data was validated in limited locations globally and the accuracy at the study area can thus not be fully guaranteed. Therefore, the uncertainty band was used for evaluation instead, aiming at reducing the impact of the observation uncertainty on the evaluation. Moreover, only the catchment average AET was computed for each month instead of also considering the AET on subcatchment or HRU level. The pre-processing of the MOD16 ET dataset aimed at amplifying the data range to reduce the possibility of creating the discrepancy between the observations and simulations due to the uncertainty of the observations. However, even with the wide data range, only 52% of the simulated AET was covered by the uncertainty band (Figure 6-13). Moreover, compared to other objectives, the rather narrower prediction uncertainty of AET indicated the insensitivity of the parameters to the AET processes. Besides AET, the baseflow and surface runoff ratio computed by the Lynne-Hollick filter was used to represent the baseflow and surface runoff aggregated to the outlet at each subcatchment cluster. These metrics were only the estimation based on an empirical approach and cannot represent the full complexity of the real conditions. Therefore, the metrics were evaluated at a monthly step by the PBIAS to reduce the impact of the data uncertainty. In comparison to the measured discharge or TN load, the multi-metric objectives belonged to the soft data (Arnold et al., 2015), which are the qualitative description of the catchment. Therefore, when evaluating the calibration results, higher tolerance of the multi-metric in terms of the uncertainty and the performance for acceptance is also recommended.

7.1.2 The evaluation of the calibration approaches

7.1.2.1 The trade-off among objectives

The trade-off of the single-objective calibration behaved to optimize the objective NSE_JK at the sacrifice of performance at the other sites and metrics (see Figure 6-12, Figure 6-13, and Figure 6-15). It only ensured the good performance of NSE_JK and NSE_MZ, due to their relatively shorter distance and similar magnitude of the observed discharge (Table 3-1). For single-objective calibration, the most sensitive process was the surface runoff computed by the SCS curve number approach, which contributed the highest portion to the discharge. As a result, this process was mathematically optimized to fit the observations and especially to the peak values, as shown in Figure 6-12. Meanwhile, the lateral flow and groundwater return flow both were relatively low when compared to the result from multi-objective calibration at both MZ and JK. The trade-off phenomenon is due to the interactive processes computed by the SWAT model, which is, however, equivalent to a grey-box model, if none of the internal processes can be calibrated. The multi-objective calibration set additional criteria to ensure the reliability of additional specific processes computed by the SWAT model, therefore, to avoid unrealistic trade-off among the objectives. In this study, the effect of the additional criteria was displayed in Figure 6-17 that multi-objective calibration derived varied surface runoff and baseflow ratio among the three subcatchment clusters.

The trade-off among objectives is inevitable in the process-based hydrological modeling system like SWAT that the relations among processes are both independent and interactive. Being independent means that the model algorithms of the individual processes, e.g. the ET and the soil water content, are independent equation systems, though in the natural system they are restraining each other. Being interactive means that these two systems impact each other mathematically by one or multiple parameters, e.g. the ESCO (see Table 6-1) to compensate the ET from soil evaporation. However, unlike the natural system, if no value boundary is defined to constrain individual processes, the iteration between processes cannot ensure the realistic simulation of the objectives. Multi-objective calibration guarantees the reliability of multiple objectives which are critical to the model application instead of only optimizing one output variable. Though trade-off is still unavoidable at the not calibrated processes, the multiple objectives ensured the good performance of the simulated discharge and TN load in terms of the magnitude and the components at the three catchment clusters for the Yuan River Catchment.

7.1.2.2 The applicability of the NSGA-II

The efficiency of deriving the Pareto Front by non-dominated sorting was determined by the size of the objectives. When the number of objectives is relatively large, the less likely a simulation can dominate the other simulations on all the objectives and it is thus less likely to generate the Pareto Front. Therefore, in this study, the three categorized objectives were proposed to substitute the 13 individual objectives. The consistency of the categorized objectives which applied the Euclidean Distance (ED) to aggregate the individual objectives was verified by its similar uncertainty range to the individual objectives in Figure 6-5. The applicability of the substitution was also indicated by the performance of the best simulation and the uncertainty range of the 31 simulations (see Figure 6-12, Figure 6-13, and Figure 6-15) that NSGA-II derived a reliable simulation of almost all individual objectives and a narrower uncertainty band of most objectives. The categorization of the objectives according to multi-site, multi-objective-function, and multi-metric enables the application of NSGA-II by reducing the number of objectives.

Population size and the number of generations are two variables to be defined before applying the NSGA-II. Population size determines the diversity of the population that if more simulations are in a population, the more varied parameter sets can be obtained, and thus a larger size of Pareto Front is also likely to be generated. The convergence of the NSGA-II indicates the percentage of the former generation to dominate the current generation, and it determines if more generations were required. Results showed that in comparison to the number of generations, the convergence was more independent from the population size. Whether population size was 50, 100, or 200, the convergence cannot be reached within 10 generations, and the larger the number of the generations is, the more likely the convergence can be reached (see Figure 6-6). Meanwhile, the stability of the model performance was also less related to the population size but to the generation size (see Figure 6-7). Therefore, if the size of the Pareto Front is not a critical factor but the computational effort, like in this study, a smaller population size, i.e. 50, should be selected for NSGA-II as more generations could be performed to ensure the convergence and the stability of the calibration results. However, since population size determines the diversity of the parameter set values, the smaller population size will also likely to derive a smaller simulation size in the Pareto Front.

7.1.2.3 Sorting algorithms comparison

When compared to the algorithms applied by NSGA-II that the parameters were optimized after each generation, ED is a post-processing approach to sort the performance of each simulation that aggregates the evaluation statistic of the individual objectives into a scalar. The value range of each objective are not identical in this study, in comparison to non-dominated sorting, it is thus a more subjective approach. The scalar function of ED is computed based on the absolute value of the evaluation statistics. However, if the prediction uncertainty ranges of the individual objectives were too distinct from each other, the emphasis of the sorting will be put naturally on the objectives with a larger range of prediction uncertainties. Due to the algorithm of the Euclidean Distance, the change of the absolute values of those objectives will have a higher impact on the value of the scalar, which is more obvious if the scalar function is the Average Weight that a higher inconsistency was observed (see Figure 6-5). However, non-dominated sorting only considers whether a value is surpassing the other instead of the magnitude of the surpassing, and thus the sorting algorithm is not impacted by the absolute value of an objective.

The best simulation derived by ED is the simulation with the lowest Euclidean Distance. However, NSGA-II with non-dominated sorting would derive multiple simulations with equivalent performances. Therefore, if only one simulation is to be selected as the calibration result, criteria should still be set to distinguish simulations at the Pareto Front. Moreover, the processing time to implement the non-dominated sorting is proportional to the population size, and applicability is decreasing as the number of objectives is increasing.

Therefore, if the objectives are of varied value ranges but they are of equivalent importance for evaluation, the non-dominated sorting is a less subjective sorting approach than Euclidean Distance. Meanwhile, if the computational cost is critical, and the objectives are not able to be categorized into a smaller number, Euclidean Distance is expected to be the more suitable approach than non-dominated sorting.

7.1.2.4 The parameter uncertainty

The prediction uncertainty is the aggregation of the input data uncertainty, model approach uncertainty, and parameter uncertainty. Under the condition of identical input data and model construction, the comparison of the uncertainty range indicates the ability of the calibration approach to narrow the parameter uncertainty. Due to equifinality, multiple parameter sets would

generate simulations that meet the requirement of the evaluation criteria. However, the higher the number of the objectives is, the fewer parameter sets could theoretically be qualified. Therefore, it resulted in a shorter range of parameter values of ED and NSGA-II (Figure 6-12), which might explain the less reliable validation in the period from 2011 to 2014.

ED screened the simulations among all the iterations and multiple parameters have kept the uniform distribution but with a narrower value range. However, the initial parameter values used by NSGA-II was only based on the population size. When selecting the population for the first generation, the parameter value ranges have already been reduced to the values that have a good performance in the initial iteration with 2000 simulations in the study (see Figure 5-13). The parameters applied by the subsequent generations were evolved based on the first generation. Though mutation and crossover were performed, the variation of the parameter values was not comparable to SUFI-2 or ED. The diversity of the population was also determined by the population size. The larger the population size is, the more diverted the parameter values can be selected. However, in this study, to guarantee the convergence and the stability of the performance of the last generation within 2000 simulations, a population size of 50 was applied. Therefore, a comparably more narrowed parameter value range is displayed and a narrower prediction uncertainty was also obtained by NSGA-II.

7.1.3 The multi-objective calibration framework

The selection of the objectives is the first step to perform the multi-objective calibration. The three groups of objectives, multi-site, multi-objective-function, and multi-metric were all obtained and applied as the objectives for the calibration. The multi-metric utilized the open-access satellite-based dataset, i.e. the MOD16 ET dataset, and the baseflow ratio extracted from the measured discharge, which was applicable in the YRC with limited observations. The aggregated impact of the objectives has proved that the consideration of the three objective groups fulfilled the purpose of selecting objectives focusing on different targets of the calibration.

Two multi-objective calibration approaches, ED and NSGA-II, were analyzed to compare the effectiveness of the algorithms. ED is the classic aggregation approach and NSGA-II is a multi-objective evolution algorithm. The applicability of the NSGA-II was first ensured by reducing the number of objectives to the three categorized objectives, and the proper population size was then derived. The consistency of the ED on the performance of categorized objectives in comparison to

the individual objectives was also proved to enable a fair comparison between the ED and NSGA-II including the effort of implementation and the algorithms.

The suitability of the two calibration approaches was then represented by obtaining a lower prediction uncertainty and a better performance of the individual objectives in comparison to single-objective calibration. The simulation of the hydrology indicated that the least uncertainty and best performance were obtained by NSGA-II, whereas ED performed better than NSGA-II on the nitrogen load simulation when the identical base hydrological model was applied to calibrate the nitrogen module.

7.1.4 Section summary

The analysis of the above aspects indicated that:

1) the trade-off is unavoidable among objectives. However, multi-objective calibration can add additional criteria to the model to ensure the reliability of the relevant internal processes for model application;

2) the applicability of NSGA-II is also determined by the proper population size and the proper number of generations; if the computational effort is critical, smaller population size is preferred;

3) ED and NSGA-II are both suitable multi-objective calibration approach, represented by the reduced parameter and prediction uncertainty.

Therefore, it is highly recommended to apply the multi-objective calibration for the process-based hydrological model, like SWAT, even to the catchment with limited observations. ED is more suitable for calibration cases requiring lower computational effort and the objectives can be standardized to a similar value ranges. If equivalent importance should be considered for each objective, it is recommended to apply the NSGA-II, which also derives multiple best simulations that enables the further selection of suitable calibration result based on the user's need.

It should also be pointed out that, though multi-objective calibration in the study was able to ensure the performance at the inner site and improve additional processes, e.g. baseflow and surface runoff. The residue between the observations to the simulations was larger than the differences among calibration approaches. Figure 6-12 showed the comparably high NSE_in at MZ. The discrepancy might due to the higher uncertainty of the discharge at MZ, which was derived from

extrapolation instead of measurements (see Section 3.5). Table 6-8 indicates the model's deficiency in the low flow simulation period, which might be resulted from the too simplified groundwater model and the deficiency of the algorithm can only be compensated to a limited extent by applying NSE_in. Figure 6-16 indicated that the calibration failed to capture the extreme peak flow value and Figure 6-13 showed the model's underestimation of the actual ET. Therefore, multi-objective calibration is able to optimize the model's performance on more than one objectives, however, further effort should also be put to overcome the deficiencies due to the uncertainty of the input data and the model's algorithms.

7.2 The interpretation of the nitrogen load simulation

The analysis of the nitrogen load in the catchment was conducted by interpreting the critical source areas (CSAs) and the annual TN load along the Yuan River. The reliability of the model was also analyzed by quantifying the prediction uncertainty of the nitrogen load. Moreover, the input data uncertainty of the nitrogen related input and output variables were also emphasized in this section.

7.2.1 The nitrogen load prediction

7.2.1.1 The CSAs in the YRC

The critical source areas (CSAs) in this study were analyzed based on the annual TN load of each HRU. The catchment was grouped into four classes and each class occupied 25% of the annual TN load. The fewer areas occupied by Class I, II, and III, the higher concentrated the CSAs were. Results have indicated that Class I and II, which contributed to 50% of the in-stream nitrogen load, occupied 26% to 37% of the entire YRC (see Figure 6-24). In YRC, three non-point sources of nitrogen were disposed mainly at the paddy field and the dry arable land. These two landuse types were distributed over the entire YRC, especially in the downstream plain region.

Results indeed indicated that the paddy field with rice contributed to over 50% of the TN load in Class I and II. In terms of spatial distribution, the paddy field at the downstream plain region was the dominant composition of Class I and Class II. Paddy field is the main receiver of the TN from anthropogenic nitrogen sources. Waste from feedlots and rural households were distributed based on the size of the rural area in a subcatchment. Therefore, the downstream plain region, where denser rural population and vast paddy field were distributed, was expected to load a higher TN to the stream. Besides the paddy field, the forest contributed to the second-highest TN load in the YRC. Forest occupied approx. 55% of the entire YRC and consists of evergreen forest and shrubs.

The growth of the forest in the model was triggered by the heat unit aggregated due to the temperature differences. 24% of the total biomass harvest was converted to the residue (Neitsch et al., 2011), which is regarded as a net increase of organic nitrogen in an HRU. Though in nature the humus layer formed by residue would also conserve the nitrogen in the forest, in the SWAT model system, the increase of residue will increase the potential of mineralization to occur. Moreover, forest in the YRC was distributed mainly at the upstream mountainous region with a deeper slope, therefore it might also result in a high contribution to the TN load from the forest.

7.2.1.2 The TN load in the stream

Previous research (Zhang et al., 2011a) on the nitrogen pollution in the Yuan River was limited to the nitrogen from point sources. They accounted for nitrogen contribution in dry season mainly from sewage and concluded a deterioration water quality due to sewage discharge. However, the nitrogen input (see Figure 6-2) had indicated the absolute higher input from non-point sources that occupied 98% of the nitrogen input, especially from the fertilizer. Though only the point sources published were calculated in the study, the domination of the NPS should not be underestimated. Though a lower contribution from NPS could be observed in the non-operational season (see Figure 6-30), which is also the dry season in the YRC, it still contributed to 37% of the TN load annually. Moreover, the scenario without point source input has indicated the nitrogen load from NPS occupied 75% of the total nitrogen load in the stream. Therefore, the quantification of the nitrogen load at individual outlets in the YRC, interpreting from the hydrological model, is an effective approach to indicate the relative contribution of the nitrogen load from both point sources and non-point sources. For the YRC, the nitrogen from the non-point sources should not be ignored due to their much higher input to the catchment and their load in the stream.

7.2.1.3 The calibration approaches

The base hydrological model applied for the nitrogen module calibration was the best simulation of ED, therefore, the performance of the simulated discharge at the internal stations was also ensured. The selection of the identical base hydrological model, as stated above, was to reduce the uncertainty introduced by the flow simulation, due to the higher sensitivity of the nitrogen load simulation to the fluctuation of the flow. Therefore, it can also be expected that the performance of the internal stations on the TN load simulation would have higher accuracy than if the best simulation of SUFI-2 were applied as the base model. Though higher NSE or NSE_in values can

be observed of SUFI-2 whereas NSGA-II and ED had a similar better performance, SUFI-2 still derived an acceptable performance at all the three water quality stations, if evaluated by the NSE criteria proposed by Moriasi (2015).

The differences in the CSAs or the in-stream TN load demonstrated by SUFI-2 and ED or NSGA-II were the consideration of TN load at the inner stations. The contrast among calibration approaches was also more obvious in the 1_Y cluster at the upstream and in the 4_F cluster at the midstream. SUFI-2 derived an overall higher annual TN load over the entire catchment (see Figure 6-29), with a more concentrated CSAs identified, and the portion of the TN from paddy field compared to the forest was also higher. Due to a more satisfying NSE value at 4_F and 1_Y derived from ED or NSGA-II (Figure 6-21), the discrepancy of the TN load was expected to be lower than SUFI-2. The differences in the model performance can be traced back to the parameter values obtained by each calibration approach (see Table 6-3). The parameters to determine the in-soil processes indicated that the values of SUFI-2 would lead to a lower content of inorganic nitrogen in the soil in 1_Y and 4_F, thus a higher nitrogen load from the land. The parameter values of ED and NSGA-II tended to decrease the nitrate transported by the surface runoff and baseflow, thus led to a lower TN load to the stream. The differences of the CSAs among the calibration approaches were reduced when the CSAs were aggregated to each sub-catchment (see Figure 6-27). Therefore, if the CSAs were identified at the subcatchment scale, the parameter uncertainty would have a lower impact on the prediction. However, TN load from individual HRUs can also not be identified.

7.2.1.4 Section summary

The distribution of the CSAs in the YRC enabled the model to trace back the in-stream nitrogen content to its load from the individual HRU. The annual nitrogen load calculated at individual outlets also indicated the dominant contribution from non-point sources. The scale that enables the CSAs to be identified to specific landuses and soil types is supportive of the further proposal of the management strategies. Models that are able to derive the CSAs should be able to account for the nitrogen related processes, which includes not only the transport and the transformation processes of different nitrogen species, but also the algorithms to compute the nitrogen inputs from individual sources and its consumption by crops and management strategies. Therefore, the SWAT model which integrates the hydrological processes with the nitrogen module including the above

abilities can be a useful tool in terms of CSAs identification and therefore to support the management strategies.

Though the SWAT model is able to derive the CSAs to individual landuse and soil types. However, the reliability of the nitrogen module is also determined by the accuracy of the input data and the internal nitrogen transformation processes. In YRC, the improvements on the simulation accuracy could also be performed if the data, e.g. the validation of the nitrogen from the forest, the denitrification rate of the paddy field, and the nitrogen consumption rate of the crops, could be obtained. Moreover, similar to flow calibration, the limitations of the calibration should also not be ignored that, the year 2008 was not well captured by calibration results and the calibration cannot compensate for the inaccuracy of the simulation results due to the lack of the calibration of individual nitrogen species.

7.2.2 The prediction uncertainty of the nitrogen load analysis

The prediction uncertainty is the aggregated effect of all the uncertainty sources on the modeling results and calibration is the procedure to reduce the uncertainty introduced by parameters. When compared to flow calibration, higher uncertainty of the nitrogen simulation exists. These uncertainties were introduced by a higher uncertainty of the input data sets spatially and temporally (see Section 3.6) and computed by the more complicated processes of nitrogen transport and transformation that relies heavily on the observations to reduce the parameter uncertainty (see Section 5.1.2). Moreover, the output variables included multiple nitrogen species and their results cannot be validated individually due to limited data availability in the YRC. Therefore, the aim of generating the prediction uncertainty for the nitrogen load was to indicate how reliable the simulation obtained by the calibration parameter sets is, under the condition of the current input data and model system.

7.2.2.1 The CSAs

As stated above, the ability to reduce the parameter uncertainty by the three calibration approaches resulted in the different scales of the prediction uncertainty. The results indicated how the prediction uncertainty would be impacted if the parameter values were changing. The range of the prediction uncertainty has shown that SUFI-2>NSGA-II>ED. The identified CSAs indicated that each of the classes was affected by the uncertainty to a different extent. When compared to ED and NSGA-II, SUFI-2 especially showed a different impact at the up to mid-stream of the

catchment where the TN load was not calibrated according to the subcatchment clusters. However, the cumulative load of NSGA-II and ED indicated that, though prediction uncertainty existed, the impact of it on the cumulative area to each class was less significant (see Figure 6-27).

Previous studies identifying the CSAs on the sub-catchment level (Liu et al., 2016) or the HRU level (Niraula et al., 2011) with the SWAT model have summarized that calibration would have no significant impact on the distribution of the CSAs. It is due to the reason that the parameterization in a lumped manner would have a similar impact on the load spatially. Thus, they concluded that the CSAs can also be identified on the ungauged catchment. The annual TN load identified on the sub-catchment level in this study (see Figure 6-27) indicated fewer spatial differences among the calibration approaches when compared to the CSAs at the HRU level, which also illustrates the lower sensitivity of the sub-catchment scale TN load to the calibration parameters. At the HRU level, the fewer variations of the areas identified of each CSAs Class obtained by ED and NSGA-II also supported the above observations. However, the prediction uncertainty was obtained from the 1000 simulations extracted from the total 3000 simulations and used the identical hydrological model, which was expected to be lower. Therefore, the more significant uneven impact of the CSAs Classes obtained by the prediction uncertainty of SUFI-2 still indicated that the CSAs were sensitive to the parameters. The disagreement in this research to the literature may be explained that the parameter sets were not identical in those studies and might result in the adjustment of different processes in the model. Moreover, the catchment conditions were not generalized by the previous studies and the sensitivity of the CSAs was also determined by the catchment condition, e.g. the portion of paddy fields in the YRC.

7.2.2.2 The in-stream nitrogen load

The r-factors of the 1000 simulations were all no higher than 0.6 (see Table 6-10) and they were much lower in comparison to the results obtained in the author's previous work (He et al., 2019) that the r-factor was 1.29. The current study screened out 1000 simulations to derive the uncertainty band from the 3000 simulations in total, whereas the previous study only derived the uncertainty from one iteration containing 1000 simulations in total. For calibration approach SUFI-2 or ED, the more the parameter sets were applied in the calibration, the higher the possibility of deriving simulations with good performance. Therefore, the current study had a significantly lower uncertainty band. Moreover, the differences in the parameters selected for calibration and whether

the base hydrological model was calibrated by multiple objectives could also affect the width of the uncertainty band.

The TN load of each HRUs was accumulated at the 8 outlets and the uncertainty of the TN was averagely 20% lower than the best simulation obtained by SUFI-2 (see Figure 6-29). The uncertainty band of SUFI-2 even covered the uncertainty range of ED and over 50% of NSGA-II. The TN load at each outlet was an indicator to demonstrate the mass of the TN discharged to the stream. The sources of the inorganic nitrogen were fertilizer, atmospheric deposition, and point source discharges; the sources of the organic nitrogen were waste from feedlots and rural households. The simulated TN was only an aggregation of multiple nitrogen species. The current observation of only TN at the three stations was not able to ensure the reliable outputs of the individual nitrogen species. Therefore, the parameter uncertainty would also unevenly impact the individual nitrogen transformation and transport processes, and thus the relative consumption of the nitrogen from each source also requires further validation.

7.2.3 The uncertainty of the input data

In terms of analyzing the uncertainty of the nitrogen module, the input data uncertainty should not be neglected in this study. The limited data of the model construction and calibration determined the YRC as a data-scarce catchment.

7.2.3.1 The uncertainty of the hydrological input

The input data for the model construction was listed in Table 3-2. It has been summarized that, in terms of assessing the hydrological model, the scarcity of the data was the precipitation and the measured discharge. The density of the precipitation gauges was relatively low that the YRC with the area of approx. 6223km^2 had only 18 rain gauges available for the study, which is lower than the standard proposed by the National Ministry of Water Resources (12 rain gauges per 3000km^2) (Ministry of Water Resources of the PRC, 2013). The available period of the hydrometric data was short and MZ station was obtained by extrapolation. The limited spatial variation of the precipitation would lead to the potential bias of the precipitation input at individual subcatchment. The higher uncertainty of the MZ station would lead to the less reliable simulation at MZ and a relatively short length of measured discharge would result in a simulation unable to capture the variation in years. Therefore, the higher the NSE_in value at MZ (Figure 6-15) and a less satisfying simulation in the validation period. (Figure 6-18) were observed.

7.2.3.2 The uncertainty of the nitrogen sources

In Section 3.6 the processing of the nitrogen sources was stated. In comparison to the data prepared for the hydrological model construction, i.e. DEM, landuse, soil property, and climate data, the processing of the nitrogen sources data was with a relatively larger uncertainty. The fertilizer per hectare applied to the paddy field was identical over the catchment without considering the variation among different regions. The method to distribute the nitrogen from feedlots and rural households was designed based on the data condition that the available data were obtained on the level of the administrative district and the magnitude of the sources was determined by the size of the rural region within a subcatchment. The identical fertilizer per hectare and the distribution of the feedlots and rural households based on the size of the rural region increased the uncertainty of the spatial distribution of the nitrogen sources. The operation date of the fertilizer or manure also introduced uncertainty temporally. Moreover, the underestimation of the point sources and the homogeneous nitrogen content in the precipitation also led to the uncertainty that cannot be quantified.

The data and methods applied were developed based on literature and the author's local survey, and thus the uncertainties cannot be avoided. The high nitrogen released from the forest cannot be fully explained. The distribution of the manure based on the size of the rural regions also assigned a higher weight to those regions in terms of CSAs identified. However, due to the size of the catchment, it is unlikely to have accurate nitrogen input for every HRUs, therefore, the uncertainty introduced by the nitrogen sources will always exist. Meanwhile, the spatial uncertainty of the data could be reduced theoretically in catchments with the above-mentioned data available at a smaller scale, e.g. at individual villages (Piniewski et al., 2015).

7.2.3.3 The observed nitrogen applied in the calibration

The observed nitrogen available in the study was the sampling data of three water quality stations. At 6_L, the data was sampled in a monthly frequency containing missing data whereas 1_Y and 4_F were sampled every two or three months. Besides the frequency, the quality of the measured nitrogen was also less reliable in comparison to the measured discharge (Harmel et al., 2006).

The calibration of TN also led to uncertainties. As mentioned above, five nitrogen pools (see Section 5.1.2) were considered in the SWAT model with their respective paths of transport and transformation. The TN is the aggregation of the five nitrogen pools including both organic and

inorganic nitrogen and it is not possible to obtain the percentage of the individual nitrogen species reversely. The nitrogen species have been specified to each nitrogen source and would be transformed and transported in the different paths based on the individual modeling approaches. However, only the TN was able to be evaluated. Therefore, no information can help to validate the transformation and transport processes respective to each nitrogen source. The parameters to determine those processes also varied between nitrogen species, e.g. the denitrification rate to determine the consumption of the nitrate in the soil, the enrichment ratio, and the in-soil carbon content to determine the mineralization process. Accordingly, the calibration only ensured the value of the TN load and the simulated load of each nitrogen species was less reliable. The water quality samples were taken based on the water quality standard (Ministry of Ecology and Environment of the PRC, 2002) in catchments like the YRC in China and they only include ammonia nitrogen and TN and even ammonia nitrogen was not available for this study. In terms of future application, if the model was implemented to simulate the impact of reducing specific nitrogen sources on the catchment's total nitrogen condition, observed nitrogen species, i.e. both nitrate and ammonia nitrogen, are recommended to be measured.

7.2.4 Section summary

The prediction uncertainty of the CSAs and the annual nitrogen load indicated the impact of the uncertainty sources on the simulation. Applying the simulations of the highest and lowest PBIAS to represent the uncertainty ranges helped to quantify the nitrogen load simulation, which is of great importance to assess the model's reliability in terms of future application. The results illustrate the advantage of NSGA-II and ED to reduce the prediction uncertainty in comparison to SUFI-2.

Based on the current data information, it is hard to quantify the uncertainty of the input hydrological or nitrogen related data. However, the potential impact of the uncertainty from those sources should also be considered when evaluating the simulation results. The higher uncertainty of the input nitrogen data, the model algorithms, and the measured data for calibration resulted in the relatively lower reliability of the nitrogen simulation with the SWAT model for the YRC.

7.3 Outlook

7.3.1 Recommendations for applying the multi-objective calibration procedure

The multi-objective calibration framework applied in the study provided a practical procedure for the calibration of catchments similar to the YRC, even with data scarcity. This practical procedure as one result of the research is pointed out by key steps as follows:

1) Preparing the observations according to the multi-site, multi-objective-function and multi-metric, and the objectives can include but not limited to those applied in this study, which are also determined by the target of the application, e.g. measured runoff rate, soil moisture from remote sensing, nitrogen content in the groundwater, etc..

   If ED or NSGA-II is considered as the calibration approach, the factors to determine the suitability of the two approaches include the processing time, number of objectives, the objectivity of the dominance, and the equifinality.

2) If ED is to be selected as the post-processing tool to evaluate the performance, then lower processing effort can be expected in comparison to NSGA-II and number of the objectives are not a limitation; however, the trade-off between objectives due to the difference in the value range of the individual evaluation statistics should also be expected. Meanwhile, the equifinality of the model system would not be emphasized, since only one best solution is obtained.
3) If NSGA-II is to be selected as the calibration approach. The applicability should first be ensured by decreasing the number of objectives, e.g. the categorization applied in this study, and by obtaining a suitable population size and a proper number of generations to ensure the convergences, e.g. evaluated by the C-function. The processing effort is expected to be higher than ED. However, the approach derives a less subjective evaluation of a simulation's dominance and meanwhile also maintains the equifinality.

7.3.2 Suggestions for future research

The multi-objective calibration is able to improve the model's performance, however, to a limited extent that the discrepancy to the observations cannot entirely vanish. If data availability is not the limitation, future research could also be conducted in the aspects of improving the simulation on certain processes, e.g. a more complex groundwater system other than the simplified approach of

SWAT to better capture the baseflow. In terms of the nitrogen load simulation, the validation of individual transformation and transportation processes could also be conducted to ensure the reliability of the prediction.

In this study, the application of the model in terms of integrating catchment management was not explored in detail. The analysis could be further conducted to assess the impact of the management strategies on the mitigation of the nitrogen load (Mittelstet et al., 2016). The CSAs could be further analyzed with additional criteria to identify the relationship to the population, water quantities, etc. (Liu et al., 2016). Besides the CSAs, the impact of the nitrogen load in the stream could be evaluated by the ecological index (Woo et al., 2019). The hydrological model could also be coupled with modules to quantify the cost of the management strategies (Liu et al., 2019). The future application of the model in terms of a holistic assessment of the catchment including the ecology and social-economic impact is of great potential to support the catchment management.

7.3.3 Proposals for the YRC model application

According to the prediction of the YRC, the critical source areas (CSAs) were identified and the annual nitrogen load along the river was quantified, which could offer an insight into the nitrogen condition in the YRC, especially under the condition without measures taken to manage the nitrogen pollution. The quantified prediction uncertainty indicates the potential improvements the model could have in terms of future application in the YRC.

1) A more detailed survey on the spatial and temporal distribution of the nitrogen sources to individual villages could theoretically decrease the input data uncertainty. Moreover, a more completed monitoring system including the measurements of the discharge, precipitation, and, in particular, the individual nitrogen species could also benefit a more reliable calibration. The construction of the monitoring system could also include groundwater measurements and surface runoff measurements to increase the reliability of the water balance observations.
2) In terms of nitrogen calibration, if the observations of the nitrate and ammonia nitrogen are available, the individual nitrogen dynamic process could also be analyzed. For example, applying the crop yield as an objective to calibrate the nitrogen consumption by the plants (Malagó et al., 2017) and applying the denitrification rate estimated from field study to calibration denitrification processes (Yen et al., 2014; Moriasi et al., 2013). In terms of the

relatively high load also from the mountainous region, an investigation to validate the vegetation and the erosions status in those regions is also recommended.

8 Conclusion

The research proved that the accuracy of the process-based hydrological model SWAT could be improved by the multi-calibration approach for the data-scarce catchment YRC and the approach to quantify the nitrogen load prediction benefits the interpretation of the simulation. The research questions proposed are then answered and summarized as follows.

1) How to derive and apply the multiple objectives in the calibration for catchments with data scarcity like the YRC?

The research proposed to apply the MOD16 ET dataset and the metrics extracted from the fundamental observations, e.g. the measured discharge, as the additional objectives for the YRC. The applicability of the above objectives has been indicated by the improved prediction of the ET processes and the baseflow ratio. Moreover, using both NSE and inverse NSE for the discharge and TN load at multiple sites showed the advantages of enhancing the fitness of the variable's magnitude. The above-mentioned objectives indicate the potential of applying the multi-objective calibration approach on the data-scarce catchment by obtaining observations from the global open-access satellite-based dataset and the existing fundamental observations.

2) How is multi-objective calibration superior to single-objective calibration?

The trade-off is unavoidable among objectives. However, multi-objective calibration can add additional criteria to the model to ensure the reliability of the critical internal processes for model application. The effect is also represented by a reduced parameter and prediction uncertainty obtained by the two multi-objective-calibration approaches, ED and NSGA-II. The consideration of the proposed three groups of objectives has ensured the model's better performance in terms of the spatial parameterization, the magnitude of the observed time-series, and the water balance components in the YRC.

3) How to interpret the nitrogen load and quantify its prediction uncertainty at the catchment scale?

The analysis of the nitrogen load was based on the prediction of the critical source areas (CSAs) and the in-stream nitrogen load along the Yuan River. The distribution of the CSAs for the YRC enabled the model to trace back the in-stream nitrogen content to its load at the individual HRU

and the annual total nitrogen load quantified the amount of nitrogen to be released to the YR. The prediction uncertainty of the CSAs was obtained by the two simulations with the highest or the lowest PBIAS to the best simulation. These two simulations have proved to be able to represent the uncertainty range of the simulated TN load in the stream and the method was thus practical to derive the prediction uncertainty for CSAs at the HRU level.

It is recommended that the multi-objective calibration framework is considered when simulating the catchment-scale hydrology and nitrogen load with the process-based hydrological model SWAT, even for catchments with data scarcity. In terms of the multi-objective calibration, ED is suggested for calibration cases preferring lower computational effort and the number of objectives is not a limitation, whereas NSGA-II is suggested that if equivalent importance is required for the individual objectives and equifinality is emphasized. In terms of analyzing the nitrogen load, the approach applied to quantify the prediction uncertainty could also be considered. The identified critical source areas and the quantified nitrogen load along the Yuan River provide references that benefit the understanding of the nitrogen condition in the Yuan River Catchment.

Summary

The study is summarized as the analysis of hydrology and nitrogen load prediction for a representative catchment with limited data availability by applying a representative process-based hydrological model in combination with a contribution to the multi-objective calibration framework and the approach to quantify the prediction uncertainty.

Process-based hydrological models play an important role in supporting catchment management to predict the nitrogen load spatially and temporally from its sources to and in the surface water body. The SWAT model was selected to analyze the nitrogen condition in the Yuan River Catchment (YRC) in China with a size of 6223.69km^2. Problems arose in the research that 1) the classic single-objective calibration has its limitations in terms of spatial parameterization and the reliability of individual hydrological processes; 2) multi-objective calibration can alleviate limitations of single-objective calibration, however, is restrained by the requirements of additional observations and a practical procedure for implementation; 3) prediction uncertainty of nitrogen load is inevitable of process-based hydrological models and should thus be quantified; 4) the YRC is a representative catchment with data scarcity, e.g. in terms of model set-up, limited data were available spatially and temporally; in terms of calibration, only observed discharge and total nitrogen load were available.

The study aimed to improve the reliability of the process-based hydrological model SWAT by applying the multi-objective calibration approach, quantifying the prediction uncertainty in terms of simulating the hydrology and nitrogen load on the data-scarce catchment.

To implement the multi-objective calibration, the study proposed to overcome the data scarcity of additional observations by the utilization of the open-access satellite-based dataset, e.g. the MODIS global terrestrial ET product, and the metrics extracted from the existing fundamental observations, e.g. measured discharge. The study also proposed to consider three groups of objectives for the calibration, which are multi-site, multi-objective-function, and multi-metric. Multi-site aimed to improve the spatial parameterization, multi-objective-function aimed to improve the simulation of the magnitude of the multi-site, and multi-metric aimed to ensure the reliability of the individual hydrological processes. To quantify the prediction uncertainty of the nitrogen load simulation, it is also proposed to use the simulations with the highest or the lowest

PBIAS, to derive the prediction uncertainty of the nitrogen load in the stream and the critical source areas (CSAs) for the YRC.

The analysis of the calibration was led by the comparison of the three calibration approaches, single-objective calibration with SUFI-2, multi-objective calibration with Euclidean Distance (ED) and NSGA-II. ED was applied as a post-processing tool for performance evaluation after each calibration iteration. The applicability of the NSGA-II was analyzed first by the convergence determined by the population size and the number of generations. 7000 simulations in total were implemented by each approach and each simulation took approx. 1.5mins (SUFI-2 and ED) or 2mins (NSGA-II) for execution by a standard PC. The hydrology and the nitrogen module were calibrated in sequence. For hydrological calibration, the multi-site calibrated the discharge at the three hydrometric stations; the multi-objective-function applied both NSE and inverse NSE (NSE_in) to evaluate the discharge, and the multi-metric considered the coverage of the MOD16 ET to the simulated actual evapotranspiration (ET) and the PBIAS of the ratio of surface runoff and baseflow to the discharge. The efficiency of the non-dominated sorting was ensured by aggregating 13 objectives with Euclidean Distance into three categorized objectives, including 1) NSE for the multi-site; 2) NSE_in for multi-site; 3) the water balance components as the multi-metric. For the nitrogen module simulation, the nitrogen load at three water quality stations were the multi-site objectives; NSE and NSE_in were the multi-objective-function.

The nitrogen sources were identified in the YRC as nitrogen fertilizer, untreated waste from feedlots and rural households, point discharge, and atmospheric deposition. The nitrogen load analysis was targeted at the in-stream nitrogen load along the Yuan River and the critical source areas (CSAs) derived at the hydrologic response unit (HRU) level for the entire YRC. The prediction uncertainty was derived from the top 1000 simulations of each calibration approach, for the CSAs and the annual total nitrogen (TN) load along the Yuan River.

The consideration of the multi-site has ensured the model's performance at all the three hydrometric stations and the water quality stations whereas an acceptable NSE at the outlet was also guaranteed. The application of multi-objective-calibration derived lower discrepancy of the baseflow and the deviation of water balance components was reduced by implementing the multi-metric at the YRC. The trade-off is unavoidable among objectives. However, multi-objective calibration added additional criteria to the model to ensure the reliability of the relevant processes.

The applicability of the NSGA-II was based on deriving the categorized objectives and the proper population size, which indicated the additional effort required for NSGA-II. The suitability of ED and NSGA-II was then represented by obtaining a lower prediction uncertainty and a better performance of the individual objectives in comparison to single-objective calibration. ED requires lower computational effort and the number of objectives is not a limitation. NSGA-II derives a less subjective evaluation by evaluating individual objectives with equivalent importance and it maintains the equifinality by obtaining multiple optimal simulations.

The nitrogen load indicated that over 50% of the TN loading to the stream was from approx. 30% of the area in the YRC and paddy field contributed over 50% of those TN loads. Meanwhile, 98% of the total nitrogen input was from non-point anthropogenic sources. The downstream plain region where denser rural population and vast paddy field were distributed, was expected to load a higher TN to the stream. The prediction uncertainty of areas loading 50% of TN ranged from 26% to 37% and the ratio of the uncertainty bandwidth to the average TN load ranged from 11% to 22%, differing among calibration approaches that single-objective function always has the highest uncertainty. The prediction uncertainty of the CSAs and the annual nitrogen load also indicated the impact of the uncertainty sources on the simulation. However, the higher uncertainty of the input nitrogen data, the model approaches, and the measured data for calibration resulted in the relatively lower reliability of the nitrogen simulation when compared with the discharge of the hydrological simulation with the SWAT model for the YRC.

The study suggested applying the calibration procedure first by generating the observations from all the three objective groups, which include multi-site, multi-objective-function, and multi-metric. For data-scarce catchments, global open-access satellite-based datasets and the best use of the existing observations should be considered to obtain the additional objectives. In terms of implementing ED or NSGA-II as the multi-objective calibration approach, their applicability and suitability should also be considered in the first place. Applying the simulations of the highest and lowest PBIAS to represent the uncertainty ranges helps to quantify the prediction uncertainty of the nitrogen load, which is of great importance to assess the model's reliability in terms of future application. A better understanding of the nitrogen condition in the YRC is also acquired by the identified CSAs and the quantified nitrogen load along the Yuan River.

References

Abbaspour, K.C., Johnson, C.A., van Genuchten, M.T., 2004. Estimating Uncertain Flow and Transport Parameters Using a Sequential Uncertainty Fitting Procedure. Vadose Zone Journal 3 (4), 1340.

Abbaspour, K.C. (Ed.), 2005. Calibration of Hydrologic Models: When is a Model Calibrated?

Abbaspour, K.C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J., Srinivasan, R., 2007. Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. Journal of Hydrology 333 (2-4), 413–430.

Abbaspour, K.C., 2015. SWAT-CUP. SWAT Calibration and Uncertainty Programs. Accessed.

Abbaspour, K.C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H., Kløve, B., 2015. A continental-scale hydrology and water quality model for Europe. Calibration and uncertainty of a high-resolution large-scale SWAT model. Journal of Hydrology 524, 733–752.

Abbaspour, K., Vaghefi, S., Srinivasan, R., 2018. A Guideline for Successful Calibration and Uncertainty Analysis for Soil and Water Assessment: A Review of Papers from the 2016 International SWAT Conference. Water 10 (1), 6.

Abiodun, O.O., Guan, H., Post, V.E.A., Batelaan, O., 2017. Comparison of MODIS and SWAT Evapotranspiration over a Complex Terrain at Different Spatial Scales. Hydrol. Earth Syst. Sci. Discuss., 1–36.

Ajami, N.K., Duan, Q., Sorooshian, S., 2007. An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resour. Res. 43 (1), 655.

Allen, R.G., 1998. Crop evapotranspiration. Guidelines for computing crop water requirements. FAO, Rome.

Alvarez-Cobelas, M., Angeler, D.G., Sanchez-Carrillo, S., 2008. Export of nitrogen from catchments: a worldwide analysis. Environmental pollution (Barking, Essex: 1987) 156 (2), 261–269.

Ansari, A.A., Gill, S.S. (Eds.), 2014. Eutrophication: Causes, Consequences and Control. Volume 2. Springer Netherlands, Dordrecht, s.l.

Arabi, M., 2010. Nutrient Modeling Overview. https://www.epa.gov/sites/production/files/2015-10/documents/day2_arabi.pdf. Accessed.

Arnold, J.G., Allen, P.M., Muttiah, R., Bernhardt, G., 1995. Automated Base Flow Separation and Recession Analysis Techniques. Groundwater 33.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: Model development. J Am Water Resources Assoc 34 (1), 73–89.

Arnold, J.G., Muttiah, R.S., Srinivasan, R., Allen, P.M., 2000. Regional estimation of base flow and groundwater recharge in the Upper Mississippi river basin. Journal of Hydrology 227 (1-4), 21–40.

Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., Santhi, C., Harmel, R.D., van Griensven, A., Van Liew, M.W., Kannan, N., Jha, M.K., 2012a. SWAT: Model Use, Calibration, and Validation. American Society of Agricultural and Biological Engineers 55 (4), 1491–1508.

Arnold, J.G., Srinivasan, R., Williams, J.R., Haney, E.B., Neitsch, S.L., 2012b. Soil & Water Assessment Tool. Input / Output Documentation, Version 2012. Accessed.

Arnold, J., Youssef, M.A., Yen, H., White, M.J., Sheshukov, A.Y., Sadegi, A.M., Moriasi, D.N., Steiner, J.L., Amatya, D.M., Skaggs, R.W., Haney, E.B., 2015. Hydrological Processes and Model Representation. Impact of Soft Data on Calibration. Trans. ASABE 58 (6), 1637–1660.

Bai, F., Wang, X.C., 2011. Study on effect on dynamics model of biodegration of organic matter in aerobic composting of human feces from temperature. In: D. Sun, W.-P. Sung, R. Chen (Editors), Frontiers of green building, materials and civil engineering. Selected, peer reviewed papers from the International Conference on Green Building, Materials and Civil Engineering (GBMCE 2011), August 22-23, 2011, Shangri-la, China / edited by Dongye Sun, Wen-Pei Sung and Ran Chen. TTP, Durnten-Zurich, pp. 2493–2496.

Barnard, J.D., 1948. Heat units as a measure of canning crop maturity. The Canner 106, 28.

Bear, J., 1988. Dynamics of Fluids in Porous Media, New York.

Bekele, E.G., Nicklow, J.W., 2007. Multi-objective automatic calibration of SWAT using NSGA-II. Journal of Hydrology 341 (3-4), 165–176.

Bergström, S., 1995. The HBV model. Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, Colorado, USA.

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Advances in Water Resources 16 (1), 41–51.

Beven, K.J., 2012. Rainfall-runoff modelling. The primer. Wiley-Blackwell, Chichester, West Sussex, Hoboken, NJ.

Bicknell, B.R., Imhoff, J.C., Kittle, J.L., Jr., Donigian, A.S., Jr., Johanson, R.C., 1997. Hydrological Simulation Program—FORTRAN: User's Manual for version 11. U.S. Environmental Protection Agency, National Exposure Research Laboratory, Athens, GA, USA.

Bingner, R.L., Theurer, F.D., Yuan, Y.P., 2011. AnnAGNPS Technical Processes Documentation, Version 5.2. Accessed.

Borah, D.K., Bera, M., 2003. Watershed-Scale Hydrologic and Nonpoint-Source Pollution Models. Review of Mathematical Bases. Transactions of the ASAE 46 (6), 1552–1566.

Bouraoui, F., Grizzetti, B., 2014. Modelling mitigation options to reduce diffuse nitrogen water pollution from agriculture. The Science of the total environment 468-469, 1267–1277.

Brown, L.C., Barnwell, JR., 1987. The enhanced water quality models QUAL2E and QUAL2E-UNCAS documentation and user manual. EPA document EPA/600/3-87/007. USEPA, Athens, GA.

Carpenter, S.R., 2008. Phosphorus control is critical to mitigating eutrophication. Proceedings of the National Academy of Sciences of the United States of America 105 (32), 11039–11040.

Confesor, R.B., Whittaker, G.W., 2007. Automatic Calibration of Hydrologic Models with Multi-Objective Evolutionary Algorithm and Pareto Optimization. J Am Water Resources Assoc 43 (4), 981–989.

Daren Harmel, R., Smith, P.K., 2007. Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. Journal of Hydrology 337 (3-4), 326–336.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Computat. 6 (2), 182–197.

Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour. Res. 28 (4), 1015–1031.

Efstratiadis, A., Koutsoyiannis, D., 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. Hydrological Sciences Journal 55 (1), 58–78.

Ercan, M.B., Goodall, J.L., 2016. Design and implementation of a general software library for using NSGA-II with SWAT for multi-objective model calibration. Environmental Modelling & Software 84, 112–120.

European Commission, 1991. Council Directive 91/676/EEC. Concerning the protection of waters against pollution caused by nitrates from agricultural sources. Accessed.

Fang, Z., 2011. The report of integrated planning of Yuan River Catchment, Jiangxi Prov. (in Chinese). Jiangxi Water Conservancy Planning Design Institute, Nanchang.

Folle, S., Dalzell, B., Mulla, D., 2007. Evaluation of Best Management Practices (BMPs) in Impaired Watersheds Using the SWAT Model. Accessed.

Grizzetti, B., Bouraoui, F., Marsily, G. de, 2008. Assessing nitrogen pressures on European surface water. Global Biogeochem. Cycles 22 (4).

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. Water Resources Research 34 (4), 751–763.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1999. Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration. J. Hydrol. Eng. 4 (2), 135–143.

Gupta, H.V., Wagener, T., Liu, Y., 2008. Reconciling theory with observations: elements of a diagnostic approach to model evaluation. Hydrological Processes 22 (18), 3802–3813. https://onlinelibrary.wiley.com/doi/full/10.1002/hyp.6989.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria. Implications for improving hydrological modelling. Journal of Hydrology 377 (1-2), 80–91.

Haas, M.B., Guse, B., Pfannerstill, M., Fohrer, N., 2016. A joined multi-metric calibration of river discharge and nitrate loads with different performance measures. Journal of Hydrology 536, 534–545.

Harmel, R.D., Cooper, R.J., Slade, R.M., Haney, R.L., Arnold, J.G., 2006. Cumulative uncertainty in measured streamflow and water quality data for small watersheds. Transactions of the ASABE 49 (3), 689–701.

Harmel, R.D., Smith, P.K., Migliaccio, K.W., 2010. Modifying Goodness-of-Fit Indicators to Incorporate Both Measurement and Model Uncertainty in Model Calibration and Validation. Trans. ASABE 53 (1), 55–63.

Harmel, R.D., Smith, P.K., Migliaccio, K.W., Chaubey, I., Douglas-Mankin, K.R., Benham, B., Shukla, S., Muñoz-Carpena, R., Robson, B.J., 2014. Evaluating, interpreting, and communicating performance of hydrologic/water quality models considering intended use. A review and recommendations. Environmental Modelling & Software 57, 40–51.

He, Q., Wendland, F., Molkenthin, F., 2019. The analysis of nitrogen load and the simulation uncertainty using SWAT in a catchment with paddy field in China. Water Science and Technology 80 (4), 806–816.

Herman, M.R., Nejadhashemi, A.P., Abouali, M., Hernandez-Suarez, J.S., Daneshvar, F., Zhang, Z., Anderson, M.C., Sadeghi, A.M., Hain, C.R., Sharifi, A., 2018. Evaluating the role of evapotranspiration remote sensing data in improving hydrological modeling predictability. Journal of Hydrology 556, 39–49. http://www.sciencedirect.com/science/article/pii/S0022169417307631.

Hu, F., Wang, M., Zhou, W., 2011. Structure of Phytoplankton Community and Evaluation of Water Quality in Yuanhe River. Journal of Hydroecology 32 (1), 27–33.

Huang, J., Yuan, R., 2003. Research on Water Quality of Yuanhe River (in Chinese). Journal of Yichun University 25 (6), 71–73.

Huang, Z., Xue, B., Pang, Y., 2009. Simulation on stream flow and nutrient loadings in Gucheng Lake, Low Yangtze River Basin, based on SWAT model. Quaternary International 208 (1-2), 109–115.

Immerzeel, W.W., Droogers, P., 2008. Calibration of a distributed hydrological model based on satellite evapotranspiration. Journal of Hydrology 349 (3-4), 411–424.

Jiangxi Statistical Bureau, 2008-2014. Jiangxi Statistical Yearbook 2008, 2009, 2010, 2011, 2012, 2013, 2014. China Statistics Press, Beijing.

Johnes, P.J., 1996. Evaluation and management of the impact of landuse change on the nitrogen and phosphorus load delivered to surface waters: the export coefficient modelling approach. Journal of Hydrology 183, 323–349.

Johnsson, H., Larsson, M., Mårtensson, K., Hoffmann, M., 2002. SOILNDB: a decision support tool for assessing nitrogen leaching losses from arable land. Environmental Modelling & Software 17 (6), 505–517.

Kan, G., He, X., Li, J., Ding, L., Hong, Y., Zhang, H., Liang, K., Zhang, M., 2017. Computer Aided Numerical Methods for Hydrological Model Calibration: An Overview and Recent

Development. Archives of Computational Methods in Engineering 26 (1), 35–59. https://doi.org/10.1007/s11831-017-9224-5.

Knisel, W.G., 1980. CREAMS: a field scale model for Chemicals, Runoff, and Erosion from Agricultural Management Systems [USA]. United States. Dept. of Agriculture. Conservation research report (USA).

Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 5, 89–97.

Ladson, A.R., Brown, R., Neal, B., Nathan, R., 2013. A standard approach to baseflow separation using the Lyne and Hollick filter. AJWR 17 (1).

Le Moine, N., 2008. Le bassin versant de surface vu par le souterrain une voie d'amélioration des performances et du réalisme des modèles pluie-débit? Doctoral thesis, Université Pierre et Marie Curie (Paris), France.

Legates, D.R., McCabe, G.J., 1999. Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35 (1), 233–241.

Lerat, J., Andréassian, V., Perrin, C., Vaze, J., Perraud, J.M., Ribstein, P., Loumagne, C., 2012. Do internal flow measurements improve the calibration of rainfall-runoff models? Water Resources Research 48.

Li, L., 2008. The analysis of the water quality of Yuan River within Yichun section (in Chinese). Anhui Agri. Sci. Bull 14 (16), 122.

Li, Q., Zheng, J., Yao, Y., Tang, J., He, J., Lin, P., Xie, L., Zhu, Q., 2013. Rural sanitation status in Jiangxi Province in 2011 (in Chinese). Journal of Environmental Hygiene 3 (4), 316–323.

Li, Y., Liu, C., 2015. The current situation of water quality in Xinyu section of Yuanhe River and its prevention (in Chinese). Journal of Xinyu University 20 (1), 13–15.

Li, Z., Yang, W., Xu, F., Ma, C., 2011. Distribution of pollutants in the sediments along the Yuan River (in Chinese). Journal of China Institute of Water Resources and Hydropower Research 9 (2), 132–136.

Lindgren, G.A., Wrede, S., Seibert, J., Wallin, M., 2007. Nitrogen source apportionment modeling and the effect of land-use class related runoff contributions. Hydrology Research 38 (4-5), 317–331.

Liu, B., Liu, H., Zhang, B., Bi, J., 2013. Modeling nutrient release in the Tai Lake basin of China: source identification and policy implications. Environmental management 51 (3), 724–737.

Liu, R., Xu, F., Zhang, P., Yu, W., Men, C., 2016. Identifying non-point source critical source areas based on multi-factors at a basin scale with SWAT. Journal of Hydrology 533, 379–388.

Liu, Y., Wang, R., Guo, T., Engel, B.A., Flanagan, D.C., Lee, J.G., Li, S., Pijanowski, B.C., Collingsworth, P.D., Wallace, C.W., 2019. Evaluating efficiencies and cost-effectiveness of best management practices in improving agricultural water quality using integrated SWAT and cost evaluation tool. Journal of Hydrology 577, 123965. http://www.sciencedirect.com/science/article/pii/S0022169419306857.

LÜ, W., Shi, Q., Pan, X., Huang, S., Shang, Y., Tan, X., Li, M., Hu, S., Zeng, Y., 2016. Changes in safe production dates and heat-light resources of double cropping rice in Jiangxi Province in recent 30 years (in Chinese). Chinese Journal of Rice Science 30 (3), 323–334.

Malagó, A., Bouraoui, F., Vigiak, O., Grizzetti, B., Pastori, M., 2017. Modelling water and nutrient fluxes in the Danube River Basin with SWAT. Science of The Total Environment 603-604, 196–218. http://www.sciencedirect.com/science/article/pii/S0048969717313505.

Mateo-Sagasta, J., Zadeh, S.M., Turral, H., 2017. Water pollution from agriculture: a global review. Executive summary. http://www.fao.org/3/a-i7754e.pdf. Accessed.

McKay, M. d., Beckman, R.J., Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21 (2), 239.

Merwade, V., Baffaut, C., Bieger, K., Boithias, L., Rathjens, H., 2017. Featured Series Conclusion. SWAT applications for emerging hydrologic and water quality challenges. JAWRA Journal of the American Water Resources Association.

Ministry of Ecology and Environment of the PRC, 1996. Integrated wastewater discharge standard (in Chinese) (GB8978-96). Ministry of Ecology and Environment of the PRC, Beijing, China. Accessed.

Ministry of Ecology and Environment of the PRC, 2002. Environmental Quality standards for surface water (in Chinese) (GB3838-2002). Ministry of Ecology and Environment of the People's Republic of China, Beijing, China. Accessed.

Ministry of Ecology and Environment of the PRC, 2015. Action Plan for Prevention and Control of Water Pollution (in Chinese). http://www.gov.cn/zhengce/content/2015-04/16/content_9613.htm. Accessed.

Ministry of Water Resources of the PRC, 2013. Technical regulations for hydrologic network design (in Chinese) (SL 34-2013). China Water & Power Press, Beijing, China. Accessed.

Mittelstet, A.R., Storm, D.E., White, M.J., 2016. Using SWAT to enhance watershed-based plans to meet numeric water quality standards. Sustainability of Water Quality and Ecology 7, 5–21.

Molina-Navarro, E., Andersen, H.E., Nielsen, A., Thodsen, H., Trolle, D., 2017. The impact of the objective function in multi-site and multi-variable calibration of the SWAT model. Environmental Modelling & Software 93, 255–267.

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.

Moriasi, D.N., Wilson, B.N., Douglas-Mankin, K.R., Arnold, J.G., Gowda, P.H., 2012. Hydrologic and water quality models: use, calibration, and validation. American Society of Agricultural and Biological Engineers 55 (4), 1241–1247.

Moriasi, D.N., Gowda, P.H., Arnold, J.G., Mulla, D.J., Ale, S., Steiner, J.L., 2013. Modeling the impact of nitrogen fertilizer application and tile drain configuration on nitrate leaching using SWAT. Agricultural Water Management 130, 36–43.

Moriasi, D.N., Gitau, M.W., Pai, N., Daggupati, P., 2015a. Hydrologic and Water Quality Models. Performance Measures and Evaluation Criteria. Trans. ASABE 58 (6), 1763–1785.

Moriasi, D.N., Zeckoski, R.W., Arnold, J.G., Baffaut, C.B., Malone, R.W., Daggupati, P., Guzman, J.A., Saraswat, D., Yuan, Y., Wilson, B.W., Shirmohammadi, A., Douglas-Mankin, K.R., 2015b. Hydrologic and Water Quality Models. Key Calibration and Validation Topics. Trans. ASABE 58 (6), 1609–1618.

Nair, S.S., King, K.W., Witter, J.D., Sohngen, B.L., Fausey, N.R., 2011. Importance of Crop Yield in Calibrating Watershed Water Quality Simulation Tools1. JAWRA Journal of the American Water Resources Association 47 (6), 1285–1297.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology 10 (3), 282–290.

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2011. Soil and Water Assessment Tool, Theoretical Documentation. Version 2009. Accessed.

Niraula, R., Kalin, L., Wang, R., Srivastava, P., 2011. Determining Nutrient and Sediment Critical Source Areas with SWAT. Effect of Lumped Calibration. Transactions of the ASABE 55 (1), 137–147.

Ongley, E.D., Xiaolan, Z., Tao, Y., 2010. Current status of agricultural and rural non-point source Pollution assessment in China. Environmental pollution (Barking, Essex: 1987) 158 (5), 1159–1168.

Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., Michel, C., 2006. Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. Water Resour. Res. 42 (7), 1467.

Ouyang, W., Wang, X., Hao, F., Srinivasan, R., 2009. Temporal-spatial dynamics of vegetation variation on non-point source nutrient pollution. Ecological Modelling 220 (20), 2702–2713.

Pfannerstill, M., Bieger, K., Guse, B., Bosch, D.D., Fohrer, N., Arnold, J.G., 2017. How to Constrain Multi-Objective Calibrations of the SWAT Model Using Water Balance Components. JAWRA Journal of the American Water Resources Association 53 (3), 532–546.

Pfannerstill, M., Guse, B., Fohrer, N., 2014. Smart low flow signature metrics for an improved overall performance evaluation of hydrological models. Journal of Hydrology 510, 447–458. http://www.sciencedirect.com/science/article/pii/S0022169413009414.

Piniewski, M., Marcinkowski, P., Kardel, I., Giełczewski, M., Izydorczyk, K., Frątczak, W., 2015. Spatial quantification of non-point source pollution in a meso-scale catchment for an assessment of buffer zones efficiency. Water 7 (12), 1889–1920.

Pushpalatha, R., Perrin, C., Le Moine, N., Andréassian, V., 2012. A review of efficiency criteria suitable for evaluating low-flow simulations. Journal of Hydrology 420-421, 171–182.

Qi, C., 2013. Investigation and Analysis of Farmers' Fertilization Status on Rice. Taking Jiangxi Province and Jiangsu Province as Examples (in Chinese). Master Thesis, Nanjing Agricultural University, Jiangsu Prov., China.

Reed, P., Minsker, B.S., Goldberg, D.E., 2003. Simplifying multiobjective optimization: An automated design methodology for the nondominated sorted genetic algorithm-II. Water Resour. Res. 39 (7), 7.

Refsgaard, J.C., Storm, B., 1995. MIKE SHE. In: Singh, V., P. (Editor), Computer Models of Watershed Hydrology. Water Resources Publications, Englewood, USA, pp. 809–846.

Rose, C., Parker, A., Jefferson, B., Cartmell, E., 2015. The Characterization of Feces and Urine. A Review of the Literature to Inform Advanced Treatment Technology. Critical reviews in environmental science and technology 45 (17), 1827–1879.

Running, S.W., Mu, Q., Zhao, M., Moreno, A., 2018. User's Guide: MOIDS Global Terrestrial Evapotranspiration (ET) Product (NASA MOD16A2/A3). NASA Earth Observing System MODIS Land Algorithm. http://files.ntsg.umt.edu/data/NTSG_Products/MOD16/. Accessed.

Schaffer, J., 1984. Some experiments in machine learning using vector evaluated genetic algorithms. PhD Thesis, Vanderbilt University, Nashville, USA.

Schwefel, H.-P., 1995. Evolution and Optimum Seeking. Wiley, New York, Chichester.

Seibert, J., McDonnell, J.J. 2002. The quest for an improved dialog between modeler and experimentalist, Duan, Gupta et al. (Ed.) 2002 – Calibration of Watershed Models.

Shafii, M., De Smedt, F., 2009. Multi-objective calibration of a distributed hydrological model (WetSpa) using a genetic algorithm. Hydrol. Earth Syst. Sci. 13 (11), 2137–2149.

Sharpley, A.N., Williams, J.R., 1990. EPIC - Erosion/Productivity Impact Calculator: 1. Model Documentation. https://agrilifecdn.tamu.edu/epicapex/files/2015/05/EpicModelDocumentation.pdf. Accessed.

Shen, Z., Liao, Q., Hong, Q., Gong, Y., 2012. An overview of research on agricultural non-point source pollution modelling in China. Technology for Sustainable Water Environment 84, 104–111.

Shrestha, M.K., Recknagel, F., Frizenschaf, J., Meyer, W., 2016. Assessing SWAT models based on single and multi-site calibration for the simulation of flow and nutrient loads in the semi-arid Onkaparinga catchment in South Australia. Agricultural Water Management 175, 61–71.

Simunek, J., van Genuchten, M.T., Sejna, M., 2012. HYDRUS: Model Use, Calibration, and Validation. Trans. ASABE 55 (4), 1263–1276.

Singh, J., Knapp, H.V., Arnold, J.G., Demissie, M., 2005. Hydrological modeling of the Iroquois River Watershed using HSPF and SWAT. J Am Water Resources Assoc 41 (2), 343–360.

Sloan, P.G., Moore, I.D., 1984. Modeling subsurface stormflow on steeply sloping forested watersheds. Water Resources Research 20 (12), 1815–1822.

Sorooshian, S., Gupta, V.K., 1995. Model calibration. In: Singh, V., P. (Editor), Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, Colorado, USA, pp. 23–67.

Tang, Y., Reed, P., Wagener, T., 2006. How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration? Hydrol. Earth Syst. Sci. 10 (2), 289–307.

Tian, Y., Wang, H., Zhang, X., Jin, Y., 2017. Effectiveness and efficiency of non-dominated sorting for evolutionary multi- and many-objective optimization. Complex & Intelligent Systems 3 (4), 247–263. https://doi.org/10.1007/s40747-017-0057-5.

Tran, V.-B., Ishidaira, H., Nakamura, T., Do, T.-N., Nishida, K., 2017. Estimation of Nitrogen Load with Multi-Pollution Sources Using the SWAT Model: a Case Study in the Cau River Basin in Northern Vietnam. J. of Wat. & Envir. Tech. 15 (3), 106–119.

Tuo, Y., Chiogna, G., Disse, M., 2015. A Multi-Criteria Model Selection Protocol for Practical Applications to Nutrient Transport at the Catchment Scale. Water 7 (6), 2851–2880.

USDA, 1972. National Engineering Handbook. U.S. Government Printing Office, Washington, D.C.

USDA-NRCS, 1993. National Engineering Handbook. U.S. Government Printing Office, Washington, D. C.

USEPA, 2017. Nutrient Pollution. The Problem. https://www.epa.gov/nutrientpollution/problem. Accessed.

Vaché, K.B., McDonnell, J.J., 2006. A process-based rejectionist framework for evaluating catchment runoff model structure. Water Resour. Res. 42 (2), 173.

van Genuchten, M.T., Simunek, J., Leij, F.J., Toride, N., Sejna, M., 2012. STANMOD: Model Use, Calibration, and Validation. Trans. ASABE 55 (4), 1355–1368.

van Griensven, A., Maskey, S., Stefanova, A. (Eds.), 2012. The use of satellite images for evaluating a SWAT model: Application on the Vit Basin, Bulgaria.

van Werkhoven, K., Wagener, T., Reed, P., Tang, Y., 2009. Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. Advances in Water Resources 32 (8), 1154–1169.

Vrugt, J.A., Gupta, H.V., Bastidas, L.A., Bouten, W., Sorooshian, S., 2003. Effective and efficient algorithm for multiobjective optimization of hydrologic models. Water Resour. Res. 39 (8), 19,481.

Whittaker, G., 2016. Non-dominated Sorting Genetic Algorithm-II (NSGA-II) in R. https://www.ars.usda.gov/research/software/. Accessed.

WHO, 2011. Nitrate and nitrite in drinking-water. Background document for development of WHO Guidelines for Drinking-water Quality. World Health Organisation.

Winsemius, H.C., Schaefli, B., Montanari, A., Savenije, H.H.G., 2009. On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information. Water Resour. Res. 45 (12), 198.

Woo, S.Y., Jung, C.G., Lee, J.W., Kim, S.J., 2019. Evaluation of Watershed Scale Aquatic Ecosystem Health by SWAT Modeling and Random Forest Technique. Sustainability 11 (12), 3397.

Wu, C., 2012. Research on Evaluation and Controlling Measure System of Agriculture Non-Point Source Pollution in Jiangxi Province (in Chinese). Master thesis, China University of Geosciences (Beijing), Beijing.

WWAP, 2015. Water for a sustainable world. https://unesdoc.unesco.org/ark:/48223/pf0000231823. Accessed.

Xie, L., Wang, D., Zheng, J., Lin, P., Tang, J., Yao, Y., Li, Q., 2015. Hygienic survey of garbage and sewage treatment in rural areas of Jiangxi Province, 2012-2013 (in Chinese) 42 (3), 557–559.

Xu, F., Dong, G., Wang, Q., Liu, L., Yu, W., Men, C., Liu, R., 2016a. Impacts of DEM uncertainties on critical source areas identification for non-point source pollution control based on SWAT model. Journal of Hydrology 540, 355–367.

Xu, F., Gao, Y., Dong, W., Hao, Z., Xu, Y., 2016b. Impact of atmospheric nitrogen and phosphorus wet deposition on nitrogen and phosphorus export and associated water quality: a case study of forest watershed in the red soil area, Southern China. Acta Ecologica Sinica 36 (20), 6409–6419.

Yalew, S., van Griensven, A., Kokoszkiewicz. L (Eds.), 2010. Parallel computing of a large scale spatially distributed model using the Soil and Water Assessment Tool (SWAT).

Yan, X., 2008. Study on Present Status of Chemical Fertilizer Application and High Efficient Utilization of Nutrition in China (in Chinese). PhD Dissertation, Chinese Academy of Agricultural Sciences, Beijing, China.

Yang, J., Reichert, P., Abbaspour, K.C., Yang, H., 2007. Hydrological modelling of the Chaohe Basin in China. Statistical model formulation and Bayesian inference. Journal of Hydrology 340 (3-4), 167–182.

Yang, J., Reichert, P., Abbaspour, K.C., Xia, J., Yang, H., 2008. Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. Journal of Hydrology 358 (1-2), 1–23.

Yang, L., Han, J., Xue, J., Zeng, L., Shi, J., Wu, L., Jiang, Y., 2013. Nitrate source apportionment in a subtropical watershed using Bayesian model. The Science of the total environment 463-464, 340–347.

Yang, X., Liu, Q., Fu, G., He, Y., Luo, X., Zheng, Z., 2016. Spatiotemporal patterns and source attribution of nitrogen load in a river basin with complex pollution sources. Water research 94, 187–199.

Yen, H., Bailey, R.T., Arabi, M., Ahmadi, M., White, M.J., Arnold, J.G., 2014. The role of interior watershed processes in improving parameter estimation and performance of watershed models. Journal of Environment Quality 43 (5), 1601–1613.

Yi, F., 2013. Estimation and Spatial Distribution of the Potential Load of Livestock and Poultry Feces in Meijiang Watershed (in Chinese). Master, Jiangxi Normal University, Jiangxi Prov. China.

Yilmaz, K.K., Gupta, H.V., Wagener, T., 2008. A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. Water Resour. Res. 44 (9), 103.

Zhang, X., Li, H., Xiao, Y., Chen, Y., 2011a. Study on Allocation Methods of the Water Pollutant Loads in Ungauged Basin (in Chinese). Journal of China Institute of Water Resources and Hydropower Research 9 (2), 136–142.

Zhang, X., Srinivasan, R., Arnold, J., Izaurralde, R.C., Bosch, D., 2011b. Simultaneous calibration of surface flow and baseflow simulations: a revisit of the SWAT model calibration framework. Hydrol. Process. 25 (14), 2313–2320.

Zhang, X., Srinivasan, R., van Liew, M., 2008. Multi-Site Calibration of the SWAT Model for Hydrologic Modeling. Transactions of the ASABE 51 (6), 2039–2049.

Zhang, Y., Shao, Q., Zhang, S., Zhai, X., She, D., 2016. Multi-metric calibration of hydrological model to capture overall flow regimes. Journal of Hydrology 539, 525–538.

Zitzler, E., Thiele, L., 1999. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. IEEE Trans. Evol. Computat. 3 (4), 257–271.

Zitzler, E., Laumanns, M., Thiele, L., 2001. SPEA2: Improving the Strength Pareto Evolutionary Algorithm. TIK-Report 103. Swiss Federal Institute of Technology, Zurich, Switzerland.

Appendix A Technical documentation

1. Hardware and software applied

Table A-1 lists the software applied and the system configured in the research.

Table A-1 Overview of model version, system configuration and software applied

Category		Description
Model	ArcSWAT2012	GUI: ArcGIS 10.4.1 Composition: SWAT Editor 2012
	SWAT-CUP2012	Version 5.1.6
System	Windows 10 Pro	Processor: Intel(R) Core(TM) i7-7700k, CPU 4.20 GHz Installed memory (RAM): 16.0GB System type: 64-bit Operation System, x64- based processor
Data processing	R-Studio	Version R-3.4.4 [32-bit]; R-Studio 1.2.5019
	ArcGIS	Version 10.4.1
	Python	Python 3.7 [32-bit]

2. R packages applied

Table A-2 lists the R packages in the data processing procedure, including the processing of the database, spatial data, data visualization, etc..

Table A-2 The applied R packages

Category	Packages
Time-series	dplyr; plyr
Spatial data	grid; raster; rgdal; gstat; sp
Database	RODBC; rlist; foreign
Data visualization	ggplot2; tmap; mapview
Hydrologic statistics	hydroGOF

3. The structure of the NSGA-II in R

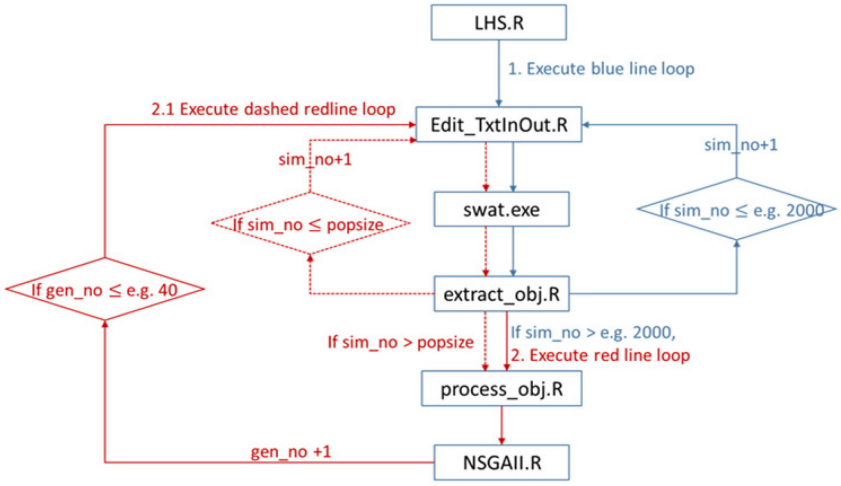

Figure A-1 The structure of the R-script to couple NSGA-II and SWAT model in exe_SWAT.R

Figure 2-1 displays the structure of the R script to couple the NSGA-II and the SWAT model. The R script was adapted from Whittaker (2016) and Ercan and Goodall (2016). The workflow is listed below and it can be executed by exe_SWAT.R.

- Launch LHS.R to generate the parameter sets for the initial iteration, the size of the parameter sets was 2000 in the research for the hydrological calibration;
- Execute the blue line loop:
    a. edit_TxtInOut.R to update all the parameters of each HRU by the values of one parameter set;
    b. launch swat.exe to execute the simulation;
    c. extract_obj.R to extract and save the results from the simulation;
    d. if the number of simulation (sim_no) is lower than 2000, continue the blue line loop with the next parameter set;
- if sim_no>2000, execute the red line loop:
    a. process_obj.R to process the simulation results from the blue line loop to obtain the evaluation statistics value for each objective within each simulation for the entire iteration or generation;
    b. launch NSGAII.R to execute the NSGA-II process:

- continue the dashed red line loop: Edit_TxtInOut, swat.exe and extract_obj.R until the number of simulations reaching the population size (popsize);

c. Continue the red line loop to execute the next generation until a certain number of generations (gen_no) is executed, e.g. gen_no=40 in the study.

The functions applied in the R scrips are listed in Table A-3.

Table A-3 Description of main functions applied in the R scripts

Function	Description
Process_obj.R	
multi_site()	Processing the simulated discharge to obtain the NSE, NSE_in, and PBIAS
multi_metric()	Aggregate the simulated monthly multi-metrics into the three sub-catchment clusters, including AET, SURQ, WYLD, GW_Q, LATQ
obj_output()	Merge the output of all objectives of a generation
NSGAII.R	
nsga2()	Obtain the Pareto Front and rank all individuals (simulations) in a generation
dominance()	Evaluate the dominance of two individuals
crowding()	Calculate the crowding distance for each individual
reverse_decode()	Convert the real value of a parameter to binary values
createMatePopFromNewandOldPops()	Create a mate population from new and old population
selection()	Select a certain number of individuals from a population
unicrossover()	The crossover between two populations at the binary level
mutation()	Mutate within a population at the binary level
decode()	Covert the parameter from binary values to the real value
Edit_TxtInOut.R	
edit_hru()	To edit files with filename extension hru, sol, rte, and mgt to update the parameter values for each HRU in folder TxtInOut
Extract_obj.R	
extract_rch()	Extract simulation results from output.rch
extract_sub()	Extract simualtion results from output.sub
LHS.R	Obtain the initial parameter sets via Latin Hypercube Sampling

The attributes of an individual in a population are listed in Table A-4.

Table A-4 The attributes of an individual

Attributes of an individual	Suffix in the code	Description
Gene	_pop	The binary values of each parameter
Parameter values	_pop_xbin	The real parameter value
Crowding distance	_pop_crowd	Crowding distance in the population
Evaluation statistics	_pop_objF	Objective function value of each objective
The rank of the performance	_pop_rank	The rank of dominance in a population

The steps of the elitism maintained by the NSGA-II among the crossover, selection, and mutation among the old population (old_pop), mate population (mate_pop), and new population (new_pop) in a generation are demonstrated in Figure A-2.

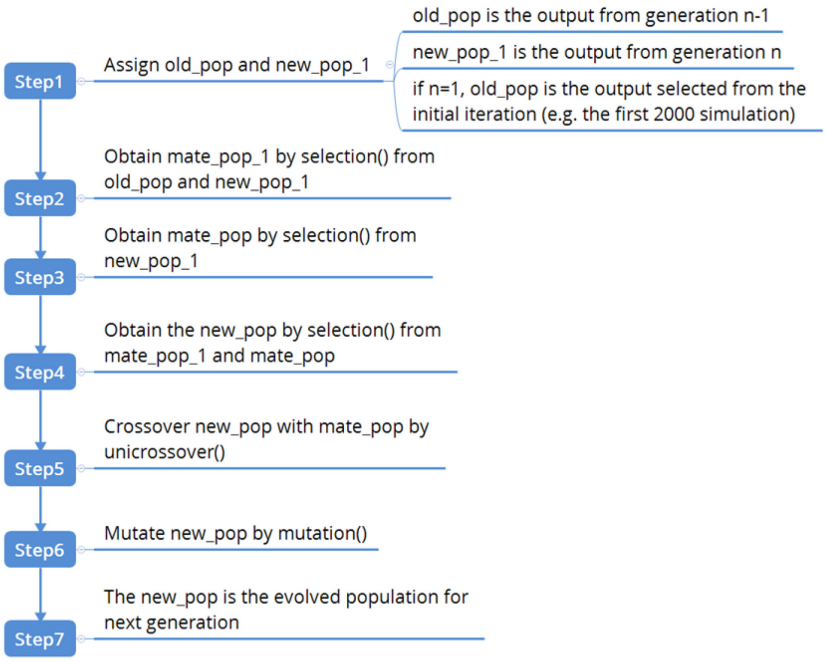

Figure A-2 The steps of maintaining the elitism implemented by NSGA-II

4. The SWAT model input and output data overview

Figure A-3 displays the input data of the time-series and spatial data prepared for the YRC. The output data extracted for analysis included discharge and total nitrogen content in the stream in time-series, and the total nitrogen load from each HRU and each subcatchment.

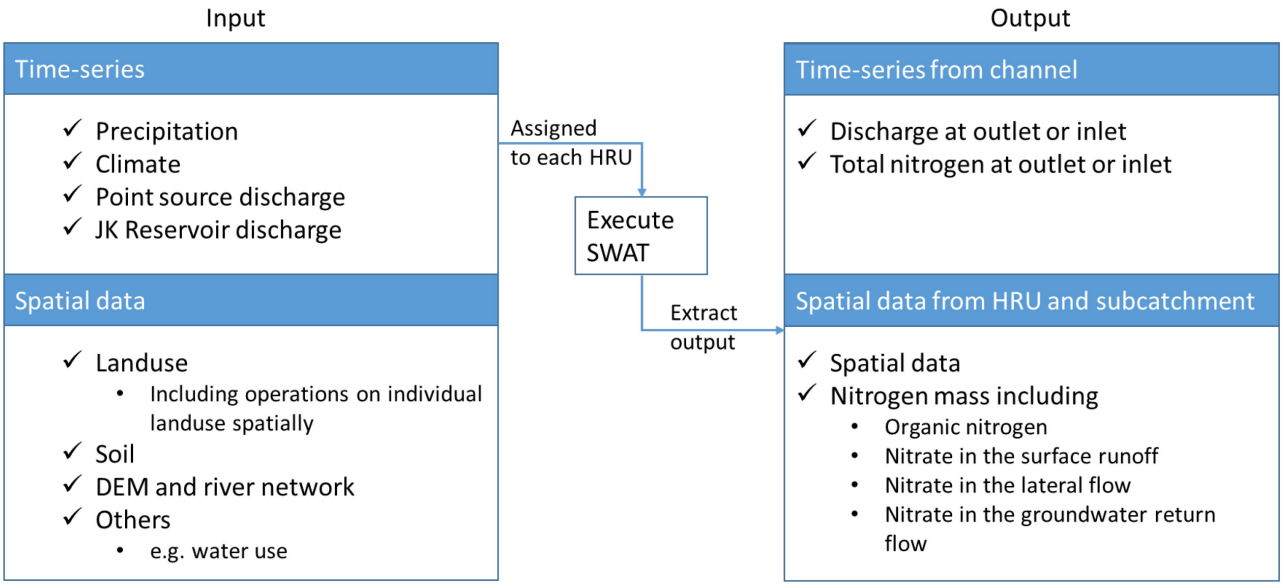

Figure A-3 The diagram displaying the input and output data of the SWAT model for the YRC

5. Overview of the digital appendix*

The input and output database of the SWAT model is listed in folder /SWAT_model_package/ and content is displayed in Table A-5.

Table A-5 SWAT model input and output database

Folder or filename	Description
Arcyuan0508.mdb	An example of a database storing the input data for the SWAT model
/TxtInOut/	The folder stored all the files, including the data of each HRU to be read or write when the SWAT.exe is executing
/Calibrated_results/	Database of the SWAT.exe output, including:
ED_hydrology.mdb	The calibrated hydrological model output by applying ED
NSGA_hydrology.mdb	The calibrated hydrological model output by applying NSGA-II
Single_hydrology.mdb	The calibrated hydrological model output by applying SUFI-2
ED_TN.mdb	The calibrated TN model output by applying ED
NSGA_TN.mdb	The calibrated TN model output by applying NSGA-II
Single_TN.mdb	The calibrated TN model output by applying SUFI-2

**The digital appendix can be requested by contacting fg-hydrologie@b-tu.de.*

The R scripts applied in Figure A-1 and the folder structure is displayed in Table A-6 and they are stored in folder /R/.

Table A-6 The folder structure of the R script

Folder name	Description
/R/	R script of Edit_TxtInOut.R, exe_SWAT.R, extract_obj.R, LHS.R, NSGAII.R, NSGAII_function.R, and process_obj.R
/R/observed/	The observed data
/R/parmaeter/	Parameter information
/R/sim_result/	The simulation results saved of each generation
/R/TxtInOut/	Data of each HRU; the folder is copied from the SWAT project

Appendix B Data of the Yuan River Catchment

Model Input data are listed in Table B-1 and stored in the folder /Model_input/. Due to the agreement with the data providers (see Chapter 3 Table 3.2), the data are not available to the public*.

Table B-1 The list of the model input data in folder Model_input

Data	Folder (/Model_input/)
Hydrology	
DEM	/DEM/
Yuan River network	/Stream_network/
Landuse	/Landuse/
Soil type and properties	/Soil/
Climate data	/Climate_input/
JK Reservoir	/Reservior/
Industry and domestic water use	/Wateruse/
Daily discharge at JK, MZ, and LX	/Discharge/
Nitrogen	
Operation at individual landuse	/Fertilizer/
Point source discharge	/Pointsource/
Monitored total nitrogen concentration at 1_Y, 4_F, and 6_L	/WQ/

**If interested, the reader can contact fg-hydrologie@b-tu.de for more information.*

DOI: https://doi.org/10.26127/BTUOpen-5233