Combining Ecohydrological Catchment Modeling and Water Quality Monitoring Data to Assess Nitrogen Pollution in the Swist River Basin, Germany

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Alexander Ahring

aus

Ostercappeln, Deutschland

Bonn, Juli 2024

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

Gutachter/Betreuer:	Prof. Dr. Bernd Diekkrüger
Gutachter:	Prof. Dr. Thomas Kistemann
Tag der Promotion:	04.11.2024
Erscheinungsjahr:	2024

Acknowledgments

Now that the end of this long and winding road finally is in clear sight, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Bernd Diekkrüger, who supported me on this journey well into his retirement, and despite some major headwind along the way. I always appreciated that he granted me a lot of freedom in pursuing my research, while at the same time being always approachable and taking his time when I needed his advice. As I know this thesis took a lot more time to complete than initially expected (or agreed upon), I would also like to thank him for his patience. I am additionally grateful towards my second supervisor, Prof. Dr. Thomas Kistemann, for many years of fruitful collaboration on numerous research projects between the Geohealth Centre in Bonn and the Erftverband. I was very privileged to participate in some of these exciting endeavours. Last but not least, I would also like to thank Prof. Dr. Jürgen Herget and Prof. Dr. Mathias Becker for kindly accepting to be part of the doctoral committee.

I am deeply thankful towards Dr. Ekkehard Christoffels, who was my first direct superior when I started working at the Erftverband and gave the very first impetus for this thesis. His decade-long work on water quality in the Erft catchment laid the foundation for my Ph.D. project, which would never have been possible without the extensive monitoring infrastructure he was instrumental in establishing. I am also indebted to my other direct superiors over the years, Dr. Christian Gattke, Dr. Daniel Bittner and Dr. Tilo Keller, who always supported and encouraged me to pursue a doctoral degree while working (first part-time, then full-time) at the Erftverband.

I would like to thank the various former members of the “old” water quality team that I was a part of during my early years at the Erftverband before it slowly but steadily dissolved. These years were incredibly formative not only for this thesis, but for my entire professional life. I probably learned the most from Dr. Andrea Brunsch, Dr. Franz Michael Mertens and Robert Krump. I will always have fond memories of our time together. Furthermore, I would like to thank Markus Rosellen from the Erftverband IT department, for setting up and maintaining the SVN repositories that helped me tremendously with organizing the work on this thesis. I also thank Dr. Stephan Lenk for providing valuable insights into the hydrogeology of the Swist catchment. There are many more Erftverband colleagues who have helped me in many ways - large and small - to complete this thesis. Unforturnately, they are too many to name them all in this short section, but I hope they know I think of them should they ever read this. I always enjoyed working with so many skilled and knowledgeable people in such a pleasant social environment, and I hope most of them will stay around for some more time to come.

I am extremely grateful to the members of the Hydrological Research Groups of the Department of Geography in Bonn for adopting me as their "Fridays Ph.D. student". Of course, pursuing a doctoral degree only on Fridays quickly turned out to be completely unfeasible, but I soon looked forward to coming to Bonn more often and enjoying the pleasant atmosphere I found there. For this, I would particularly like to thank Dr. Claudia Schepp, Dr. Mouhamed Idrissou, Dr. Kristian Näschen, Katharina Höreth and Emmanuel Nkundimana. I would also like to thank Prof. Dr. Julian Klaus for his various acts of support, especially for providing me with a desk in his offices in Bonn during the last third of my Ph.D. project. I would furthermore like to thank Dr. Clarissa Glaser for her valuable feedback on the introduction section of this thesis.

Coming to the end, I am deeply grateful for the love and support of my family and friends. Thank you for the privilege of knowing there will always be a place I can return to. Special mentions go to my brother Andreas Ahring, for accommodating me every once in a while in his small apartment in Bonn when I spent long evenings or weekends working on this thesis, and to Dr. Johannes Klassen, for his years-long friendship since school time and his emotional support from afar to finally "get this done".

- Bonn, July 1, 2024

Abstract

Legislation like the EU Water Framework Directive demands the development as well as implementation of catchment-scale monitoring and management plans to control pollution of surface waters and ensure their good status. Water pollution with excess amounts of reactive nitrogen (N) - for example nitrate (NO_3^-) - can cause environmental problems like eutrophication while also posing potential health risks to humans. Knowledge of pollution sources and their relative importance in a given catchment is a prerequisite for developing targeted monitoring strategies and efficient management plans. The process of quantifying the contribution of each pollution source to overall pollution in a catchment is called *source apportionment*. Due to the abundance of pollution sources as well as their heterogeneity in time and space, source apportionment solely based on water quality monitoring is not feasible in large catchments, especially with regard to non-point source (diffuse) pollution. Ecohydrological computer models can be used to complement water quality monitoring data in space and time, identify critical source areas and carry out source apportionment studies. The overall research aim of the present thesis is to assess N pollution of surface waters in the Swist river basin, Germany, from 2012 to 2018 using the ecohydrological catchment model SWAT (Soil and Water Assessment Tool).

Calibration of SWAT using Monte Carlo analysis generated a SWAT model ensemble allowing for analysis of modeling uncertainty. Furthermore, two independent reference data sets on diffuse N pollution in the Swist basin were available for comparison with the SWAT simulation results: (1) An emission inventory from upscaled *edge-of-field* monitoring data and (2) raster data generated by the AGRUM modeling system. The calibrated SWAT models predicted a mean diffuse N emission load of 8.1 kg/(ha a) for the entire Swist catchment area in the study period (median: 6.9 kg/(ha a)). Agricultural tile drainage was identified as the most important emission pathway in the catchment, shaping much of the spatial variability in diffuse N pollution. In contrast, due to missing groundwater contact in large parts of the catchment (a consequence of lignite mining north of the study area), surface waters received only minimal N from groundwater, possibly indicating a dominance of faster (tile drains) over slower (groundwater) emission pathways in the catchment. The highest diffuse N loads were predicted near the headwaters of the Schiessbach tributary by all SWAT models, making this region a likely critical source area for N in the catchment. On average, wastewater treatment plants contributed approximately one fourth (26.8 %) of overall N pollution in the simulations. As all these findings were basically consistent between the SWAT simulations and the two reference data sets, they can be used with relative confidence in the future to focus monitoring efforts, devise emission control strategies and implement effective mitigation measures.

However, comparison of the available data also revealed some meaningful discrepancies, highlighting the remaining uncertainties in the results. Diffuse N emission loads simulated by the SWAT ensemble ranged from 4.7 to 11.4 kg/(ha a) (excluding one outlier at 15.9 kg/(ha a)). Since the two reference data sets mostly agree with SWAT in the order of magnitude of these values, the results are considered a realistic appraisal of diffuse N pollution in the Swist catchment. Still, with a factor of more than two (or three when including the outlier) between the upper and lower end of the uncertainty interval, the exact amount of N released to the Swist and its tributaries remains difficult to quantify. Although the general ranking of the individual pathways is mostly stable between the SWAT models and likewise confirmed by the reference data, the relative uncertainty intervals associated with their N contributions are even wider than for the overall emission loads. Apart from the Schiessbach headwaters, no other obvious critical source area candidate emerged among the SWAT simulation results and the reference data, with regions of high N emission loads mostly fluctuating between the different model results. The simulated contributions from wastewater treatment plants ranged from 14.7 to 36.7 % of overall N emissions in the catchment.

For a future continuation of SWAT modeling in the Swist catchment to assess N pollution, there are several possibilities to potentially reduce the uncertainty in the simulation results. First, including additional data and/or objective functions in model calibration probably helps to better constrain the SWAT parameters. Second, the correction of some deficiencies in the model setup (i.e. input data and model structure) - for example in the amounts of applied N fertilizer or the omission of combined sewer overflows - makes model calibration presumably more efficient and may eliminate potential bias in the simulation results. In summary, ecohydrological catchment modeling with SWAT in the present thesis was successful in generating novel insights regarding N pollution in the Swist catchment. Here, model-based source apportionment benefited immensely from the extraordinary wealth of monitoring data available for the Swist catchment and the possibility to compare the SWAT modeling results to two independent reference data sets for the study area. This underscores the importance of comprehensive monitoring and knowledge of relevant local conditions to reach a sound understanding of surface water pollution on the catchment scale.

Zusammenfassung

Rechtliche Vorgaben wie die EU-Wasserrahmenrichtlinie verlangen die Entwicklung und Umsetzung von Überwachungs- und Bewirtschaftungsplänen auf Einzugsgebietsebene, um Schadstoffeinträge in Oberflächengewässer zu kontrollieren und den guten Gewässerzustand zu gewährleisten. Gewässerbelastungen durch reaktiven Stickstoff (N) - zum Beispiel in Form von Nitrat (NO_3^-) - können Umweltprobleme wie Eutrophierung verursachen und stellen zudem ein potenzielles Gesundheitsrisiko für den Menschen dar. Die Kenntnis der Belastungsquellen und ihrer relativen Bedeutung in einem bestimmten Gewässereinzugsgebiet ist eine Voraussetzung für die Entwicklung gezielter Monitoring-Strategien sowie effizienter Bewirtschaftungspläne. Der Prozess der Quantifizierung des Beitrags einzelner Belastungsquellen zur Gesamtbelastung eines Gewässers wird als *Source Apportionment* bezeichnet. Aufgrund der Vielzahl möglicher Belastungsquellen sowie ihrer zeitlichen und räumlichen Heterogenität ist ein Source Apportionment allein aufgrund von Monitoring-Maßnahmen in großen Einzugsgebieten nicht durchführbar, insbesondere im Hinblick auf diffuse Gewässerbelastungen. Ökohydrologische Computermodelle können hier verwendet werden, um Monitoring-Daten in Raum und Zeit zu ergänzen, wichtige Quellgebiete (*critical source areas*) zu identifizieren und ein Source Apportionment durchzuführen. Das übergeordnete Forschungsziel dieser Arbeit ist es, die N-Belastung der Oberflächengewässer im Einzugsgebiet der Swist (Nordrhein-Westfalen) in den Jahren zwischen 2012 und 2018 mit Hilfe des ökohydrologischen Einzugsgebietsmodells SWAT (Soil and Water Assessment Tool) zu untersuchen.

Die Modellkalibrierung per Monte-Carlo-Analyse lieferte ein SWAT-Modell-Ensemble, das zur Analyse der Modellunsicherheit genutzt werden kann. Daneben standen zwei unabhängige Referenzdatensätze zur diffusen N-Belastung im Swist-Einzugsgebiet für Vergleichszwecke zur Verfügung: (1) Ein Emissionsinventar aus hochskalierten *Edge-of-Field*-Monitoring-Daten und (2) gebietsweite Rasterdaten, die mit dem AGRUM-Modellsystem erzeugt wurden. Die kalibrierten SWAT-Modelle sagten eine mittlere diffuse N-Emissionsfracht von 8,1 kg/(ha a) für das gesamte Swist-Einzugsgebiet im Untersuchungszeitraum voraus (Median: 6,9 kg/(ha a)). Als wichtigster Emissionspfad im Einzugsgebiet wurden landwirtschaftliche Felddrainagen identifiziert, die für einen Großteil der räumlichen Variabilität in den diffusen N-Emissionen verantwortlich sind. Aufgrund von oftmals fehlendem Grundwasserkontakt (eine Folge des Braunkohletagebaus nördlich des Untersuchungsgebiets) erhielten die Oberflächengewässer nur wenig N über das Grundwasser, was insgesamt eine Dominanz schnellerer (Felddrainagen) gegenüber langsameren (Grundwasser) Emissionspfaden im Einzugsgebiet nahelegt. Die höchsten diffusen N-Frachten wurden über alle Modellläufe hinweg für den Oberlauf des Schiessbaches vorhergesagt, was darauf hindeutet, dass sich in diesem Bereich höchstwahrtscheinlich ein maßgebliches Quellgebiet für N an der Swist befindet. Die Kläranlagen trugen im Durchschnitt ein Viertel (26,8 %) zur ge-

samten N-Belastung in den Simulationen bei. Da all diese Ergebnisse im Wesentlichen zwischen den SWAT-Simulationen und den beiden Referenzdatensätzen übereinstimmen, können sie diesbzüglich als belastbare Grundlage dienen um Monitoring-Maßnahmen zu fokussieren, Emissionskontrollstrategien zu entwickeln und wirksame Maßnahmen zur Eintragsminderung umzusetzen.

Der Vergleich der Simulationsdaten untereinander sowie mit den verfügbaren Referenzdaten förderte jedoch auch einige bedeutsame Unterschiede zutage, welche die verbleibenden Unsicherheiten in den Ergebnissen aufzeigen. So reichten die diffusen N-Emissionsfrachten aus dem SWAT-Modell-Ensemble von 4,7 bis 11,4 kg/(ha a) (ein Ausreißer von 15,9 kg/(ha a) ausgenommen). Da die beiden Referenzdatensätze in der Größenordnung hier weitgehend mit SWAT übereinstimmen, wird dieser Wertebereich als realistische Einschätzung der diffusen N-Gewässerbelastung im Swist-Einzugsgebiet angesehen. Mit einem Faktor von mehr als zwei (bzw. drei mit Ausreißer) zwischen der oberen und der unteren Grenze dieses Unsicherheitsintervalls bleibt es jedoch schwer die genauen N-Eintragsmengen in die Swist und ihre Nebenflüsse zu quantifizieren. Obwohl die Reihenfolge der einzelnen Eintragspfade zwischen den SWAT-Modellen weitgehend stabil ist und auch durch die Referenzdaten bestätigt wird, sind die relativen Unsicherheitsbereiche für die pfadspezifischen N-Emissionsfrachten noch breiter als für die Gesamtfracht. Abgesehen vom Oberlauf des Schiessbachs ging unter den SWAT-Modellen kein weiterer klarer Kandidat für ein kritisches Quellgebiet aus den SWAT-Simulationsergebnissen oder den Referenzdaten hervor: Die Lage von Belastungsschwerpunkten im Untersuchungsgebiet schwankte zumeist stark von einem Modelldurchlauf zum nächsten. Bezüglich der Belastungsbeiträge aus Kläranlagen reichte der Unsicherheitsbereich aus dem Modell-Ensemble von 14,7 bis 36,7 % aller N-Emissionen im Einzugsgebiet.

Für eine künftige Fortführung der SWAT-Modellierung im Swist-Einzugsgebiet zur Untersuchung der N-Gewässereinträge ergeben sich mehrere Ansätze, um die Unsicherheit der Simulationsergebnisse zu verringern. Erstens würde die Einbeziehung zusätzlicher Daten und/oder Zielfunktionen in die Modellkalibrierung voraussichtlich helfen, die Schwankungsbreiten der unsicheren SWAT-Parameter besser einzugrenzen. Zweitens kann die Korrektur einiger Unzulänglichkeiten in den Modelleingangsdaten (z.B. die Menge ausgebrachten Stickstoffdüngers im Untersuchungsgebiet) sowie in der Modellstruktur (wie die Nicht-Berücksichtigung von Mischwasserabschlägen) helfen, die Modellkalibrierung effizienter zu gestalten und etwaige systematische Fehler in den Simulationsergebnissen zu beseitigen. Zusammenfassend lässt sich sagen, dass die Einzugsgebietsmodellierung mit SWAT in dieser Arbeit erfolgreich war, neue Erkenntnisse über die N-Belastung im Einzugsgebiet der Swist zu gewinnen. Dabei hat das modellbasierte Source Apportionment von der außerordentlichen Fülle an Messdaten im Swist-Einzugsgebiet sowie der Möglichkeit, die Modellergebnisse mit unabhängigen Referenzdaten zu vergleichen, profitiert. Dies unterstreicht die Bedeutung eines umfassenden Gewässermonitorings und der genauen Kenntnis wichtiger Gebietseigenschaften, um ein Verständnis von Gewässerbelastungen auf Einzugsgebietsebene zu erhalten.

Contents

List of Figures

List of Tables

List of Abbreviations

a	Year
AAT	All-at-a-time
AGRUM	Agrar- und Umweltmaßnahmen (Agro-Environmental Measures)
ARS	USDA Agricultural Research Service
ATKIS	Amtliches Topographisch-Kartographisches Informationssystem (Authorative Topographic-Cartographic Information System)
Basis-DLM	Digitales Basis-Landschaftsmodell (Digital Basic-Landscape Model)
BFD50	Bodenflächendaten für Rheinland-Pfalz im Maßstab 1:50000 (Authorative Soil Map Data for Rhineland-Palatinate in the Scale of 1:50 000)
BK50	Bodenkarte von Nordrhein-Westfalen im Maßstab 1:50 000 (Authorative Soil Map of North Rhine-Westphalia in the Scale of 1:50 000)
CAN	Calcium ammonium nitrate
CLC	CORINE Land Cover
CN	Curve Number
CORINE	Coordination of Information on the Environment
CSA	Critical Source Area
CSO	Combined Sewer Overflow
CWA	US Clean Water Act
DEM	Digital Elevation Model
DENUZ	Denitrifikation in der ungesättigten Zone (Denitrification in the Vadose Zone)
DHI	Dansk Hydraulisk Institut (Danish Hydraulic Institute)
DWD	Deutscher Wetterdienst (German Meteorological Service)
EE	Elementary Effect

EOF	Edge-of-Field
ESRI	Environmental Systems Research Institute
ET	Evapotranspiration
EU	European Union
FAO	United Nations Food and Agriculture Organization
GIS	Geographic Information System
H_2O	Molecular formular for water
HAD	Hydrologischer Atlas von Deutschland (Hydrological Atlas of Germany)
HRU	Hydrologic Response Unit
IGB	Leibniz Institut für Gewässerökologie und Binnenfischerei (Leibniz Institute of Freshwater Ecology and Inland Fisheries)
KA	Bodenkundliche Kartieranleitung (German Pedological Mapping Guidelines)
KDE	Kernel Density Estimation
KGE	Kling-Gupta-Efficiency
KIT	Karlsruher Institut für Technologie (Karlsruhe Institute of Technology)
kNN	k-Nearest Neighbor
LAI	Leaf Area Index
LANUV	Landesamt für Natur, Umwelt und Verbraucherschutz NRW (State Office for Nature, Environment and Consumer Protection NRW)
LBM-DE	Digitales Landbedeckungsmodell für Deutschland (Digital Land Cover Model Germany)
LHS	Latin Hypercube Sampling
LOQ	Limit of Quantification
LSU	Landscape Unit
LU	Land Use
LWK	Landwirtschaftskammer (Chamber of Agriculture)
m.a.s.l.	meters above sea level
MAE	Mean Absolute Error
MC	Monte Carlo
MEPhos	Modell zur Ermittlung des Phosphoreintrags (Model for the Determination of Phosphorus Inputs)

mGROWA ..... Monatliches Großräumiges Wasserhaushaltsmodell (Monthly Large-Scale Water Balance Model)
MONERIS ..... Modelling of Nutrient Emissions in River Systems
MoRE ..... Modeling of Regionalized Emissions
MSE ..... Mean Squared Error
MUSLE ..... Modified Universal Soil Loss Equation
N ..... Elemental symbol of nitrogen
N_2 ..... Molecular formular for nitrogen gas
NAC ..... Trade name of some CAN fertilizer products
NH_3 ..... Molecular formular for ammonia
NH_4^+ ..... Molecular formular for ammonium
NH_4^+-N ..... Ammonium-N
NO_2^- ..... Molecular formular for nitrite
NO_2^--N ..... Nitrite-N
NO_3^- ..... Molecular formular for nitrate
NO_3^--N ..... Nitrate-N
NRW ..... North Rhine-Westphalia
OAT ..... One-at-a-time
P ..... Elemental symbol of phosphorus
PBIAS ..... Percent Bias
PE ..... Population Equivalents
pH ..... pondus Hydrogenii
R ..... R Programming Language for Statistical Computing
R^2 ..... Coefficient of Determination
RAUMIS ..... Regionalisiertes Agrar- und Umweltinformationssystem (Regionalized Agro-Environmental Information System)
RF ..... Random Forest
RLP ..... Rhineland-Palatinate
RMSE ..... Root Mean Squared Error
RSE ..... Residual Standard Error
RSS ..... Residual Sum of Squares
SA ..... Sensitivity Analysis
SAFER ..... Sensitivity Analysis for Everyone, R version
SCS ..... USDA Soil Conservation Service
SD ..... Standard Deviation

SHE	Système Hydrologique Européen (European Hydrological System)
SUFI-2	Sequential Uncertainty Fitting, version 2
SWAT	Soil and Water Assessment Tool
SWAT-CUP	SWAT Calibration Uncertainty Program
TMDL	Total Maximum Daily Load
TNb	Total Nitrogen bound
UBA	Umweltbundesamt (German Federal Environment Agency)
US	United States
USDA	US Department of Agriculture
USEPA	US Environmental Protection Agency
VIF	Variance Inflation Factor
WEKU	Verweilzeiten und reaktiver N-Transport im Grundwasser (Residence Times and Reactive N-Transport in Groundwater)
WFD	EU Water Framework Directive
WWTP	Wastewater Treatment Plant

1. Introduction

1.1. Problem Statement

1.1.1. Surface Water Pollution

Freshwater resources are essential for human and environmental well-being. Most prominently, they provide drinking water, ensure food production, enable various industrial activities and form the basis for a multitude of freshwater ecosystems. On the global scale, freshwater makes up only 2.5 % of the hydrosphere, nearly 70 % of which is perennially frozen (Carpenter et al. 2011). Freshwater habitats such as rivers, lakes and reservoirs contain only 0.01 % of the world's water, yet harbor almost 6 % of estimated global biodiversity. 40 % of all fish species and one third of all vertebrate species live in freshwater or are tied to freshwater habitats (Dudgeon et al. 2006). Anthropogenic pollution of freshwater (i.e. the excessive release and accumulation of adverse physicochemical influences due to human activities) is a major driver of both a worldwide decline in biodiversity (Dudgeon et al. 2006, Reid et al. 2019) and increasing water scarcity (Boretti & Rosa 2019). Notable pollutants include nutrients like nitrogen (N) and phosphorus (P), heavy metals, pathogens, plastics, pesticides and pharmaceuticals (Carpenter et al. 2011, Reid et al. 2019, Walker et al. 2019). Limiting emissions of these pollutants is crucial for protecting aquatic life and securing long-term freshwater availability for human use.

Legislation aimed at controlling water pollution and protecting water quality includes the *Clean Water Act* (CWA) in the United States (US) and the *Water Framework Directive* (WFD) in the European Union (EU). The CWA in its modern form was enacted in 1972 and primarily applies to surface ("navigable") water bodies. The WFD was put into effect in the year of 2000 and seeks to achieve a good status for all surface and groundwater water bodies in the EU member states until 2027 at the latest. Comparative descriptions of the CWA and the WFD are provided by Coffey et al. (2014) as well as McDowell et al. (2016). As water bodies transcend regular administrative units, the WFD establishes *river basin districts* (i.e. catchments) as the primary spatial entities for water management. Originating from the CWA, the *Total Maximum Daily Load* (TMDL) program of the US Environmental Protection Agency (USEPA) similarly encourages a catchment-scale approach to restore water quality in polluted US surface waters (Radcliffe

et al. 2009, Daniel et al. 2011). In order to meet water quality standards, the WFD demands the development and implementation of management plans for all river basin districts. In the US, TMDLs need to be developed for all impaired water bodies (McDowell et al. 2016).

Knowledge of pollution sources and their relative importance in a given river basin is a prerequisite for designing efficient management plans as well as developing targeted monitoring strategies, prioritizing mitigation efforts and implementing the *polluter pays principle* (Kern et al. 2007). Table 1.1 gives an overview of different surface water pollution sources.[1] With respect to their spatial characteristics, they are commonly classified as point and non-point sources. Whereas point sources may be assigned a specific coordinate at which pollutants are released into the water body, emissions from non-point sources occur over a wider area and may enter the water body at multiple locations that are usually unknown. As they cannot be reduced to a single spatial coordinate, non-point sources are often synonymously called "*diffuse*" sources. Considering temporal aspects, pollution sources may further be classified as continuous, episodic and event-specific. Emissions from continuous pollution sources occur steadily over long periods (months, years). They may vary in time, but rarely cease entirely. In contrast, emissions from event-specific sources are a direct response to precipitation events. As such, they are short-lived (minutes, hours) and virtually unpredictable. Episodic pollution sources are also bound to precipitation events, but pollutant release is often delayed and stretched out over longer time periods (days, weeks).

The process of quantifying the contribution of each pollution source to overall pollution in a catchment is called *source apportionment* (Bøgestrand et al. 2005). Due to the abundance of pollution sources as well as their heterogeneity in time and space, monitoring-based source apportionment is not feasible in large-scale catchments (Shirmohammadi et al. 2006). While continuous point sources like the effluents of wastewater treatment plants (WWTP) are easily accessible for sampling, monitoring of other pollution sources is considerably more difficult. Comprehensive data on pollutant quantities emitted by event-specific point sources like combined sewer overflows (CSO) is usually scarce due to the short duration and unpredictability of emission events (Christoffels 2008). The same is true for all event-specific or episodic non-point pollution sources. Establishing and maintaining *edge-of-field* (EOF) monitoring systems for runoff sampling is laborious and costly. Even more importantly, EOF monitoring is typically

[1] There is some ambiguity in the exact meaning of the term 'pollution source" in the literature. A pollution source may either refer to an entire domain of human activity that pollutants originate from (e.g. agriculture, industry, road traffic) or to the specific process or structure that releases pollutants into a water body (erosion, atmospheric deposition, wastewater treatment plants). Imagining an emission pathway that encompasses pollutant mobilization, pollutant transport and pollutant input into a water body (Kern et al. 2007), the first interpretation places the source at the start of this pathway, whereas the second interpretation places it at the end. This thesis primarily makes use of the second interpretation.

Table 1.1.: Overview of different surface water pollution sources classified with respect to their spatial and temporal characteristics (Christoffels et al. 2016, expanded and modified). From left to right and from top to bottom, the listed pollution sources get increasingly inaccessible to source monitoring.

	Temporal characteristics		
Spatial characteristics	Continuous source	Episodic source	Event-specific source
Point source (known coordinates)	Wastewater treatment plants Direct discharge from industry or trade		Combined sewer overflow Separate sewer outlets
Point source (unknown coordinates)		Tile drainage flow	
Non-point source	Groundwater flow Atmospheric deposition	Interflow	Surface runoff Erosion

restricted to individual hillslopes, fields or tile drain outlets[2], but does not provide information for the greater catchment area. Similarly, monitoring sites for continuous non-point pollution sources like groundwater flow and atmospheric deposition only allow for point-scale measurements.

Extrapolating monitoring results from non-point sources to the catchment scale is subject to considerable uncertainties. The density of monitoring networks is usually not sufficient to draw reliable conclusions on overall emission loads. Often, it is not entirely clear how representative the selected monitoring sites are for the complete catchment area (e.g. in terms of agricultural management practices). Furthermore, for many pollutants such as nutrients, the catchment-wide distribution of emission loads per unit-area is highly skewed, meaning that a disproportionate share of non-point source emissions originates from a very limited number of source areas. Such areas are commonly referred to as *critical source areas* (CSA) (White et al. 2009). If overall emissions of a pollutant are dominated by only a few CSAs, projecting the outcomes of point- or hillslope-scale measurements to the entire catchment area bears the risk of introducing substantial bias to the resulting estimates for the total emission load. Indirect estimation of diffuse emission loads from observed in-stream loads relies on comprehensive knowledge of pollution from point sources as well as reliable information on in-stream reten-

[2] In theory, each individual tile drain outlet can be seen as a point source. In practice however, tile drain outlets effectively act as non-point sources, since their number and exact position is usually unknown. Mapping tile drained fields with available data like aerial photographs is a complex task that usually comes with considerable uncertainties (Tetzlaff et al. 2009, Møller et al. 2018, Tetzlaff 2021, Kothe et al. 2021).

tion and transformation, which is unavailable in most cases. To overcome the various limitations of monitoring-based assessment of non-point source pollution in large river basins, often catchment-scale computer models are being used to support the implementation of TMDLs and river basin management plans in the US and Europe, respectively (Borah et al. 2006, Radcliffe et al. 2009, Daniel et al. 2011). These models are able to simulate non-point source pollution and its effect on water resources. Section 1.2 will give an overview of different types of catchment-scale computer models, their application for simulating water quality and water pollution as well as the associated uncertainties and limitations.

1.1.2. Nitrogen in the Environment

As a macronutrient, nitrogen (N) is essential for every living organism on Earth. Despite its abundance in the atmosphere, triple-bonded nitrogen gas (N_2) is chemically inaccessible for most organisms and has to be converted into various forms of *reactive nitrogen* to be available for biomass production. Reactive nitrogen is created when the triple bond of N_2 is broken and the single N atoms are incorporated into other inorganic molecules containing either hydrogen or oxygen, including for example ammonia (NH_3) and nitrate (NO_3^-). In nature, this process of nitrogen *fixation* is mainly mediated by specialized microorganisms (biological N fixation), some of them living in close symbiotic association with the roots of certain plants like legumes. To a lesser extent, nitrogen fixation also occurs in the atmosphere with help of lightning. Inorganic forms of reactive N can subsequently be converted to organic forms through assimilation by plants and other autotrophic organisms, forming the basic nitrogen supply for all ecosystems. Biogeochemical processes continuously interconvert the various forms of reactive N in the environment, before it is eventually returned to the atmosphere as N_2 by denitrification. See Figure 1.1 for a comprehensive overview of the processes involved in environmental nitrogen cycling illustrated for an agricultural ecosystem.

Since the beginning of the 20th century, human activities such as production of mineral fertilizers (via the Haber-Bosch process) and fossil fuel combustion have led to a surge in reactive N inputs to the environment compared to natural background levels (Galloway et al. 2004, Galloway et al. 2021). As denitrification is unable to fully compensate for the increasing inputs (Galloway & Cowling 2002), reactive N accumulates in various environmental compartments (Van Meter et al. 2016, Ascott et al. 2017). The benefits of rising supplies of reactive N - first and foremost increased agricultural production - are contrasted by various adverse effects reactive N can have on human health and the environment, including surface waters (Galloway & Cowling 2002, Erisman et al. 2013). Ammonia and nitrite (NO_2^-) are directly toxic to many aquatic organisms and may also pose health risks for humans. As the primary forms of reactive

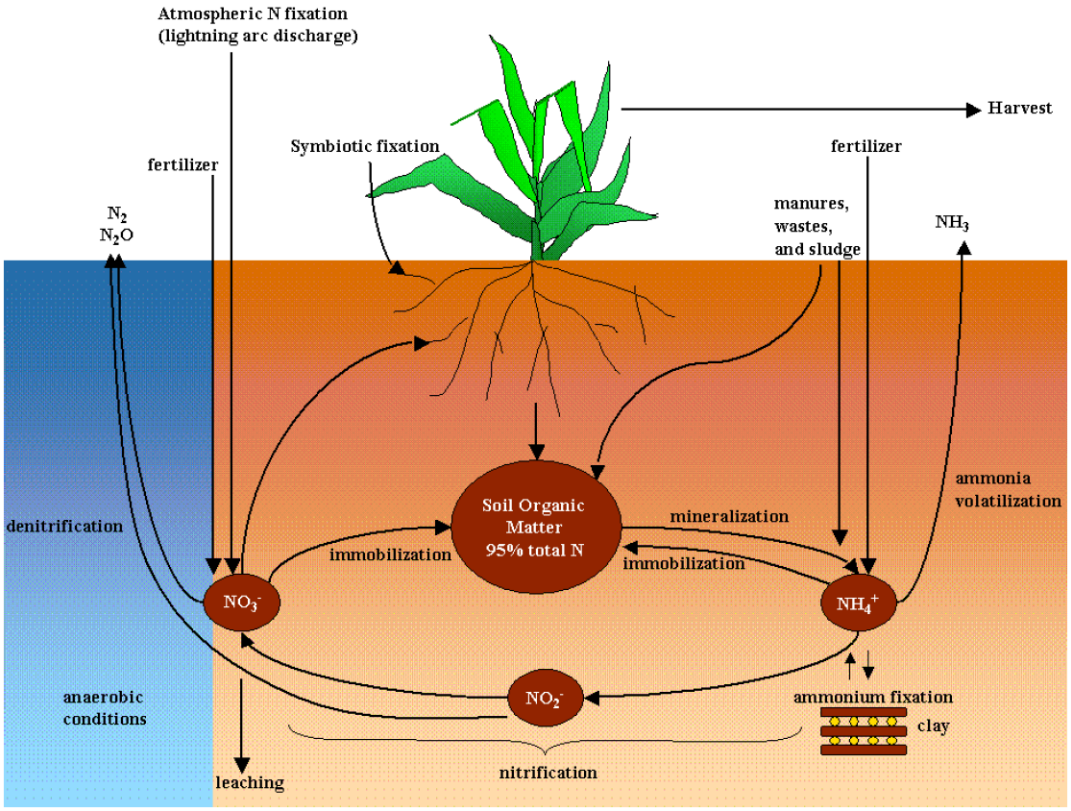


Figure 1.1.: Major components of the nitrogen cycle in agricultural soils. N inputs with fertilization usually are the dominant source of reactive N in agricultural environments. Some of the N added with manure or fertilizer may be lost again via ammonia volatilization. Natural processes that introduce reactive N to soils are biological (symbiotic) fixation and atmospheric fixation with subsequent deposition. Human activities like livestock farming and fossil fuel combustion can increase the natural background levels of atmospheric N deposition. Most plants primarily use nitrate (NO_3^-) to meet their N demand, and only resort to ammonium (NH_4^+) if supply with the more accessible nitrate is insufficient. Decomposition of dead biomass and subsequent mineralization leads to the re-release of inorganic N compounds in the soil, where the balance between N mineralization and immobilization is controlled by the C:N ratio of the soil organic matter. Nitrification is the two-step oxidative conversion of ammonium to nitrite (NO_2^-) and nitrate by chemoautotrophic bacteria to meet their energy demand. Nitrate (and other forms of oxidized N) may subsequently be reduced to N_2 by certain heterotrophic bacteria that switch to alternative electron acceptors for their respiratory chain in the absence of oxygen (denitrification under anaerobic conditions). High mobility of nitrate due to anion exclusion at the interface between soil solution and soil matrix makes the NO_3^- molecule prone to leaching with percolating water. The figure is taken from Neitsch et al. (2011).

N available to plants, algae and cyanobacteria, excess release of ammonium (NH_4^+) and nitrate into surface waters can cause eutrophication, harmful algae blooms, hypoxia and "*dead zones*" (areas with minimal aquatic life due to oxygen depletion) (Galloway & Cowling 2002, Withers et al. 2014). This is particularly true for coastal waters, where N is usually the limiting factor for algal growth, although it may often play a role in eutrophication of inland surface waters as well (Paerl et al. 2016). In the long term, elevated levels of reactive N alter aquatic biocenoses and reduce biodiversity.

Nitrogen enters surface waters from both point and non-point sources, the former typically associated with wastewater and the latter primarily associated with agriculture. Past decades have seen a steady decline in N emission loads to surface waters where regulations have been put in practice to counteract pollution of water resources, with an overall pattern of a more rapid decline in N emissions from point sources (as an example, see Fuchs et al. (2010) for the progression of N emissions in the major German river catchments since the early 1980s). Despite such advances, reactive N levels in surface waters have often failed to decline correspondingly, which is explained with the effects of historical nitrogen accumulation in environmental compartments causing long delays between the implementation of mitigation measures and visible effects in monitored surface water bodies ("*legacy pollution*") (Withers et al. 2014, Ascott et al. 2021, Basu et al. 2022). Legacy pollution and the associated lag times in catchment response have been identified as a key research question in hydrology (Kirchner 2003, Blöschl et al. 2019), with calls to illuminate the underlying processes and adequately account for them in computer simulation models (Hrachowitz et al. 2016, Lutz et al. 2022, Golden et al. 2023). Globally, creation of new reactive N and its use in agriculture as well as other sectors is still increasing. N pollution of surface waters will almost certainly follow that trajectory, as will N accumulation in other environmental compartments and the accompanying legacy effects (Galloway et al. 2021, Bijay-Singh & Craswell 2021).

1.2. State of the Art

1.2.1. Ecohydrological Catchment Modeling

Past decades have seen the emergence of a variety of computer models capable of simulating water quality and surface water pollution on the river catchment scale. Such models are applied for instance to complement water quality monitoring data in space and time, identify CSAs for a specific pollutant, carry out source apportionment studies, conduct scenario analysis and assess management impacts (Borah & Bera 2003, Radcliffe et al. 2009, Fu et al. 2019). To accomplish all this, the models need to integrate a wide array of catchment processes includ-

ing runoff generation and streamflow, erosion, biogeochemical nutrient cycling and transport, plant growth as well as land and water management. Therefore, they are often referred to as *ecohydrological* catchment models (Krysanova & Arnold 2008). Ecohydrological models are very diverse in their capabilities and representation of the relevant processes. For comprehensive overviews of important ecohydrological models with emphasis on different applications and modeling aspects, see Borah & Bera (2003), Horn et al. (2004), Borah et al. (2006), Radcliffe et al. (2009), Schoumans et al. (2009), Daniel et al. (2011), Pandey et al. (2016), Hrachowitz et al. (2016) and Fu et al. (2019).

Figure 1.2 proposes a concept for systematic classification of ecohydrological models along three dimensions: model complexity as well as spatial and temporal scale. Model complexity describes the level of detail a model applies for process representation. At the lower end of the complexity spectrum, *empirical* models derive their results exclusively from observed data (e.g. via regression functions). This purely data-driven approach is often referred to as a *black box* because it relates model inputs to outputs with little to no insight into the actual physical quantities and processes that govern the catchment behavior. On the opposite side of the spectrum, *physically-based* models make explicit assumptions about the processes involved using established physical descriptions. In theory, this completely process-driven approach allows the underlying assumptions to be scrutinized for their ability to reproduce observed system behavior and - since the associated parameters have direct physical meaning - the model to be parameterized independently from any calibration exercise (e.g. through field measurements). The process-driven approach is also thought to better extrapolate to new catchments and environmental conditions than the data-driven approach. In practice however, these alleged advantages are usually compromised by a lack of available data at the needed spatial and temporal resolution as well as scale-dependencies of the employed process descriptions (Beven 1989, Beven 2001, Refsgaard et al. 2010). Furthermore, the models require considerable computational resources, which can become prohibitive for large catchment areas or research designs that involve repeated simulations (e.g. for uncertainty assessment) (Borah & Bera 2003, Hrachowitz et al. 2016, Herrera et al. 2022). *Conceptual* models take a middle-ground between the purely data-driven and process-driven approaches described above. To a varying degree, they combine empirical methods with explicit process descriptions that are often based on simplifying assumptions (e.g. the subdivision of the entire environment into discrete compartments and generalized storage elements).

Regarding spatial scale, it is common to distinguish *lumped*, *semi-distributed* and *distributed* models. The output from lumped models is calculated for the entire catchment area at once, working with area-aggregated input data and neglecting all spatial differences. Semi-distributed

models account for spatial heterogeneity by dividing the catchment area into smaller entities. While some authors already consider the division of the catchment area into several sub-catchments to be a semi-distributed approach (Daniel et al. 2011), others merely interpret that as the iterated application of a lumped approach (Fu et al. 2019). In the latter case, (sub-)catchments are instead viewed to consist of even smaller spatial units, defined either by similarity in important landscape properties such as land use and soils, or by their position in the topographic continuum between upland areas and valley bottoms (Fu et al. 2019). Fully distributed models operate on arbitrarily fine spatial resolutions that can be defined by the model user based on the available data and computational power. Typically, they superimpose the entire catchment area with a network of hydrologically connected entities (typically cells in a grid) that form the basic processing units.

With respect to temporal scale, models can be *static* or *dynamic*. The output of static models consists of average values representative of longer time periods (years to decades). In contrast, dynamic models are capable of generating time series output of varying resolution (annual, monthly, daily, hourly, etc.). At low temporal resolutions, there usually is a smooth transition from dynamic to static models and it becomes increasingly difficult to distinguish both model types. Looking at the three classification criteria in combination, there is a tendency of empirical models working on low and physically-based models operating on high spatial and temporal resolutions. This general pattern is inherent to the operating principles of the different model types. However, it does not apply to all ecohydrological models as illustrated by the small selection of models included in Figure 1.2. These models are:

1. *MIKE SHE* (Refsgaard & Storm 1995) is a commercial hydrological model developed at the *Dansk Hydraulisk Institut* (DHI) originating from the *Système Hydrologique Européen* (SHE) (Abbott et al. 1986*a*, Abbott et al. 1986*b*). MIKE SHE can be coupled with the root zone model DAISY to include emissions of nitrogen or pesticides in the simulations (Styczen et al. 1999, Hoang et al. 2014).

2. *SWAT* (Soil and Water Assessment Tool) (Arnold et al. 1998) is an open source catchment model initiated by the US Department of Agriculture (USDA). It can simulate emissions of nutrients (N, P), pesticides, heavy metals and other pollutants.

3. *MONERIS* (Modelling of Nutrient Emissions in River Systems) (Venohr et al. 2011) is a model for assessment of nutrient emissions (N, P) developed at the Berlin-based Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB).

4. *MoRE* (Modeling of Regionalized Emissions) (Fuchs et al. 2017) is an open source modeling system developed at the Karlsruhe Institute of Technology (KIT) in Germany. MoRE

is an independent branch of MONERIS with added functionality and a broader range of pollutants that can be simulated (nutrients, heavy metals, different organic micropollutants).

5. The *AGRUM* (Agro-Environmental Measures) (Zinnbauer et al. 2023) model suite encompasses several models for estimation of non-point source emissions of nutrients (N, P):

   a) *RAUMIS* (Regionalized Agro-Environmental Information System)

   b) *mGROWA* (Monthly Large-Scale Water Balance Model)

   c) *DENUZ* (Denitrification in the Vadose Zone)

   d) *WEKU* (Residence Times and Reactive N-Transport in Groundwater)

   e) *MEPhos* (Model for the Determination of Phosphorus Inputs)

   (RAUMIS is developed at the Thünen Institute in Braunschweig, Germany, while the other four models are developed at the Forschungszentrum Jülich, Germany. They are complemented by MONERIS to account for nutrient emissions from urban systems and point sources as well as in-stream nutrient retention. Note that this set of models is referred to as AGRUM only in part of its application contexts, for example Zinnbauer et al. (2023). For the sake of simplicity and brevity, this thesis adopts the name AGRUM.)

As evident from Figure 1.2, model complexity as well as spatial and temporal scale positively correlate for MONERIS/MoRE, SWAT and MIKE SHE+DAISY, whereas AGRUM - as a conceptual distributed modeling system operating on an annual time step - is set apart to another region of the model space.

The classification system outlined in Figure 1.2 does not claim to be comprehensive. Apart from model complexity, spatial scale and temporal scale, the literature mentions other classification criteria for ecohydrological models not included in the figure. For instance, it can also be distinguished between *deterministic* and *stochastic* modeling techniques. Strictly deterministic models have definite parameters, inputs and outputs, whereas stochastic models work with distributions that can account for environmental variability and input uncertainty. Models furthermore differ in their general scope. For example, some catchment models are limited to the terrestrial domain and neglect in-stream processes like pollutant retention and reactive transport, while typical river water quality models focus on the aquatic domain and are unable to simulate pollutant inputs into surface waters from the surrounding environment (Horn et al. 2004, Fu et al. 2019). The range of different pollutants a model is able to simulate is another

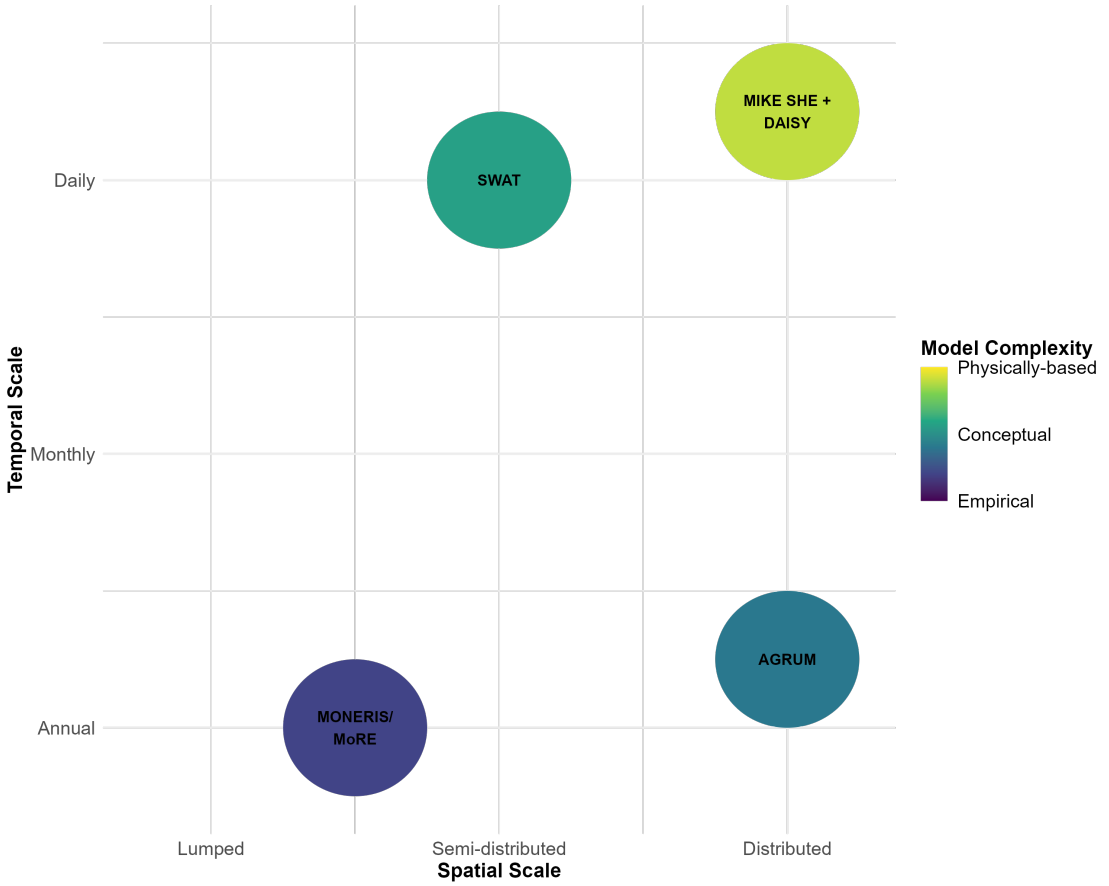


Figure 1.2.: Illustrative selection of different ecohydrological catchment models classified in terms of model complexity as well as spatial and temporal scale. As classification is always ambiguous to a degree, all model characteristics have been deliberately plotted on a continuous scale. The models are positioned in accordance with their highest possible (spatial or temporal) resolution. Note however that the simulation results can always be downscaled for each individual model (e.g. SWAT is also capable of generating monthly and annual simulation results).

possible classification criterion. While some models (e.g. MoRE) take a universal approach that can accommodate many different pollutants, other models (e.g. MONERIS) are specifically designed to simulate a certain pollutant class and are thus restricted to this set of substances.

1.2.2. Modeling Uncertainty and Equifinality

Although catchment modeling is arguably indispensable for assessing and ultimately mitigating non-point source pollution on a larger scale, it is also met with criticism especially when models demonstrably fail to reproduce observed catchment behavior or predict the system response to specific mitigation measures (Rissman & Carpenter 2015). Such prediction failures are the result of the spatiotemporal complexity of catchments (Beven 2000) and their - by definition - imperfect representation in computer models, making all model output inevitably subject to considerable uncertainty. Model uncertainty, as well as the need to adequately assess and communicate it, has long been recognized in the hydrologic scientific community (Beven & Binley 1992, Reckhow 1994, Pappenberger & Beven 2006, Beven & Binley 2014, Harmel et al. 2014, Beven 2016, Blöschl et al. 2019). In its broadest sense, uncertainty may be defined as "*any deviation from the unachievable ideal of completely deterministic knowledge*" (Walker et al. 2003). Daggupati et al. (2015) describe model uncertainty as the lack of *precision* in the simulation results (as opposed to model error, which is a lack of *accuracy* in the results). Uncertainty in hydrological modeling originates from several interdependent sources, the most important of which being measurement error, spatiotemporal scale, the model parameters and the model structure (Shirmohammadi et al. 2006, Guzman et al. 2015):

Measurement Error Random as well as systematic measurement error in environmental data largely depends on the employed measuring instrumentation and the environmental conditions under which it operates. In principle, all monitoring data is subject to measurement error that may affect simulation results when used in a modeling study. In the context of hydrological modeling, error in precipitation data is a particularly important source of uncertainty.

Spatiotemporal Scale Field measurements are usually unavailable on the spatiotemporal scale needed to adequately represent the spatial heterogeneity and temporal variability of real catchments. Instead, measured data is often aggregated or interpolated, adding a further layer of uncertainty to the modeling process. Closely related is the issue of *incommensurability* between model inputs and field measurements. Both are often treated as directly comparable, even if there is a mismatch in the underlying spatiotemporal scales (Beven 2012, pp. 18, 245). A typical example are point measurements of soil proper-

ties or precipitation being up-scaled to larger areas like grid cells or Thiessen polygons, respectively.

Model Parameters Parameter-based uncertainty stems from incomplete knowledge of the model parameter values, their plausible ranges and variability. Empirical parameters have no physical meaning and thus cannot be measured or estimated *a priori*. Although the parameters of physical process descriptions are theoretically measurable, their spatial representation in even the most sophisticated distributed models is often much coarser than the scale at which they actually vary (e.g. hydraulic conductivities of soils). Instead, they are usually interpreted to integrate over and be representative of all the small-scale variability within each model spatial element. However, the non-linear nature of the underlying equations will make the model results deviate from the actual system response (Beven 2001) even if this assumption was justified (which it is not, see the issue of *incommensurability* mentioned above).

Model Structure A model is always an imperfect representation of an actual catchment. Structural uncertainty arises from the inevitable discrepancies between model and reality. These include oversimplification, misrepresentation or complete omission of important processes in the model equations. Moreover, structural uncertainties may originate from the spatiotemporal setup and discretization of the model as well as technical issues like numerical instabilities in the routines that solve model differential equations. Structural uncertainty is particularly difficult to assess quantitatively, but its contribution to overall model uncertainty is probably substantial (van Griensven & Meixner 2006). In the context of ecohydrological models, an interesting example of structural uncertainty is the often neglected difference between pressure wave *celerities* and flow *velocities*, which is likely responsible for the long time spans many catchments need to adjust to a new water quality regime following pollution control measures (Hrachowitz et al. 2016).

Due to the numerous uncertainties and potential errors that propagate through the models, their parameters are usually calibrated against observed data (typically streamflow and pollutant data at the catchment outlet) to give plausible simulation results. However, model calibration does not reduce uncertainty. In essence, calibration is an inverse problem which for hydrological catchment models is typically *ill-posed* (Herrera et al. 2022) in that different parameters may compensate for each other to give the same (or very similar) calibrated model output. As a consequence, calibration of catchment models is unable to identify a single unique parameter set that fits the calibration data significantly better than all others. This problem is commonly referred to as the *non-uniqueness* or *equifinality* of parameter sets (Beven 2006). It is further aggravated in distributed and semi-distributed models, where each spatial element may be pa-

rameterized differently, resulting in a proliferation of model parameters (Beven 2001). Each additional parameter provides an extra dimension of flexibility for model output adjustment, exacerbating parameter equifinality (Schoups et al. 2008) and effectively creating a "*mathematical marionette*" (Kirchner 2006) that can easily be overfit to the calibration data. Although calibrated models are able to capture the integrated catchment response (i.e. the calibration data at the catchment outlet), it remains unclear if this is based on a realistic representation of the internal processes generating that response ("*getting the right answers for the right reasons*") (Kirchner 2006).

Model parameters do not only compensate for each other, but may equally compensate for measurement errors in the input data, spatiotemporal inaccuracies or structural model deficiencies. They may furthermore compensate for problems with the calibration data, either measurement errors (Harmel et al. 2006, Harmel & Smith 2007) or incommensurability with the model output data (Beven 2012, p. 18). Hence, the concept of equifinality extends beyond just the model parameters. Instead it encompasses all possible sources of modeling uncertainty, and parameter calibration implicitly accounts for all of these uncertainties at once, which questions the existence of a single "true" parameter set altogether (Beven & Binley 1992, Beven 2006). Equifinality is a major challenge for catchment modeling in general, and for simulation of non-point source pollution in particular. When using an ecohydrological model to assess catchment-wide water pollution, the values of the parameter values that govern pollution have been obtained in a model calibration process. However, due to equifinality, there are many different parameter sets that can reproduce the calibration data equally well, with no indication which parameter set is the best approximation to reality. Consequently, the results of model-based source apportionment or identification of critical source areas are subject to incalculable uncertainty.

There are some approaches and strategies that, if not completely solve the fundamental problem of equifinality, then at least help to mitigate its most severe effects. Of course, calibrated ecohydrological models are routinely validated against independent data, although there are objections that the commonly used *temporal* split-sample test is too weak to assess model performance in conditions that deviate greatly from the calibration period (Kirchner 2006). In these cases, Klemeš (1986) advocated the use of a *differential* split-sample approach, where calibration and validation data are deliberately selected to reflect diverging environmental conditions. Differential model validation could also be a viable option for controlling equifinality, as it likely poses a higher barrier for models with an unrealistic representation of catchment processes to pass the performance test. Another approach to account for equifinality is uncertainty analysis, which provides a means to explicitly assess uncertainty in the model outputs. It entails the variation of uncertain model elements - in most cases (a subset

of) the model parameters - across multiple model runs, leading to simulation results bracketed by an uncertainty band. As argued by Abbaspour et al. (2017), calibration and uncertainty are "intimately linked". Hence, several techniques have been developed that directly couple model calibration and uncertainty analysis, e.g. *Generalized Likelihood Uncertainty Estimation* (GLUE) (Beven & Binley 1992, Beven & Binley 2014) or *Sequential Uncertainty Fitting, version 2* (SUFI-2) (Abbaspour et al. 2004, Abbaspour et al. 2007), which is a GLUE descendant. These calibration techniques identify multiple parameter sets (instead of only one) that satisfy some form of performance criterion.

A further strategy for limiting the impact of equifinality is the use of *parsimonious* models, meaning models with a minimum number of parameters to be determined via calibration. It has been repeatedly demonstrated that around five model parameters[3] are usually enough to reproduce essential parts of typical hydrological calibration data (e.g. a streamflow time series), and that additional parameters will likely introduce parameter non-uniqueness and identifiability problems into the calibration process (Jakeman & Hornberger 1993, Schoups et al. 2008). In this sense, most complex, fully distributed or semi-distributed catchment models are *overparameterized*, meaning that the quantity and quality of calibration data does not support unambiguous parameter identification. However, many tasks in catchment-scale pollution modeling (e.g. CSA identification) require exactly these complex, spatially distributed models. In such cases, further information can be used to better constrain the model output in a multi-objective calibration approach (Gupta et al. 1998, Efstratiadis & Koutsoyiannis 2010), possibly incorporating qualitative or semi-quantitative *soft data* to complement the quantitative *hard data* used in the calibration process (Seibert & McDonnell 2002, Arnold et al. 2015). Alternatively, additional information or data may also be consulted in the model validation process to apply a stricter and more incisive test regime.

1.3. Research Aim and Study Design

The overall research aim of this thesis is to quantitatively assess surface water nitrogen (N) pollution in the Swist catchment (290 km^2), with a focus on non-point source pollution using the ecohydrological catchment model SWAT (Arnold et al. 1998). The Swist is a tributary of the Erft river in North Rhine-Westphalia, Germany. The necessary data basis has largely been

[3] Of course, the minimum number of model parameters may deviate from this value depending on the modeling aim and the level of detail required in the process descriptions. For example, reproducing data on in-stream pollutant loads is likely to require additional model parameters to account for the increased number of processes involved. However, a value of five may give a rough indication for the order of magnitude in the necessary number of parameters.

made available by courtesy of the Erftverband local water board (see Section 2.5), which is commissioned with wastewater disposal and surface water maintenance in most of the Erft catchment area as well as groundwater monitoring in the Rhenish lignite mining region. Nitrogen pollution encompasses all emissions of reactive N compounds with potentially adverse effects on the aquatic environment. These include all forms of organically bound N, as well as the inorganic N compounds nitrate (NO_3^-), nitrite (NO_2^-) and ammonium (NH_4^+). If not stated otherwise, quantitative statements on N concentrations or loads throughout this thesis will be made in terms of the mass fraction of N contained in these compounds (i.e. as organic N, NO_3^--N , NO_2^--N or NH_4^+-N). The EU Nitrates Directive from 1991, which complements the WFD, defines a general NO_3^- concentration limit of 50 mg/l (corresponding to approximately 11.3 mg/l of NO_3^--N). Several water bodies within the Swist river basin have repeatedly exceeded this value during multiple WFD monitoring cycles in the more recent past (State Ministry for Climate Protection, Environment, Agriculture, Nature Conservation and Consumer Protection NRW 2015, State Ministry for Environment, Agriculture, Nature Conservation and Consumer Protection NRW 2021).

The study period encompasses the years from 2012 to 2018 (seven years in total), which is long enough to allow for comprehensive model calibration and validation as well as examination of N emissions over several years. At the same time, this period is short enough to consider it approximately stationary, in particular with regard to the number and location of wastewater point sources that influence the N loads transported in the Swist river network. For the largest part of the study period, a total of four wastewater treatment plants discharged wastewater into the surface waters of the Swist catchment (see Section 2.3). All of them were explicitly accounted for in the simulations. In contrast, wastewater discharges from combined sewer overflows and separate sewer outlets were not considered explicitly, as the available data was insufficient to accurately represent them in the model. However, SWAT implicitly accounts for these additional sources of water in its simulation routines for urban runoff, albeit without considering the additional pollutant release from combined sewer overflow events. In order to assess the SWAT simulation results, they are compared to two reference data sets on (diffuse) N emissions in the Swist catchment: The first of these data sets is based on upscaled *edge-of-field* (EOF) monitoring data collected during previous research projects in the Swist catchment and representing the years from 2005 to 2009 (Kistemann et al. 2007, Mertens et al. 2012). The second reference data set is based on raster data of N emissions from diffuse sources covering the entire German federal state of North Rhine-Westphalia. This raster data represents the years 2014 to 2016 and has been generated in a research project using the static AGRUM modeling system briefly described in Section 1.2 (Wendland et al. 2021a).

1.4. Research Questions

As a dynamic and conceptual catchment model, SWAT requires calibration. Equifinality implies that calibration will yield multiple SWAT model parameterizations (i.e. a model ensemble) for the Swist catchment that can recreate the available observed data in the study period similarly well. The various N emission scenarios associated with the SWAT model ensemble can be compared among each other as well as with the two reference data sets mentioned in Section 1.3, which serves several purposes: First, to identify the *similarities* between the various emission estimates, i.e. the conclusions about non-point source N pollution in the Swist river basin that are substantiated across different data sets. More specifically, this entails the following research questions:

1. How high are the (average) annual non-point source nitrogen emission loads in the Swist catchment during the study period, and what are the contributions of the different emission pathways?
2. How important are point source nitrogen emissions from wastewater treatment plants in the Swist catchment compared to non-point source nitrogen emissions?
3. Which spatial patterns of non-point source nitrogen emissions in the Swist catchment emerge in the modeling results? In particular, do the results indicate the existence of critical source areas for diffuse nitrogen pollution in the Swist catchment, and where are they located?

Second, to explore the *discrepancies* between the various emission estimates, i.e. assess the uncertainty in the simulation results. This addresses the following research questions:

4. How wide is the range of simulated non-point source nitrogen emissions in the Swist catchment between different calibrated SWAT models (i.e. how large is the parameter-based model uncertainty)?
5. How much do the simulated non-point source N emissions as well as critical source areas differ between SWAT and AGRUM (i.e. how large is the structural model uncertainty)?

Third, to examine the *origins* of the differences in the various emission estimates, which leads to the practically relevant final research question:

6. What are the relative strengths and weaknesses of the different approaches to assess catchment-wide non-point source pollution (*edge-of-field* monitoring, static AGRUM modeling or dynamic SWAT modeling)? Which level of effort is necessary and feasible to draw reliable conclusions particularly on nitrogen pollution?

1.5. Thesis Outline

The present thesis is structured as follows: Chapter 2 provides a description of the study area, the Swist river basin. Chapter 3 gives insights into the functionality and inner workings of the SWAT model, its capabilities (as well as limitations) and why it is chosen as the model to pursue the research aims of this thesis. Furthermore, Chapter 3 portrays the methods and procedures used to evaluate the SWAT simulation results. The Chapters 4 to 6 present and discuss the research results of this thesis: Chapter 4 describes the model setup and examines the effect of different model setup choices on the uncalibrated SWAT model output. Chapter 5 gives account of the SWAT model calibration as well as validation results and explores the model parameter sets yielding the best model performances. Chapter 6 analyzes the SWAT-simulated N emission scenarios associated with the best parameter sets identified in Chapter 5 and compares them to the two reference data sets based on EOF monitoring and AGRUM modeling. Finally, Chapter 7 provides a synthesis of the main research results of this thesis and discusses them in an overarching manner, addressing the individual research questions of this thesis stated in Section 1.4. Moreover, Chapter 7 provides an outlook on potential improvements for a future continuation of the modeling studies in the Swist catchment as well as open research needs and perspectives.

2. Study Area and Data Basis

2.1. General Geography

At a length of 43.6 km, the Swist is the longest tributary of the Erft river in North Rhine-Westphalia, Germany, and part of the greater Rhine river system (Mertens et al. 2012). Originating from the northern edge of the Eifel mountain range, the largest part of the Swist watercourse traverses the loess regions of the Lower Rhine Embayment west of Bonn from south to north. In this area, it displays the typical features of a loess-clay (middle reaches) and gravel (lower reaches) dominated lowland brook (Christoffels et al. 2016). The Eifel foothills and the Ville mountain ridge form the major divides of the Swist river basin to the south and east, respectively (Koch 2004). Elevation ranges from 330 m.a.s.l. at the source of the Swist down to 108 m.a.s.l. near the mouth to the Erft river. The average slope along the Swist river bed is 1.3 % in the headwater region upstream of the village of Vettelhoven and 3 ‰ for the rest of the watercourse (Mertens et al. 2012).

Following the watercourse downstream, major Swist tributaries are Altendorfer Bach, Morsbach, Eulenbach, Wallbach, Steinbach[1] and Schiessbach, all of which originate from the hilly regions of the Eifel foothills and join the Swist from the left hand-side. The whole stream network including all tributaries is about 250 km long (Schreiber et al. 2016). Among its most distinctive features are two dammed reservoirs located within the Steinbach subbasin, the Steinbach reservoir and the considerably smaller Madbach reservoir. While the stream beds of the uppermost Swist reaches and many of the Swist tributaries exist in a nearly natural condition, the middle and lower reaches have been considerably affected by engineering measures: The largest part of the Swist main watercourse has been straightened and framed in a trapezoidal cross section profile, particularly within residential and commercial areas (Schreiber 2011, Christoffels et al. 2016).

[1] Being the largest tributary of the Swist, only the upper reaches are known as “Steinbach”. The middle reaches downstream of Schweinheim are named “Ohrbach”, while the lower reaches just before entering the Swist are called “Jungbach”. For the sake of simplicity, the entire water body will be exclusively referred to as “Steinbach” in this thesis.

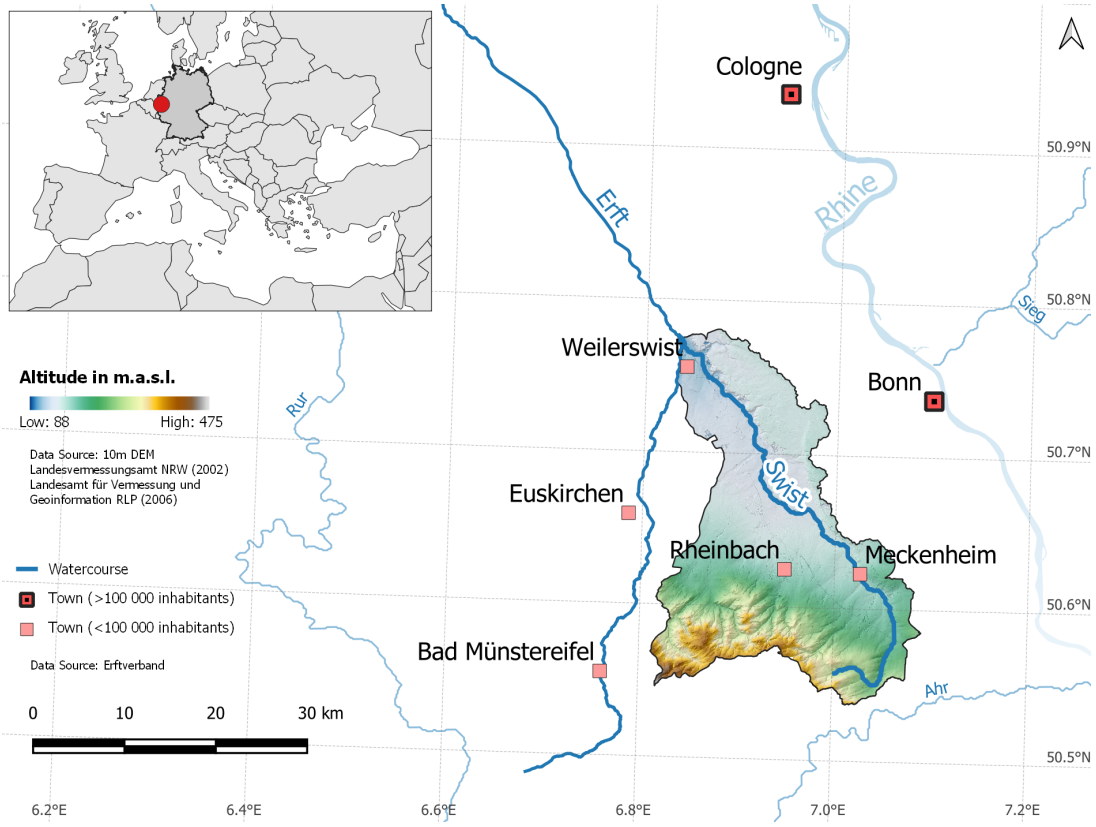


Figure 2.1.: Location of the Swist river basin in Europe and Germany, in relation to the surrounding larger stream network and major towns in its vicinity. The catchment borders contain a digital elevation model (DEM) with a 10 m resolution and a superimposed hillshade depiction of that same DEM. For reasons of scale and space, the Swist river network and tributaries are omitted in this map. Instead, they are included in the more detailed map on page 21 (Figure 2.2).

Most of the Swist river basin is located in North Rhine-Westphalia (NRW). A smaller portion of the catchment area (around 10 %) near the Swist headwaters in the southeast belongs to the federal state of Rhineland-Palatinate (RLP). In its entirety, the catchment area covers approximately 290 km^2 and harbors around 92 000 inhabitants (Christoffels et al. 2016). This amounts to a population densitiy of 317 inhabitants per km^2, which is well below the general population density of NRW (Kistemann et al. 2007). Compared to the adjacent Cologne and Bonn agglomeration, the Swist river basin can be characterized as rural. Major administrative units are the Rhein-Sieg-Kreis district in the east and the Euskirchen district in the west, with the former containing the largest towns within the catchment, Rheinbach and Meckenheim. The municipalities of Rheinbach, Meckenheim, Swisttal and Bornheim in the Rhein-Sieg-Kreis district as well as Euskirchen and Weilerswist in the Euskirchen district cover the majority of the Swist catchment area in NRW. The municipality of Grafschaft is located in the southeast and belongs

to the Ahrweiler district in Rhineland-Palatinate. Minor area shares belong to the municipalities of Alfter, Altenahr, Bad Münstereifel and Wachtberg. Figure 2.1 shows the location of the Swist river basin in Europe and Germany, the surrounding major water bodies and adjacent towns as well as elevation within the catchment. Figure 2.2 on page 21 provides a more detailed depiction of the stream network within the catchment, including the major Swist tributaries.

2.2. Soils and Land Use

In the Eifel headwater regions of the Swist catchment, the parent material for soil formation is mostly clay- and siltstone, whereas the downstream regions along the Swist main watercourse are dominated by unconsolidated siliciclastic sediments of the Rhine fluvial terrace sequence (Koch 2004, Gumm et al. 2016). Predominantly, these sediments are covered with widespread periglacial layers of loess, which in some parts can reach down to a depth of 2 m (Kistemann et al. 2001). The influence of loess contributed to the formation of exceptionally fertile soils in the area, which in some places - however - can be affected by local perched water tables over highly compacted Rhine terrace deposits (Mertens et al. 2012). The soils in the Eifel foothills that formed over consolidated parent rock tend to be shallower than those in the lower parts of the catchment (Geological Survey NRW 2019). Following German soil systematics[2], Pseudogley[3] (about 25 %), Parabraunerde[4] (about 21 %) and Braunerde[5] (about 13 %) are the most dominant soil types in the Swist catchment, followed by the transitional soil types Pseudogley-Parabraunerde (about 10 %), Pseudogley-Braunerde (about 7 %) and Parabraunerde-Pseudogley (about 7 %)[6] (Geological Survey NRW 2019, Kothe et al. 2021).

Due to the excellent conditions for crop cultivation, most agricultural holdings in the region have specialized on crop farming. Livestock farming plays only a very subdued role and is mostly restricted to horses and dairy cattle (Koch 2004, Schreiber 2011). Around 40 % of the entire catchment area is used as arable land, with orchards constituting an additional 5 %. Together with tree nurseries (less than 2 %), these orchards are a special land use trait of the Swist river basin and concentrate in the southeastern part around Meckenheim. About 95 % of the

[2] The German soil systematics classifies soils based on pedogenesis and is laid out in the official German Pedological Mapping Guidelines (Bodenkundliche Kartieranleitung, KA) (Ad-hoc AG Boden 2005).

[3] Pseudogley describes mineral soils affected by stagnating surface water over a horizon with destroyed or clogged soil pores (e.g. by relocated clay particles from upper horizons).

[4] Parabraunerde describes mineral soils of four horizons which display redistribution of clay minerals and iron oxides under an upper humus-rich surface horizon.

[5] Braunerde describes mineral soils with an upper humus-rich surface horizon overlaying a horizon rich in weathering products like clay minerals and iron oxides (with no visible redistribution of these components in separate horizons).

[6] Transitional soil types are termed with the more important soil type named last.

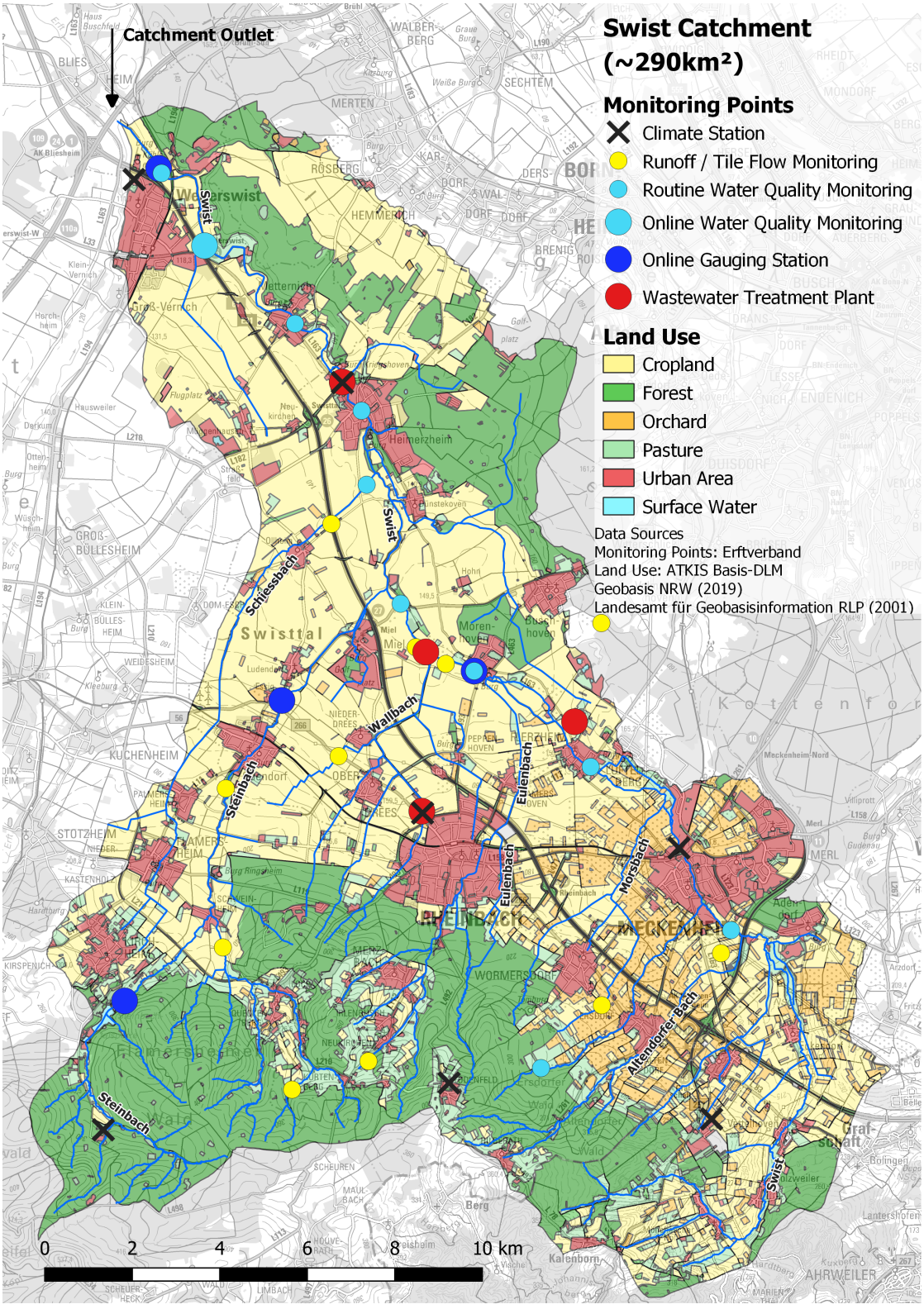


Figure 2.2.: Monitoring activities and land use inside the Swist catchment area.

entire agricultural land in the basin is located within a distance of 500 m away from the nearest water body (including ditches), and 10 % lies within a distance of 50 m (Christoffels et al. 2016). Approximately 10 % of the catchment area is grassland, which is located mainly in the hilly Eifel regions, or where stagnating water renders soils unfavorable for crop cultivation. In many places, stagnating soil water or perched water tables may also affect cropland soils, which is the main reason for relatively widespread tile drains: About one fifth of the entire basin area and one third of the agricultural area in the river basin have been tile drained (Kothe et al. 2021). Forests make up around 29 % of the catchment area, which concentrate on the Ville and Eifel moutain ridges in the northeast and southwest, respectively. Urban areas constitute around 10 % of the catchment area. Figure 2.2 gives a detailed overview of the spatial distribution of land uses within the catchment borders.

2.3. Climate and Hydrology

The climate in the Swist river basin can be characterized as temperate oceanic (climate class Cfb according to the Köppen-Geiger system), with mild winters and moderately warm summers. Average annual air temperature is approximately 10 °C. Precipitation is distributed evenly over the entire year, with increases typically during midsummer and winter (Kistemann et al. 2007). Being located on the leeward side of the Eifel mountains, rainfalls in the region are limited in comparison: Average annual precipitation is between 600 and 700 mm and thus below the German average. Figure 2.3 shows monthly air temperature averages and precipitation heights for Weilerswist close to the Swist mouth. The data from the Weilerswist weather station represents the climate in the lower parts of the Swist catchment. The elevated Eifel foothills in the very south of the catchment exhibit significantly higher precipitation, with average annual precipitation exceeding 650 mm. Due to relatively steep slopes and shallow soils, the rainfalls in this area are of particular importance for the occurrence of streamflow peaks in the Swist main watercourse and its tributaries (Koch 2004).

The hydrology of the Swist catchment is affected by large scale groundwater extraction due to lignite mining. The lignite seams of the Lower Rhine Embayment are embedded in extensive layers of unconsolidated sediments of marine and fluviatile origin. The oldest of these sediments date back to the Oligocene (Lenk 2008). In the center of the embayment, they exhibit a thickness of more than 1000 m and contain a complex sequence of sedimentary aquifers. As part of a larger rift system, several geological faults intersect the Lower Rhine Embayment in the northwest to southeast direction and divide the region into distinct tectonic blocks. Although some faults may connect aquifers vertically (Gumm et al. 2016), they mostly inhibit

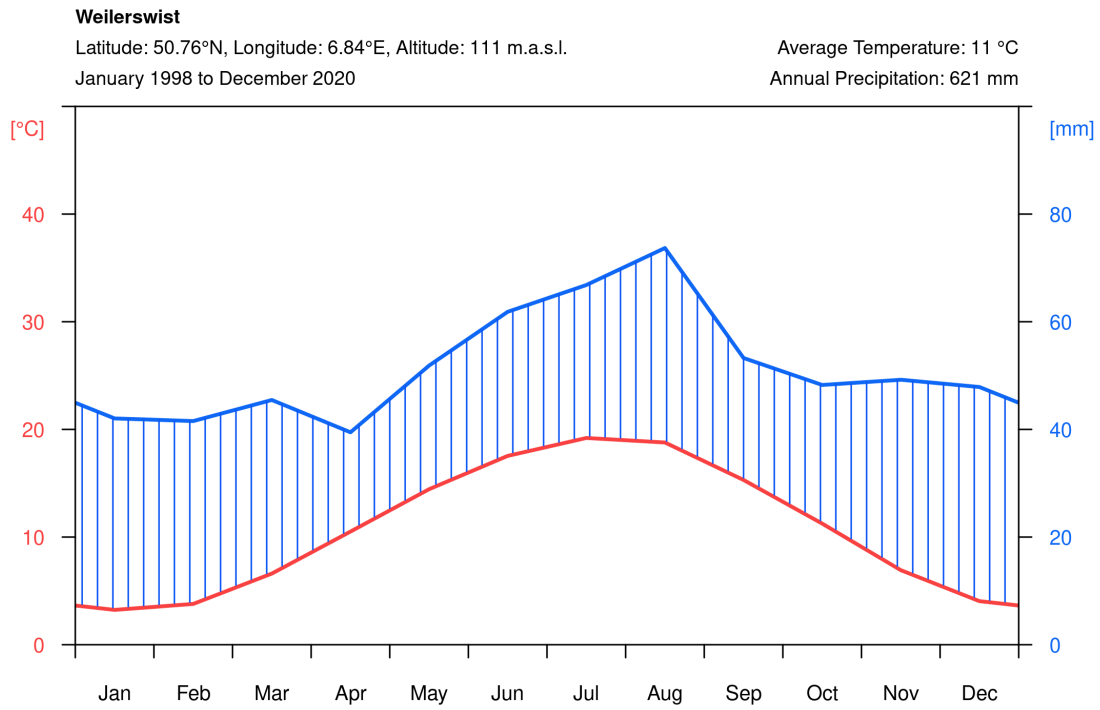


Figure 2.3.: Walter and Lieth climate diagram for Weilerswist. The underlying air temperature and precipitation data stem from measurements at the Erftverband weather station in Weilerswist and span the calendar years from 1998 to 2020.

lateral groundwater transfer between blocks (Lenk 2008). The northern part of the Swist catchment belongs to the Erft block, which is separated from the Cologne block in the east by the Erft fault.[7] Although being situated several kilometers away from the active opencast mining pits of Hambach, Inden and Garzweiler, the water table draw down within the Swist catchment reaches as far south as Meckenheim and has in parts completely drained the uppermost aquifer to a depth of approximately 100 m (Erftverband personal communication).

As a consequence, large parts of the Swist river network have lost permanent groundwater contact, and several Swist tributaries display intermittent or ephemeral discharge behavior. Local geological faults contribute to this by often preventing the formation of large and cohesive groundwater bodies where water tables may otherwise be sufficiently high (Kistemann et al. 2007). In the southern headwater region of the catchment, the Eifel fractured rock aquifers provide a more steady source of baseflow. However, many surface waters are affected by water

[7] A southward continuation of the Erft fault - the Swist fault - is visibly excavated in the Flerzheim gravel pit northeast of Rheinbach.

loss and seepage further downstream. Stream water loss occurs - for example - along the Steinbach: Discharges at the upstream Kirchheim gauge often exceed corresponding discharges at the downstream Essig gauge. Water flow in many streams and ditches regularly ceases during extended dry weather periods, and severe droughts in the summers of 2018 to 2020, as well as 2022, even caused the Swist main watercourse to fall dry upstream of the wastewater treatment plant (WWTP) in Flerzheim.

The long-term Swist mean discharge near the mouth to the Erft river in Weilerswist is 0.8 $m^3\ s^{-1}$, which amounts to a mean specific discharge of 2.8 $l\ s^{-1}\ km^{-2}$. On average, 13.3 % of annual precipitation reached the Swist catchment outlet from 2012 to 2018 (see Table 2.2 on page 27). Estimates of average annual evapotranspiration for the Swist catchment are about 80 % of annual precipitation, albeit with a notable degree of uncertainty (see Section 6.2). Depending on the actual height of annual evapotranspiration, seepage and underground water loss due to groundwater extraction at the lignite mining pits may constitute a meaningful part of the water balance in the Swist catchment. However, underground flow paths and the ultimate fate of percolated surface water are highly uncertain and difficult to verify. As a consequence of the relatively low streamflow quantities, the principal watercourses inside the catchment transport a markedly high municipal wastewater load. On average, wastewater constitutes around 40 % of annual streamflow at the catchment outlet. Under low flow conditions, this percentage can rise to values above 80 % (Christoffels et al. 2016). Wastewater is discharged into the stream network either continuously from the wastewater treatment plants inside the catchment, or discontinuously during combined sewer overflow events. Table 2.1 gives an overview of the wastewater treatment plants along the Swist and its tributaries as of 2018 (see Figure 2.2 and Appendix A for their positions inside the catchment).

Table 2.1.: Wastewater treatment plants in the Swist river basin and their operational characteristics (as of 2018). From left to right, the wastewater treatment plants are ordered by size and - coincidentally - from south to north or upper to lower stream reaches (compare with Figure 2.2 on page 21). Capacity and connection state are specified in population equivalents (PE). The data is taken from the Erftverband annual review 2018 as well as Brunsch et al. (2018).

WWTP	Flerzheim	Rheinbach	Miel	Heimerzheim
Water Body	Swist	Wallbach	Swist	Swist
Capacity [PE]	50 000	27 000	11 000	10 700
Connected [PE]	42 060	25 184	10 758	7948
Annual Wastewater [m^3]	3 555 490	1 809 659	767 646	488 954
Base Flow Rate [l/s]	78	37	18	12
Maximum Flow Rate [l/s]	400	200	110	107

2.4. Hydrological Conditions in the Research Period

Figure 2.4 gives an overview of annual precipitation measured at the meteorological station in Weilerswist during the research period from 2012 to 2018. Again, precipitation data from the Weilerswist station was chosen to illustrate the conditions in the lower part of the catchment. It is important to remember that the depicted data may not be representative of precipitation in other locations, especially the elevated headwater regions of the Eifel foothills where average annual precipitation is significantly higher. However, interannual variation of total precipitation heights is mostly comparable with that measured at other weather stations within the Swist catchment, the year of 2016 being a notable exception (see below). Moreover, total annual precipitation from Weilerswist generally correlates well with measured annual mean discharges given in Table 2.2, which integrate over different gauge catchments and thus practically the entire Swist basin.

Comparing total annual precipitation in Weilerswist for each year with the long-term average value (i.e. comparing the endpoints of the blue cumulative curves in Figure 2.4 with the endpoint of their grey counterparts), the years 2013 to 2015 may be described as wetter than average, whereas the remaining four years may be attributed to be drier than average. However, deviations from the long-term average are relatively small for most study years. Assuming relative differences smaller than 10 % to indicate average conditions, only two years can be characterized as markedly wet or dry. In 2014, a series of heavy rainfall events during summer compensates for an unsusally dry spring and drives overall precipitation to exceed the long-term average by approximately 13 % at the end of the year. In contrast, 2018 is marked by an extended period of almost no rainfalls during summer which causes annual total precipitation to stay below the long-term average by approximately 23 %. Data from the other weather stations (not shown) suggests that annual precipitation in Weilerswist considerably underestimates precipitation over the entire Swist catchment area for 2016, especially for the elevated Eifel foothills in the south. Annual mean discharges in Table 2.2 are highest in 2016 at all gauging stations, indicating that 2016 probably was at least as wet as 2014 when considering the entire catchment area.

For three of the four stream gauges listed in Table 2.2 (Essig, Morenhoven, and Weilerswist), mean discharges for the entire research period are below the official long-term values.[8] Therefore, according to measured stream discharges, the research period as a whole may be characterized as slightly drier than the long-term average. However, this can largely be attributed to

[8] As retrieved from `https://www.erftverband.de/hochwasser-aktuelle-messwerte/` on January 4, 2023.

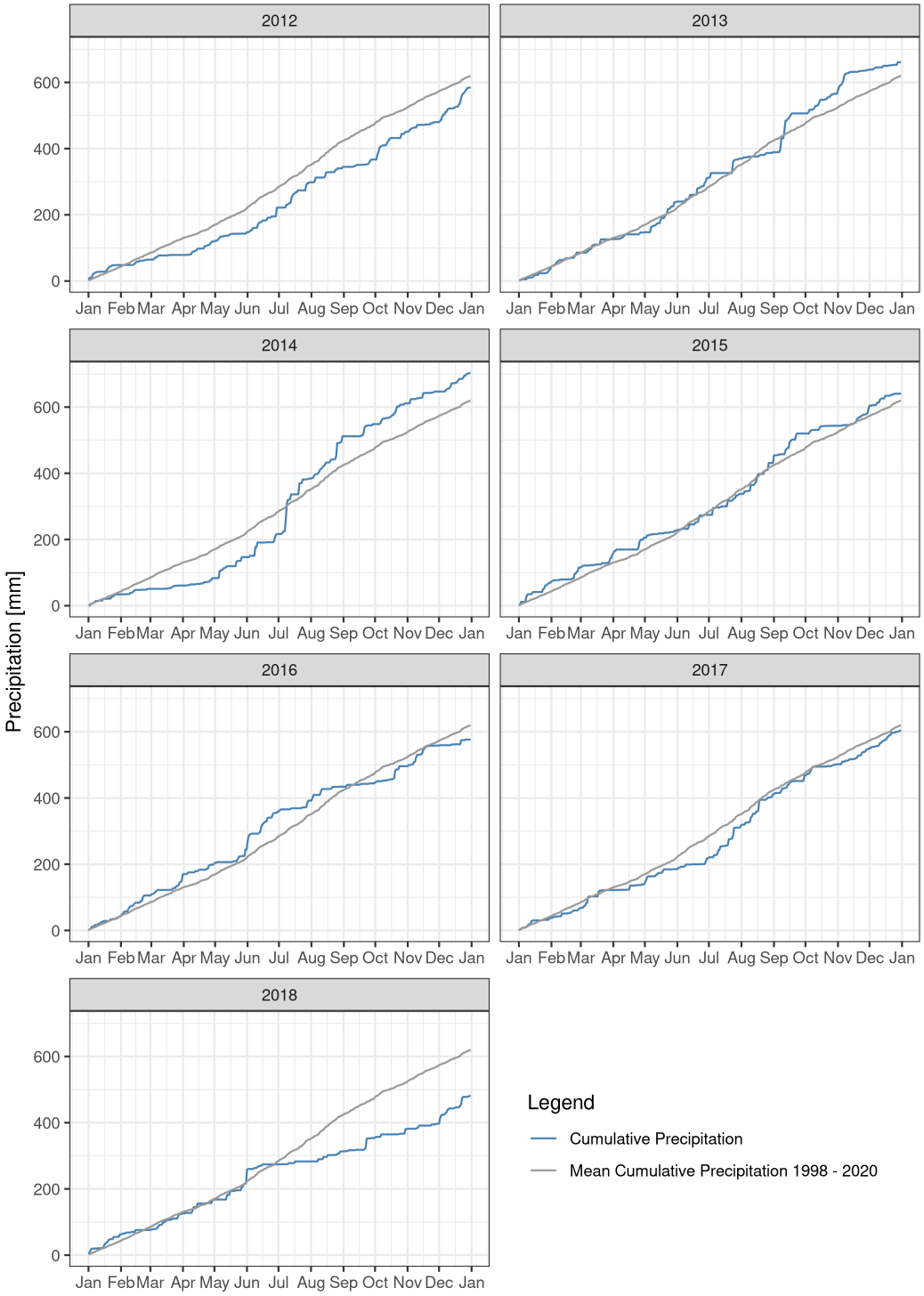


Figure 2.4.: Annual precipitation measured in Weilerswist in the research period from 2012 to 2018 (calendar years). For comparison, cumulative precipitation for each year is depicted next to mean cumulative precipitation from 1998 to 2020 (calendar years). See for reference the climate diagram for Weilerswist in Figure 2.3 based on data from the same time period.

Table 2.2.: Streamflow characteristics at the gauging stations along the Steinbach (top) and Swist (bottom) from 2012 to 2018. Areas of the individual gauge catchments are given in parantheses. $\bar{Q}$ denotes annual mean discharge, $\bar{q}$ denotes corresponding specific discharge and r denotes the overall ratio of streamflow to precipitation for the respective year. r was estimated by constructing Thiessen polygons around each precipitation station and calculating the area weighted mean of annual total precipitation for each gauge catchment. See Figure 2.2 on page 21 for gauge positions inside the catchment. See Figure 2.5 for a visual impression of observed daily streamflow time series in Weilerswist near the Swist catchment outlet. Long-term mean values are official values made available by the gauge operators.

	Kirchheim (14 km^2)			Essig (40 km^2)		
Year	$\bar{Q}$ [m^3/s]	$\bar{q}$ [l/(s km^2)]	r [%]	$\bar{Q}$ [m^3/s]	$\bar{q}$ [l/(s km^2)]	r [%]
2012	0.033	2.4	11.8	0.062	1.6	7.7
2013	0.070	5.0	23.4	0.129	3.2	15.2
2014	0.062	4.5	18.6	0.111	2.8	12.0
2015	0.078	5.6	26.7	0.137	3.4	16.3
2016	0.097	6.9	28.9	0.160	4.0	17.4
2017	0.029	2.1	9.7	0.045	1.1	5.2
2018	0.053	3.8	22.9	0.056	1.4	8.9
Arithmetic Mean	0.060	4.3	20.3	0.100	2.5	11.8
Long-term Mean	0.06	4.3	-	0.12	3.0	-
	Morenhoven (108 km^2)			Weilerswist (288 km^2)		
Year	$\bar{Q}$ [m^3/s]	$\bar{q}$ [l/(s km^2)]	r [%]	$\bar{Q}$ [m^3/s]	$\bar{q}$ [l/(s km^2)]	r [%]
2012	0.335	3.1	15.4	0.611	2.1	10.8
2013	0.476	4.5	20.7	0.857	3.0	14.3
2014	0.425	4.0	17.8	0.823	2.9	13.1
2015	0.484	4.5	21.0	0.892	3.1	14.8
2016	0.525	4.9	22.9	0.974	3.4	16.6
2017	0.320	3.0	13.8	0.629	2.2	10.4
2018	0.269	2.5	17.2	0.574	2.0	13.4
Arithmetic Mean	0.405	3.8	18.4	0.766	2.7	13.3
Long-term Mean	0.43	4.0	-	0.80	2.8	-

the very dry year of 2018. Mean discharges computed only for the years 2012 to 2017 match the official values very closely. For the fourth gauge in Kirchheim, the research period average (including 2018) and the long-term value are almost exactly the same. However, the gauging station is situated directly downstream of the Steinbach reservoir and thus probably gives a distorted impression of upstream wetness conditions. Taking a closer look at Table 2.2, it is notable that the downstream gauges along the Steinbach (Essig) and Swist (Weilerswist) exhibit considerably lower specific mean discharges and streamflow to precipitation ratios than their upstream counterparts in Kirchheim and Morenhoven, respectively. Regarding the specific discharges, this pattern may in part be an effect of higher annual precipitation in the elevated headwater regions. The lower streamflow to precipitation ratios however cannot be explained

with differences in the spatial distribution of precipitation. Instead, they may be a result of the water table draw down in the lignite mining district extending into the northern part of the Swist catchment, lowering groundwater contributions to overall streamflow and facilitating seepage of stream water.

2.5. Prior Research and Monitoring Activities

The Swist river basin has been chosen as the study area for this thesis due to the vast pool of environmental data made available by the Erftverband, as well as the wide array of prior research activities carried out within the catchment. The Erftverband provided access to streamflow data from four online gauging stations along the Swist and Steinbach watercourses (including the Weilerswist gauging station near the basin outlet), meteorological data from several weather stations (seven of which being located inside the Swist catchment area) and operational data from the four wastewater treatment plants emitting wastewater into the Swist and its tributaries. Furthermore, the Erftverband maintains a water quality monitoring network for the surface water bodies in the Erft catchment. This monitoring network includes an online water quality station in Metternich near the catchment outlet recording continuous time series of NO_3^--N and NH_4^+-N concentrations in the Swist with a measurement interval of five minutes. Figure 2.5 gives an impression of daily streamflow observed in Weilerswist and daily NO_3^--N concentrations observed in Metternich during the study period. In addition to the continuous water quality monitoring data from Metternich, the Erftverband determines nutrient concentrations (including NO_3^--N, NH_4^+-N, NO_2^--N as well as organic N) via grab sampling at several routine monitoring points along the Swist main watercourse and some of its tributaries. Grab sampling at these monitoring points takes place up to four times a year.

In the course of the past decades, several research projects with emphasis on water quality and surface water pollution have been conducted in the Swist catchment (Christoffels et al. 2016, Schreiber et al. 2016). These research projects spawned a variety of specific literature on the Swist river basin, as well as additional water quality data supplementing routine monitoring records by expanding the number of data points as well as adding completely new sampling sites. For the specific purposes of the research projects, qualified grab sampling and 24 h composite sampling in part replaced grab sampling at the monitoring points. Moreover, some of the projects examined non-point source pollution at the field and hillslope scale, collecting *edge-of-field* data on pollutant concentrations in surface and subsurface runoff as well as tile flow. This includes concentrations of nutrients like NO_3^--N and organic N (Kistemann et al. 2007, Mertens et al. 2017, Kothe et al. 2021). Catchment-scale emission loads projected

from parts of this data are presented in Mertens et al. (2012). Figure 2.2 on page 21 gives an overview of Erftverband monitoring activities within the Swist catchment. It depicts the location of all sampling points that in some way are relevant for the research presented in this thesis.

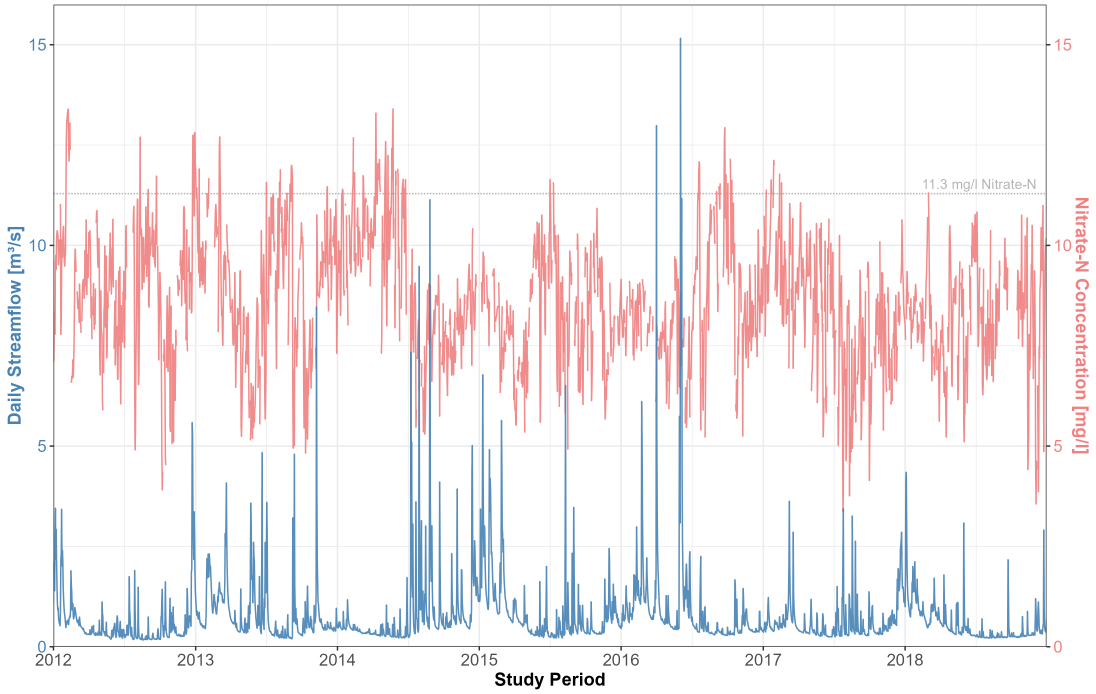


Figure 2.5.: Observed daily time series of streamflow (in Weilerswist) and NO_3^--N concentrations (in Metternich) near the Swist catchment outlet during the study period (2012-2018). The horizontal dashed line in grey indicates the limit concentration of NO_3^--N set by the EU Nitrate and Water Framework Directives (corresponding to 50 mg/l of NO_3^-). Note that under German legislation, this limit concentration applies to the annual average NO_3^--N concentration (not the maximum concentration).

3. Modeling Approach

3.1. The Soil and Water Assessment Tool

3.1.1. General Overview

The *Soil and Water Assessment Tool* (SWAT) is a conceptual, dynamic and deterministic eco-hydrological catchment model. It was originally developed by the Agricultural Research Service (ARS) of the US Department of Agriculture (USDA) in the early 1990s, with the purpose to evaluate the impact of land management decisions on water supply and pollution (Arnold et al. 1998). At its heart, SWAT draws on functionalities of several ARS field scale models developed in response to the US *Clean Water Act* in the 1970s and 1980s, but adds channel routing and other routines to apply them to arbitrary large river basins and simulate all major hydrological processes (Gassman et al. 2007, Krysanova & Arnold 2008, Arnold et al. 2012). Over the years, several updated SWAT versions with new features have been released and model capabilities have been continuously reviewed, refined and expanded (Arnold & Fohrer 2005, Douglas-Mankin et al. 2010, Tuppad et al. 2011, Gassman et al. 2014, Krysanova & White 2015, Bieger et al. 2017).

SWAT is a semi-distributed model in that it disaggregates the entire basin area into so-called hydrologic response units (HRU). These are the basic processing units all calculations are performed on, as far as catchment response prior to in-stream routing is concerned. Each HRU is defined as being approximately uniform in terms of land use, soil and often slope characteristics. However, HRUs are not spatially identified and need not even represent continuous stretches of land. Instead, they are expressed as area fractions of stream subbasins (Gassman et al. 2007). Being attached to a specific segment of the stream network, a subbasin constitutes the smallest georeferenced landscape unit in SWAT. Runoff and pollutant exports calculated on the HRU level are aggregated proportionately for each subbasin and subsequently passed on to the corresponding stream segment (Arnold et al. 2010).

Based on this spatial setup, SWAT is able to dynamically simulate catchment hydrology and streamflow, plant growth, land management, erosion, nutrient cycling and water quality on a

daily time step. Supported water quality parameters include sediment, nitrogen, phosphorus, pesticides, heavy metals and bacteria. Due to its vast scope and versatility, SWAT has emerged as one of the most wideley used tools worldwide to simulate catchment hydrology and water quality, and a plethora of peer-reviewed SWAT studies have been published over the years (Gassman et al. 2007, Refsgaard et al. 2010, Fu et al. 2019). Reported SWAT studies have been conducted on a multitude of spatial scales, ranging from headwater catchments smaller than 1 km^2 (Musyoka et al. 2021, Musyoka et al. 2023) to transnational multi-catchment study areas covering 1 000 000 km^2 and more (Abbaspour et al. 2015, Poméon et al. 2018).

In general, SWAT is designed for long-term simulations spanning several years and decades, enabling researchers to assess the effects of land management on water supply and water quality. SWAT has frequently been used for modeling studies assessing the regional impact of land use and climate change on water resources and the hydrological cycle. Recent examples include Näschen et al. (2019) for land use change and Schaffhauser et al. (2023) for climate change analysis. Additionally, a major focal point of SWAT applications is the simulation of non-point source water pollution, with a large number of published SWAT studies dealing with source apportionment and CSA identification for different pollutants. A collection of SWAT source apportionment studies includes Salvetti et al. (2006), Lam et al. (2010), Malagó et al. (2017) and De Girolamo & Lo Porto (2020). Examples of CSA studies conducted with SWAT includes Tripathi et al. (2003), Gitau et al. (2004), Srinivasan et al. (2005), White et al. (2009), Ghebremichael et al. (2010), Niraula et al. (2012), Niraula et al. (2013), Winchell et al. (2015), Lee et al. (2018) and Evenson et al. (2021).

3.1.2. Model Components

This section will give a broad summary of the SWAT model components which have the largest relevance for the present thesis. A detailed and comprehensive documentation of all process representations and the underlying equations is given in Neitsch et al. (2011). SWAT distinguishes between land phase and routing phase hydrology. Land phase hydrology is simulated at the HRU level and depicted schematicaly in Figure 3.1. Taking this figure as a basis, the daily HRU water balance as simulated by SWAT may be formulated as:

$$S_t = S_0 + \sum_{i=1}^{t}(R_i - E_{a,i} - Q_{s,i} - Q_{l,i} - Q_{g,i} - L_i) \tag{3.1}$$

where S_t is the final amount of water stored after t days, S_0 is the initial amount of water stored at the start of the simulation, R_i is the amount of precipitation on day i, $E_{a,i}$ is the actual evapo-

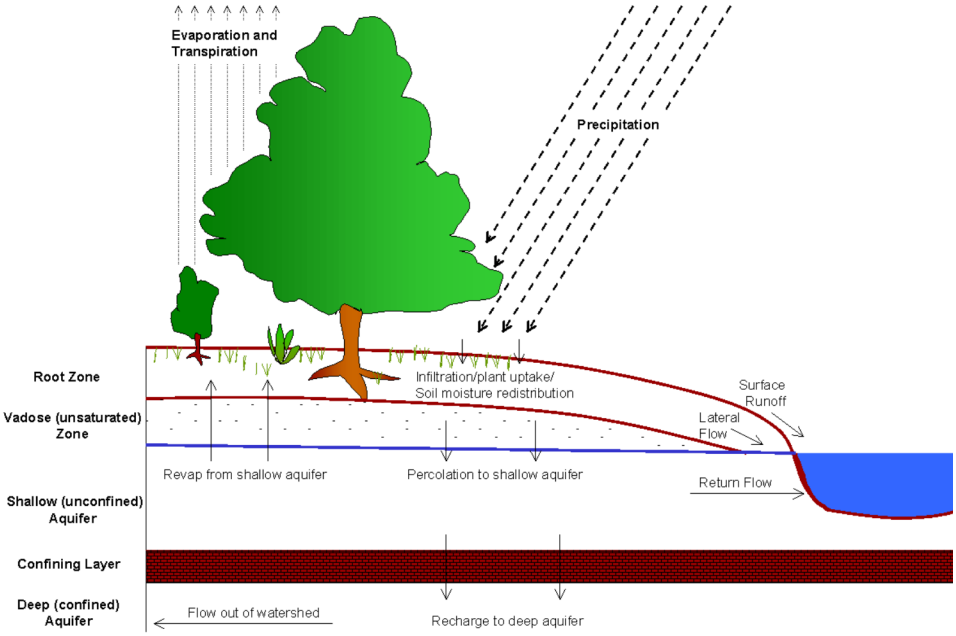


Figure 3.1.: Schematic summary of land phase hydrology as represented in SWAT. Management practices that affect the water balance but are not depicted include irrigation, agricultural tile drainage as well as consumptive water use. The figure is taken from Neitsch et al. (2011).

transpiration on day i, $Q_{s,i}$ is the surface runoff generated on day i, $Q_{l,i}$ is the lateral subsurface flow generated on day i, and $Q_{g,i}$ is the return flow from groundwater generated on day i. L_i is the amount of water lost from the system on day i because of percolation to a confined aquifer. Water storage S encompasses soil water and groundwater, as well as other more transient water reservoirs like snow cover or rainfall intercepted by the vegetation canopy. The quantities in brackets represent the daily inputs and outputs of water that result in a net change of storage. In addition, SWAT is also capable of simulating management practices that increase or deplete water storage, like irrigation, tile drains or consumptive water use (for example groundwater extraction). SWAT assumes all the quantities in Equation 3.1 to be expressed in mm H_2O.

Daily precipitation R_i is provided as measured time series data. Alternatively, SWAT comes with the built-in WXGEN weather generator which is able to simulate daily precipitation. The weather generator is a stochastic model parameterized with statistical data of the regional climate. It may also be used to impute gaps in the measured daily precipitation time series. Further meteorological inputs are needed for estimating daily evapotranspiration. SWAT provides several routines for calculating potential evapotranspiration $E_{p,i}$, the most data demanding of which being the Penman-Monteith method (Monteith 1965). Calculation of $E_{p,i}$ with the Penman-Monteith method requires daily time series of air temperature, relative humidity,

solar radiation and wind speed. As with daily precipitation, these may also be simulated by the built-in weather generator. SWAT estimates daily actual evapotranspiration $E_{a,i}$ by first evaporating intercepted rainfall held in the vegetation canopy. Depending on the amount of available water, the remainder of the atmospheric demand defined by $E_{p,i}$ is satisfied by plant transpiration and direct evaporation from the soil. Additionally, sublimation may occur in case a snow cover is present in the HRU.

The remaining water may either infiltrate into the soil or contribute to surface runoff. SWAT provides several routines for estimating surface runoff. The most commonly used routine is the curve number (CN) approach developed by the USDA Soil Conservation Service (SCS). The CN approach is empirical in that it relates direct runoff to previously fallen precipitation based on monitoring data from small stream catchments and hillslope plots (Mockus 1972). The formulation used in SWAT is

$$Q_{s,i} = \frac{(R_i - 0.2S)^2}{(R_i + 0.8S)} \tag{3.2}$$

where S is called the retention parameter (in mm H_2O). Runoff will only occur if $R_i > 0.2S$, otherwise $Q_{s,i}$ is set to zero. The retention parameter calculates as follows:

$$S = 25.4 \left(\frac{1000}{CN} - 10 \right) \tag{3.3}$$

The dimensionless curve number CN is an empirical parameter. As Equation 3.3 has originally been derived to calculate the value of S in inches, the constant 25.4 serves to convert the result to mm.

Tabulated CN values have been published for different groups of soils, land uses and management practices (Cronshey 1986). These values can be used to let the SWAT curve number parameter vary in space according to the local conditions represented in each HRU. SWAT provides two different modes to adjust the effective curve number for each simulation day. The traditional approach is to let the daily CN vary with soil moisture. As this approach tends to overpredict runoff from shallow soils, the option to vary the daily CN with evapotranspiration has been introduced to the model (Yen et al. 2015). By using the second option, the effective CN is supposed to become more dependent on antecedent climate rather than current soil water conditions. As an optional feature, the effective curve number may additionally be adjusted by slope. Although there have been attempts to link CN runoff predictions to hydrological theory, it is important to note that the curve number approach by itself is purely empirical

and generalizes over different runoff generating mechanisms such as inflitration or saturation excess (Garen & Moore 2005). As an alternative to the curve number approach, SWAT provides the process-based Green-Ampt method to estimate infiltration and surface runoff. The Green-Ampt method requires subdaily instead of daily precipitation time series as input data, however.

Water neither evaporating nor contributing to surface runoff infiltrates into the soil profile, which may consist of up to ten different layers. For each layer, water not frozen and in excess of the field capacity may either percolate to the layer below or enter the stream network; the latter either as lateral subsurface flow or via tile drainage pipes (if these are present in the HRU). Water can only percolate if the receiving soil layer below is not saturated. The amount of water that leaves a soil layer by percolation is calculated as

$$w_{perc,ly} = SW_{ly,excess} \cdot \left(1 - e^{\left\lfloor \frac{-\Delta t}{TT_{perc}} \right\rfloor}\right) \tag{3.4}$$

where $w_{perc,ly}$ is the amount of water percolating out of the respective soil layer on a given day and $SW_{ly,excess}$ is the drainable volume of water held in the soil layer (both in mm H_2O). TT_{perc} is the travel time for percolation (in h) and Δt is the simulation time step (24 h). The travel time is given by

$$TT_{perc} = \frac{SAT_{ly} - FC_{ly}}{K_{sat}} \tag{3.5}$$

where SAT_{ly} and FC_{ly} are the water contents of the layer (in mm H_2O) at saturation and field capacity, respectively (the difference of which being the maximum drainable volume of water). K_{sat} is the saturated hydraulic conductivity of the soil layer (in mm/h). Water that percolates past the lowermost soil layer enters the vadose zone and ultimately becomes groundwater recharge. SWAT allows defining an impervious zone below the soil profile, in which case percolation out of the soil profile is reduced depending on the distance of the impervious zone to the bottom of the soil profile. As a consequence, a perched water table may build up in the profile, inhibiting percolation of water from above by successively saturating soil layers from bottom to top.

Lateral subsurface flow (or synonymously interflow) occurs when percolating water meets an impermeable or nearly impermeable layer in the soil profile. Water ponding above the impermeable layer may subsequently move along a hillslope gradient and enter the stream network.

In SWAT, each individual soil layer may contribute lateral subsurface flow. SWAT calculates the daily amount of subsurface flow from a specific layer as

$$Q_{ly,i} = \frac{0.024 \cdot SW_{ly,excess} \cdot K_{sat} \cdot slp}{\phi_{ly,d} \cdot L_{hill}} \tag{3.6}$$

where $Q_{ly,i}$ is the daily lateral subsurface flow calculated for the soil layer. The hillslope gradient slp of the interface between the permeable and impermeable soil layer is usually assumed to be equal to the gradient of the soil surface. It is expressed as elevation per unit length of lateral distance (m/m). $\phi_{ly,d}$ is the drainable porosity of the soil layer (in mm/mm), which is equivalent to the fraction of porosity filled with air when the soil layer is wetted to field capacity. L_{hill} is the hillslope length in m and the constant 0.024 serves to convert the result from hourly to daily lateral flow in mm H_2O. The overall daily lateral subsurface flow $Q_{l,i}$ for the entire HRU is obtained by summing $Q_{ly,i}$ over all soil layers: $Q_{l,i} = \sum_{ly=1}^{n} Q_{ly,i}$.

Additionally, water may be removed from the soil via tile drainage pipes. Several different routines to estimate tile flow have been incorporated in SWAT and tested over the years (Du et al. 2005, Moriasi et al. 2012, Guo et al. 2018). The simplest tile flow routine detailed in Du et al. (2005) calculates daily tile flow as

$$Q_{t,i} = \frac{h_{wtbl} - h_{drain}}{h_{wtbl}} \cdot SW_{excess} \cdot \left(1 - e^{\frac{-\Delta t}{t_{drain}}}\right) \tag{3.7}$$

where $Q_{t,i}$ is the daily tile flow from the HRU and $SW_{excess} = \sum_{ly=1}^{n} SW_{ly,excess}$ is the drainable soil water content for the entire soil profile. Both quantities are measured in mm H_2O. t_{drain} is the time needed to drain the soil from saturation to field capacity (in h) and Δt is the simulation time step (24 h). h_{wtbl} is the current height of the water table above an impervious zone below the soil profile, and h_{drain} is the height of the tile drains above the impervious zone (both in mm). Tile flow will only occur if $h_{wtbl} > h_{drain}$, otherwise $Q_{t,i}$ is set to zero. h_{wtbl} calculates as follows:

$$h_{wtbl} = \frac{SW_{excess}}{(SW_{Sat} - SW_{FC})\,(1 - \phi_{air})} \cdot d_{imp} \tag{3.8}$$

SW_{Sat} is the water content in the entire soil profile at saturation, SW_{FC} is the water content of the entire soil profile at field capacity (both in mm H_2O), ϕ_{air} is the porosity of the entire soil profile currently filled with air (in mm/mm) and d_{imp} is the depth of the impervious zone from the soil surface (in mm).

Groundwater recharge from percolating water reaches the groundwater zone with a delay and is subsequently directed into one of two aquifers. Partitioning of recharge between the two aquifers is governed directly by a model parameter. The shallow aquifer is in contact with surface water bodies and allows return flow of groundwater into the stream network. Return flow from the shallow aquifer is dependent on a threshold water level. If the water level in the shallow aquifer is below this threshold parameter, return flow is set to zero. Otherwise it calculates as

$$Q_{g,i} = Q_{g,i-1} \cdot e^{-\alpha_{gw} \cdot \Delta t} + w_{rchrg,sh} \cdot \left(1 - e^{-\alpha_{gw} \cdot \Delta t}\right) \tag{3.9}$$

where $Q_{g,i}$ is the return flow from the shallow aquifer on day i, $Q_{g,i-1}$ is the return flow from the shallow aquifer on the preceding day $i-1$ and $w_{rchrg,sh}$ is the recharge that reaches the shallow aquifer on day i (all in mm H_2O). α_{gw} is the base flow recession constant for the shallow aquifer (in 1/d) and Δt is the simulation time step (1 d). Apart from return flow, the shallow aquifer also allows "revap" of groundwater to the soil profile to compensate soil moisture deficiencies during dry periods. In contrast to the shallow aquifer, the deep aquifer is assumed to be confined and does not allow any water exchange with the stream network or soil profile.[1] Water directed to the deep aquifer is thus considered lost from the system and consitutes the loss term L_i in Equation 3.1.

Besides land phase hydrology, other processes simulated on the HRU level are plant growth, management operations, nutrient cycling, erosion and pollutant mobilization. SWAT applies the heat unit approach to simulate maximum daily plant growth. Heat units are defined as the number of degrees the daily average air temperature is above a species-dependent threshold value. The threshold value is called the plant's base temperature, below which plant growth is not possible. The heat unit approach assumes that specific stages of plant development can be linked to the number of heat units a plant has accumulated during the current growth cycle, provided there is sufficient access to water and nutrients. SWAT uses inidcators like biomass and leaf area index (LAI) to track plant growth. Optimal development of these indicators is directly governed by the amount of received heat units. Their actual development may be inhibited due to water, temperature, nitrogen or phosphorus stress. Management operations can be scheduled either by heat units or by calender date. In the first case, the management operation is triggered once a specified threshold value of accumulated heat units is exceeded. In the second case, the management operation is performed on a specified date. SWAT supports

[1] Later SWAT revisions actually allow return flow from the deep aquifer. Computation of return flow from the deep aquifer is analogous to Equation 3.9. Deep return flow is governed by a seperate base flow recession constant $\alpha_{gw,d}$. In contrast to shallow return flow, deep return flow is not restricted by a threshold water level.

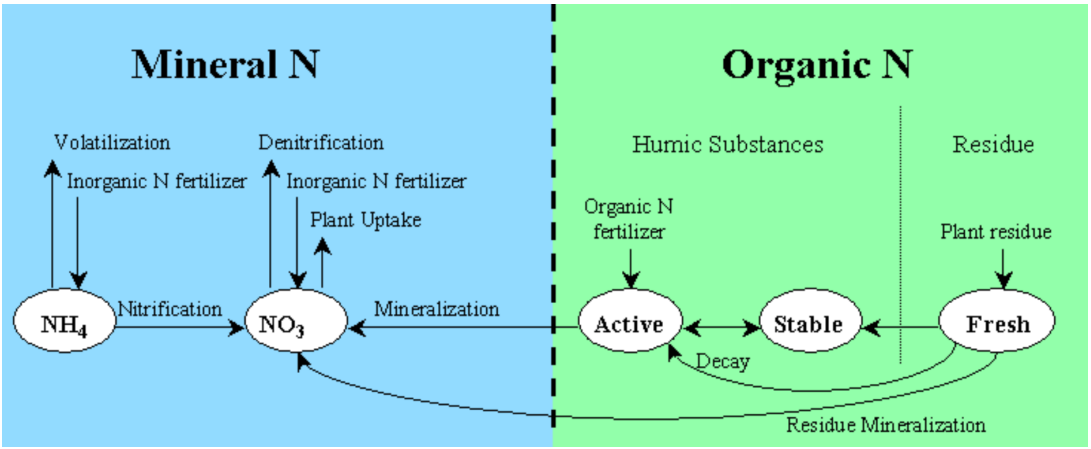


Figure 3.2.: Soil nitrogen pools and conversion processes in SWAT. Partitioning between active and stable organic N depends on the current mass ratio between both pools. Decay of fresh organic N and mineralization require soil temperatures above 0 °C. Note that SWAT adds all mineralized N directly to the NO_3^--N pool. Volatilization and nitrification occur only at temperatures above 5 °C. Denitrification requires wet (anaerobic) conditions. Uptake from the root zone by growing vegetation is governed by plant N demand, which is dependent on species and growth stage. Mineral N may also be added to the soil profile by atmospheric deposition (not depicted). The figure is taken from Neitsch et al. (2011).

a wide range of management operations for different land uses. Agricultural management operations include planting and harvesting of crops, plowing as well as fertilizer or manure application (among others).

Considering nutrients, SWAT models both nitrogen and phosphorus. The nitrogen cycle is simulated in the soil profile and the shallow aquifer. SWAT assumes instantaneous mixing for all soil and groundwater compartments. Figure 3.2 gives an overview of the five soil nitrogen pools included in SWAT as well as the major conversion processes between them. Details are given in the corresponding caption. Nitrogen bound as nitrate (NO_3^--N) and ammonium (NH_4^+-N) constitute the two mineral nitrogen pools. The other three N pools represent different forms of organic nitrogen: The fresh organic nitrogen pool is associated with plant and microbial biomass, whereas the active and the stable organic nitrogen pool are associated with soil humus. Soil nitrogen may be transported into the aquatic environment either via erosion or with moving soil water:

1. SWAT uses the Modified Universal Soil Loss Equation (MUSLE) to model soil erosion and sediment delivery (Williams 1975). Eroded nitrogen is taken from the organic N pool at the soil surface and transported to the nearest stream segment. Corresponding organic N loads are calculated directly from the amount of transported sediment and the

concentration of organic nitrogen in the top 10 mm of the soil profile, modified by an enrichment ratio accounting for the preferential transport of clay particles.

2. As nitrate is highly mobile in soils due to anion exclusion at the interface between soil solution and soil matrix, nitrogen transported with moving water is taken from the NO_3^--N pool. As such, it may either move downwards with percolation or laterally with surface runoff, interflow or tile flow. In the former case, nitrogen may move past the bottom of the soil profile and leach to the groundwater, whereas in the later case it is transported to surface water bodies.

Transported NO_3^--N loads are computed by simply multiplying NO_3^--N concentrations in the mobile soil water fraction with flow rates, where mobile soil water is the sum of moving soil water volumes calculated for the respective simulation day. For example, NO_3^--N transported with surface runoff calculates as

$$L_{s,i} = \beta \cdot C_{s,i} \cdot Q_{s,i} \tag{3.10}$$

where $L_{s,i}$ is the NO_3^--N load lost from the HRU with surface runoff on day i (in kg) and $C_{s,i}$ is the NO_3^--N concentration in the mobile soil water of the top 10 mm on day i (in kg/mm). β is the dimensionless nitrate percolation coefficient, a conceptual model parameter ranging from 0.01 to 1 governing the share of soil NO_3^--N available for transport with surface runoff. The NO_3^--N loads transported with the remaining fractions of moving soil water are calculated analogously. However, note that β is not considered for NO_3^--N loads in lateral flow below 10 mm from the soil surface, tile flow and percolation.

Additionally, NO_3^--N may enter surface water bodies with groundwater flow from the shallow aquifer. In order to account for nitrate fluctuations in groundwater over time, SWAT explicitly models nitrate dynamics in the shallow aquifer. The NO_3^--N pool there grows with nitrate reaching the aquifer in shallow groundwater recharge and shrinks with nitrate leaving the aquifer in groundwater flow to surface water bodies or "revap" to the soil profile. Furthermore, NO_3^--N levels in the shallow aquifer may be reduced in response to depletion processes like denitrification. Instead of modeling different depletion processes individually, SWAT lumps them together by means of a generic nitrate half life for groundwater which may be adjusted by the user as a model parameter. Hence, groundwater NO_3^--N depletion in SWAT is governed by first order kinetics.

Once the different runoff components as well as sediments, nutrients and other pollutants have been directed to the receiving stream segment, they are subject to the in-stream processes modeled as part of the routing phase. Two alternative methods to model streamwater routing are provided: The variable storage and the Muskingum method. Both essentially treat the stream network as a cascade of discrete storage elements - the individual stream segments. Downstream routing of water is dependent on current storage in the segments. Additionally, the model accounts for evaporation and transmission losses as well as bank storage. Transmission losses can only occur on simulation days on which the respective reach does not receive any groundwater; their magnitude is directly dependent on the hydraulic conductivity of the channel bed. If they occur, they are partitioned between bank storage, from where they can re-enter the stream with a certain delay, and the deep aquifer, where they are permanently lost from the modeled system.

Sediment and pollutants are routed along with the stream water. Nutrients are tracked both dissolved in the stream water and adsorbed to the transported sediment. The latter fraction is allowed to be deposited to the channel bed. In-stream nutrient cycling is modeled based on equations adapted from the stream water quality model QUAL2E (Brown & Barnwell 1987). These include the stepwise transformation of organic N to NH_4^+-N and then to NO_3^--N (via NO_2^--N), as well as incorporation of mineral N into algal biomass and re-release of organic N as a consequence of decomposition. Relevant state variables like water depth are estimated on the assumption of a trapezoidal cross section profile. In-stream denitrification is not represented, neither in the water column nor in the channel bed. Hence, deposition with settling particles and subsequent burial is the only permanent N loss process included in the SWAT routing phase. Importantly, this is true both for the stream segments and for ponded water bodies like reservoirs.

3.1.3. Model Capabilities and Limitations

The research aim pursued in this thesis (see Section 1.3) requires a computer model that is capable of simulating long-term catchment hydrology, agricultural land management, plant growth as well as nitrogen cycling and transport. SWAT combines all of these features and enables the user to examine the impact of different management practices on the simulation results. As a dynamic model, SWAT explicitly accounts for temporal effects like elevated nitrogen emissions from rainfall events directly after fertilizer applications (Schürz et al. 2017). With regard to usability, there is extensive model documentation in the form of the SWAT Theoretical Documentation (Neitsch et al. 2011) and the SWAT Input/Output Documentation (Arnold et al. 2013). In addition, the open source model code makes SWAT completely transparent in

terms of model equations and simulation routines. As the model is free to use for everyone, a large and active community of scientists and practitioners has grown around SWAT, together with a rich software ecosystem encompassing i.a. interfaces for model setup or environments for model calibration. Finally, the widespread use of the model in catchments around the world has generated a large body of published SWAT studies that the present thesis can draw from.

However, these advantages of the SWAT model are counterbalanced by several drawbacks, some of which being of particular concern with regard to non-point source pollution modeling. First, HRUs in SWAT possess no exact position within each subbasin and do not interact with each other. More specifically, there is no explicit representation of water flow and pollutant transport across different HRUs. Instead, water and pollutants mobilized in any HRU are instantly transferred to the stream segment and outlet of the respective SWAT subbasin. Their transport, deposition and retention in the landscape are not explicitly accounted for (Arnold et al. 2010, Rathjens et al. 2015, Bieger et al. 2017). The model is therefore unable to directly represent how runoff and pollutant transport are affected by the small-scale topography within each subbasin (e.g. upslope-downslope relationships, divergence and convergence, surface depressions, hydrologic connectivity, variable source areas). This however is important for modeling the spatiotemporal variability of non-point source surface water pollution, precisely identifying CSAs inside individual subbasins and assessing the impact of local mitigation measures such as riparian buffer zones and filter strips (Arnold & Fohrer 2005, Gassman et al. 2007, Krysanova & Arnold 2008, Krysanova & White 2015, Bieger et al. 2017).

Second, the use of the CN method in most SWAT studies futher complicates the interpretability of the simulation results (Garen & Moore 2005). As stated before, the CN method is a purely empirical approach to estimate direct runoff for individual flood events in small catchments. It makes no assumptions and gives no indication about the runoff generating processes involved in the predicted stormflow response. Hence, the CN method is unable to distinguish between different runoff source areas and flow paths, which is particularly important to predict erosion, pollutant mobilization and pollutant transport in the landscape (Garen & Moore 2005). Furthermore, SWAT misapplies the CN method to estimate (hortonian) surface runoff that individual HRUs contribute to daily streamflow. Other forms of direct runoff like interflow and tile flow are calculated separately, although they should already be subsumed in the runoff calculated by the CN method. As a result, SWAT tends to overpredict surface runoff and calibration of SWAT models often leads to the downward correction of curve numbers (see for example Douglas-Mankin et al. (2010) or Tuppad et al. (2011)). Additionally, it is unclear if the CN method as a flood event model is suited to estimate ordinary daily flow volumes as it is used in SWAT (Garen & Moore 2005).

Finally, SWAT possesses a plethora of model parameters. The SWAT Input/Output Documentation (Arnold et al. 2013) lists hundreds of parameters that can be modified in SWAT, for example affecting the simulated water balance, soil and groundwater processes, plant growth, land and water management, erosion, nutrient cycling, pollutant transport and transformation as well as streamflow routing and water quality. In addition, many of these parameters are distributed within the modeled catchment and vary between different SWAT subbasins, HRUs as well as soil depth zones, amplifying the effective number of model parameters that need to be calibrated. Hence, SWAT is often considered to be overparameterized (Beven 2012, p. 222), making it difficult to disentangle the effects of the unknown model parameters during model calibration and subsequently identify the parameter sets that yield realistic emission scenarios for the examined catchment (equifinality). As argued in Section 1.2, equifinality can be detrimental to many typical aims of catchment-scale pollution modeling. Whittaker et al. (2010) as well as Her & Chaubey (2015) argue that SWAT offers enough physical constraints to sufficiently regularize the model parameters. However, in the opinion of the author of this thesis, this remains doubtful, particularly considering the fundamentally empirical nature of the CN method that is central to runoff simulation in most SWAT applications.

3.2. Model Evaluation

3.2.1. Performance Criteria

Quantitative performance (or "*goodness-of-fit*") criteria summarize model agreement with observed data in a single number. By means of this number, they can be used to establish a ranking of different models. In comparison with purely qualitative or graphical methods, they allow for a less subjective assessment of model performance and may also be used to construct objective functions for automated model calibration (see Subsection 3.2.4). It has to be noted, however, that performance criteria differ in how exactly they measure agreement with observed data. Different criteria may thus yield different or even contradictory assessments of model performance. It is therefore good practice to make complementary use of multiple performance criteria when evaluating modeling results (Krause et al. 2005). Moriasi et al. (2007) and Moriasi et al. (2015) give an overview of many - albeit not all - performance criteria commonly applied in hydrological modeling. This section lists the criteria used throughout the present thesis, states the equations to compute them and briefly summarizes their individual strengths and weaknesses. It looks at them mainly through the lense of hydrological modeling, although many of them are well-known in other application contexts as well. For more details, please refer to the literature cited in the following paragraphs.

In hydrological modeling, the coefficient of determination R^2 is commonly used as a measure of collinearity between the predicted and observed data:

$$R^2 = r^2 = \left[\frac{\sum_{i=1}^{n}(O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2}\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2}} \right]^2 \tag{3.11}$$

r is Pearson's correlation coefficient between predicted and observed data. O_i denotes a specific data point in the time series of observed values and $\bar{O}$ is the mean of all observed values. Correspondingly, P_i denotes a specific data point in the time series of predicted values and $\bar{P}$ is the mean of all predicted values. n is the number of data points in either of the two time series (observed as well as predicted). R^2 ranges between 0 and 1, with the former indicating no correlation and the latter indicating perfect (linear) correlation between observed and predicted values. Legates & McCabe (1999) point out that as a measure of collinearity, R^2 is unable to detect "*additive and proportional differences*" between observed and predicted values (i.e. differences in magnitude as well as variability). Consequently, even sharp disagreement in the means and variances of the observed and predicted data may result in a value of $R^2 = 1$ as long as there is perfect linear correlation between both data sets. Futhermore, R^2 is sensitive to outliers and shows considerable sampling uncertainty when applied to typical hydrological time series data with high levels of variability and skewness (Legates & McCabe 1999, Barber et al. 2020). Despite these shortcomings, R^2 is widely used as a performance metric in hydrological modeling studies of the past and present (Moriasi et al. 2015). A good R^2 score usually indicates co-occurrence of peaks in observed and predicted time series of streamflow or pollutant loads, which is an aspect of particular concern in many hydrological modeling studies.

The percent bias (PBIAS) metric is a relative measure of over- or underestimation (Gupta et al. 1999). It is expressed in % and calculates as:

$$PBIAS = \frac{\sum_{i=1}^{n} P_i - O_i}{\sum_{i=1}^{n} O_i} \cdot 100\% \tag{3.12}$$

PBIAS may be interpreted as the ratio of residual mean (i.e. bias) to the observed mean or - equivalently - the relative deviation of the predicted mean from the observed mean. Hence, a value of 0 % represents the optimal result for PBIAS. In the formulation presented in Equation 3.12, a positive PBIAS indicates a tendency of the predictions to overestimate the observations, whereas a negative value indicates a tendency to underestimate the observations. As R^2, PBIAS is a widely used and frequently reported performance measure in hydrological modeling stud-

ies. It is particularly suited to quickly identify bias in continuous long-term simulations. As a relative measure of model error, PBIAS may also be used to compare model performance across different catchments or time periods. As it only accounts for differences in the means of observed and predicted data, PBIAS is unable to assess how well a model reproduces temporal patterns like timing of peaks, seasonality or long-term trends in the observations. Furthermore, PBIAS is insensitive to differences in variance and may give misleading results if the model overpredicts as much as it underpredicts. In this case $PBIAS \approx 0\%$ in spite of model residuals which may be large in absolute value.

The mean squared error (MSE) overcomes this limitation by squaring the residuals.

$$MSE = \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n} \tag{3.13}$$

The MSE is an absolute (rather than relative) measure of the average deviance of the predictions from the observations. Hence, its value depends on the unit of the considered quantity and ranges between 0 and ∞, with 0 being the optimal result. It does not distinguish between over- and underprediction. Due to the squaring of the residuals, MSE exhibits a marked sensitivity to outliers (Legates & McCabe 1999, Krause et al. 2005). For the often right-skewed time series in hydrology (e.g. streamflow), prediction outliers typically co-occur with large values. As a result, time series peaks tend to have a disproportionately strong effect on MSE scores, which may however be a desirable property for many purposes in hydrology.

Two related criteria are the mean absolute error (MAE) and root mean squared error (RMSE).

$$MAE = \frac{\sum_{i=1}^{n}|P_i - O_i|}{n} \tag{3.14}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \tag{3.15}$$

Both MAE and RMSE share many of the properties described above for the MSE. However, they express model error in the unit of the observed quantity and thus provide an interpretation that is more intuitive as well as easier to communicate. Since the MAE applies absolute instead of squared residuals, MAE scores usually are below their RMSE counterparts and are less sensitive to outliers (Legates & McCabe 1999). To a certain degree, the spread between MAE and RMSE may serve as an indicator for the presence of outliers in the data (Ritter & Muñoz-Carpena 2013).

The Nash-Sufcliffe-Efficiency (NSE) has emerged as one of the most widely applied performance criteria in hydrology (Moriasi et al. 2015) and has originally been proposed specifically for use in hydrological modeling (Nash & Sutcliffe 1970).

$$NSE = 1 - \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \tag{3.16}$$

A popular interpretation of Equation 3.16 is that NSE is a normalized version of MSE designed to vary between $-\infty$ and 1 (Gupta et al. 2009).[2] A negative NSE score indicates that the model MSE is larger than the variance in the observed data meaning that the mean of the observations outperforms the model as a predictor. This inherent benchmark is one of the appealing features of the NSE as a performance metric. A value of 1 is the best possible NSE score, indicating a perfect fit to the observed data. As a normalization of MSE, the NSE inherits the drawbacks coming with squared residuals, e.g. sensitivity to outliers, high sampling uncertainty and disproportionate emphasis on model error in peak values (Legates & McCabe 1999, Krause et al. 2005, Ritter & Muñoz-Carpena 2013, Lamontagne et al. 2020, Clark et al. 2021).

NSE (and by extension MSE) may be decomposed into a bias, variance and correlation component (Gupta et al. 2009). NSE thus gives a comprehensive performance rating by considering all possible error components at once. However, Gupta et al. (2009) show that the different components interact when calclulating NSE scores and that this interaction may yield misleading results. First, NSE puts different weight on the bias component depending on the variance in the observed data. Second, NSE systematically favours models that underestimate the variance in the observed data. Gupta et al. (2009) propose the Kling-Gupta-Efficiency (KGE) as an alternative performance metric that also includes all error components but avoids interaction between them.

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \tag{3.17}$$

r is Pearson's correlation coefficient and represents the correlation component. $\alpha = \frac{var(P)}{var(O)}$ is the variance component expressed as the ratio of predicted variance to observed variance, whereas $\beta = \frac{\bar{P}}{\bar{O}}$ is the bias component expressed as the ratio of predicted mean to observed

[2] Alternatively, (positive) NSE values may be interpreted as the fraction of explained variance in the observations. This is equivalent to the generalized definition of the coefficient of determination R^2 (see for example James et al. (2013), p. 70). NSE may therefore be seen as the R^2 score of the hydrological model, measuring model fit to the observed data. In contrast, the R^2 defined in Equation 3.11 represents the coefficient of determination for the simple linear regression model between the observed and predicted data fit by minimizing the residual sum of squares.

mean. Imagining the space spanned by the three components, the square root in Equation 3.17 is the Euclidian distance to the point of perfect agreement between observations and predictions at $r = \alpha = \beta = 1$. By substracting the Euclidian distance from 1, the KGE is made to fall on the interval from $-\infty$ to 1, where a value of 1 indicates the best possible result. Despite the analogy implied by their identical ranges, KGE and NSE are not directly comparable, meaning that high KGE may correspond to low NSE scores and *vice versa*. In particular, the KGE does not share the inherent benchmark property of the NSE at value 0 (Knoben et al. 2019). Negative KGE scores do not necessarily indicate "*unacceptable*" model performance, whereas some distinctly positive KGE scores may be associated with distinctly negative NSE scores. Consequently, modelers should be cautious not to transfer their habitual interpretation of NSE scores to KGE scores (Knoben et al. 2019). The complex relationship between NSE and KGE is investigated in more detail within Knoben et al. (2019) as well as Lamontagne et al. (2020).

3.2.2. Latin Hypercube Sampling

Global sensitivity analysis (Subsection 3.2.3) and calibration of hydrological models (Subsection 3.2.4) often require sampling of large parameter spaces to perform a Monte Carlo (MC) analysis. For the present thesis, SWAT parameter spaces are sampled using Latin Hypercube Sampling (LHS) (McKay et al. 1979). LHS is a so-called quasi-random sampling technique. In order to obtain a Latin Hypercube sample of size n, the entire range of each model parameter is subdivided into n non-overlapping intervals with equal probability of occcurence. The interval boundaries are equidistant if model parameters are assumed to follow a uniform probability distribution. For all parameters, exactly one parameter value is drawn from each of the n intervals - either at random or by using the interval midpoints (Rajabi et al. 2015). The drawn values for the different model parameters are randomly combined to form n parameter sets.

LHS is more efficient than random MC sampling in that it generates a more stable result with the same number of sample points n (equivalently, it requires a smaller number of sample points to reach a stable result) (McKay et al. 1979, Helton & Davis 2003).[3] The superior efficiency of LHS comes along with several desirable properties of the associated sampling design. First, LHS tends to fill the parameter space more evenly than simple random sampling, which is prone to clumping of individual sample points (Helton & Davis 2003). Second, LHS guarantees that sampling points are spread over the entire range of each model parameter and do not conincide

[3] Sampling of the model parameter space leads to a set of simulation results that may be evaluated using one of the performance criteria described in Subsection 3.2.1 (e.g. NSE or KGE). Then each sample of parameter sets corresponds to a sample of values for the performance criterion. A "stable result" is reached when repeated sampling from the parameter space results in similar estimates for the mean, variance and possibly other distributional properties of the performance criterion.

when projected into a subspace with lower dimensionality (Pronzato & Müller 2012). Several variants of LHS have been developed to optimize its space filling properties (see for instance Rajabi & Ataie-Ashtiani (2014)). This thesis makes exclusive use of LHS in its original form as described above.

3.2.3. Sensitivity Analysis

Sensitivity analysis (SA) refers to the process of investigating how model output variability can be attributed to the variation of input factors (Pianosi et al. 2016). It is closely related to uncertainty analysis which deals with error propagation in the opposite direction - from the input factors to the model output. Quite often, the input factors of interest are the model parameters (either all of them or a subset). However, any other model element that can be changed before model execution may equally serve as an input factor for sensitivity analysis. This includes the meteorological forcing data used to run the model, the applied temporal as well as spatial resolution and even the model equations themselves. Due to the complexity of most catchment models and the diversity of potential input factors, it is generally not possible to assess model sensitivity analytically. Instead, sensitivity analysis usually relies on repeated model runs to approximate the model response surface across a multitude of different input factor combinations. Model response is commonly measured either as an aggregate output variable like mean streamflow (called a "*prediction function*") or a performance metric like NSE or KGE (called an "*objective function*") (Pianosi et al. 2016). Typically, SA results for hydrological models do not generalize well. Rather, they are strongly dependent on the specific catchment and time period under investigation, as well as design choices like included input factors and model parameters, applied parameter ranges and choice of evaluated model response (Shin et al. 2013).

Pianosi et al. (2016) name three main purposes that may be pursued with sensitivity analysis (note that this list does not claim to be exhaustive). *Screening* aims at identifying those input factors that have no measurable effect on the model output, *ranking* attempts to order input factors according to their relative contribution to output variability, and *mapping* tries to determine regions in the input variability space that produce a particular kind of output behavior. Each of these settings favors different methods of sensitivity analysis that require different numbers of model evaluations to reach a stable result. Yuan et al. (2015) and Pianosi et al. (2016) provide an overview of SA methods applied in the context of environmental modeling. Very broadly, they distinguish qualitative and quantitative as well as local and global methods. Qualitative methods often rely on visualizations, for example scatter plots ("*dotty plots*") of model response against parameter values. Quantitative methods assign numerical values ("*sensitivity indices*") to the input factors that provide a basis for direct comparison. Local SA

evaluates sensitivity by varying input factors *one-at-a-time* (OAT) around a fixed reference point in the input variability space. Results strongly depend on the chosen reference point and do not account for input factor interactions. In contrast, global SA assesses sensitivity across the entire input variability space, making it better suited to deal with factor interactions and avoid crucial type II error ("*false negatives*") in identification of influential factors than local SA (Saltelli et al. 2008, pp. 35-36). For global SA, input factors are often - but not always - varied *all-at-a-time* (AAT).

The Elementary Effect (EE) test is an efficient global SA method often used for parameter screening (Morris 1991). Compared with other global SA methods, it requires only a small number of model evaluations by using an OAT (instead of AAT) approach for parameter sampling. It is thus particularly suited for models with high computational cost or a large number of parameters (Saltelli et al. 2008, Pianosi et al. 2016). Briefly, the EE test repeats local sensitivity analyses at multiple reference points in the parameter space. An elementary effect EE is the quotient of finite differences that measures deviation in the model response to a single parameter perturbation from the reference point. The EE with regard to the i-th model parameter at the j-th reference point is defined as

$$EE_i^j = \frac{y(x_j^i, x_2^j, \ldots, x_i^j + \Delta, \ldots, x_m^j) - y(x_1^j, x_2^j, \ldots, x_i^j, \ldots, x_m^j)}{\Delta} \cdot c_i \tag{3.18}$$

y is the model response for the inserted parameter values $x_1^j, x_2^j, \ldots, x_m^j$, while Δ is a predefined perturbation of the i-th model parameter and m is the overall number of considered model parameters. c_i is a scaling factor to accommodate different units of measurement. Computing the EE at n different reference points in the parameter space and averaging the results provides a global sensitivity index μ_i for parameter i:

$$\mu_i = \frac{\sum_{j=1}^{n} EE_i^j}{n} \tag{3.19}$$

A second index σ_i computes the standard deviation of elemental effects and is usually interpreted as a measure of interaction of the i-th model parameter with all other included parameters:

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^{n} (EE_i^j - \mu_i)^2}{n}} \tag{3.20}$$

As the model response has to be evaluated once for every reference parameter set and again for every perturbed parameter, the EE test requires $n \cdot (m+1)$ model evaluations. A common recommendation in the literature is to set $n \approx 10$ (Saltelli et al. 2008, p. 273), although $n \approx 100$ probably is a safer option to guarantee stable sensitivity estimates in a screening setting (Sarrazin et al. 2016). Several variations of the EE test have been proposed since its initial formulation (Campolongo et al. 2007, Campolongo et al. 2011). These variations affect the choice of the n reference points, the sampling design around the reference points and the computation of the indices given in Equations 3.19 and 3.20 (for example, the use of absolute *EE*s in Equation 3.19 prevents that positive and negative values compensate for each other, see Campolongo et al. (2007)).

3.2.4. Model Calibration and Validation

Calibration of a catchment model is the process of adjusting the model so that the model output matches observed catchment behavior sufficiently well. Typically, this adjustment procedure is restricted to the model parameter values. However, it may equally extend to other input factors like the model equations and structure (Beven & Binley 1992). There are many different approaches to model calibration, with no standardized or generally accepted optimum procedure. Daggupati et al. (2015) criticize that model calibration often is "*ad hoc*, piecemeal, incomplete and inadequate" for meeting the requirements of the indended model use. Broadly, one can distinguish manual and automated calibration approaches. Manual calibration involves the incremental adjustment of model parameters or input factors by hand, guided by the modelers expertise and knowledge of the modeled catchment. As such, it allows the direct incorporation of qualitative ("*soft*") data into the calibration process as well as the intuitive examination of the model response to specific input adjustments. However, manual calibration is inherently subjective and irreproducible if not thoroughly documented. Furthermore, it is often time-consuming and labor-intensive for complex models with many possible combinations of parameter values or input factors (Arnold et al. 2012, Daggupati et al. 2015).

Advances in computer performance in the past decades have led to the increasing use of automated calibration approaches (Daggupati et al. 2015). The classical strategy in automated model calibration is the computer-aided use of optimization algorithms to improve the fit between simulated and observed data. Whereas manual calibration is achievable by visual inspection of the model fit, automated model calibration necessitates the definition of an objective function using quantitative performance criteria like those described in Subsection 3.2.1. At this point, it is important to note that while automated calibration provides a more objective and reproducible course of action than the manual approach, the results still are strongly de-

pendent on the applied optimization procedure (including the starting parameter values) and objective function (Abbaspour et al. 2017). Furthermore, the classical strategy for automated calibration typically yields exactly *one* optimum set of model parameters or input factors to fit the observed data[4], implying the existence of a "*true*" model that can be identified by rigorous optimization. In contrast, Beven & Binley (1992) argue that this assumption is wrong due to the multiple sources of uncertainty and potential model error that may interfere with the calibration process. In accordance with the idea of equifinality (see Section 1.2), they instead make the case for accepting many different models as potential predictors for the catchment response of interest.

To that effect, an alternative type of automatic model calibration makes use of MC analysis to identify entire sets of acceptable catchment models (Beven 2012, pp. 249 - 252). A prominent example is the GLUE technique proposed by Beven & Binley (1992) (see also Section 1.2). Model calibration based on MC analysis searches the parameter space for regions and parameter combinations that yield viable model performance. As long as no evidence to the contrary emerges, all of these models are considered to be more or less accurate predictors of the real system. In this manner, MC analysis also provides a means of assessing the uncertainty in the parameter values or the model output (assuming the respective uncertainty bands cover the "*real*" catchment behavior). Within this setting, there are basically two options to distinguish acceptable and unacceptable model performance. The first option is the application of a *limit of acceptability*, a performance threshold a model needs to exceed to be deemed *behavioral*, i.e. sufficiently representative of the observed catchment response (Beven 2006). The threshold value should be set in accordance with the intended use of the model and completely independent of any prior assessment of model performance. The second option for model selection is the identification of the Pareto-optimal set of parameterizations in a multi-objective calibration strategy (Gupta et al. 1998). The Pareto-optimal set includes those models that cannot improve on one performance criterion without deterioration in another criterion. Both options can be combined and may then complement each other, as illustrated by Efstratiadis & Koutsoyiannis (2010): Whereas the size of the Pareto-optimal set increases with each considered calibration objective (i.e. performance criterion), the combination of multiple limits of acceptability (one for each performance criterion) works towards a reduction in the number of included models.

Model validation is the process of testing the performance of a calibrated model against independent data not used during calibration. There are countless possibilities to allocate the available data to model calibration and validation. A systematic and exhaustive approach, that is however computationally too expensive for many purposes in catchment modeling, is *k*-

[4]The same is true for manual model calibration.

fold cross validation (James et al. 2013, Schoups et al. 2008). As models are typically calibrated on time series data (e.g. streamflow), it is common to instead simply split the available data by time periods (what Klemeš (1986) called the *split-sample* approach). There are no universal rules for the relative lengths of the calibration and validation periods in this setting (Daggupati et al. 2015). A commonly-held believe is that both the calibration and validation periods should cover the full range of conditions (e.g. wet and dry) under which the model is supposed to operate (Arnold et al. 2012). Although this may be a viable option if the intended model use is *prediction* alone, it is most likely insufficient in other cases. If the modeling purpose is gaining an *understanding* of the catchment response (particularly within scenario analysis), it may be more promising to deliberately validate the model in conditions not covered by the calibration data (what Klemeš (1986) called the *differential split-sample* approach).

As an alternative, data may be split in space instead of time to reserve data for model validation, e.g. between different stream gauges within the same river basin. For streamflow simulation in ungauged catchments, Klemeš (1986) proposed the *proxy-basin* approach, where calibration and validation are performed on two gauged catchments in the vicinity of the ungauged catchment of interest (assuming comparable conditions across all of these catchments). The validated parameters are subsequently transferred to the model for the ungauged catchment. Finally, a model may also be validated on types of model output that remained completely unconsidered during calibration. For example, for their large-scale study on multiple West African catchments, Poméon et al. (2018) used remote sensing data on evapotranspiration, soil moisture and total water storage to complement *split-sample* validation of SWAT on observed streamflow data.

4. Model Setup and Sensitivity to Different Model Specifications

4.1. Introductory Remarks

This chapter examines the effect of different input data sets and model subroutines (i.e. model structures) on the SWAT model output for the Swist river basin. It has originally been motivated by a recommendation of Abbaspour et al. (2017) to test all available model input data sets prior to model calibration. The tested input data include different georeferenced land use and soil data sets as well as alternative versions of the applied N emission time series for wastewater treatment plants (WWTPs). The tested model structures include different procedures to update the effective curve number (CN) in each HRU (either with soil moisture or evapotranspiration, see Subsection 3.1.2). With respect to agricultural management, they furthermore omit or include simulation of tile drains and crop rotation. The present chapter serves several purposes: First, it aims to make model setup transparent and reproducible. Second, it strives to test the robustness (i.e. sensitivity) of the model output towards various decisions made during model setup and raise awareness for their potential impact on the simulation results. Third, it attempts to make an informed decision on a particular model variant to achieve the wider goals of this thesis (see Chapter 1) and build confidence in the modeling results.

The chapter is organized as follows: Section 4.2 details the model setup and initial parameterization. It does so in a tiered manner, with each tier representing a further level of sophistication (and effort) in the model setup process. Alternating between the different land use and soil data sets as well as subroutines for CN adjustment (Subsection 4.2.1) requires no additional effort during model setup. In contrast, building a regression model to impute missing data in WWTP N emissions (Subsection 4.2.2) as well as identifying tile drained areas and introducing representative crop rotation (Subsection 4.2.3) are considerably more laborious. Section 4.3 compares the (uncalibrated) model performances originating from the different land use and soil data as well as CN adjustment routines. It does so for the base model setups (Subsection 4.3.2), for the model setups with regression-based imputation of WWTP N emissions (Subsection 4.3.3) as

well as for the model setups that include tile drains and crop rotation (Subsection 4.3.4). Section 4.4 provides a concluding discussion of the results. All considered SWAT models were built with help of the ArcSWAT interface, which is implemented as an extension of the ArcGIS geographic information sytem (GIS). Specifically, ArcSWAT version 2012.10_5.22 was used with ArcMap 10.5.1. As for the rest of the thesis, all simulations were run using SWAT2012, revision 670.

4.2. Model Setup and Parameterization

4.2.1. Base Models

Although all model variants differ in various aspects, their fundamental design is identical and summarized in this section. SWAT model setup requires the definition of a stream network along with the associated subbasins. An ESRI shapefile of the Swist river network was provided by the Erftverband. The shapefile consisted of line features usually running from one confluence or bifurcation point to the next. Additionally, features were split at the location of wastewater treatment plants, reservoirs and monitoring points like stream gauges. In some cases, lines were additionally split at their middle point, where otherwise they would have been disproportionately long. As a result, 153 line features were generated, each of them representing a stream segment in the SWAT models. The corresponding 153 subbasins were delineated using the ArcHydro toolbar extension for ArcGIS and a 10 m raster DEM. The DEM was put together from two distinct laserscanning data sets for North Rhine-Westphalia and Rhineland-Palatinate, respectively (State Office for Land Surveying NRW 2002, State Office for Land Surveying and Geoinformation RLP 2006). In order to ensure hydrologic connectivity of the calculated flow direction grid as well as flow accumulation along the actual stream locations, sinks in the DEM were filled and the stream shapefile was “burned” into the DEM applying the AGREE method (Hellweger 1997). The resulting subbasin areas were highly variable and skewed towards small values, ranging from 3 to 2052 ha, a mean of 189 ha and a standard deviation of 240 ha. Appendix A provides a map depicting the 153 stream segments and subbasins. Both the stream segments and their subbasins were fed directly into ArcSWAT to establish the river network topology and to calculate subbasin and stream parameters for SWAT.

In ArcSWAT, HRU definition within these subbasins is based on similarity in land cover, soil characteristics and slope. Two geodata sets for land cover in vector format were available for HRU definition and land use dependent model parameterization:

1. The first data set is the European CORINE[1] Land Cover (CLC) data set for Germany (German Federal Agency for Cartography and Geodesy 2012). The reference point in time is the vegetation period of 2012. The data set is derived from the Digital Land Cover Model Germany (LBM-DE), which in turn is derived from the Digital Basic-Landscape Model (Basis-DLM) of the German Authorative Topographic-Cartographic Information System (ATKIS). The underlying data has been generalized so that the minimum area for the mapped polygon features is 10 ha (in contrast to 25 ha for the Europe-wide data set). Each feature is assigned one of 37 different CLC classes used in Germany, which were identified with SWAT land cover classes using an ArcSWAT look up table.

2. The second data set is the ATKIS Basis-DLM for North Rhine-Westphalia with 2019 as the reference year (Geobasis NRW 2019). In order to account for the parts of the Swist catchment not located in North Rhine-Westphalia, the Basis-DLM was complemented with its 2001 counterpart for Rhineland-Palatinate (State Office for Land Surveying and Geoinformation RLP 2001). The Basis-DLM describes the spatial relationship of different topographical objects based on land register data, aerial photographs and other official information. The declared maximum spatial error is 3 m. As CLC data for Germany is in effect aggregated from Basis-DLM data, the latter contains considerably more detailed information on land cover and land use. In order to obtain a SWAT land use data set with complete coverage for the Swist river basin, selected Basis-DLM object types from the categories vegetation (including agriculture), settlement, traffic and water bodies were combined. These were identified with SWAT land use classes by constructing a look up table for ArcSWAT. Figure 2.2 on page 21 depicts the resulting land use map.

Similarly, two digital geodata sets for soils in vector format were available to define HRUs and determine SWAT soil parameters. Both data sets are different versions of the Authorative Soil Map for North Rhine-Westphalia in the scale of 1:50 000 (BK50):

1. The first data set is the BK50 from 2015 (Geological Survey NRW 2015). The BK50 is based on semiquantitative data on soil type, soil composition and texture reaching down to a depth of 2 m from the soil surface. For the map sheet covering the Swist river basin, this data has originally been collected in 2002. From this, the BK50 makes a quantitative assessment of soil parameters (e.g. contents of sand, silt and clay, field capacity as well as saturated hydraulic conductivity) in accordance with the third edition of the German Pedological Mapping Guidelines (AG Bodenkunde 1982). As the BK50 did not extend beyond the administrative border to Rhineland-Palatinate, soil types in this part of the

[1] Coordination of Information on the Environment

Swist catchment were derived from geological data. For the area directly adjacent to the border, the most common BK50 soil type for each parent rock material was determined and subsequently transfered to areas of comparable geology in Rhineland-Palatinate. All parameter values for these soils were estimated by assuming uniformity over the entire soil depth and assigning soil type specific averages over all depth zones from the BK50 data within the Swist catchment.

2. The second data set mostly consists of the BK50 for North Rhine-Westphalia from 2019 (Geological Survey NRW 2019). This BK50 version is based on the same soil data as that from 2015, however with updated data for the map sheet covering the Swist river basin. Furthermore, soil parameters have been estimated in accordance with the fifth (instead of third) edition of the German Pedological Mapping Guidelines (Ad-hoc AG Boden 2005), resulting in larger differences for some soil parameters like saturated hydraulic conductivity. In order to avoid having to extrapolate large parts of the NRW soil data to Rhineland-Palatinate (as for the 2015 soil data), the 2019 BK50 was complemented with the 2019 Soil Map Data for Rhineland-Palatinate in the scale of 1:50 000 (BFD50) (State Office for Geology and Mining RLP 2019). Still, while providing qualitative information on soil texture, the BFD50 was missing quantitative information on sand, silt, clay and rock contents needed for SWAT. These were instead estimated by averaging sand, silt and clay contents for each textural class in the BK50 data within the Swist catchment and subsequently applying these averages to the respective textural class of each depth zone in the BFD50 data. As sand, silt and clay contents (particles with a diameter of up to 2 mm) as well as rock contents (fragments with a diameter above 2 mm) are assumed to make up the entire soil dry matter in the BK50 data, rock contents in the BFD50 data were estimated by substracting the contents of sand, silt and clay from 100 %. All other important soil parameters provided by the BK50 were also present in the BFD50 data. Soil bulk density classes given in the BFD50 were translated to numerical estimates by by applying the class middle points as specified by Ad-hoc AG Boden (2005).

Most SWAT soil parameters could be derived from these two soil data sets, either by adopting them directly (barring unit conversions) or by calculating them based on the data. Sand, silt and clay contents were rescaled to add up to 100 %. Soil hydrologic groups as described in Neitsch et al. (2011) to set SCS curce number values were assigned based on mean sand and clay contents as well as soil profile depth, following a procedure described by Gijsman et al. (2007). As the 2015 version of the BK50 does not contain any information on soil bulk densities, they were estimated from sand, silt and clay contents based on tabulated values given by Bachmann et al. (2014, p.37).

Two SWAT soil parameters could not be inferred from the available data and had to be estimated based on other information. Soil albedo was set universally to a value of 0.13, as suggested for moist topsoils of brown and grey color by Gijsman et al. (2007). Organic carbon content (in that portion of soil dry matter that passes through a 2 mm sieve) was set as a function of land use, depth and rock content. The basis for these estimates was soil monitoring data from the second German Forest Soil Inventory and the German Agricultural Soil Inventory. The related monitoring campaigns were conducted from 2006 to 2008 and from 2011 to 2018, respectively. Results have been published for example in Grüneberg et al. (2014) or Wellbrock et al. (2017) for forest soils and in Poeplau et al. (2020) for agricultural soils[2]. Data on organic carbon in forest soils of North Rhine-Westphalia has been made available for this thesis by courtesy of the State Office for Nature, Environment and Consumer Protection in NRW (LANUV), whereas organic carbon contents in agricultural soils for all of Germany have been taken directly from Poeplau et al. (2020). Mean soil organic carbon contents were calculated for each combination of land use and depth zone included in these data sets and subsequently redistributed according to the depth zones defined in the BK50 data. As organic carbon contents were sampled not deeper than 90 cm for forest soils and 100 cm for agricultural soils, organic carbon contents were assumed to be zero below these thresholds. Similarly, soil organic carbon contents were assumed to be zero for all land uses other than forest, cropland, grassland, orchards and tree nurseries (essentially all urban land uses in the Swist catchment).

The two land use data sets and the two soil data sets outlined above give rise to a total of four different base models, each resting on a different combination of land use and soil data. Furthermore, since HRU definition is based on land use and soils, each of these base models corresponds to one of four distinct sets of HRUs. ArcSWAT additionally allows for slope to be considered in the HRU definition process. Using slope for HRU definition requires identification of areas with comparable slope characteristics. For this purpose, ArcSWAT provides a tool to classify the obtained slope gradients. Two slope classes were created for the present study: the first class includes slopes not larger than 2 % and the second class includes all steeper slopes. As slope gradients were calculated from one and the same DEM (also used for subbasin delineation), they were identical for all base model setups and did not contribute additional model variants with different underlying sets of HRUs.

ArcSWAT identifies each unique combination of land use, soil and slope class inside a subbasin as a seperate HRU, where minor area shares may be eliminated from the analysis by setting

[2]Here, agriculutural soils encompass (i) cropland soils, (ii) grassland soils, (iii) soils under permanent crops like orchards or vineyards, and (iv) organic soils of agricultural use. Since there are no organic soils present in the Swist catchment, organic carbon data for these soils has not been used in this thesis.

threshold values. Land uses, soils or slope classes that cover a percentage of subbasin area less than the chosen threshold values are ignored. Increasing the threshold values thus leads to a smaller number of HRUs in each subbasin, which in turn results in increased simulation speed at the expense of accuracy in spatial representation. For this study, the R package *topHRU* (Strauch et al. 2016) was used to identify threshold combinations with minimal spatial aggregation error for any given number of HRUs. For all four model variants, exclusion thresholds of 5 % for land use, 10 % for soil and 20 % for slope were found to give acceptable results both in terms of aggregation error and number of HRUs. As for subbasin areas, the distribution of HRU areas is of considerable width and highly skewed towards small values. Table 4.1 gives an overview over the four sets of HRUs and the associated statistics. The ATKIS land use data produces a distinctly larger number of HRUs than the CLC data, reflecting the higher level of spatial detail in the ATKIS data. Accordingly, the ATKIS data set also results in smaller mean (as well as median and minimum) HRU areas. According to Tuppad et al. (2011), the typical size of an HRU in SWAT ranges from about 50 to 500 ha. Both median as well as mean HRU areas are well below this range for all four model variants.

The meteorological time series that was applied to run the SWAT simulations covered eleven years in total, including a model warm-up period of four years (2008 to 2011) in addition to the seven simulation years actually evaluated in this thesis (2012 to 2018). Daily amounts of precipitation were obtained from the seven Erftverband weather stations that are located within the borders of the Swist catchment (Weilerswist, Heimerzheim, Rheinbach, Meckenheim, Steinbach, Todenfeld, Gelsdorf). The exact positions of the weather stations are mapped in Figure 2.2 on page 21. The applied rain gauges measure precipitation automatically by weighing of collected rain- and snowfall. The orifice area of the collecting unit is exactly 200 cm^2 (in accordance with standards of the World Meteorological Organization). Daily minimum and maximum air temperatures as well as daily averages of relative humidity were obtained from a

Table 4.1.: Statistics for the HRU sets asscociated with each of the four base model setups. Base model setups are defined by different combinations of underlying land use and soil data. The table column "HRUs" gives the total number of HRUs in the respective model setup. HRU areas have been calculated from subbasin areas and the percentages of area each HRU covers in its respective subbasin. SD denotes the standard deviation.

				HRU Area [ha]				
Model	Land Use Data	Soil Data	HRUs	Mean	SD	Min	Median	Max
1	CLC 2012	BK50 2015	1423	20.3	37.8	0.12	8.40	399
2	CLC 2012	BK50 2019	1462	19.8	36.9	0.16	8.12	386
3	ATKIS 2019	BK50 2015	1730	16.7	33.6	0.04	6.58	433
4	ATKIS 2019	BK50 2019	1824	15.8	32.7	0.06	6.09	417

subset of five of the weather stations (Weilerswist, Rheinbach, Meckenheim, Steinbach, Todenfeld, Gelsdorf). Capacitance-based sensors are applied to measure relative humidity, with air temperature measurement usually integrated in the same device.

Daily averages of wind speed and solar radiation are not measured at any of the seven Erftverband weather stations in the Swist river basin and had to be taken from other stations operated by the German Meteorological Service (DWD) outside of the catchment borders. Wind speed data was obtained from DWD weather stations in Bonn-Roleber and at Cologne Bonn airport. Solar radiation was estimated based on daily sunshine duration data collected at the DWD stations in Weilerswist-Lommersum, Bonn-Roleber and at Cologne Bonn airport. Both daily wind speed and sunshine duration time series contained a small number of days with no data: fourteen days for sunshine duration measured in Bonn-Roleber and well below ten days for all other time series. None of these data gaps was longer than a single day. They were filled with the respective values measured at the nearest other station. For sunshine duration, a single day was missing data at all three DWD stations and was instead estimated by linear interpolation between the preceding and successive day.

The applied method to convert from sunshine duration to solar radiation is taken from guidelines of the United Nations Food and Agriculture Organization (FAO) on estimating crop evapotranspiration from meteorological data (Allen et al. 1998). It is based on the Ångström equation:

$$R_s = \left(a_s + b_s \frac{n}{N}\right) \cdot R_a \tag{4.1}$$

where R_s is the solar radiation reaching earth's surface on a specific day in the year (in MJ/m^2) and n is the number of sunshine hours on the same day. N is the maximum possible number of sunshine hours (daylight hours) on that day in the year and calculates as $N = \frac{24}{\pi}\omega_s$, with ω_s being the respective sunset hour angle day expressed in radians. a_s and b_s are constants assumed to be 0.25 and 0.5, respectively. R_a is the extraterrestrial radiation on the evaluated day, which calculates as

$$R_a = \frac{1440}{\pi} \cdot G_{SC} \cdot d_r \cdot (\omega_s \sin(\phi) \sin(\delta) + \cos(\phi) \cos(\delta) \sin(\omega_s)) \tag{4.2}$$

where G_{SC} is the solar constant ($0.0820\,MJ\,m^{-2}\,min^{-1}$) and d_r is the inverse relative distance between sun and earth on the evaluated calendar day. ϕ denotes latitude and δ is the solar declination on the evaluated calendar day (both expressed in radians). See Allen et al. (1998) for more details.

In order to account for the four[3] wastewater treatment plants (WWTP) in the Swist catchment, Erftverband data on water quantity and quality in the WWTP effluents was used. Daily discharge of wastewater volumes into the streams was measured continuously and thus was available for almost every single day in the simulation period. Remaining gaps in the records were filled by fitting a random forest (RF) regression model (Breiman 2001) to the available discharge data for each WWTP. The random forest implementation in the *ranger* package (Wright et al. 2022) for the R programming language (R Core Team 2022) was used for this purpose. The applied predictor variables for the discharged wastewater volume on a specific day were the discharged volume on the preceding day as well as precipitation (measured at the treatment plant) on the same day and the seven preceding days. The number of discharge and precipitation values on preceding days used as predictor variables was chosen by optimizing the out-of-bag RMSE for each of the four treatment plants to avoid overfitting. The models were fit by constructing 500 bootstrapped samples from the training data and building regression trees on these samples with $\lfloor \sqrt{p} \rfloor = 3$ randomly chosen predictor variables at each node split ($p = 9$ is the overall number of predictor variables). The R^2 values associated with the out-of-bag model predictions (which may be understood as standardized measures of out-of-bag error in this context) were 0.68 for WWTP Flerzheim, 0.74 for WWTP Rheinbach, 0.71 for WWTP Miel and 0.66 for WWTP Heimerzheim. Built on an example from WWTP Miel, Figure 4.1 illustrates that the RF model is capable of filling even longer gaps in the discharge time series with plausible estimates.

Records of wastewater quality in the WWTP effluents were available for roughly one day per week (in accordance with the prevailing self-monitoring regulations in NRW) and encompassed mean concentrations of NO_3^--N , NH_4^+-N , NO_2^--N as well as total nitrogen (*Total Nitrogen bound*, TNb). The weekday the measurement was taken varied from one week to the next. Since the gaps between the measured values in the time series were considerably larger, more frequent and of variable length, an auto-regressive modeling approach as with discharged wastewater volumes could not be applied for N concentrations. In a first attempt, the missing values in the concentration times series were simply imputed with the mean of the known concentrations. This procedure is incapable of properly representing long-term trends, seasonality and other temporal patterns. Furthermore, it probably introduces bias to the data. Long-term average concentrations as well as average emission loads calculated from these concentrations are only preserved under the strong assumption that the unknown values are missing com-

[3] Initially, a fifth treatment plant - WWTP Loch - was operational in the catchment, treating the wastewater of approximately 800 residents (Kistemann et al. 2001). It was decommissioned in 2013. Due to its only small contribution to the overall wastewater load in the Swist catchment, WWTP Loch was considered negligible for the research aims of this thesis and thus not included in the SWAT simulations.

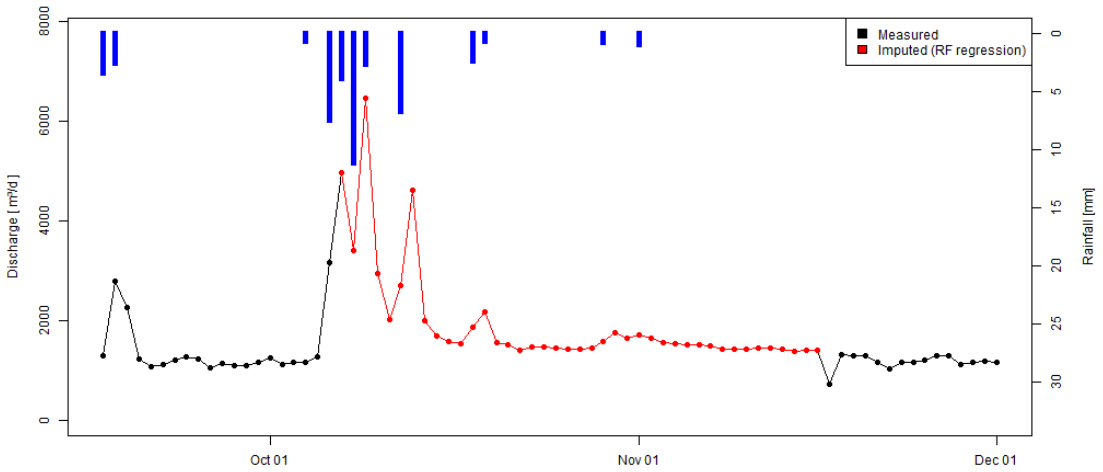


Figure 4.1.: Imputed values in the daily wastewater discharge time series for WWTP Miel. The figure depicts a section of the time series with an exceptionally long stretch of missing data in fall 2011. The black line and dots indicate measured wastewater discharge values while the red line and dots indicate imputed values estimated by random forest (RF) regression. Note how the estimated discharges react to daily rainfall inputs depicted as blue bars at the top (measured at the WWTP).

pletely at random. Subsection 4.2.2 describes a more sophisticated regression-based approach to impute the missing data. In both cases, the different N loads were calculated by multiplying measured or imputed daily mean concentrations with the respective daily wastewater discharges. Organic N concentrations were estimated by substracting measured or imputed mineral N concentrations (NO_3^--N , NH_4^+-N , NO_2^--N) from the measured or imputed total N concentrations. In the rare cases where this procedure yielded negative values, they were set to zero.

Of the two reservoirs in the Swist catchment, only the Steinbach reservoir was built into the SWAT models used in this study, whereas the smaller and unmanaged Madbach reservoir was deemed negligible. There are no binding operational rules in place to control outflow from the Steinbach reservoir, other than a regulatory limit for discharge from the principal spillway at 0.5 m^3/s (Erftverband personal communication). Consequently, it is generally difficult to account for the reservoir in hydrological models. For this study, daily discharge data recorded at the online gauging station in Kirchheim downstream of the dam was used to read outflows from the Steinbach reservoir directly into the model. The observed discharge time series contained a small amount of gaps (mostly shorter than ten days) which were filled by linear interpolation. Predicting reservoir outflow from precipitation in the Steinbach catchment with a random forest model (as with wastewater discharges from treatment plants) did not give satisfactory

results and could possibly have benefitted from including reservoir water levels as a predictor variable. However, reservoir water levels were not available for large parts of the simulation period. As a consequence of reading in observed outflow for the Steinbach reservoir, subbasins upstream of the reservoir only have a limited impact on simulated discharge in the downstream reaches (exceptions occur when the reservoir runs empty or water volumes exceed the storage capacity, in which cases the reservoir outflow either ceases entirely or increases when stored water is released via the emergency spillway, in both cases overriding the observed outflow time series). See Chapter 6 for a discussion of reservoir representation in the SWAT models and its implications for the research aims of this thesis.

Several "*switch*" parameters determine the use of multiple alternative sub-models to simulate specific hydrological processes in SWAT. For the simulation of infiltration and surface runoff generation, the Curve Number (CN) approach was preferred to the Green-Ampt method because it is less demanding in terms of data requirements and computational cost. The effective CN values can be adjusted for either soil moisture or evapotranspiration (ET) in the modeled HRUs. Both methods were tested in separate model runs for this thesis. Further CN adjustment with slope was disabled, as was the simulation of preferential flow in cracked soil structures. Tile flow from agricultural fields was set inactive in the base models, but was included in the more refined model versions described in Subsection 4.2.3. The variable storage instead of Muskingum method was used for streamflow routing. In-stream transformation of nutrients using the QUAL2E equations was enabled, whereas variable channel geometry due to degradation and sediment dynamics was switched off.

For the purposes of this chapter, the majority of all other model parameters were left unchanged from the default values set by the ArcSWAT interface. Exceptions were made where better knowledge was available or the specific characteristics of the Swist catchment required a deviation from the default settings. The respective parameters and their changes are listed in Table 4.2. The FFCP.bsn parameter was set manually to make the initial soil water storage independent of the contents in the weather generator data files. Moreover, the channel cross sections were set to resemble an isosceles trapezium with banks inclined by 45°. The default half-life of nitrate in the shallow aquifer was considered unrealistically low and set to the maximum possible value. Both the initial NO_3^--N concentration in the shallow aquifer and the water storage capacity of the vegetation canopy (to estimate rainfall interception) were set as a function of land use. The NO_3^--N concentrations in the shallow aquifer were estimated in accordance with Erftverband groundwater monitoring data in the Swist catchment. Depending on the vegetation type and structure, empirical canopy storage capacities are between 0 mm and 5 mm (de Jong & Jetten 2007, Zhong et al. 2022). To a varying degree, canopy

storage capacity is related to LAI (de Jong & Jetten 2007). Considering the default LAI parameters in the ArcSWAT plant database, a value of 2 mm for forest and 1 mm for all other land uses seems to be roughly realistic and consistent with general SWAT plant parameterization. Both NO_3^--N groundwater concentrations and canopy storage capacities are associated with considerable uncertainties and are essentially meant to be further adjusted during model calibration. Note that proportions between land use specific NO_3^--N groundwater concentrations as well as canopy storage capacities are preserved when varying them by a constant fraction relative to their base values given in Table 4.2.

Most of the remaining groundwater parameters were set to reflect the effects of the water table draw down around the lignite mining pits in the Lower Rhine Embayment. Streambed elevations of the Swist main watercourse and its largest tributaries were compared to groundwater levels from October 2018 to provide an assessment of surface water contact to groundwater (Erftverband personal communication). Simplistically, it was assumed that in regions with sufficiently high water tables there is unhindered direct exchange between surface waters and groundwater over the entire simulation period, whereas in all other regions surface waters have permanently lost groundwater contact. These regions were subsequently mapped onto the SWAT stream segments and associated subbasins. For all subbasins without groundwater contact, initial groundwater storage and seepage to the deep aquifer were set to 0, while water use for both the shallow and deep aquifer (i.e. groundwater extraction) was set to the maximum possible value. By this means, it was effectively guarenteed that groundwater storage in the shallow aquifer never reaches the threshold levels that trigger groundwater flow into the stream segments in these subbasins. Additionally, groundwater flow from the deep aquifer to the stream segments was prohibited over the entire catchment area by setting the deep aquifer baseflow recession constant to 0. For a map depicting the spatial distribution of stream reaches and subbasins with and without groundwater contact, see Appendix A.

The default ArcSWAT management schedules for each land use are controlled by the number of accumulated heat units during the current growth cycle. In order to get more control over the exact timing of each management operation, the scheduling-by-date mode was applied instead. Additionally, for almost all land uses except urban areas, the SWAT auto-fertilizing operation was replaced by fixed date application of fertilizer. For the sake of simplicity, all fertilizer applications added 100 % NO_3^--N to the HRU soils (no NH_4^+-N or organic N). As this study deals exclusively with N pollution, application of P fertilizer was omitted from the simulations (as plants did not experience phosphorus stress during the entire simulation period with or without P fertilizer, plant growth remained unaffected). For cropland HRUs, fertilizer was given in three applications of 60 kg NO_3^--N per ha in mid to late March, April and May. All

Table 4.2.: Modified SWAT model parameters for the base model setup. All parameters that deviate from the ArcSWAT default values are listed. If the changes only apply to a subset of HRUs or subbasins, this is indicated in parantheses.

Parameter	Value	Unit	Meaning
		Basin-wide Parameters	
FFCB.bsn	0.6		Initial soil water storage as fraction of field capacity
HLIFE_NGW_BSN.bsn	500	d	Groundwater NO_3^--N half-life
		Groundwater Parameters (HRU level)	
SHALLST.gw	0	mm	Shallow aquifer initial storage (subbasins without groundwater contact)
DEEPST.gw	0	mm	Deep aquifer initial storage (subbasins without groundwater contact)
RCHRG_DP.gw	0		Deep aquifer percolation fraction (subbasins without groundwater contact)
ALPHA_BF_D.gw	0	1/d	Deep aquifer baseflow recession coefficient
SHALLST_N.gw	100	mg/l	Initial NO_3^--N groundwater concentration (cropland, orchard)
SHALLST_N.gw	20	mg/l	Initial NO_3^--N groundwater concentration (grassland)
SHALLST_N.gw	10	mg/l	Initial NO_3^--N groundwater concentration (all other land uses)
		Water Use Parameters	
WUSHAL.wus	10 000	10^4 m^3/d	Shallow aquifer water removal for all months (subbasins without groundwater contact)
WUDEEP.wus	10 000	10^4 m^3/d	Deep aquifer water removal for all months (subbasins without groundwater contact)
		HRU Parameters	
CANMX.hru	2	mm	Maximum canopy storage (forest)
CANMX.hru	1	mm	Maximum canopy storage (all other land uses)
		Routing Parameters	
CH_SIDE.rte	1		Tangent of channel bank angle (trapezoidal cross section profile with 45° bank)
		Soil Parameters	
SOL_ALB.sol	0.13		Moist soil albedo (according to Gijsman et al. (2007))
		Plant Parameters	
BIO_LEAF	0		Fraction of tree biomass converted to residue for dormancy (forest, orchard, tree nursery)
ALAI_MIN	0		Minimum LAI for plant during dormancy (deciduous forest, orchard, tree nursery)
ALAI_MIN	1		Minimum LAI for plant during dormancy (mixed forest)
ALAI_MIN	4.5		Minimum LAI for plant during dormancy (evergreen forest)

cropland HRUs were parameterized to grow winter wheat (*Triticum aestivum*), which is planted in mid to late March and harvested in late July. For grassland and orchard HRUs, fertilizer was given as 50 kg NO_3^--N per hectare in mid March and as 100 kg NO_3^--N per hectare in mid to late April, respectively. Grassland HRUs were parameterized to grow tall fescue (*Festuca arundinacea*), which is mowed once a year in late May. All orchard HRUs were parameterized to grow apple trees (*Malus domestica*) with a harvest in late September.

The default management schedules for forest land use include annual felling and replanting of tree biomass. The rationale for this setup could be to simulate leaf growth and senescence in a steady-state mature forest. However, this explanation is inconsistent with…

1. …biomass development in evergreen forests, which do not exhibit seasonal leaf senescence,
2. …the SWAT plant growth model, which explicitly accounts for both leaf shedding and biomass accumulation over several years for tree vegetation.

An attempt to simulate fully-developed, high-biomass forest stands failed, as the mature tree vegetation suffered from excessive nitrogen stress, which in turn inhibited biomass growth, LAI development and plant transpiration. Varying soil parameters to increase nitrogen supply did not sufficiently improve the simulation results. Neither did modifying initial tree biomass and plant parameters, e.g. as suggested by Yang & Zhang (2016), Yang et al. (2018) or Haas et al. (2022). Consequently, the ArcSWAT default settings for forests were retained, albeit with date scheduling and the parameter changes for trees specified in Table 4.2 to better reconcile the annual felling and replanting operations with the notion of modeling seasonal development of leaf biomass in mature forests. Inadequate representation of forest processes as well as deficient parameterization of tree growth has been a reoccurring concern in several peer-reviewed SWAT publications of the more recent past (Yang & Zhang 2016, Yang et al. 2018, Guo et al. 2019, Yang et al. 2019, Haas et al. 2022).

4.2.2. Refinement of Point Source Emissions

Building on the procedure of imputing gaps in the daily N concentration time series for the WWTP effluents with the averages of the known values (see the preceding Subsection 4.2.1), a second string of SWAT models were setup that incorporated a more sophisticated imputation approach. In summary, the missing values in the concentration time series were predicted from additional daily time series of the WWTP operational data records. These included daily measurements of…

1. ...precipitation,
2. ...minimum and maximum air temperatures,
3. ...total wastewater volumes at the WWTP influent and effluent,
4. ...daily maximum flow rates at the WWTP influent and effluent,
5. ...mean wastewater temperatures at the biological treatment effluent,
6. ...minimum and maximum pH-values at the WWTP effluent and the biological treatment effluent.

In a preliminary test of different regression approaches, the results for linear regression were on a par with (if not better than) non-parametric methods such as k-nearest neighbor (kNN) or random forest regression. Thus, multiple linear regression was the method of choice for imputing the missing N concentrations in the context of this thesis.

For each of the sixteen different combination of WWTPs (Flerzheim, Rheinbach, Miel, Heimerzheim) and N parameters (NO_3^--N , NH_4^+-N , NO_2^--N, TNb), an individual set of predictor variables from the above list was chosen. Completeness of the respective daily time series was the first criterion for predictor selection. The second criterion was lack of multicollinearity between the predictor candidates, which was assessed using the variance inflation factor (VIF). See James et al. (2013, pp.101-102) for a basic explanation of the VIF calculation, as well as its interpretation and underlying concept. In repeated iterations, the candidate with the highest VIF score (i.e. highest degree of multicollinearity) above a value of 4 was removed, until either all VIF scores remained below 4 or only one candidate was left. The remaining predictors were used to fit an ordinary least squares linear regression model to each of the sixteen sets of N concentration data, applying the lm() function from base R's *stats* package (R Core Team 2022). These models covered all possible subsets that could be combined from the predictor variables, including their two-way interactions (meaning multiplicative combinations of two individual predictors in the model equation, ignoring quadratic terms). Models containing two-way interactions as predictors had to follow the hierarchical principle, i.e. each predictor included in an interaction term also had to be present as an independent predictor in the model.

Among all of the fitted models, for each WWTP and N parameter the one with the best predictive power was chosen for filling the gaps in the concentration time series. Predictive power was evaluated by means of Mallow's C_p as specified in James et al. (2013, p. 211). C_p is calculated as

$$C_p = \frac{1}{n}\left(RSS + 2d\hat{\sigma}^2\right) \tag{4.3}$$

where RSS denotes the sum of squared residuals (also called residual sum of squares, hence RSS), d is the number of predictors of the considered model and $\hat{\sigma}^2$ is a sample estimate of the model's error variance σ^2. The squared residual standard error RSE^2 of the full model is used to estimate σ^2 for all sub-models.[4] In the context of this study, the full model is the one which includes all predictors that survived VIF assessment of multicollinearity, along with all their two-way interactions. The sub-models are those that include a (strict) subset of all predictor variables used to fit the full model. Provided that $\hat{\sigma}^2 = RSE^2$ is an unbiased estimate of σ^2, C_p is an unbiased estimate of the MSE for model predictions on out-of-sample data (and thus a measure of predictive power).

For all sixteen combinations of WWTPs and N parameters, the model with the lowest C_p score was chosen to predict the missing values in the respective concentration time series. In order to meet the common linear regression assumptions (in particular linearity and independent, homoscedastic residuals), in some cases it was necessary to perform a logarithmic transformation or a square root transformation of the N concentration data (i.e. the response variable) before model fitting. Predictions from these models were subsequently backtransformed with the respective inverse functions, applying the corrections suggested by Miller (1984) to handle backtransformation bias. The basic properties of the sixteen best models are listed in Table 4.3. It should be noted that the procedure described for identifying the predictors that yield the models with the best predictive power is a purely data-driven, brute force approach. As such, the models are not expected to generalize well beyond the specific WWTP and N parameter combination for which they were fitted. For this reason, a detailed account of the selected predictors and the associated coefficients for each individual model is omitted. The most important predictors that were reoccurring across the best performing models were those related to wastewater quantity and temperature, governing dilution and residence times in the treatment plants as well as biological activity and the kinetics of chemical reactions, respectively. In contrast, pH values rarely emerged as strong predictors in the regression models, probably

[4] The squared residual standard error is given as

$$RSE^2 = \left(\sqrt{\frac{RSS}{n-p-1}}\right)^2 = \frac{RSS}{n-p-1} \tag{4.4}$$

with n and p being the number of data points and the number of predictors used to fit the model, respectively. Regarding Equation 4.3 and the difference between d and p, note that $d = p$ for the full model and $d < p$ for all sub-models.

Table 4.3.: Summary of the linear regression models used to impute missing N concentrations in the WWTP effluents. For each model, the type of response transformation (if any) and the number of predictors used to fit the model are specified. R^2 assesses the quality of the model fit, while C_p is the criterion used to identify the model fit with the best predictive power. Since C_p is an estimate of the model's MSE for out-of-sample data, the in-sample MSE is provided for reference. C_p scores are generally plausible, as $C_p > MSE$ for all models, which follows from Equation 4.3 (note that $MSE = RSS/n$). In case of a response transformation, please be aware that R^2, MSE and C_p are given for the model fit to the transformed data. A meaningful comparison of their values between models with different response transformations is not possible. Furthermore, while MSE and C_P are measured in mg^2/l^2 for untransformed model responses, this is no longer the case for transformed responses. This needs to be considered when interpreting the given MSE and C_p values.

WWTP	Parameter	Transformation	Predictors	R^2	MSE	C_p
Flerzheim	NO_3^--N	-	8	0.145	6.82	7.09
	NO_2^--N	-	11	0.143	1.76×10^{-4}	1.86×10^{-4}
	NH_4^+-N	-	6	0.079	1.07×10^{-2}	1.10×10^{-2}
	TNb	-	7	0.111	7.93	8.25
Rheinbach	NO_3^--N	Square Root	14	0.237	3.33×10^{-1}	3.58×10^{-1}
	NO_2^--N	Square Root	4	0.073	1.09×10^{-2}	1.11×10^{-2}
	NH_4^+-N	Square Root	10	0.111	1.04×10^{-2}	1.09×10^{-2}
	TNb	Square Root	8	0.192	2.82×10^{-1}	2.95×10^{-1}
Miel	NO_3^--N	-	6	0.235	4.05	4.17
	NO_2^--N	Logarithm	8	0.487	3.92×10^{-1}	4.08×10^{-1}
	NH_4^+-N	Square Root	10	0.340	5.80×10^{-2}	6.09×10^{-2}
	TNb	-	6	0.210	5.77	5.96
Heimerzheim	NO_3^--N	-	14	0.318	5.00	5.35
	NO_2^--N	Square Root	12	0.207	1.20×10^{-2}	1.27×10^{-2}
	NH_4^+-N	-	6	0.076	5.74×10^{-2}	5.91×10^{-2}
	TNb	-	12	0.360	4.97	5.32

because pH values were relatively constant in most WWTPs and almost never diverged from their optimum ranges to impair biological treatment conditions.

Figure 4.2 gives a visual impression of how the time series of repredicted TNb concentrations in the four WWTP effluents compare to their measured counterparts. It is noteworthy that the regression models are capable of reproducing the basic seasonality in the concentration data for the WWTPs in Rheinbach, Miel and Heimerzheim. The seasonality in the predicted time series originates from the temperature time series used as predictors in most of the regression models. Interestingly, the measured concentrations for WWTP Flerzheim do not exhibit the same degree of seasonality, which is reflected in the reduced importance of temperature-related predictors for the Flerzheim regression models. As a result, the overall model fit for Flerzheim does not appear to be of the same quality as for the other WWTPs, which is confirmed by the relatively low R^2 value of 0.111 reported in Table 4.3. Although a direct comparison of the R^2 scores in Table 4.3 is not always possible due to different response transformations, a

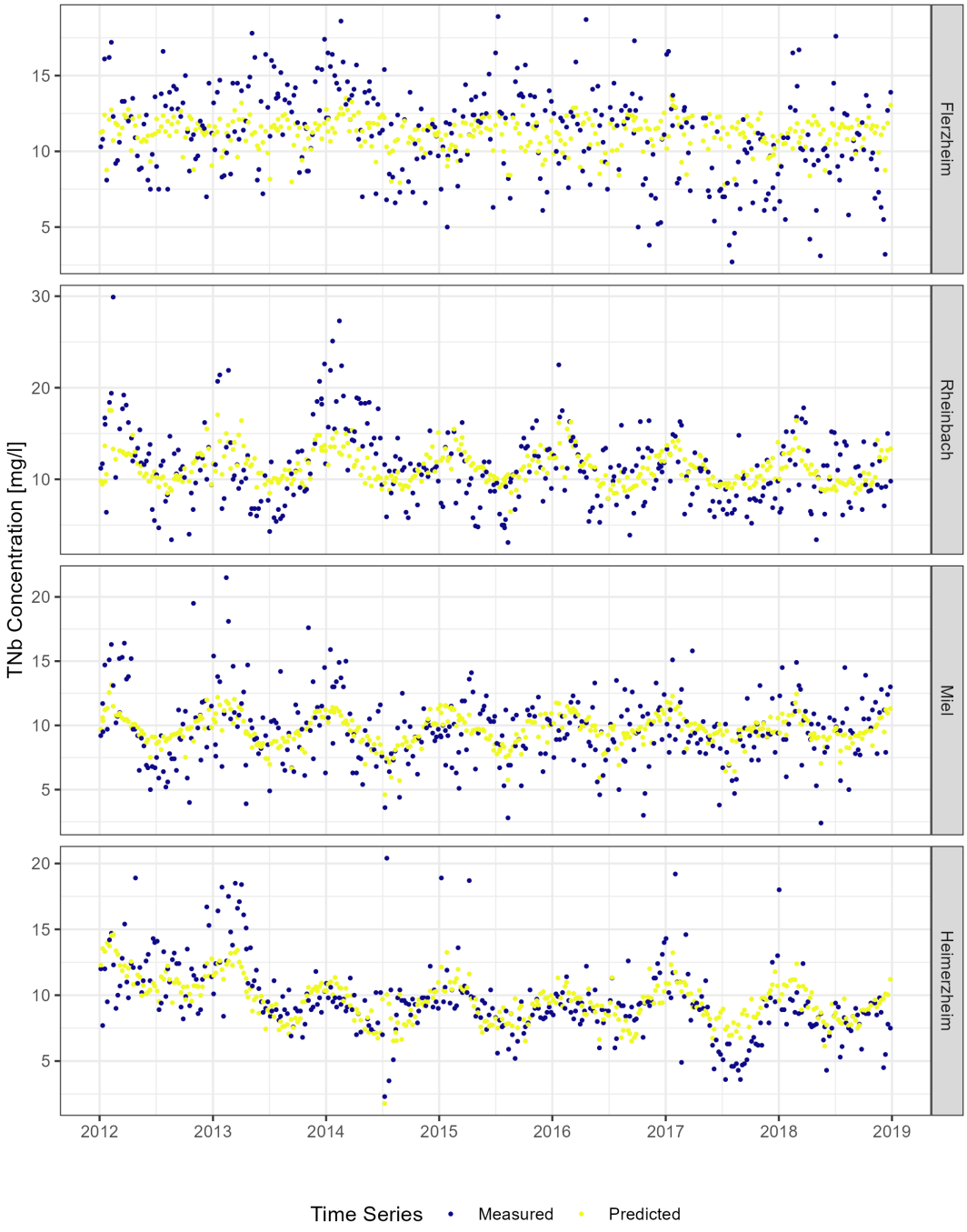


Figure 4.2.: Time series of measured and predicted TNb concentrations in the WWTP effluents from 2012 to 2018. The plots are ordered with decreasing WWTP size from top to bottom. For comparison, the predicted time series encompasses only those data points for which measured values exist. They have been repredicted with help of the linear regression models for TNb detailed in Table 4.3. For further use of the time series in SWAT, the measured data points were retained and only the gaps between them were filled with regression estimates.

general pattern of poorer model fits in Flerzheim is also evident for the other N parameters. For TNb concentrations, WWTP Heimerzheim exhibits the best model fit, according to both the reported R^2 value and visual inspection of Figure 4.2. In this case, the regression model is clearly capable of reproducing temporal trends, as shown by the notable downward shift of measured concentrations in 2013, which is captured well by the prediction.

TNb concentrations in WWTP effluents were generally dominated by NO_3^--N concentrations, whereas NO_2^--N and NH_4^+-N concentrations were negligibly low in most cases (the WWTP in Miel being an exception). As in the base model setups (see Subsection 4.2.1), concentrations of organic N were estimated by substracting the concentrations of the inorganic N compounds from the TNb concentrations. This procedure was identical for both measured and predicted data points. Occasional negative values for organic N were set to 0. As expected, organic N was the second most abundant nitrogen fraction in the WWTP effluents after NO_3^--N. The daily WWTP emission loads were calculated by multiplying the measured or imputed mean concentrations for each day with the corresponding wastewater discharge volumes. In order to generate data for the warm-up period from 2008 to 2011, a four-year constant sequence of average daily emission loads was added to the beginning of each time series. The annual N emission loads for each WWTP during the simulation period from 2012 to 2018 are presented and discussed in Chapter 6.

4.2.3. Refinement of Agricultural Management and Crop Growth

The third set of model variants built for this thesis includes refined model parameterization with respect to agricultural management, in addition to the linear regression time series imputation of WWTP nitrogen emissions described in Subsection 4.2.2. One of the features of the new models is tile drainage of agricultural areas. A digital map of tile drained fields, pastures, orchards and tree nurseries in the Swist catchment was obtained from Kothe et al. (2021), who digitized and assembled historical maps of tile drainage systems from engineering consultancies, government agencies and farmers. Most of these historical maps originate from construction or land consolidation measures. They were complemented by an extensive analysis of aerial photographs. It should be noted that the Kothe et al. (2021) data set is not guaranteed to be complete or up-to-date. The underlying map material dates as far back as the 1950s and may include drainage systems that are now damaged or heavily modified. In effect, entries in the data set neither are necessary nor sufficient for the current existence of functional drainage systems at the mapped location and should therefore be interpreted as potentially (rather than actually) drained areas. Further information on the compilation and interpretation of the digital tile drainage map can be found in Kothe et al. (2021).

Of the various approaches to tile flow simulation available in SWAT, the most parsimonious (in terms of model parameters) has been chosen (Du et al. 2005, see Subsection 3.1.2). In order to determine the regions inside the Swist catchment that are able to contribute tile flow in the models, a GIS overlay of tile drained areas with agricultural areas in the Swist catchment was created. The drained areas were taken from the digital map described above. The agricultural areas encompassed cropland, orchards, pastures and tree nurseries from the ATKIS 2019 land use data set described in Subsection 4.2.1. From the GIS overlay, the proportion of the total agricultural area that is tile drained was determined for each individual SWAT subbasin. Tile flow simulation was activated for all agricultural HRUs in subbasins that exceeded 40 % tile drained area by setting the DDRAIN.mgt parameter to a value of 1000 mm (see Table 4.5 on page 74). Appendix A shows a map with the location of all subbasins with active tile drainage simulation, together with the map of tile drained areas compiled by Kothe et al. (2021) for reference.

Apart from tile flow simulation, the refined management models also include rotation of different field crops from one simulation year to the next. For this purpose, the management schedules for each cropland HRU were heavily modified compared to their counterparts in the first two sets of model variants. A total of 35 different management schedules were prepared, based on conversations with local farmers as well as the annual best practice recommendations published by the Chamber of Agriculture (LWK) in NRW (Paffrath et al. 2021). Each management schedule represents one archetypal crop rotation cycle of three to five years before it is repeated, and each year represents a defined sequence of planting, fertilizing, harvesting and plowing operations for the specific crop cultivated that year. Appendix C details all management schedules and associated crop rotation cycles prepared for this thesis in tabular form. Table 4.4 gives an overview of the rotated field crops, together with the estimated quantities of annual N fertilizer applied in the simulations for each field crop. Depending on the crop, the total amount of fertilizer is given in several applications at different times during the respective crop year. Note that that the fertilizing routine for winter wheat is the same as in the default cropland management schedules of the base model setups described in Subsection 4.2.1, meaning three application of 60 kg NO_3^--N /(ha a) each in March, April and May. Details on the timing of all management operations can be inferred from the tables in Appendix C. As in the base models, date scheduling was applied.

In addition to the crops listed in Table 4.4, “Polish” canola (*Brassica rapa*, or synonymously *Brassica campestris*) is used as a winter catch crop whenever a main crop harvested in fall is followed by another main crop planted in spring. The catch crops do not receive any N fertilizer and freeze before the start of the next growing season (“kill” operation in SWAT).

Table 4.4.: Rotated field crops with annual N fertilizer quantities. Management schedules for corn and corn silage only differ in the timing of the harvest operation. Annual N fertilizer quantities have been estimated based on the best practice recommendations provided by Paffrath et al. (2021). Field peas are not fertilized due to N fixation.

SWAT Crop Code	Crop Name	Taxonomic Name	N Fertilizer [kg NO_3^--N /(ha a)]
CANA	Canola	*Brassica napus*	170
CORN	Corn	*Zea mays*	150
CSIL	Corn Silage	*Zea mays*	150
FPEA	Field Peas	*Pisum arvense*	0
OATS	Oats	*Avena sativa*	80
POTA	Potato	*Solanum tuberosum*	130
RYE	Rye	*Secale cereale*	130
SGBT	Sugar Beet	*Beta vulgaris*	110
WBAR	Winter Barley	*Hordeum vulgare*	140
WWHT	Winter Wheat	*Triticum aestivum*	180

Their residues are either incorporated into the soil in a springtime plowing operation or left at the soil surface before planting of the following crop for soil conversation purposes ("mulch tillage"). It should be noted that *Brassica rapa* is not suitable to be grown directly before sugar beet and potatoes, due to the promotion of nematodes, a major soil-borne pest of root crops. However, parameterizations for nematode-resistant catch crops (e.g. oilseed radish, *Raphanus sativus*) were not available in the ArcSWAT database. As pests and their effects on crop growth are not simulated in SWAT and *Brassica rapa* was assumed to be reasonably similar to oilseed radish in terms of height, habitus and physiology, it was grown prior to sugar beet and potatoes nonetheless.

The management schedules and rotation cycles were assigned to the different cropland HRUs so that the resulting distribution of area shares for each crop in the total arable area of the Swist catchment resembles the actual distribution as closely as possible. For this purpose, the following algorithm for assigning management schedules to the cropland HRUs in the SWAT models was implemented in R. First, randomly select a cropland HRU. Second, applying every available management schedule and associated starting crop configuration to the current HRU, calculate the area share of each crop in each simulation year in the total area of all HRUs evaluated so far (as a real number between 0 and 1). The area share for each simulation year is calculated by adding the area of the current HRU to the combined area of all HRUs the respective crop has already been assigned to for the same year, and dividing this value by the total area of all previously considered HRUs (obviously, for the very first HRU the resulting area share will be exactly 1 for the crop grown each year and exactly 0 for all other crops). The starting crop configuration of a management schedule determines which crop of the associated rotation cycle is grown in the first simulation year (e.g. for a management schedule that covers five years, there

are five possible starting crop configurations). Third, compare the area shares obtained by this procedure to observed area shares by calculating their MSE over all considered field crops for every year. Fourth, sum the MSE over all simulation years and pick the management schedule (and the respective starting crop) that yields the lowest MSE sum. Repeat this procedure with all cropland HRUs until all of them have been considered.

Figure 4.3 illustrates this process using the model setup that is based on the ATKIS 2019 land use data and the BK50 2019 soil data as an example. In part A of the figure, a heat map shows the progression of the MSE calculated for each simulation year (including the warm-up period) with each newly evaluated cropland HRU. This heat map directly corresponds to part B of the figure, which tracks the optimization criterion - the MSE sum over all years - over all evaluated HRUs. Both part A and B of Figure 4.3 show the results for the best possible (i.e. smallest) value of the optimization criterion. As can be seen from part B, the drop of the optimization criterion is extremely steep (note the logarithmic scale), but not monotonous. At multiple points along the downward progression of the MSE sum, the curve jumps back up, indicating stages in the process when the algorithm was unable to improve the fit to the observed area shares with any of the available crop rotation cycles. The large number of options (35 different rotation cycles, along with their respective starting crop configurations) facilitated rapid alignment of modeled with observed area shares by reducing the number of upward spikes in the depicted curve. Part C of Figure 4.3 shows the final distribution of crop area shares for each simulation year in the model compared to the reference data set of observed values used for MSE calculation. As can be seen, the procedure recreated the reference data extremely well, with relative errors being particularly low for the three most common field crops in the catchment: winter wheat (WWHT), winter barley (WBAR) and sugar beet (SGBT). Taken together, these crops cover more than three quarters of the entire arable land area in the Swist catchment.

The reference data set stems from the German agricultural structure survey of 2016 for the federal states of NRW and RLP (State Office for Information and Technology NRW 2016, State Office for Statistics RLP 2016). The survey data includes municipality-level information on the acreages of various field crops and other agricultural plants cultivated by farmers in 2016. The reference area shares were calculated by summing the acreages for each crop over all municipalities with substantial shares of arable land inside the Swist catchment and dividing them by the total arable area. The crops listed in Table 4.4 constitute the vast majority of cultivated field crops in the Swist catchment according to the survey data. As triticale was not included in the ArcSWAT crop database, triticale acreages were reclassified as rye acreages for the model setup. The associated area shares were generally small (see rye area shares in the stacked bar chart of Figure 4.3 C). Furthermore, the survey data does not distinguish between

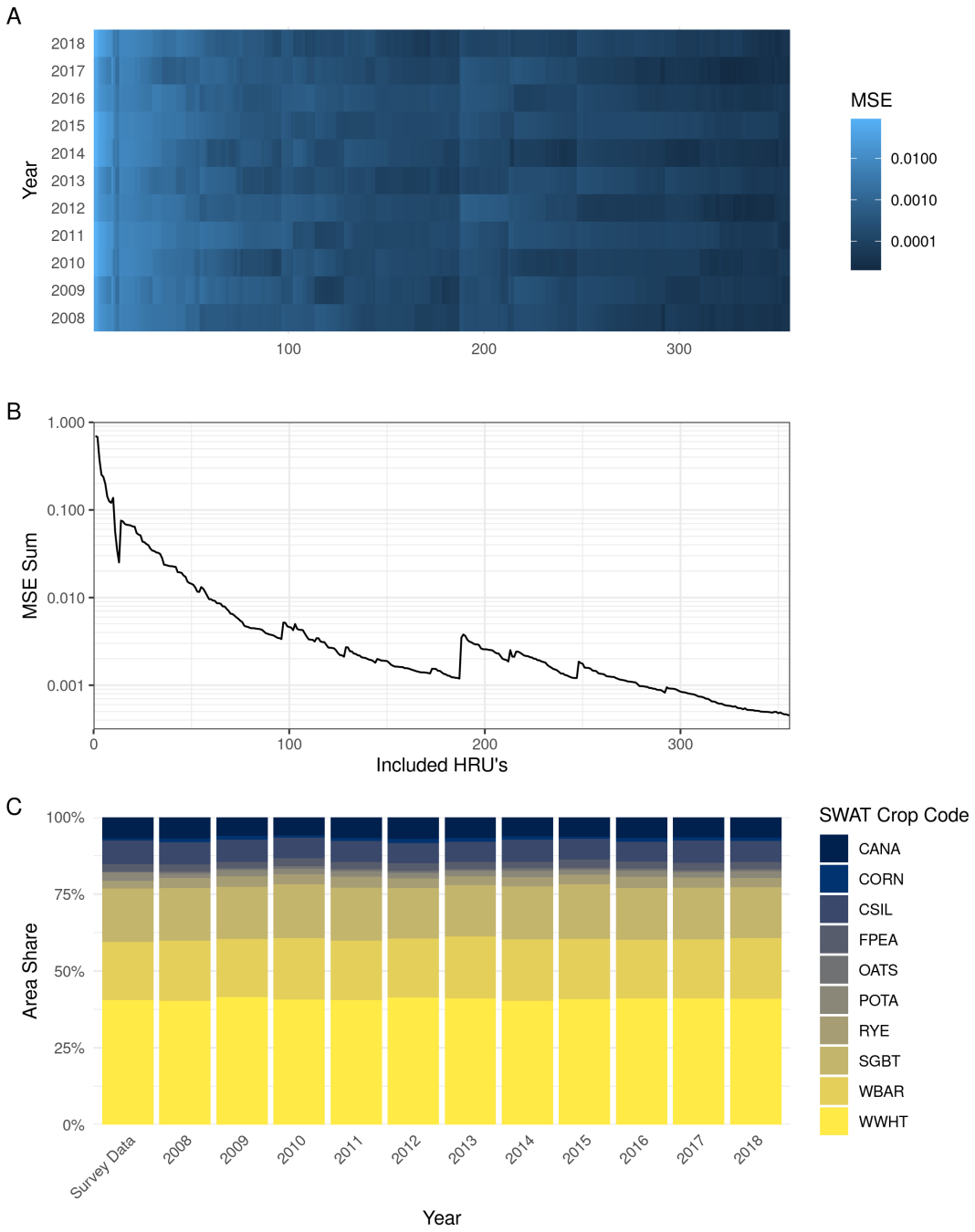


Figure 4.3.: Assignment of crop rotation cycles to cropland HRUs. Only results for the SWAT model variant built upon the ATKIS 2019 land use and the BK50 2019 soil data are shown (356 cropland HRUs). (A) Progression of MSE between surveyed and assigned crop area shares for each simulation year and included cropland HRU. (B) Progression of the optimization criterion - the MSE sum over all simulation years - with each included cropland HRU. (C) Resulting area shares of field crops for each simulation year compared to surveyed data. For the meaning of the SWAT crop code please refer to Tables 4.4 and 4.6. Survey data stems from the agricultural structure survey of 2016 for NRW as well as RLP and is an aggregate for the Swist municipalities.

different legumes. Therefore, acreages used to grow legumes according to the survey data were unanimously identified with field pea acreages in the SWAT models.

The described procedure guaranteed that the refined SWAT models included realistic rotation cycles of field crops typically grown in the region. Moreover, as the dates of specific management operations varied between different field crops and rotation cycles, the assignment process quasi-randomized the timing of management operations throughout the study area, ensuring that operations are not carried out simultaneously in all cropland HRUs (excessive stream nutrient peaks due to fertilizer application on the same day in the entire catchment are a known issue with SWAT raised for instance by Schürz et al. (2017)). However, these upsides of the method are contrasted with some imperfections. First, as the modeled crop area shares are fit exclusively to data for 2016, the procedure is unable to generate temporal changes in crop cultivation patterns (see the bar charts in Figure 4.3 C for confirmation). Second, the agricultural structure survey allocates acreages to municipalities based on where the farm that cultivates the respective fields is situated, not the fields themselves, as pointed out by Gömann et al. (2021). The reference data for a specific municipality may therefore include acreages lying outside the municipality's boundaries, while omitting other acreages situated within but farmed by an agricultural holding located elsewhere. Although this most probably leads to spatial misallocations at the municipality-level, it is considered unlikely to introduce unacceptable distortions in the regional crop area shares. Keep in mind that crop area shares were aggregated over several different municipalities and that spatial variability within the Swist catchment remained unconsidered. Furthermore, some misallocations already are inherent to the mismatch between administrative and catchment boundaries, whenever municipalities are not completely enclosed by the catchment boundaries.

Average annual crop yields simulated with the default crop parameterization were implausibly low for some of the rotated field crops (for the base model setup described in Subsection 4.2.1, harvest yields were not investigated). This was of particular concern for winter wheat, sugar beet, potato, corn and corn silage. Although SWAT is known to have difficulties in simulating interannual variation in crop yields, it should be capable of reproducing long-term average yields (Srinivasan et al. 2010, Abbaspour et al. 2015, Bauwe et al. 2019). The inability to do so may indicate problems in simulating either plant growth or biomass withdrawal with harvest operations, both of which possibly being detrimental to the simulated soil nitrogen balance. Partly following the suggestions of Bauwe et al. (2019), the crop parameters BIO_E and HVSTI were set to the values listed in Table 4.5, affecting parameterization for winter wheat, sugar beet, potatoes, corn and corn silage. In addition, harvest parameters for apple trees were changed to better reflect the notion of tree growth simulation only representing annual de-

Table 4.5.: Modified SWAT management and crop parameter values. All parameters that deviate from the ArcSWAT default values, in addition to those in Table 4.2, are listed. If the changes only apply to a subset of HRUs, subbasins or field crops, this is indicated in parantheses.

Parameter	Value	Unit	Meaning
		Management Parameters	
DDRAIN.mgt	1000	mm	Depth to tile drain from soil surface (cropland, grassland, orchard and tree nursery HRUs in subbasins with tile drainage)
		Crop Parameters	
BIO_E	36	1×10^{-4} kg/MJ	Radiation-use efficiency (winter wheat)
BIO_E	45	1×10^{-4} kg/MJ	Radiation-use efficiency (corn, corn silage)
BIO_E	35	1×10^{-4} kg/MJ	Radiation-use efficiency (sugar beet, potato)
HVSTI	0.57		Harvest index for optimal growing conditions (winter wheat)
HVSTI	0.95		Harvest index for optimal growing conditions (corn silage)
HVSTI	1.25		Harvest index for optimal growing conditions (potato)
HVSTI	0.7		Harvest index for optimal growing conditions (apple)
WSYF	0.35		Harvest index for highly stressed growing conditions (apple)

velopment of leaves, flowers and fruits (see Subsection 4.2.1). The modifications resulted in significant increases in plant growth and harvested yields for the respective crops, for example with annual dry weight yields for winter wheat almost doubling.

Table 4.6 gives an overview of long-term average crop yields that were simulated with the modified crop parameterization. Across the different combinations of land use and soil data underlying the various model setups, differences in simulated dry weight yields stay well below 1000 kg/(ha a) for all field crops, indicating low sensitivity of the simulated yields to the applied land use and soil data sets. In purely qualitative terms, the simulated yields are largely considered to be reasonable, with yields for sugar beet and corn silage being the highest, followed by winter barley, winter wheat and corn. From a quantitative point of view, most yields align well with their expected values, albeit with a small subset of crops (potatoes and corn silage) undercutting their expected yields by implausibly wide margins. Potato yields are particularly concerning, remaining approximately three times below their expected value despite the modified growth and harvest parameterization. Under normal conditions, potato yields would be expected to exceed those of cereals such as wheat and barley. Furthermore, barley

Table 4.6.: Average simulated crop yields over all cropland HRUs (2012-2016). Expected values have been estimated on the basis of typical fresh weight yields and dry matter contents. Note that these are meant only as a rough guide to assess the plausibility of the simulated values. Simulated yields are reported for the four combinations of soil and land use data sets used for model setup (see Subsection 4.2.1) and each field crop included in the refined SWAT models. They exclusively refer to the model versions with CN adjustment based on evapotranspiration and refined management parameterisation, including the modified crop parameters detailed in Table 4.5. For completeness, yields for agricultural land uses other than cropland (apples for orchards and fescue for pastures) are added at the end of the table.

		Average Dry Weight Yields [kg/(ha a)]				
SWAT Code	Crop Name	Expected (approx.)	CLC 2012 BK50 2015	CLC 2012 BK50 2019	ATKIS 2019 BK50 2015	ATKIS 2019 BK50 2019
CANA	Canola	4000	4231	4197	4354	4278
CORN	Corn	8000	6799	7183	7332	7250
CSIL	Corn Silage	15 000	9719	9470	9425	8924
FPEA	Field Peas	3500	4578	4502	4582	4449
OATS	Oats	4000	4776	4470	4639	4516
POTA	Potato	9000	3466	3306	3344	3242
RYE	Rye	5000	5502	5424	5431	5531
SGBT	Sugar Beet	15 000	12 922	12 834	12 877	12 839
WBAR	Winter Barley	6000	7319	7296	7168	7188
WWHT	Winter Wheat	7000	7290	7166	7383	7187
ORCH	Apple	5000	7275	7134	7155	7197
WPAS	Tall Fescue	3500	3360	3169	3462	3378

yields are probably slightly overestimated and should not exceed long-term yields for winter wheat. Similarly, simulated apple yields are higher than expected, which may be attributed to the modified harvest parameters that have been re-estimated without a sound scientific basis. In general, the simulated yields are considered acceptable, especially since the three most important crops in the Swist river basin - winter wheat, winter barley and sugar beet - exhibit only relatively low yield deviations from their expected values. Keep in mind that the expected yields given in Table 4.6 only constitute general estimates, and that actual yields are dependent on the specific growing conditions in the research area during the simulated time period.

4.3. Uncalibrated Model Performance

4.3.1. Performance Evaluation

In the following subsections, the different (uncalibrated) model setups described in the first part of this chapter are compared in terms of model performance. Model performance was evaluated by assessing the capability of the models to reproduce various observed time series of

streamflow and in-stream NO_3^--N loads in the Swist river basin. Performance evaluation with regard to streamflow was carried out using daily streamflow time series measured at three of the four online gauging stations in the river basin (names of the associated streams are given in parantheses): Essig (Steinbach), Morenhoven (Swist) and Weilerswist (Swist). Streamflow measurements in Kirchheim (Steinbach) were unavailable for calibration since they had already been used to set the model outflow from the Steinbach reservoir. Model performance with respect to in-stream NO_3^--N loads was evaluated using daily NO_3^--N load time series observed at the online water quality monitoring station in Metternich. See the map in Appendix A for the position of the gauging stations and the water quality monitoring station in the river basin.

As the results of the model comparison were also used to select an appropriate model version for parameter calibration (and thus can be considered part of the calibration process), the performance evaluation was limited to five years from 2012 to 2016, whereas the remaining two simulation years 2017 and 2018 were reserved for model validation (see Chapter 5). The observed daily discharge time series were complete (i.e. without gaps) at all three discharge gauges. The observed daily NO_3^--N loads were based on five-minute NO_3^--N concentration measurements exhibitung considerable gaps due to maintenance and repair interruptions. If gaps in the five-minute concentration data did not exceed one hour, they were filled using linear interpolation. NO_3^--N loads transported in the river were calculated by multiplying the five-minute NO_3^--N concentrations data with five-minute streamflow data measured at the gauging station in Weilerswist, with remaining gaps in the concentration time series translating into gaps in the resulting load time series. The five-minute NO_3^--N loads were subsequently summed up to daily values, with remaining gaps (all longer than one hour) also generating a gap in the daily time series.

The water quality monitoring station in Metternich lies several river kilometers upstream of the gauge in Weilerswist, which corresponds to a temporal offset between concentration and streamflow measurements of approximately 1 to 2 h, depending on the current flow conditions (Erftverband personal communication). Therefore, load calculations were repeated for NO_3^--N concentrations that were shifted relative to the streamflow data to represent water travel times between Metternich and Weilerswist of 5, 10, …, 120 min. For each time interval, the relative error to the daily loads based on unshifted concentration data was calculated (see Figure 4.4). For days on which the relative error exceeds a value of 10 %, the respective daily load was removed from the time series and thus remained unconsidered in performance evaluation. Similarly, for days on which a disproportianate amount of rainfall is concentrated on the region between Metternich and Weilerswist (i.e. high daily precitation measured at the

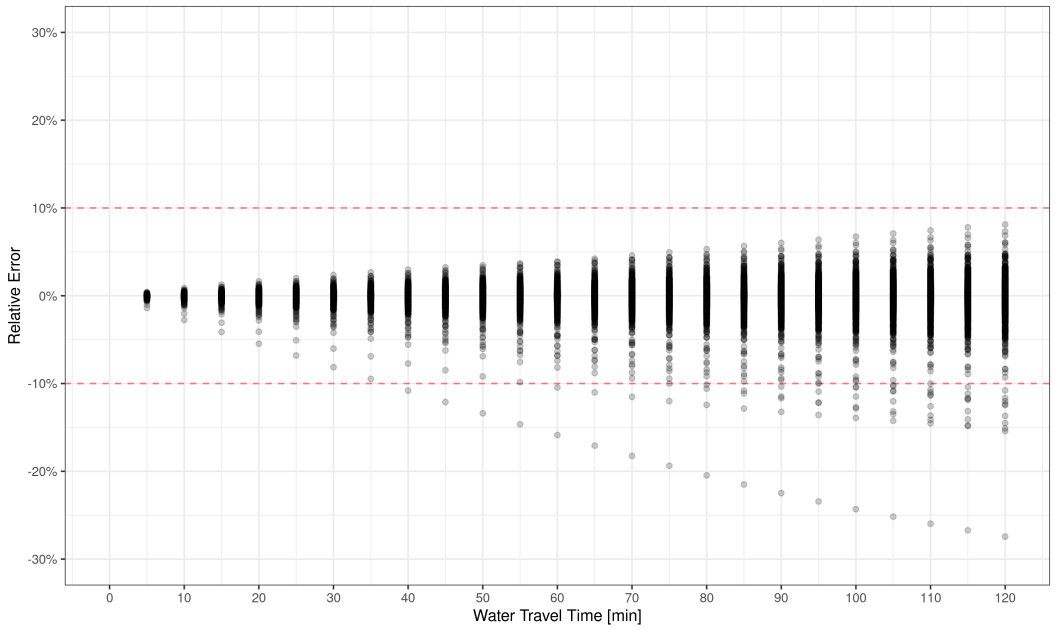


Figure 4.4.: Relative error of daily NO_3^--N loads due to different assumed water travel times between Metternich and Weilerswist. Each data point represents the relative error of the calculated in-stream NO_3^--N load transported on one day in the simulation period (2012-2018), assuming different water travel times. For days on which the absolute relative error exceeds 10 % for any assumed water travel time, the respective load is removed from the daily load time series and thus remains unconsidered in model performance evaluation.

weather station in Weilerswist compared to low daily precipitation in the rest of the Swist catchment), the respective daily load was removed from the time series as well. The observed and simulated daily time series were used to calculate four of the performance criteria introduced in Subsection 3.2.1 for each of the four monitoring sites. The coefficient of determination R^2 was calculated by determining Pearson's correlation coefficient r using the cor() function in base R's *stats* package (R Core Team 2022) and squaring the result. The PBIAS, NSE and KGE metrics were calculated using the various functions implemented in the R package *hydroGOF* (Zambrano-Bigiarini 2020), applying the equations given in Subsection 3.2.1.

4.3.2. Performance of Base Models

Table 4.7 lists the performance ratings at the four different monitoring sites achieved by different versions of the base model setup described in Subsection 4.2.1. These model versions encompass the four different combinations of land use (LU) and soil data used to build the models, as well as two different methods of dynamic CN adjustment, either via daily soil moisture or daily evapotranspiration (ET, see Subsection 3.1.2). Consequently, performance was

tested for eight different model versions at each monitoring site (Essig, Morenhoven, Weilerswist, Metternich). In general, the choice of the CN adjustment method produces the largest performance differences between the model setups, whereas performance differences originating from varying land use and soil data are mostly minuscule. Considering streamflow, models adjusting the CN with evapotranspiration are clearly superior to those using CN adjustment based on soil moisture. The performance differences are particularly pronounced for the Essig gauge, where there is a considerable increase of all performance metrics when applying the evapotranspiration approach. To a lesser, but still significant degree, this effect is also visible at the other stream gauges in Morenhoven and Weilerswist, with NSE and KGE scores unanimously turning from negative to positive. Considering NO_3^--N loads in Metternich, the evapotranspiration based method for CN adjustment tends to perform better as well, albeit not across all metrics. While R^2 and NSE scores are slightly higher for the evapotransiration method, differences between the PBIAS and KGE scores are ambiguous.

Given that the models are not yet calibrated, some R^2, NSE and KGE scores already reach reasonably high. NSE values for Essig in particular are distinctly positive for all model setups involving evaporation based CN adjustment. For the same model setups, the Essig R^2 scores reach as high as 0.5, indicating reasonable good correlation of simulated and observed streamflow peaks (however, note that the gauge in Essig receives a large part of its water from the Steinbach reservoir, for which the Kirchheim discharge time series is used to fix daily reservoir outflow in the SWAT models). KGE scores for simulated streamflow in Morenhoven and Weilerswist are consistently above 0.4 (for Morenhoven sometimes even above 0.5) when evaporation based CN adjustment is used. However, the respective NSE scores are considerably lower, in Weilerswist even staying below 0 for all but one model setup. With respect to in-stream NO_3^--N loads in Metternich, NSE scores are positive for all model setups, albeit unanimously staying well below 0.2. The corresponding KGE scores are in a similar numerical range.

With regard to the PBIAS scores listed in Table 4.7, there is a clear pattern of overpredicted streamflow at all stream gauges. This general pattern is independent of the CN adjustment method, although streamflow overprediction is less pronounced for CN adjustment based on evapotranspiration. The PBIAS scores never fall below 20 % at any of the three stream gauges. For the Essig gauge, streamflow overprediction is particularly strong, with PBIAS scores ranging from 83.2 to 149.1 %. Overprediction of streamflow is contrasted by systematic underprediction of NO_3^--N loads in Metternich. For all model setups, PBIAS scores for NO_3^--N load prediction are below −30 %. There are many possible explanations for these systematic deviations from the observed data. NO_3^--N load underprediction may be caused by misrepresentation of N inputs (in agricultural catchments particularly via fertilization), soil nitrogen cycling (e.g.

Table 4.7.: Perfomance criteria for the simulation results of the eight uncalibrated base model variants. The different model variants are defined by their underlying land use (LU) and soil data as well as CN adjustment procedure. Model performance is evaluated for streamflow at the gauges in Essig, Morenhoven and Weilerswist as well as NO_3^--N loads at the water quality monitoring station in Metternich (simulation years 2012-2016). The best value for each performance criterion and monitoring site is typeset in bold.

Monitoring Site	LU Data	Soil Data	CN Adjustment	R^2	PBIAS	NSE	KGE
Essig	CLC 2012	BK50 (old)	Soil Moisture	0.09	124.4	−7.99	−1.52
Essig	CLC 2012	BK50 2019	Soil Moisture	0.09	140.0	−8.68	−1.69
Essig	ATKIS 2019	BK50 2015	Soil Moisture	0.08	132.8	−9.11	−1.71
Essig	ATKIS 2019	BK50 2019	Soil Moisture	0.09	149.1	−9.96	−1.89
Essig	CLC 2012	BK50 2015	Evapotranspiration	**0.50**	96.3	0.27	−0.03
Essig	CLC 2012	BK50 2019	Evapotranspiration	0.49	**83.2**	**0.31**	**0.09**
Essig	ATKIS 2019	BK50 2015	Evapotranspiration	**0.50**	100.7	0.26	−0.07
Essig	ATKIS 2019	BK50 2019	Evapotranspiration	0.47	**83.2**	0.30	**0.09**
Morenhoven	CLC 2012	BK50 2015	Soil Moisture	0.24	44.1	−3.40	−0.52
Morenhoven	CLC 2012	BK50 2019	Soil Moisture	0.24	46.4	−3.95	−0.66
Morenhoven	ATKIS 2019	BK50 2015	Soil Moisture	0.24	44.2	−3.21	−0.48
Morenhoven	ATKIS 2019	BK50 2019	Soil Moisture	0.25	46.9	−3.62	−0.59
Morenhoven	CLC 2012	BK50 2015	Evapotranspiration	**0.34**	29.6	0.13	0.49
Morenhoven	CLC 2012	BK50 2019	Evapotranspiration	0.33	**24.9**	0.10	**0.51**
Morenhoven	ATKIS 2019	BK50 2015	Evapotranspiration	**0.34**	30.4	**0.17**	0.48
Morenhoven	ATKIS 2019	BK50 2019	Evapotranspiration	0.33	25.8	0.09	0.50
Weilerswist	CLC 2012	BK50 2015	Soil Moisture	0.13	51.5	−6.91	−1.14
Weilerswist	CLC 2012	BK50 2019	Soil Moisture	0.13	54.3	−7.11	−1.19
Weilerswist	ATKIS 2019	BK50 2015	Soil Moisture	0.13	51.6	−6.76	−1.12
Weilerswist	ATKIS 2019	BK50 2019	Soil Moisture	0.13	54.3	−6.84	−1.14
Weilerswist	CLC 2012	BK50 2015	Evapotranspiration	0.26	29.6	−0.07	0.43
Weilerswist	CLC 2012	BK50 2019	Evapotranspiration	0.25	25.2	−0.12	0.43
Weilerswist	ATKIS 2019	BK50 2015	Evapotranspiration	**0.28**	29.1	**0.06**	0.44
Weilerswist	ATKIS 2019	BK50 2019	Evapotranspiration	0.26	**23.9**	−0.02	**0.45**
Metternich	CLC 2012	BK50 2015	Soil Moisture	0.15	**−31.3**	0.08	**0.11**
Metternich	CLC 2012	BK50 2019	Soil Moisture	0.13	−33.4	0.05	0.09
Metternich	ATKIS 2019	BK50 2015	Soil Moisture	0.16	−31.6	0.09	0.10
Metternich	ATKIS 2019	BK50 2019	Soil Moisture	0.14	−35.6	0.05	0.07
Metternich	CLC 2012	BK50 2015	Evapotranspiration	0.20	−32.4	**0.13**	0.10
Metternich	CLC 2012	BK50 2019	Evapotranspiration	0.21	−35.6	0.11	0.07
Metternich	ATKIS 2019	BK50 2015	Evapotranspiration	**0.22**	−32.8	**0.13**	0.10
Metternich	ATKIS 2019	BK50 2019	Evapotranspiration	**0.22**	−38.9	0.10	0.05

rates of mineralization and denitrification) and in-stream retention. Subsection 4.3.4 provides a broader discussion of this issue in connection with simulated agricultural management. A general explanation for streamflow overprediction could be systematic error in the measurement of precipitation heights and/or river discharge. However, considering the applied measurement approaches, the multitude of precipitation and discharge gauges and their instrumentation, input data errors are deemed an unlikely cause for streamflow overestimation of more than 10 % at all monitoring sites.

Instead, two other explanations for streamflow overprediction are more plausible, especially considering the characteristics of the Swist catchment. First, the option to update the CN with evapotranspiration has been introduced to SWAT to address overprediction of surface runoff from shallow soils by the classical soil moisture method (Yen et al. 2015). Shallow soils are prevalent in the elevated Eifel foothills on the southern edge of the Swist catchment (Geological Survey NRW 2019) and may thus contribute excess surface runoff in simulations with soil moisture based CN adjustment. Indeed, overpredicted streamflow is mostly concentrated in the hydrograph peaks directly after rainfall events in these simulations. In contrast, hydrograph peaks are substantially reduced for evapotranspiration based CN adjustment (see Figure 4.5 for illustration). Given that the largest part of the Essig gauge catchment is situated in the Eifel foothills, this may furthermore explain the particular strong effect the choice of the CN adjustment method has on performance metrics for this specific gauge. Second, the uncalibrated base models do not account for transmission losses from the SWAT main channels, as the hydraulic conductivity of the main channel bed (model parameter CH_K2.rte) has been left at its general default value of 0 mm/h.[5] As described in Chapter 2, transmission losses from the streams very likely play an important role in the hydrology of the Swist catchment. Neglecting them in the simulations may therefore lead to severe overprediction of the actual stream discharges, especially along the Steinbach watercourse, for which transmission losses are well documented.

Although performance differences between models based on different land use and soil data are to a large part negligible, there is a slight tendency towards better simulation results for the land use and soil data sets from 2019. The updated soil data set (BK50 2019) in particular seems to mitigate overestimation of streamflow when applying CN updating based on evapotranspiration, with PBIAS scores dropping markedly compared to the models built on the older soil data (BK50 2015). On the one hand, seeing better model performances ratings for the more

[5] Although this was considered unrealistic, better knowledge of hydraulic channel bed conductivities and their spatial distribution was unavailable. Determining better values for CH_K2.rte was therefore left to subsequent parameter calibration (see Chapter 5).

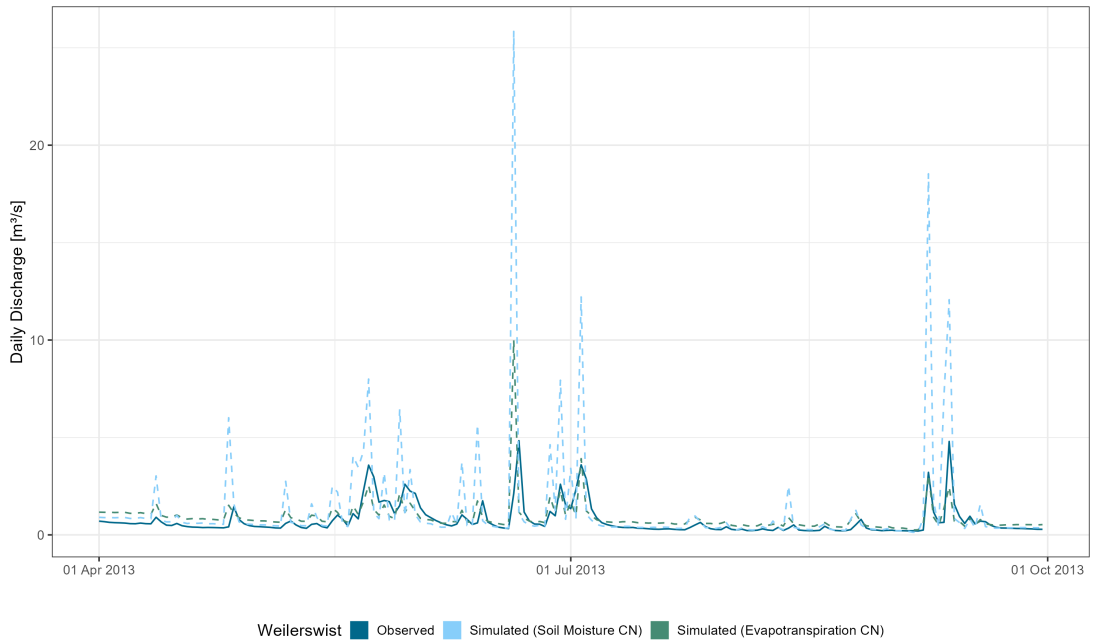


Figure 4.5.: Observed and simulated hydrographs at the Weilerswist gauge for the second and third quarter of 2013. The two simulated hydrographs belong to the uncalibrated base model setup that dynamically updates the CN either with soil moisture or with evapotranspiration. The method based on soil moisture substantially overpredicts the hydrograph peaks. Both model setups have been built on the ATKIS 2019 land use data set and BK50 2019 soil data set.

recent land use and soil data sets is reassuring (even with models being still uncalibrated at this stage and model error being difficult to attribute), as they were expected to be superior to the older data. The ATKIS 2019 land use data set incorporates more spatial detail than the CLC 2012 data, and the BK50 2019 soil data set is based on a more recent soil mapping for the Swist river basin, more reliable data for the soils located in RLP as well as updated methods to determine important soil parameters. On the other hand, the older land use and soil data sets mostly produce similar performance ratings, which is not necessarily surprising given that both share structural similarities and a common origin with their counterparts from 2019 (see Subsection 4.2.1).

4.3.3. Performance of Models with Refined Point Source Emissions

Replicating the structure of Table 4.7, Table 4.8 gives an overview of model performances calculated for the refined imputation of point source N emissions (see Subsection 4.2.3). Since the revised time series of N emission loads from the four WWTP effluents in the catchment do not affect water quantities, performance metrics for the three streamflow gauges in Essig,

Morenhoven and Weilerswist are exactly the same as those listed in Table 4.7. However, the revised time series do affect water quality. Considering in-stream NO_3^--N loads at the water quality monitoring station in Metternich, the unexpected result is that model performance deteriorates compared to the base model results provided in Table 4.7. Although performance ratings generally differ only by a small amount, they unanimously worsen for each performance metric and model setup. In part, this result can be explained with the overall point source N quantities released into the streams in both sets of models. For the time period for which the performance ratings in Tables 4.7 and 4.8 are calculated (2012-2016), the regression based imputation approach results in total N emissions from WWTPs that are approximately 6 % lower than those resulting from simply filling emission time series gaps with the mean of the known values. As the base models already underestimate in-stream NO_3^--N loads, reduction of point source emissions obviously pushes the PBIAS scores towards even more negative values. Similarly, bias in predicted NO_3^--N loads may also be responsible for lower NSE and KGE scores.

However, R^2 scores decrease as well, although they exclusively measure correlation between observed and predicted NO_3^--N loads in Metternich. This means that the lower model performance is not only an expression of stronger bias, but is also due to poorer representation of the NO_3^--N in-stream load dynamics. Although the effect is only weak, it nevertheless is the exact opposite of what was originally hoped to be achieved with the revised time series. Even if differences to the base model performance ratings are considered to be insignificant, the results still show that the comparatively elaborate and time consuming regression approach offers no benefit over the extremely simple mean substitution approach in this application case. Regression models with stronger fits to the WWTP effluent data (see the relatively low R^2 values in Table 4.3) may improve these results, but the available data did not allow to build such models as long as they assume linearity between predictor and response variables. Non-linear models may be able to provide better fits and possibly more predictive power, although they are also more likely to overfit the available data. In this regard, remember that some non-parametric regression approaches that can accommodate non-linear relationships have already been tested for this study but did not appear to provide better results (see Subsection 4.2.2).

Although the regression approach for imputing the time series gaps was unable to improve simulation of daily NO_3^--N in-stream dynamics, Figure 4.2 suggests that it may be better at capturing seasonal variations. To confirm this, the performance metrics for Swist NO_3^--N loads in Metternich were recalculated based on the monthly simulation results both for the base and the refined model setups. However, while overall performance ratings expectedly increased due to the coarser temporal aggregation level, the base models again slightly outperformed the

Table 4.8.: Performance criteria for the simulation results of the eight uncalibrated model variants extended by regression-based point source inputs. The different model variants are defined by their underlying land use (LU) and soil data as well as CN adjustment procedure. Model performance is evaluated for streamflow at the gauges in Essig, Morenhoven and Weilerswist as well as NO_3^--N loads at the water quality monitoring station in Metternich (simulation years 2012 - 2016). The best value for each performance criterion and monitoring site is typeset in bold.

Monitoring Site	LU Data	Soil Data	CN Adjustment	R^2	PBIAS	NSE	KGE
Essig	CLC 2012	BK50 2015	Soil Moisture	0.09	124.4	−7.99	−1.52
Essig	CLC 2012	BK50 2019	Soil Moisture	0.09	140.0	−8.68	−1.69
Essig	ATKIS 2019	BK50 2015	Soil Moisture	0.08	132.8	−9.11	−1.71
Essig	ATKIS 2019	BK50 2019	Soil Moisture	0.09	149.1	−9.96	−1.89
Essig	CLC 2012	BK50 2015	Evapotranspiration	**0.50**	96.3	0.27	−0.03
Essig	CLC 2012	BK50 2019	Evapotranspiration	0.49	**83.2**	**0.31**	**0.09**
Essig	ATKIS 2019	BK50 2015	Evapotranspiration	**0.50**	100.7	0.26	−0.07
Essig	ATKIS 2019	BK50 2019	Evapotranspiration	0.47	**83.2**	0.30	**0.09**
Morenhoven	CLC 2012	BK50 2015	Soil Moisture	0.24	44.2	−3.40	−0.52
Morenhoven	CLC 2012	BK50 2019	Soil Moisture	0.24	46.5	−3.94	−0.66
Morenhoven	ATKIS 2019	BK50 2015	Soil Moisture	0.24	44.3	−3.21	−0.48
Morenhoven	ATKIS 2019	BK50 2019	Soil Moisture	0.25	47.0	−3.62	−0.59
Morenhoven	CLC 2012	BK50 2015	Evapotranspiration	**0.34**	29.7	0.14	0.49
Morenhoven	CLC 2012	BK50 2019	Evapotranspiration	**0.34**	**25.0**	0.10	**0.51**
Morenhoven	ATKIS 2019	BK50 2015	Evapotranspiration	**0.34**	30.5	**0.17**	0.48
Morenhoven	ATKIS 2019	BK50 2019	Evapotranspiration	0.33	25.9	0.10	0.50
Weilerswist	CLC 2012	BK50 2015	Soil Moisture	0.13	51.5	−6.90	−1.14
Weilerswist	CLC 2012	BK50 2019	Soil Moisture	0.13	54.3	−7.11	−1.19
Weilerswist	ATKIS 2019	BK50 2015	Soil Moisture	0.13	51.6	−6.76	−1.12
Weilerswist	ATKIS 2019	BK50 2019	Soil Moisture	0.13	54.4	−6.84	−1.14
Weilerswist	CLC 2012	BK50 2015	Evapotranspiration	0.26	29.6	−0.07	0.43
Weilerswist	CLC 2012	BK50 2019	Evapotranspiration	0.25	25.3	−0.11	0.44
Weilerswist	ATKIS 2019	BK50 2015	Evapotranspiration	**0.28**	29.1	**0.06**	0.44
Weilerswist	ATKIS 2019	BK50 2019	Evapotranspiration	0.26	**24.0**	−0.02	**0.45**
Metternich	CLC 2012	BK50 2015	Soil Moisture	0.13	**−33.4**	0.05	**0.07**
Metternich	CLC 2012	BK50 2019	Soil Moisture	0.11	−35.6	0.02	0.06
Metternich	ATKIS 2019	BK50 2015	Soil Moisture	0.13	−33.8	0.06	**0.07**
Metternich	ATKIS 2019	BK50 2019	Soil Moisture	0.12	−37.8	0.02	0.04
Metternich	CLC 2012	BK50 2015	Evapotranspiration	0.17	−34.5	0.09	**0.07**
Metternich	CLC 2012	BK50 2019	Evapotranspiration	0.18	−37.8	0.07	0.04
Metternich	ATKIS 2019	BK50 2015	Evapotranspiration	**0.19**	−35.0	**0.10**	0.06
Metternich	ATKIS 2019	BK50 2019	Evapotranspiration	**0.19**	−41.1	0.06	0.01

refined models across all performance metrics, including R^2 (data not shown). Thus, the refined models are unable to improve simulation of seasonal variability as well. Multiple factors work together to produce this outcome. First, WWTP emission loads are dominated by discharge volumes, which are the same in both sets of model variants. While load seasonality was more pronounced in the refined models due to the increased variability of N concentrations, the basic seasonal pattern was discernible in both the base models and the refined models. Second, the largest WWTP in Flerzheim exhibits the smallest amount of seasonality in N emissions, possibly reducing the seasonal variability of the total emission loads from all WWTPs across the catchment. Finally and most importantly, the effect of seasonality in WWTP emissions on in-stream loads in Metternich is probably overridden by other catchment processes, like seasonal differences in in-stream N retention (Schwientek & Selle 2016) as well as increased nitrate wash-out from soils during winter in response to plant dormancy (Bauwe et al. 2020, Kothe et al. 2021).

4.3.4. Performance of Models with Refined Agricultural Management

Table 4.9 summarizes the performance ratings for the models with refined agricultural management, which includes active tile drains in parts of the agricultural area, as well as updated management schedules with crop rotation. Focusing on the simulation results for streamflow, there is a clear tendency of improved model performance at all three gauging stations compared to Tables 4.7 or 4.8, barring some exceptions: In Essig, the already too high PBIAS scores further increase for the models that adjust the CN with evapotranspiration, however only by small and probably insignificant amounts. The corresponding KGE scores show no clear tendency in either direction. All other performance ratings for Essig improve slightly but noticeably. At the gauges in Morenhoven and Weilerswist, performance ratings improve by small amounts across almost all model variants and performance metrics (only the change in NSE scores for Weilerswist remains inconclusive when ET based CN adjustment is used). Although altered plant growth due to modified management schedules and crop parameters may affect catchment hydrology, the improved model performance for streamflow is probably caused mostly by the inclusion of tile flow in the models. Tile drains are common under many agricultural areas in the Swist catchment, but were neglected completely in all preceding model variants. Omission of tile flow in catchments with widespread tile drainage can severely compromise SWAT simulation results, as shown for instance by Koch et al. (2013).

In contrast to the streamflow simulation results, the effects of the modifications on simulated NO_3^--N loads in Metternich are more ambiguous. While there is a clear increase in R^2 scores, the already distinctly negative PBIAS scores from Tables 4.7 and 4.8 further decrease to values

Table 4.9.: Performance criteria for the simulation results of the eight uncalibrated model variants with refined management parameterization. The different model variants are defined by their underlying land use (LU) and soil data as well as CN adjustment procedure. Model performance is evaluated for streamflow at the gauges in Essig, Morenhoven and Weilerswist as well as NO_3^--N loads at the water quality monitoring station in Metternich (simulation years 2012 - 2016). The best value for each performance criterion and monitoring site is typeset in bold.

Monitoring Site	LU Data	Soil Data	CN Adjustment	R^2	PBIAS	NSE	KGE
Essig	CLC 2012	BK50 2015	Soil Moisture	0.09	120.7	−7.57	−1.44
Essig	CLC 2012	BK50 2019	Soil Moisture	0.10	136.3	−8.23	−1.61
Essig	ATKIS 2019	BK50 2015	Soil Moisture	0.08	128.9	−8.64	−1.63
Essig	ATKIS 2019	BK50 2019	Soil Moisture	0.09	145.5	−9.46	−1.82
Essig	CLC 2012	BK50 2015	Evapotranspiration	0.52	96.7	0.29	−0.02
Essig	CLC 2012	BK50 2019	Evapotranspiration	0.52	**84.6**	**0.35**	**0.09**
Essig	ATKIS 2019	BK50 2015	Evapotranspiration	**0.53**	101.0	0.28	−0.07
Essig	ATKIS 2019	BK50 2019	Evapotranspiration	0.51	84.9	0.33	0.08
Morenhoven	CLC 2012	BK50 2015	Soil Moisture	0.25	36.0	−2.86	−0.39
Morenhoven	CLC 2012	BK50 2019	Soil Moisture	0.26	37.4	−3.07	−0.45
Morenhoven	ATKIS 2019	BK50 2015	Soil Moisture	0.26	36.7	−2.60	−0.33
Morenhoven	ATKIS 2019	BK50 2019	Soil Moisture	0.27	39.8	−2.89	−0.42
Morenhoven	CLC 2012	BK50 2015	Evapotranspiration	0.37	24.4	0.17	0.54
Morenhoven	CLC 2012	BK50 2019	Evapotranspiration	0.37	**21.3**	0.13	**0.55**
Morenhoven	ATKIS 2019	BK50 2015	Evapotranspiration	0.37	25.9	**0.21**	0.53
Morenhoven	ATKIS 2019	BK50 2019	Evapotranspiration	**0.38**	23.5	0.16	**0.55**
Weilerswist	CLC 2012	BK50 2015	Soil Moisture	0.14	43.4	−5.80	−0.94
Weilerswist	CLC 2012	BK50 2019	Soil Moisture	0.14	45.8	−5.81	−0.95
Weilerswist	ATKIS 2019	BK50 2015	Soil Moisture	0.14	43.3	−5.54	−0.89
Weilerswist	ATKIS 2019	BK50 2019	Soil Moisture	0.15	47.3	−5.66	−0.92
Weilerswist	CLC 2012	BK50 2015	Evapotranspiration	0.28	26.3	−0.08	0.46
Weilerswist	CLC 2012	BK50 2019	Evapotranspiration	0.28	23.5	−0.12	0.46
Weilerswist	ATKIS 2019	BK50 2015	Evapotranspiration	**0.30**	26.0	**0.06**	0.48
Weilerswist	ATKIS 2019	BK50 2019	Evapotranspiration	**0.30**	**23.4**	0.02	**0.49**
Metternich	CLC 2012	BK50 2015	Soil Moisture	0.21	**−45.0**	0.07	0.05
Metternich	CLC 2012	BK50 2019	Soil Moisture	0.16	−47.7	0.01	0.01
Metternich	ATKIS 2019	BK50 2015	Soil Moisture	0.21	−45.9	0.06	0.04
Metternich	ATKIS 2019	BK50 2019	Soil Moisture	0.20	−46.2	0.05	0.03
Metternich	CLC 2012	BK50 2015	Evapotranspiration	0.27	−46.2	0.09	**0.06**
Metternich	CLC 2012	BK50 2019	Evapotranspiration	0.24	−48.5	0.06	0.02
Metternich	ATKIS 2019	BK50 2015	Evapotranspiration	0.28	−47.9	0.08	0.04
Metternich	ATKIS 2019	BK50 2019	Evapotranspiration	**0.31**	−47.9	**0.10**	0.05

below −40 % for all model setups. On the one hand, the higher R^2 scores hint at improved simulation of in-stream NO_3^--N dynamics. On the other hand, the drop in PBIAS signals greater underestimation of overall NO_3^--N loads transported in the Swist and its tributaries. NSE and KGE scores for the different model setups change only by small amounts in both directions, which is plausible given that the two metrics incorporate various error components - including both bias and lack in collinearity. The improved simulation results in terms of NO_3^--N dynamics are most likely due to the introduction of tile drains to the models and the overall better performance in simulating streamflow. In contrast, the increased underprediction of overall NO_3^--N loads can mostly be attributed to the changes in crop growth and agricultural management that were adopted in the refined models. First, rising agricultural yields caused by the modified crop parameter values detailed in Table 4.5 lead to increased N withdrawal from cropland HRUs during harvest and consequently a long-term reduction of the N quantities available for transport into surface waters. Second, the introduction of crop rotation in the refined models comes with smaller overall inputs of N fertilizer, as winter wheat receives the greatest amount of N fertilizer of all field crops (see Table 4.4 on page 70) but is grown only on approximately 40 % of the arable land area in the Swist catchment (remember that wheat is grown on 100 % of the arable land area in the base model setups while receiving the same amount of N fertilizer).

Expanding on this, there is some uncertainty as to whether the N fertilizer quantities applied in the SWAT models are high enough to realistically represent the actual fertilizer use by farmers. Anecdotal evidence from conversations with farmers suggests this is not the case and that the values given in Table 4.4 are in fact an underestimate. In this respect, it should be noted that the basis for fertilizer quantities applied in the models are merely best practice recommendations by the Chamber of Agriculture in NRW (Paffrath et al. 2021) with no guarantee these recommendations resemble actual agricultural management practices in the study area. Furthermore, Paffrath et al. (2021) base their recommendations on the 2020 amendment of the German Fertilizer Application Ordinance, which imposes stricter requirements on the application of N fertilizers to agricultural land. These stricter requirements were not yet active during the study period from 2012 to 2018, providing further indication for underestimated levels of applied N fertilizer in the models. In order to assess the numbers listed in Table 4.4, additional sources with information on applied N fertilizer quantities were consulted for comparison. These sources encompass official data for all administrative districts in NRW compiled by the Chamber of Agriculture (“nutrient reports”), management records for a small number of agricultural fields in the Swist river basin obtained from farmers, as well as small-scale SWAT studies that had access to agricultural management records from farmers in other parts of Germany and Europe.

The nutrient report of the Chamber of Agriculture NRW analyzes data on organic and mineral fertilizer in NRW and is updated approximately every three years. The nutrient report of 2017 estimates average annual N fertilization for the agricultural area in NRW to be 206 kg N/ha in 2016 (Chamber of Agriculture NRW 2018). Applying a different methodology, the 2021 report estimates average annual N fertilization for the agricultural area in NRW to be 196 kg N/ha in the years between 2015 and 2017 (Chamber of Agriculture NRW 2021). The local estimates for the two administrative districts that cover the majority of the Swist catchment area are substantially lower than the NRW-wide averages. For the Rhein-Sieg-Kreis district, the reports assume the total N fertilizer inputs to be approximately 175 kg N/ha in 2016 (Chamber of Agriculture NRW 2018) and 165 kg N/ha in the three years from 2015 to 2017 (Chamber of Agriculture NRW 2021). For the Euskirchen district, the reports assume the total N fertilizer inputs to be approximately 175 kg N/ha in 2016 (Chamber of Agriculture NRW 2018) and 169 kg N/ha from 2015 to 2017 (Chamber of Agriculture NRW 2021). Note that these numbers include fertilizer application to agricultural land uses other than cropland (e.g. grassland and orchards), which makes it difficult to directly compare them with the fertilizer quantities implemented in the SWAT models. Still, they most likely exceed N fertilization in both the base and refined model setups, especially since cropland (which typically receives more fertilizer than other forms of agricultural land) is over-represented in the Swist river basin when compared with both NRW as a whole as well as the Rhein-Sieg-Kreis and Euskirchen districts (Chamber of Agriculture NRW 2021).

As part of a broader research project, Kothe et al. (2021) requested management records from farmers to interpret nutrient and pesticide concentrations measured in three agricultural tile drains in the Swist catchment. These management records encompass five individual fields that cover a total area of around 28 ha. Four of the five fields are used for cultivation of grain, root crops, and other annual field crops, while the fifth field is an orchard planted with apple trees. The fields are farmed by three different agricultural holdings. The management records span the time period from 2016 to 2020 for three of the four fields for crop cultivation. Management data for the fourth cropland field covers the years 2019 and 2020. The cultivated crops during these years include winter wheat, winter barley, sugar beet, winter canola, winter rye and soybean. Data for the orchard was only available for 2020. All management records give detailed account on all applied fertilizer and pesticide quantities during these years. For most years, N is exclusively supplied as mineral fertilizer like calcium ammonium nitrate (CAN), whereas organic fertilizers like manure or fermentation residues are used only occasionally. Annual N fertilization for all fields and years is roughly 80 to 160 kg N/ha, depending on the cultivated crop. Soybeans did not receive any N fertilizer in the year they were grown. The apple trees received approximately 110 kg N/ha in 2020. All these numbers are in very good agreement

with the N fertilizer quantities listed in Table 4.4 for the refined model setups. Of course, they represent only a small portion of all cropland areas and cultivation practices within the Swist catchment and cannot be extrapolated to the entire river basin.

Musyoka et al. (2023) conducted a SWAT study in a small agricultural catchment (66 ha) in Petzenkirchen, Lower Austria. Detailed agricultural management records were available for each of the 37 agricultural fields in the catchment from 2008 to 2018. These fields cover around 87 % of the entire catchment area and are cultivated mostly with corn, winter barley, winter wheat and canola. The authors provide a two-year excerpt (2012 and 2013) of a management schedule applied in their SWAT model including applied fertilizer quantities.[6] According to this schedule, winter wheat was supplied with 183 kg N/ha in 2012. After cultivation of clover as a winter catch crop, corn was supplied with 127 kg N/ha in 2013. Average annual N input on each field via mineral fertilizers was 121 kg N/ha from 2008 to 2018 (Musyoka et al. 2023). Seen in isolation, these numbers are in good agreement with the N fertilizer quantities listed in Table 4.4 for the specified field crops. However, they do not include organic fertilizers like pig slurry applied to some of the fields in the Petzenkirchen catchment (Musyoka et al. 2023). In an earlier SWAT study, Musyoka et al. (2021) estimate overall annual N fertilization in the catchment to be approximately 200 kg N/ha from 2008 to 2016, which is considerably higher than N fertilizer quantities applied in both the base and the refined SWAT models for the Swist river basin. Of course, conditions for crop cultivation (e.g. in terms of climate and soils) as well as spatial scale differ between both study areas, which has to be kept in mind when comparing and interpreting the given numbers.

Bauwe et al. (2019) used SWAT to investigate the effect of N fertilization on exported NO_3^--N in-stream loads and crop yields in a small agricultural catchment (167 ha) in Mecklenburg-Western Pomerania near the German Baltic Sea coast. During the research period from 2004 to 2015, the entire cropland in the catchment (four different fields covering about 95 % of the catchment area) was farmed by one agricultural holding providing detailed management records for the study. The principal field crops grown in the catchment were winter wheat, winter barley, canola and corn silage, with occasional planting of other crops like sugar beet and peas. Average annual N fertilization was 180 kg N/ha, which is in very good agreement with the base model setups but higher than the fertilizer quantities in the SWAT models with refined agricultural management. Crop-specific quantities of N fertilizer applied in the SWAT model of Bauwe et al. (2019) are 221 kg N/ha for winter wheat, 176 kg N/ha for winter barley, 204 kg N/ha for

[6] Musyoka et al. (2023) specify N fertilizer quantities as kg NAC/ha, stating NAC to mean N-acetyl cysteine. However, N-acetyl cysteine is a medical drug. NAC also is a trade name for some CAN fertilizer products, which contain approximately 27 % elemental N. Accordingly, fertilizer quantities provided by Musyoka et al. (2023) were converted to kg N/ha assuming a nitrogen mass fraction of 27 %.

canola and 117 kg N/ha for corn silage. For three of the four crops, these values are markedly higher than those given in Table 4.4 for the refined SWAT models of the Swist river basin (with the notable exception of corn silage, which receives considerably more N fertilizer in the present study). Again, all numbers need to be interpreted with care, as conditions in both study areas may not be directly comparable.

In summary, although parts of the reference data support the N fertilizer inputs derived from Paffrath et al. (2021) for the Swist catchment, most of the consulted sources suggest higher fertilization rates to be more realistic. Given the relatively wide spread of fertilizer quantities in the reference data, the best practice recommendations of Paffrath et al. (2021) have the advantage of providing a clear and reproducible method to derive catchment-wide N fertilization rates. But even if the applied fertilization rates are considered to be realistic, it is clear that they most likely constitute the lower end of the plausible range of values, especially taking into account the stricter regulatory framework enforced after the end of the study period. Therefore, underestimation of fertilization rates is a reasonable explanation for systematic underprediction of in-stream NO_3^--N loads. Fertilization rates however work in complex interaction with many other factors that influence the resulting NO_3^--N loads. In principle, these factors may thus be equally responsible for model error at the catchment outlet. They include processes of (soil-borne) nitrogen cycling, like mineralization and denitrification, as well as in-stream N retention. Their representation and relative contribution to the overall N burden in the Swist catchment will be examined in more detail in subsequent chapters.

4.4. Concluding Discussion

On the one hand, the results in this chapter demonstrate the significant impact that individual model setup decisions can have on SWAT model output and performance. Specifically, in this study, the choice of CN adjustment procedure and the representation of agricultural management had particularly strong effects on the SWAT simulations. On the other hand, certain decisions that were anticipated to be influential had only marginal effects. This notably includes the imputation method for filling gaps in N emission data from WWTPs — whether through simple mean substitution or utilizing multivariate regression models to generate more dynamic estimates. From a practical point of view, providing recommendations from these results for future SWAT studies regarding the importance of specific model setup choices would be intriguing. However, due to the unique nature of surface water catchments and the plethora of potential SWAT model specifications and environmental data sets available for parameterization, it is not deemed feasible to draw generalizable conclusions on model setup from the

results in this chapter. Instead, it is advised to thoroughly document all major decisions taken during model setup and evaluate the impact of diverging model setup alternatives prior to calibration (as done in this thesis chapter). In this respect, it is crucial to remember that all results of subsequent model calibration are conditioned on the model setup, i.e. the specification of the model input data and model structure along with the associated errors.

The question remains which of the twenty-four model variants scrutinized in this chapter should be chosen for calibration and subsequent assessment of N emissions in the Swist river basin. In principle, the decision on a specific model could be made part of the calibration process, varying the model setup along with the parameter values and thereby accounting for error interaction between parameter values, input data and model structure. This would be in the spirit of Beven & Binley (1992), who envisioned the application of GLUE to evaluate different model structures on top of the model parameters. However, this procedure was deemed to be too computationally expensive[7] and a decision was made to base calibration and all subsequent analysis on only one of the examined model setups. This model setup is based on:

1. ATKIS 2019 land use data.
2. BK50 2019 soil data.
3. Daily CN adjustment with evapotranspiration.
4. Imputation of WWTP N emission data with linear regression estimates.
5. Tile drained agricultural areas and crop rotation.

Although only slightly superior in terms of the performance metrics listed in Tables 4.7 to 4.9, the 2019 land use and soil data sets were preferred to their older counterparts because they are based on updated information and - in the case of land use - exhibit a greater level of detail. The decision for these two data sets corresponds to the fourth spatial configuration in Table 4.1 comprising a total of 1824 HRUs (in 153 subbasins). Clearly, evapotranspiration was preferred to soil moisture as the basis for CN adjustment due to the vastly enhanced model performance associated with this option (also see Tables 4.7 to 4.9). In contrast, the decision for linear regression imputation of WWTP emission data as well as tile drainage and crop rotation in agricultural management is not unequivocally backed by the model performances evaluated in this chapter. Regression-based imputation of WWTP N emissions leads to slight deterioration of the model fit to observed NO_3^--N in-stream loads across all performance metrics as

[7] Parameter calibration as described in Chapter 5 took about a week of model runtime using the available hardware. In an exhaustive approach that incorporates all model setups, each of the examined parameter sets would have to be combined with every model setup variant, amounting to approximately twenty-four weeks of overall processing time.

evident from Table 4.8. Inclusion of tile drains and crop rotation in the models leads to slightly improved streamflow simulation as well as better correlation of simulated NO_3^--N in-stream loads with observed data, while simultaneously causing substantially increased underestimation of NO_3^--N in-stream loads.

Despite these shortcomings, both regression-based imputation of WWTP data and refined agricultural management are likely to provide a more realistic representation of important factors influencing N pollution in the Swist catchment. In this respect, keep in mind that model errors may equally arise from the model parameters, which at this stage are mostly unchanged from their ArcSWAT default values. Hence, they do not yet sufficiently reflect the characteristics of the Swist catchment. Again, all results of subsequent model calibration are conditioned on the model setup, and a more realistic model setup may thus facilitate parameter identifiability and lead to better results in model calibration and validation. The likely underestimation of N fertilization may pose a problem in this regard, but to some degree is probably present in all model setups examined. The impact of N fertilizer inputs on the calibrated parameter values and model output will be given special attention in the forthcoming thesis chapters.

5. Model Calibration and Validation

5.1. Introductory Remarks

This chapter deals with the identification of SWAT parameter values that significantly improve the preliminary SWAT simulation results for the Swist river basin presented in Chapter 4. For this purpose, the SWAT model setup selected in Chapter 4 is calibrated and validated. In order to account for equifinality and assess the effect of parameter uncertainty on the simulation results, a calibration approach based on Monte Carlo (MC) analysis is applied to identify multiple viable parameter sets. To be considered viable, parameter sets need to satisfy two different requirements: acceptable simulation results for both streamflow and in-stream NO_3^--N loads (multi-objective model calibration). Limiting the number of calibrated parameters, a global sensitivity analysis serves to screen model parameters with substantial impact on the simulation results prior to calibration. Considering model validation, a *differential split-sample* strategy (Klemeš 1986) is pursued in that the validation period (2017 - 2018) features extraordinarily dry conditions not present in the calibration period (2012 - 2016). The calibrated modeling results are additionally checked for plausibility by comparison with monitoring data not used in split-sample calibration and validation. Only those parameter sets that yield satisfactory model performance in both the calibration and validation period are considered in subsequent analysis of simulated N emissions in Chapter 6.

Model performance was evaluated based on the same observed data as in Chapter 4 (names in parentheses indicate the surface water stream each monitoring site is located at): Daily streamflow time series at the gauging stations in Essig (Steinbach), Morenhoven (Swist) and Weilerswist (Swist), as well as daily in-stream NO_3^--N loads at the water quality monitoring station in Metternich (Swist). See Subsection 4.3.1 for a more detailed description of these data sets. The fit of the simulation results to each set of observed data was assessed calculating the four performance metrics $R2$, PBIAS, NSE and KGE with R's *stats* (R Core Team 2022) and *hydroGOF* (Zambrano-Bigiarini 2020) packages (see Subsection 4.3.1). As a holistic metric that integrates over several independent facets of model error (Gupta et al. 2009) and exhibits relatively low sampling uncertainty (Lamontagne et al. 2020), KGE was the performance metric

used to construct objective functions for sensitivity analysis as well as model calibration and validation. The other performance metrics were used to complement KGE where appropriate to gain a deeper understanding of the simulation results. Section 5.2 presents the results of global sensitvity analysis, while Sections 5.3 and 5.4 summarize and discuss the results of model calibration and validation, respectively. As for the entire thesis, all SWAT simulations were run using SWAT2012, revision 670.

5.2. Sensitivity Analysis

Sensitivity analysis (SA) is used in this study to distinguish influential from non-influential model parameters and thus narrow the set of parameters considered in model calibration. As streamflow and NO_3^--N loads are supposed to be calibrated simultaneously (see Section 5.3), the ultimate goal is to identify a combined set of influential parameters for both streamflow and nitrate simulation. This is a typical screening setting for sensitivity analysis. As it enables efficient global screening for parameter-rich models like SWAT, the Elementary Effect (EE) test (Morris 1991) described in Subsection 3.2.3 was the SA method of choice for this purpose. The R package *SAFER* (Sensitivity Analysis for Everyone, R version) (Pianosi et al. 2015) was used to generate parameter samples for the EE method, calculate sensitivity as well as interaction indices and assess their sampling uncertainty using bootstrapping. The R package *SWATplusR*[1] (Schürz 2019) was used to automatically vary parameter values in the SWAT input files, run the SWAT2012 executable file and extract the relevant simulation results from SWAT output files for subsequent analysis in R.

In large parts, the present study follows the procedure and recommendations of Sarrazin et al. (2016) in their application of the EE test to SWAT. This entails the sampling of $n = 100$ reference points using ordinary Latin Hypercube Sampling (LHS), which led to a stable SWAT screening result in Sarrazin et al. (2016). The reference points were sampled using a uniform probability distribution for each model parameter. Around each reference point, a radial sampling design was employed, setting each parameter back to its reference value before perturbing the next parameter (this corresponds to the repeated *one-at-a-time* sampling approach described in Subsection 3.2.3). The perturbation Δ is randomly chosen for each parameter and reference point. Consequently, perturbations are unequal between the tested parameters at each refer-

[1] For most of this thesis, *SWATplusR* was preferred to the widely applied SWAT-CUP (SWAT Calibration Uncertainty Program) and the SUFI-2 calibration scheme (Abbaspour et al. 2004, Abbaspour et al. 2007) as it freely enables parallel processing, unlimited simulation runs and more flexible output analysis. However, access to the SWAT output files is more restricted in *SWATplusR*. Specifically, access to reservoir output files is unavailable in *SWATplusR* as of the writing of this thesis.

ence point. Campolongo et al. (2011) showed that the radial sampling design is more efficient than the trajectory-based design originally proposed by Morris (1991). In order to calculate the sensitivity indices, the absolute values of the EEs were used to avoid compensation between values with opposite signs as suggested by Campolongo et al. (2007). The sensitivity threshold used to distinguish influential and non-influential parameters was set to 5 % of the maximum sensitivity index.

SWAT features hundreds of individual model parameters (Arnold et al. 2013). Including all of them in sensitivity analysis was not feasible in terms of computational cost and defining meaningful parameter ranges. Instead, the published SWAT literature was reviewed to identify commonly calibrated SWAT parameters with respect to streamflow and nitrogen simulation. Compilations of such parameters can be found for instance in Douglas-Mankin et al. (2010), Tuppad et al. (2011), Arnold et al. (2012) and Yuan et al. (2015). To limit the dimensionality of the investigated uncertainty space, QUAL2E parameters for in-stream water quality simulation were generally excluded from analysis. This approach led to a set of $m = 45$ parameters listed in Table 5.1 along with their uncertainty ranges (lower and upper bounds). The ranges were set as wide as necessary to reflect the associated uncertainty. Spatially distributed parameters (marked with an astersik in Table 5.1) were varied relative to their base values, and uncertainty ranges were expressed as fractions of change mostly between ±50 %. Soil bulk densities (SOL_BD.sol) had a considerably narrower uncertainty range (between ±10 %) to stay within physically meaningful limits throughout the catchment. SWAT curve numbers (CN2.mgt) were limited to negative fractions of change (from −50 to 0 %) to adress overprediction of streamflow peaks in the uncalibrated model. Additional information on the parameters in Table 5.1 can be found in Appendix B.

With $n = 100$ reference points and $m = 45$ parameters, the EE test in this study encompasses $n \cdot (m + 1) = 4600$ parameter combinations that need to be evaluated in separate model runs. To quantify the model response with regard to streamflow, the (unweighted) arithmetic mean of KGEs across the three gauges in Essig, Morenhoven, and Weilerswist was used as an objective function. With regard to nitrate simulation, the objective function was the KGE for NO_3^--N loads at the water quality station in Metternich. Figures 5.1 and 5.2 give a visual impression of the resulting sensitivity indices μ_i (EE Mean) and interaction indices σ_i (EE Standard Deviation) for each parameter i regarding streamflow and in-stream NO_3^--N loads, respectively. Both figures show a considerable degree of correlation between sensitivity and interaction indices. The bars in Figures 5.1 and 5.2 indicate the widths of the 95 % confidence

Table 5.1.: Results of the elementary effects sensitivity analysis. The 45 considered model parameters have been varied within the specified bounds. The EE Mean columns list the obtained sensitivity indices for streamflow and nitrate loads. Parameters are sorted by descending sensitivity indices for nitrate. Ranks with respect to nitrate are specified in the first column, ranks with respect to streamflow are added in reduced font size. Asterisks behind parameter names indicate variation relative to preset base values. Units refer to the parameter bounds, not necessarily to the parameter values in SWAT (for parameters with relative variation). See Appendix B for more details on the listed parameters.

			Bounds			EE Mean	
Rank		Parameter	Lower	Upper	Unit	Streamflow	Nitrate Load
1	1	TRNSRCH.bsn	0	1	-	6.6×10^{-1}	4.7×10^{-1}
2	2	CH_K2.rte	0	127	mm H_2O/h	5.8×10^{-1}	4.0×10^{-1}
3	4	DEPIMP_BSN.bsn	2000	6000	mm	1.8×10^{-1}	2.0×10^{-1}
4	5	ALPHA_BNK.rte	0	1	-	1.8×10^{-1}	1.6×10^{-1}
5	20	SDNCO.bsn	0.9	1.3	-	3.1×10^{-3}	1.2×10^{-1}
6	31	ANION_EXCL_BSN.bsn	0	1	-	1.2×10^{-3}	1.2×10^{-1}
7	3	CN2.mgt*	−0.5	0	-	2.0×10^{-1}	9.8×10^{-2}
8	8	SOL_AWC.sol*	−0.5	0.5	-	4.1×10^{-2}	8.9×10^{-2}
9	25	CDN.bsn	0	3	-	2.7×10^{-3}	8.6×10^{-2}
10	41	NPERCO.bsn	0.01	1	-	3.0×10^{-5}	8.4×10^{-2}
11	6	CNCOEF.bsn	0.5	2.0	-	1.6×10^{-1}	6.9×10^{-2}
12	45	DIS_STREAM.hru	4	50	m	0	4.9×10^{-2}
13	10	SOL_BD.sol*	−0.1	0.1	-	2.3×10^{-2}	4.6×10^{-2}
14	15	EPCO.bsn	0	1	-	1.3×10^{-2}	4.2×10^{-2}
15	9	SOL_K.sol*	−0.5	0.5	-	2.9×10^{-2}	4.1×10^{-2}
16	27	CMN.bsn	0.0003	0.003	-	2.3×10^{-3}	4.0×10^{-2}
17	12	ESCO.bsn	0.6	1.0	-	1.9×10^{-2}	3.4×10^{-2}
18	7	CANMX.hru*	−0.5	0.5	-	5.0×10^{-2}	3.0×10^{-2}
19	33	SOL_CBN.sol*	−0.5	0.5	-	1.1×10^{-3}	2.9×10^{-2}
20	14	TDRAIN.mgt	0	72	h	1.4×10^{-2}	2.5×10^{-2}
21	24	SOL_ALB.sol	0	0.3	-	2.7×10^{-3}	8.1×10^{-3}
22	11	CH_N2.rte	0.014	0.065	s/$m^{1/3}$	2.2×10^{-2}	7.7×10^{-3}
23	21	SURLAG.bsn	1	24	-	3.0×10^{-3}	6.3×10^{-3}
24	17	SMTMP.bsn	−2	2	°C	4.9×10^{-3}	6.0×10^{-3}
25	13	RCHRG_DP.gw	0	1	-	1.4×10^{-2}	5.9×10^{-3}
26	22	TIMP.bsn	0.01	1	-	2.9×10^{-3}	5.7×10^{-3}
27	38	N_UPDIS.bsn	1	20	-	2.3×10^{-4}	4.8×10^{-3}
28	19	SFTMP.bsn	−2	2	°C	3.6×10^{-3}	3.3×10^{-3}
29	16	GWQMN.gw	500	1500	mm H_2O	7.1×10^{-3}	2.9×10^{-3}
30	18	SNO50COV.bsn	0.01	0.9	-	4.7×10^{-3}	2.4×10^{-3}
31	26	SNOCOVMX.bsn	0	5	mm H_2O	2.6×10^{-3}	2.4×10^{-3}
32	30	SMFMN.bsn	1.4	4.2	mm H_2O/°C	1.4×10^{-3}	2.1×10^{-3}
33	23	GW_REVAP.gw	0.02	0.2	-	2.7×10^{-3}	1.9×10^{-3}
34	28	REVAPMN.gw	500	1500	mm H_2O	2.2×10^{-3}	1.3×10^{-3}
35	35	OV_N.hru*	−0.5	0.5	-	7.0×10^{-4}	1.2×10^{-3}
36	37	SMFMX.bsn	4.1	6.9	mm H_2O/°C	4.5×10^{-4}	1.1×10^{-3}
37	32	EVRCH.bsn	0	1	-	1.1×10^{-3}	9.8×10^{-4}
38	44	HLIFE_NGW_BSN.bsn	5	500	d	0	8.7×10^{-4}
39	34	DRAIN_CO_BSN.bsn	10	51	mm H_2O/d	7.9×10^{-4}	8.6×10^{-4}
40	29	GW_DELAY.gw	0	500	day	1.6×10^{-3}	8.3×10^{-4}
41	39	CH_N1.sub	0.014	0.065	s/$m^{1/3}$	2.2×10^{-4}	3.9×10^{-4}
42	42	RCN.bsn	0	2	mg N/l	3.2×10^{-6}	1.8×10^{-4}
43	40	ALPHA_BF.gw	0	1	1/d	1.2×10^{-4}	9.2×10^{-5}
44	43	SHALLST_N.gw*	−0.5	0.5	-	0	5.0×10^{-5}
45	36	RES_K.res	0.0	0.5	mm H_2O/h	5.2×10^{-4}	3.5×10^{-5}

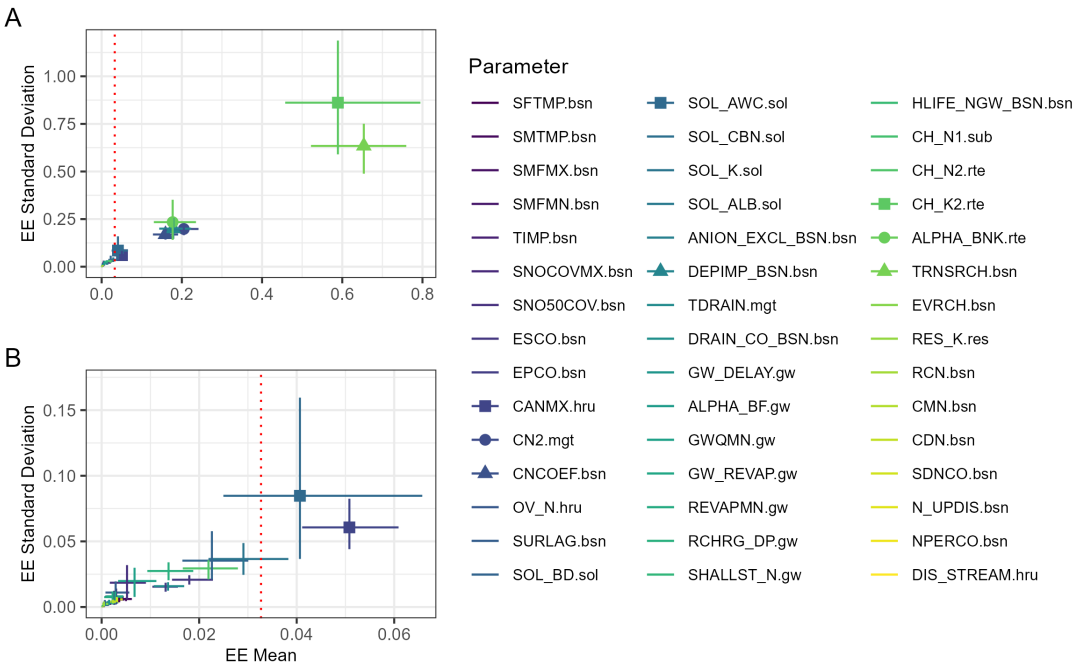


Figure 5.1.: Results of the elementary effects sensitivity analysis with respect to streamflow. Model response is measured as the average KGE for the three streamflow gauges in Essig, Morenhoven and Weilerswist. Part (A) shows the space spanned by the sensitivity indices (EE Mean) and interaction indices (EE Standard Deviation) for all considered parameters. Part (B) zooms in on the region around the chosen sensitivity threshold marked by the red dashed vertical line (5 % of the maximum sensitivity index). The bars indicate 95 % confidence intervals for both indices obtained via bootstrapping. The influential parameters (on the right of the threshold line) are marked by a symbol in the plots and legend to make them more recognizable compared to the non-influential parameters (on the left of the threshold line).

intervals for both index types obtained via bootstrapping[2] taking 100 subsamples from the original sample. There is a general pattern towards wider confidence intervals with larger sensitivity and interaction indices. The dashed red lines indicate the chosen threshold value for the sensitivity index to distinguish influential and non-influential parameters. On that basis, the EE test finds that streamflow is insensitive to a total of 37 parameters and that NO_3^--N loads are insensitive to a total of 25 parameters. Accordingly, there remain eight parameters that are deemed influential for streamflow and twenty parameters that are considered influential for NO_3^--N loads.

[2]See for example James et al. (2013). Briefly, the bootstrap technique uses iterative resampling from the original sample *with replacement* to approximate the sampling distribution of a specific quantity (here the sensitivity and interaction indices).

Focusing on the influential streamflow parameters in Figure 5.1, the prominent role of parameters governing transmission losses from the main channels to the deep aquifer is particularly noteworthy. This includes the parameters TRNSRCH.bsn and CH_K2.rte that exhibit by far the largest sensitivity indices. In this regard, they even surpass the curve number parameter CN2.mgt in position three; one of the most influential and frequently calibrated SWAT parameters according to the literature (Arnold et al. 2012, Yuan et al. 2015, Abbaspour et al. 2015).[3] Another channel parameter sits in position five: ALPHA_BNK.rte controls residence times and re-release of water in bank storage. It is just barely less influential on simulated streamflow than DEPIMP_BSN.bsn in fourth place. This parameter specifies the depth to a subsurface impervious layer. It is crucial for the buildup of perched water tables in the soil profile and is therefore closely linked to the occurrence of tile flow in SWAT. Notably, for reasons discussed in Chapter 6, there are no groundwater parameters that significantly affect streamflow. As expected, N cycling and transport parameters are practically irrelevant for streamflow simulation. The same is true for snow parameters. The list in Table 5.1 gives detailed account of the sensitivity indices for each parameter with regard to streamflow. The ranks of the parameters with respect to their influence on streamflow simulation are given in reduced font size.

A closer look at the twenty influential nitrate parameters in Figure 5.2 reveals that they include all of the eight influential streamflow parameters. At the top, they even mirror the SA results for streamflow to a large extent: Again, the channel parameters TRNSRCH.bsn and CH_K2.rte lead the field by a wide margin, followed by DEPIMP_BSN.bsn and ALPHA_BNK.rte in positions three and four (the curve number parameter CN2.mgt has a lower relative importance for NO_3^--N loads than for streamflow simulation and occupies position seven). Additional parameters only affecting nitrate simulation include those that specifically control soil-borne N cycling and transport, most importantly ANION_EXCL_BSN.bsn and SDNCO.bsn. While the former governs nitrate mobility in the soil profile, the latter determines the soil moisture threshold for denitrification. Distributed soil parameters (with suffix ".sol") also play a more prominent role for nitrate simulation than for streamflow simulation: While only one of them (SOL_AWC.sol, the available water capacity) has a meaningful impact on streamflow, the SA results show that four of them are influential regarding the simulated NO_3^--N loads. Interestingly, the evapotranspiration parameters EPCO.bsn and ESCO.bsn (modulating availability of soil water for plant transpiration and evaporation, respectively) emerge as influential for nitrate simulation although considered irrelevant for streamflow simulation. Again, groundwater parameters are generally deemed unimportant. Table 5.1 lists the sensitivity indices for each parameter with regard to nitrate simulation and ranks the parameters in descending order

[3] The comparatively small impact of CN2.mgt on streamflow in this study may in part be explained with the restriction to negative fractions of change from −50 to 0 %.

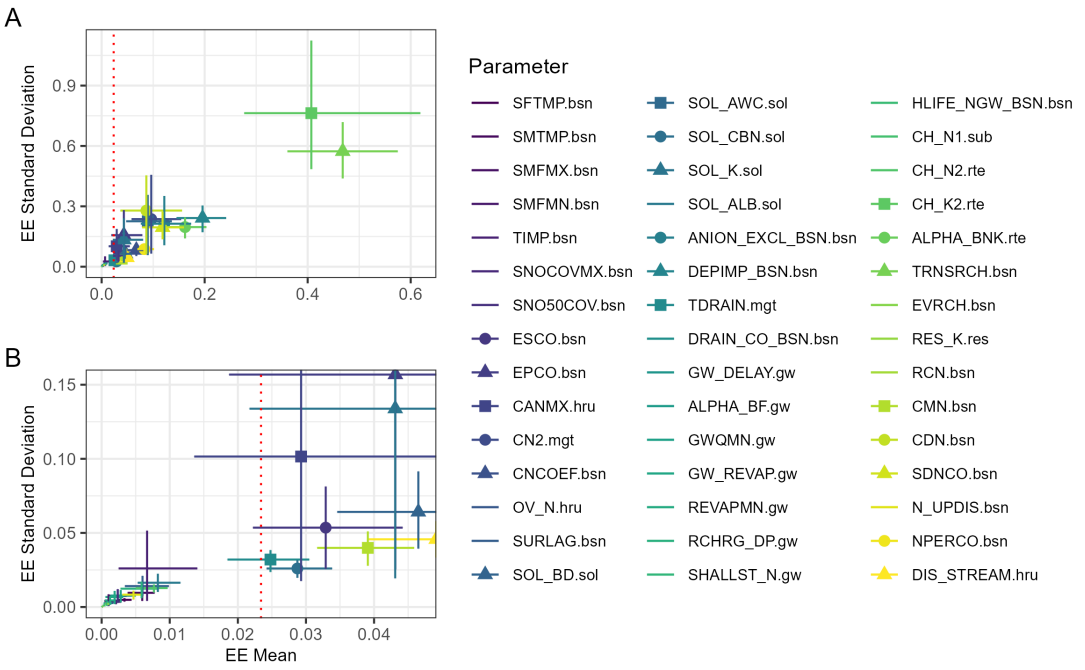


Figure 5.2.: Results of the elementary effects sensitivity analysis with respect to NO_3^--N load. Model response is measured as the KGE score for NO_3^--N loads at the water quality monitoring station in Metternich. Part (A) shows the space spanned by the sensitivity indices (EE Mean) and interaction indices (EE Standard Deviation) for all considered parameters. Part (B) zooms in on the region around the chosen sensitivity threshold marked by the red dashed vertical line (5 % of the maximum sensitivity index). The bars indicate 95 % confidence intervals for both indices obtained via bootstrapping. The influential parameters (on the right of the threshold line) are marked by a symbol in the plots and legend to make them more recognizable compared to the non-influential parameters (on the left of the threshold line).

of nitrate sensitivity. All parameters with a rank of 21 or lower are considered non-influential for both streamflow and nitrate simulation (as indicated by the horizontal line in Table 5.1). The twenty parameters with higher ranks represent the combined set of influential parameters for both streamflow and NO_3^--N loads this sensitivity analysis aimed for.

The question remains whether this screening result is stable, meaning whether the number of evaluated model runs is high enough to come to the same conclusion about influential and non-influential parameters when the EE test is repeated as described above but with a different set of sampled reference points and parameter perturbations. To answer this question, Sarrazin et al. (2016) developed a test for screening convergence which - in the context of this analysis - is equivalent to verify if all confidence intervals for the sensitivity indices left of the threshold lines in Figures 5.1 and 5.2 are narrower than the interval from 0 up to the threshold line. Visual

inspection of part B in both figures reveals that this is indeed the case and the obtained screening result (i.e. the position of the horizontal line in Table 5.1 that separates the non-influential parameters from the influential parameters) would probably be stable across multiple reiterations of the EE test. Note that this is not necessarily the case for the ranks given in Table 5.1, although the results of Sarrazin et al. (2016) suggest that screening convergence in many cases also entails ranking convergence. The stability of the parameter ranks has not been independently tested, however, as parameter ranking was not the primary goal of sensitivity analysis in this thesis. Convergence of the sensitivity indices themselves towards stable values is probably not achieved in this thesis, as that usually requires a considerable larger number of model evaluations (Sarrazin et al. 2016).

Another issue that needs to be discussed is the dependence of the screening results on specific SA design choices. Note that this is different from convergence as convergence exclusively addresses uncertainty associated with the underlying sampling process within one and the same SA design. Two design choices are of particular importance in the context of the EE test as applied in this thesis: First the choice of the objective function that quantifies the model response (Shin et al. 2013), and second the choice of the sensitivity threshold to filter out the non-influential parameters (Sarrazin et al. 2016). Considering the objective function, the EE test as described above was repeated applying three performance criteria other than KGE: NSE, PBIAS and R^2 (data not shown). Rerunning the analysis with the NSE to quantify the model response identifies soil hydraulic conductivity SOL_K.sol as an additional influential parameter for streamflow simulation. In contrast, for nitrate simulation it only detects seventeen of the twenty parameters that are considered influential using the KGE. While PBIAS generally identifies a considerably smaller subset of parameters as influential than KGE and NSE for both streamflow and nitrate simulation, R^2 detects more influential parameters. However, the only completely new influential parameter outside the set of twenty listed in Table 5.1 is Manning's roughness coefficient for the main channel, CH_N2.rte.

The choice of the sensitivity threshold is usually associated with a considerable degree of subjectivity, given the continuous scale of sensitivity indices. The sensitivity threshold selected in this study corresponds to the preliminary value used by Sarrazin et al. (2016) for their test of screening convergence (0.05 applied to a scale normalized by the maximum sensitivity index). As such, the threshold is - by design - relatively high compared to the general recommendations and validation results provided by Sarrazin et al. (2016). While a smaller threshold would probably not lead to diverging screening results for nitrate simulation, it might identify a considerable larger number of parameters to be influential for streamflow simulation (consider the pronounced gap between the threshold line and the sensitivity indices left of it in Figure 5.2 B

as well as the complete absence of such a gap in Figure 5.1 B). However, most of these parameters are already included in the set of twenty specified in Table 5.1 due to there influence on nitrate simulation. A notable exception again is Manning's roughness coefficient for the main channel, CH_N2, making this parameter the most likely candidate for a false negative screening error. As there are no other obvious candidates for additional parameters with significant influence on the simulation results, the choice of sensitivity threshold was deemed reasonable for the purposes of this study without further simulations to validate the threshold value as proposed by Sarrazin et al. (2016).

5.3. Model Calibration

Model calibration in this study is based on Monte Carlo (MC) analysis encompassing a total of 10 000 parameter sets evaluated in individial model runs. The calibration period covers the simulation years 2012 to 2016, following a four-year warm-up period. The parameter space was sampled using the R package *lhs* (Carnell 2022) to conduct ordinary Latin Hypercube Sampling (LHS). Of the twenty influential parameters identified in sensitivity analysis, only eighteen were considered for Monte Carlo analysis. The parameters CNCOEF.bsn and CDN.bsn were left out due to their interaction in the model equations (Neitsch et al. 2011) with the even more influential[4] parameters CN2.mgt and SDNCO.bsn, respectively. The eighteen remaining parameters were assumed to follow a uniform distribution within the bounds given in Table 5.1. Again, *SWATplusR* (Schürz 2019) was used to automatically vary parameter values in the SWAT input files, run the SWAT2012 executable file and extract the simulation results from the SWAT output files. For multi-objective calibration, the same two objective functions as in sensitivity analysis were applied to assess model performance: The average KGE score across the three gauges in Essig, Morenhoven and Weilerswist with regard to streamflow ($KGE_{Streamflow}$), and the KGE score at the water quality station in Metternich with regard to NO_3^--N in-stream loads ($KGE_{Nitrate}$). For both objective functions, a behavioral threshold of 0.5 was applied to determine satisfactory model performance.

An important consequence of this calibration strategy is that the model response is evaluated simultaneously at all considered monitoring sites and that the model parameters are varied homogeneously over the entire catchment area. In particular, there is no nested calibration of streamflow at individual gauges, where parameter values may differ in different gauge catchments and calibrated parameter values at downstream gauges depend on pre-calibrated param-

[4] According to the ranks specified in Table 5.1, which of course do not necessarily represent a stable ranking result and depend on the objective function used to quantify the model response.

eters at upstream gauges. Furthermore, there is also no sequential calibration of streamflow and NO_3^--N in-stream loads, where calibrated NO_3^--N parameter values depend on pre-calibrated streamflow parameters. Such calibration strategies - that involve what Daggupati et al. (2015) termed a stepwise "*staging design*" for calibration - are often used to guide manual or automatic parameter optimization (Arnold et al. 2012). When applied to MC-based calibration, they restrict sampling to specific regions of the parameter space and thus tend to underappreciate the implications of equifinality. In contrast, by evaluating the model performance simultaneously for all considered model outputs and monitoring sites, the present study explores the parameter space across the entire range of its 18 dimensions in a single sampling stage (albeit with a comparatively low sampling density).

The scatter plot in Figure 5.3 depicts the objective space each sampling point in the parameter space is mapped into. It gives an overview of the vast majority of model runs that show performance ratings from −1 to 1 for both objective functions (a handful of runs with performance ratings below −1 are not included). Of the 10 000 evaluated model runs, a total of 133 surpass the behavioral threshold for streamflow and a total of 105 exceed the behavioral threshold for NO_3^--N in-stream loads. Only 32 model runs exceed the behavioral threshold for both objectives. The maximum objective function scores are roughly 0.60 for streamflow and 0.72 for NO_3^--N loads. Although deemed satisfactory for the purposes of this thesis, these calibration performances leave ample room for improvement, especially regarding performances for streamflow simulation which exhibit a distinct ceiling near the maximum score around 0.60. This ceiling most likely is an effect of the averaging between three different KGE ratings and hints at a trade-off between the involved gauges (keep in mind that the model parameters are varied homogeneously across all gauge catchments). The maximum KGE scores at each individual gauge are roughly 0.63 in Essig and 0.70 in Morenhoven as well as Weilerswist. Only 27 of the 133 model runs that show behavioral streamflow performance yield KGE scores above the applied threshold of 0.5 at each individual gauge, with only 5 of them additionally surpassing the threshold for nitrate simulation. The option of assessing the model performance in four dimensions (with one objective function for each stream gauge) was dismissed, as the higher dimensionality would have reduced the density of solutions in the objective space and diminished the role of NO_3^--N loads in the calibration process.

The triangular shape and orientation of the point cloud in Figure 5.3 reveals a significant degree of correlation between both objective functions, with a Pearson's correlation coefficient of $r = 0.87$. The causality of this relationship seems to be mostly unidirectional: Behavioral simulation of NO_3^--N in-stream loads strictly requires values for $KGE_{Streamflow}$ that are well above 0, whereas behavioral streamflow simulation is possible even for some distinctly negative

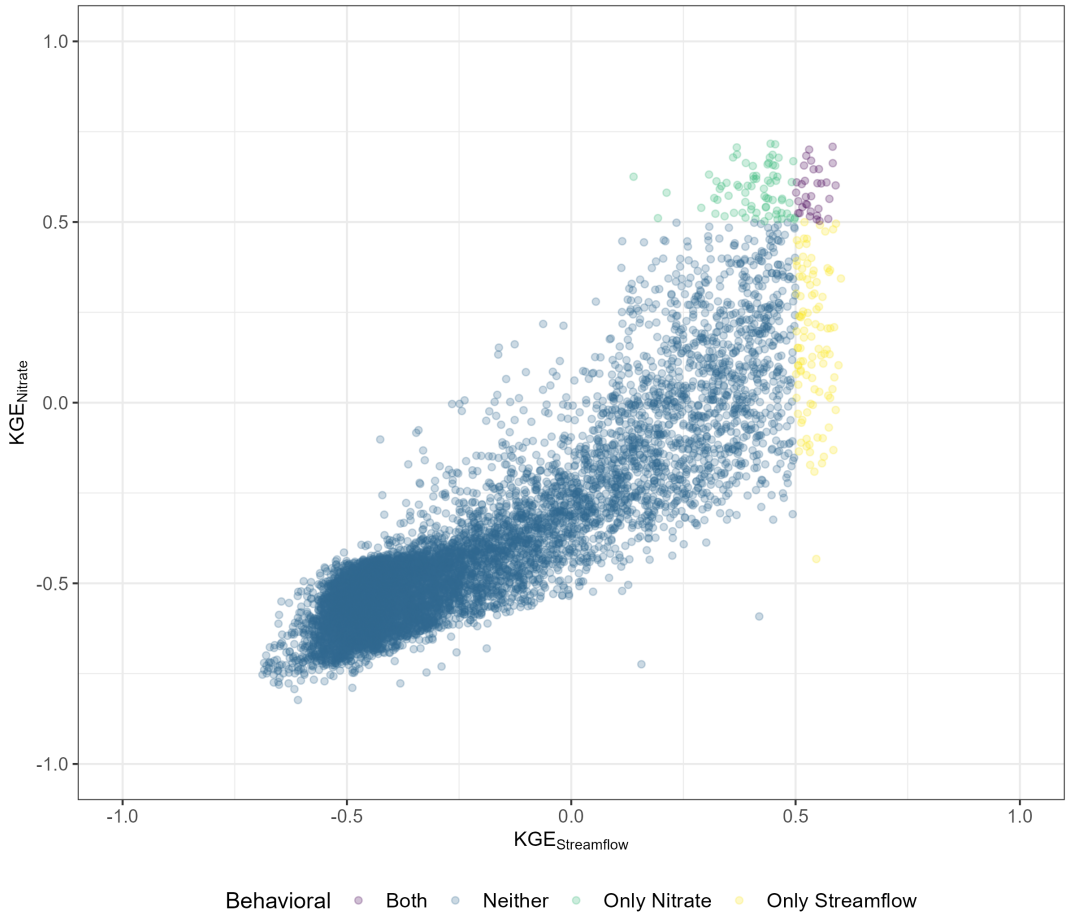


Figure 5.3.: Scatter plot of model performances during the calibration period (2012-2016) in the two-dimensional objective space. The objective function for streamflow simulation is the average KGE score across the three gauges in Essig, Morenhoven and Weilerswist. The objective function for nitrate simulation is the KGE score for NO_3^--N loads at the water quality monitoring station in Metternich. The behavioral threshold for both objectives is 0.5. Keep in mind that the KGE metric does not share the same benchmark property as the NSE metric at value 0.

values of $KGE_{Nitrate}$. Consequently, reasonable model performance regarding streamflow seems to be a prerequisite for reasonable model performance with respect to nitrate, but not *vice versa*. This is plausible, as hydrology is the main driver of N emissions into surface waters as well as the height and dynamics of the resulting in-stream loads (Bauwe et al. 2020). The set of Pareto-optimal solutions (the Pareto front) at the right and upper edges of the point cloud exhibits a more or less "*rectangular*" shape, indicating little trade-off between both objective functions for the very best model runs. However, the comparatively low density of points in this area suggests an only imprecise approximation of the real Pareto front. Care should therefore be

taken when drawing conclusions from the apparent shape of the Pareto front in Figure 5.3. The most densely populated area of the point cloud is located opposite of the Pareto front around a performance rating of approximately −0.5 for both objective functions.

As cautioned by Abbaspour et al. (2017), calibration results are highly dependent on the applied objective functions. Figure 5.4 assesses the effect of analogous objective functions built on other performance metrics than KGE - R^2, PBIAS and NSE. These objective functions clearly would have identified different sets of behavioral model runs compared to KGE, as evident from the wide scatter of the original behavioral solutions in Figure 5.4 A, D and G.[5] At the same time, Figure 5.4 B, E and G as well as C, F and I reveal a certain degree of alignment and correlation between the alternative objective functions and their KGE-based counterparts. Regarding PBIAS, note that KGE selects behavioral solutions that are centered around the perfect PBIAS score of 0 % for both streamflow simulation and nitrate simulation, albeit with a relatively large degree of variability and a preference for negative PBIAS scores for nitrate simulation. Interestingly, KGE selects behavioral solutions for streamflow that concentrate on R^2 scores roughly between 0.4 and 0.5, with many non-behavioral model runs reaching notably higher values for R^2 (panel B in Figure 5.4).

Keep in mind that R^2 and PBIAS exclusively assess collinearity and bias in the model fits, respectively. Used on their own, both are inadequate objective functions, as they ignore important forms of model error and thus only provide an incomplete performance assessment (see Subsection 3.2.1). Similar to KGE, NSE considers all possible forms of model error, and therefore is the best suited alternative to KGE as an objective function. The shape of the point cloud in the NSE objective space (Figure 5.4, G) to a large extent mirrors the point cloud in the KGE objective space of Figure 5.3, although with an even more pronounced ceiling for streamflow performances at a NSE of approximately 0.5. NSE scores for nitrate simulation reach markedly higher, with a maximum value of around 0.6. Panel H in Figure 5.4 reveals high variability of NSE scores for streamflow among the KGE-based behavioral solutions, with some NSE scores even falling narrowly below the critical benchmark of 0. In contrast, although also highly variable, NSE scores for nitrate simulation are strictly positive among the KGE-based behavioral solutions, as shown in panel I of Figure 5.4. The very best NSE scores for nitrate (> 0.5) are almost always classified as behavioral by the KGE criterion. The general conclusion is that NSE and KGE - despite some discrepancies regarding their assessment of streamflow performance - exhibit an overall acceptable degree of alignment assessing the performance of nitrate

[5] In the following, the use of the term "behavioral" exclusively refers to model runs that surpass one or both of the behavioral thresolds for the KGE-based objective functions. Note that behavioral thresholds remain undefined for the alternative objective functions examined in Figure 5.4, so there is no definitive set of behavioral solutions for these objective functions.

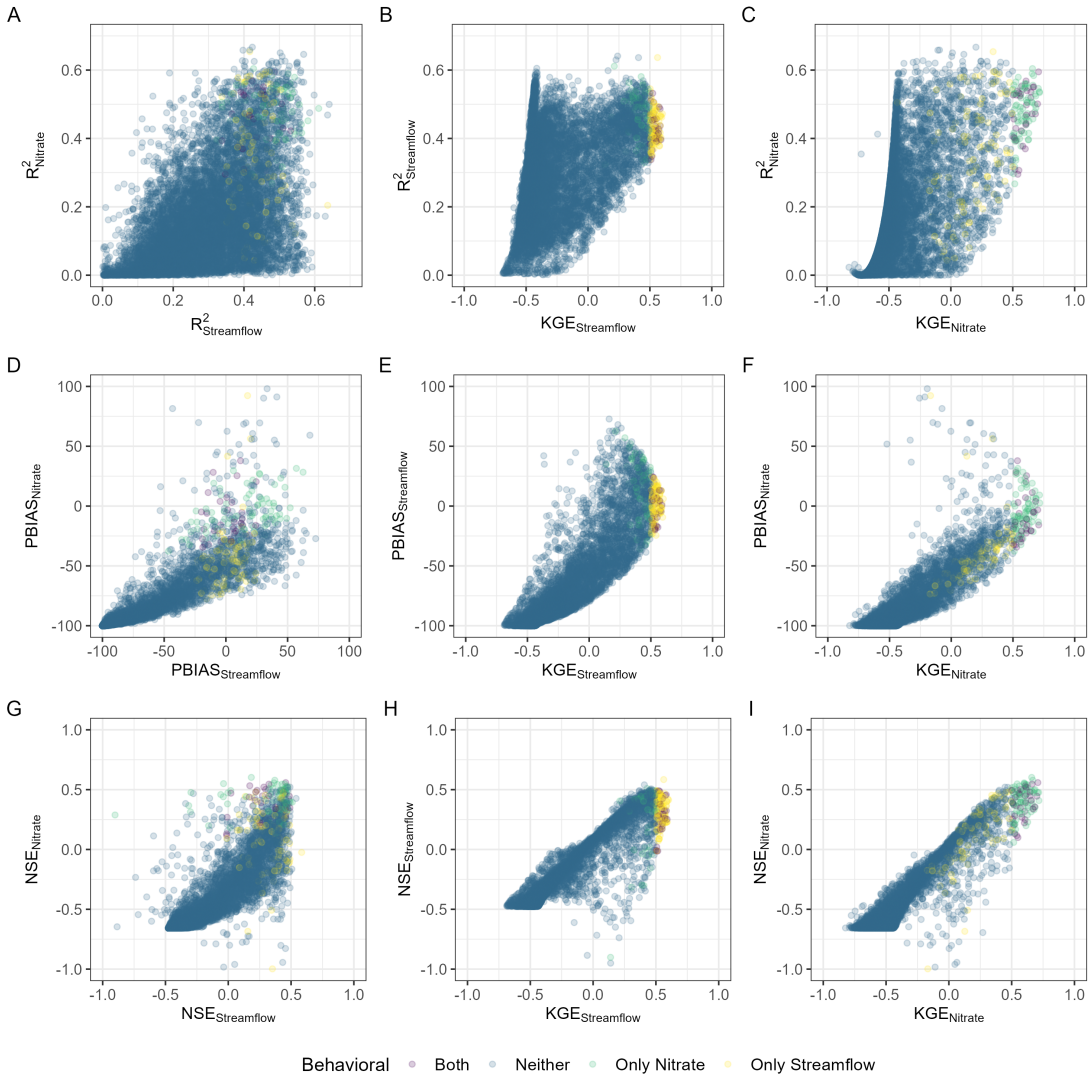


Figure 5.4.: Scatter plots of model performances during the calibration period (2012-2106) for alternative objective functions. The alternative objective functions are based on the performance metrics R^2 (top row, panels A, B and C), PBIAS (middle row, panels D, E and F) and NSE (bottom row, panels G, H and I) instead of KGE, but constructed analogously. The scatter plots in the left column (panels A, D and G) mimic Figure 5.3 in depicting the objective space for the alternative performance metrics. The middle column (panels B, E and H) compares each alternative objective function for streamflow to the KGE-based objective function for streamflow. The right column (panels C, F and I) compares each alternative objective function for NO_3^--N in-stream loads to the KGE-based objective function for NO_3^--N in-stream loads. Color-coding indicates the behavioral status according to the KGE-based objective functions (as in Figure 5.3).

simulation. As the ultimate goal of this thesis revolves around the simulated N emissions, the behavioral set of model runs identified using the KGE was deemed adequate for all further analysis.

For each included SWAT parameter, Figure 5.5 depicts the posterior distributions for the different classes of behavioral calibration runs, compared to the prior distributions for all 10 000 runs. The distributions are approximated using kernel density estimation (KDE) as implemented in base R's *stats* package (R Core Team 2022, see figure caption for details). The density estimates for the prior distributions resemble horizontal lines (apart from KDE artefacts at the interval boundaries), confirming uniform coverage of the parameter space during Latin Hypercube Sampling.[6] The degree to which the posterior distributions deviate from the prior distribution is a visual indicator of parameter influence on the simulation results (i.e. model sensitivity). The posterior distributions furthermore give insights into parameter identifiability. Guse et al. (2020) distinguish three cases in a multi-objective calibration setting: First, low parameter identifiability when all posterior distributions deviate only insignificantly from the prior distribution. Second, low parameter identifiability when (a subset of) posterior distributions deviate greatly from the prior distribution, but with multiple diverging maxima that may imply a trade-off between different calibration objectives. Third, high parameter identifiability when (a subset of) posterior distributions deviate greatly from the prior distribution and their maxima are aligned on a relatively narrow range of values. In this sense, parameter influence may be interpreted as a necessary but not sufficient precondition for identifiability.

For most parameters, at least one of the single-objective posterior distributions - for streamflow or nitrate alone - deviates significantly from the uniform prior distribution in Figure 5.5, confirming parameter influence on the simulation results. Furthermore, when the single-objective distributions are both strongly non-uniform, their maxima are mostly well aligned. Still, many parameters remain rather loosely identified by single-objective calibration: The posterior distributions for only one of the two objectives often have maxima that are stretched out over relatively wide intervals, and in some cases their shape may even be distinctly bimodal. The multi-objective posterior distributions in Figure 5.5 - based on model runs that satisfy both calibration objectives - can frequently narrow and/or shift the feasible parameter ranges to a moderate extent (see for example CN2.mgt). Moreover, when one of the single-objective distributions is bimodal, multi-objective calibration sometimes manages to alleviate that bimodality in the combined distribution (see SDNCO.bsn). These findings are consistent with the com-

[6] Stable reproduction of the uniform shape in the prior parameter distributions is a consequence of the large overall number of model evaluations used for calibration. Given the far lower number of behavioral model runs available for estimating the posterior distributions, their shape in Figure 5.5 is much more likely to be affected by sampling uncertainty. Keep that in mind when interpreting the KDE results for the posterior distributions.

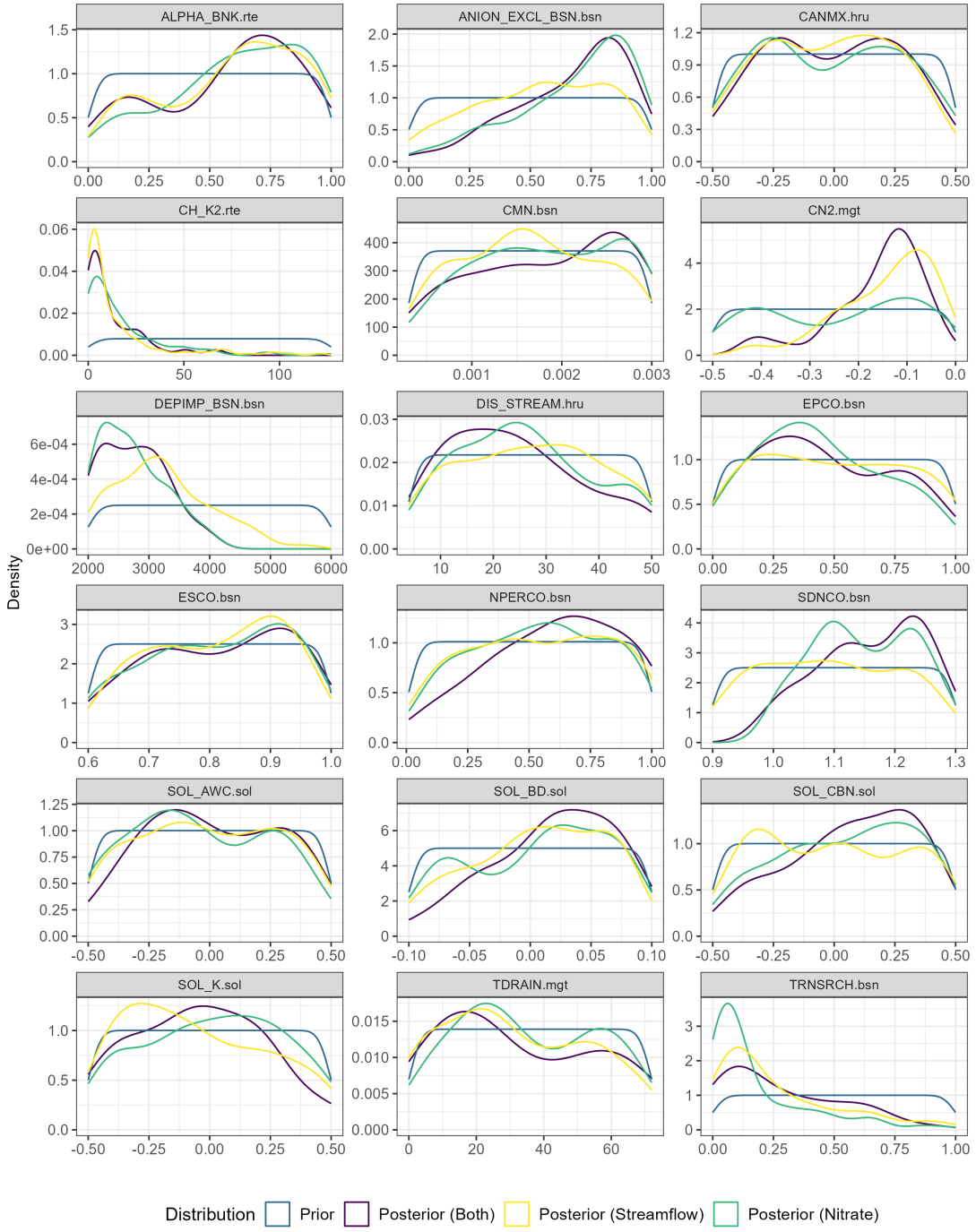


Figure 5.5.: SWAT parameter distributions for the calibration period (2012-2016). The parameter distributions were approximated using kernel density estimation (KDE) with Gaussian kernels as implemented in base R's *stats* package (R Core Team 2022). The density estimates are based on all 10 000 model runs for the prior distributions, as well as 133, 105 and 32 model runs for the posterior distributions regarding streamflow, nitrate and both objectives, respectively. The downward bends at either side of the distributions are KDE artefacts that originate from uncorrected boundary bias. While more visible for the uniform prior distributions, they equally affect the posterior distributions.

mon wisdom that multi-objective calibration tends to better constrain model parameters and improve their identifiability (Guse et al. 2020, Sprenger et al. 2022). However, in contrast to this notion, there are also cases where the combined posterior distribution more or less resembles the single-objective distributions and does not seem to add much information in terms of parameter identifiability (e.g. CANMX.hru and ESCO.bsn).

Looking at the few parameters with divergent maxima in the single-objective distributions, there are further examples for both enhanced and unaffected parameter identifiability after multi-objective calibration. For parameters CMN.bsn and SOL_K.sol, the combined distributions offer a clear indication on how to reconcile the conflicting results of single-objective calibration. While for CMN.bsn the maximum of the combined distribution picks a side and aligns with the maximum of the nitrate posterior distribution, it takes a middle ground for SOL_K.sol, falling on a relatively narrow interval between the two single-objective maxima. In contrast, the combined distribution for DEPIMP_BSN.bsn has a plateaued maximum that spans the entire distance between the two single-objective maxima, providing no additional information about the most feasible range of parameter values. Such failure of multi-objective calibration to further constrain individual model parameters is not unusual and has been reported for example by Holmes et al. (2020). In these cases, parameter identifiability would likely benefit from alternative and/or additional calibration objectives. Efstratiadis & Koutsoyiannis (2010) suggest a ratio of 1:5 or 1:6 between calibration objectives and calibrated parameters as a criterion for parsimony in catchment modeling, extending an empirical rule established and repeatedly confirmed in various studies (Jakeman & Hornberger 1993, Schoups et al. 2008, among others). Following this argument, calibration of SWAT in the present thesis would need to include at least one additional objective besides streamflow and NO_3^--N in-stream loads to adequately accommodate the 18 considered model parameters.

The channel parameters TRNSRCH.bsn and especially CH_K2.rte exhibit by far the highest identifiability, with pronounced maxima of all posterior distributions that align at or very close to values of 0. This means that the calibration process strongly favors model parameterizations that result in only small transmission losses of water from the SWAT main channels to the deep aquifer. Figure 5.6 shows various dotty plots for CH_K2.rte for closer examination of the model response to this parameter. The two rightmost panels in the bottom row depict overall model response as measured by the KGE-based objective functions, with a clear pattern of increasing model performances when CH_K2.rte approaches 0 both for streamflow and nitrate simulation. This basic pattern is replicated for all alternative objective functions based on R^2, PBIAS and NSE (see the remaining panels in the two rightmost columns), underscoring the robustness of the identified parameter range. Furthermore, the same pattern emerges when looking at the

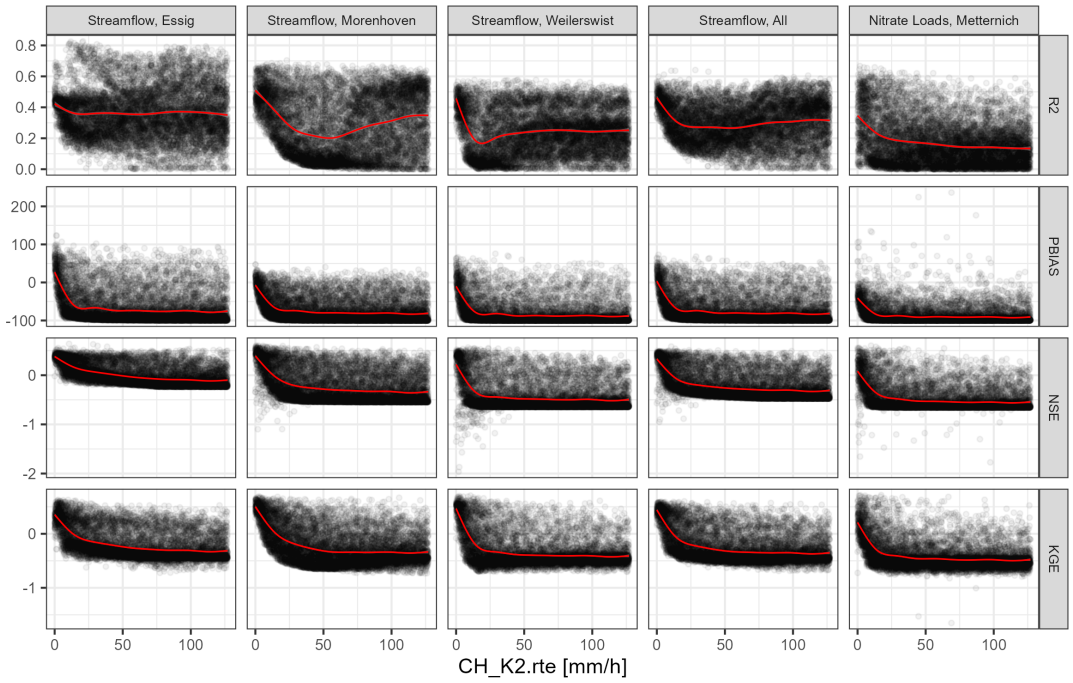


Figure 5.6.: Dotty plots for the SWAT channel parameter CH_K2.rte (hydraulic conductivity of the main channel alluvium) in the calibration period (2012-2016). Each row shows the dotty plots for another performance metric (R^2, PBIAS, NSE or KGE). Each column shows the dotty plots for another model output ("Streamflow, All" denotes the mean value of the respective performance metric across the three gauging stations in Essig, Morenhoven and Weilerswist). The regression fits depicted as red lines were obtained using local polynomial regression as implemented in the *loess* function of base R's *stats* package (R Core Team 2022).

performance ratings for streamflow at the individual gauges (i.e. the three leftmost columns in Figure 5.6), with one notable exception: The regression line for the PBIAS scores at the Essig gauge reveals a slight tendency to overpredict streamflow when CH_K2.rte is too close to 0, reflecting the problem of extended "*no-flow*"-periods in Essig despite constant discharge from the upstream Steinbach reservoir fed into the model. For SWAT, the only possibility to reproduce the streamflow cessation observed in Essig is setting CH_K2.rte to values above 0 to remove water from the main channel before it reaches the gauging station. As CH_K2.rte and TRNSRCH.bsn are varied uniformly over the entire catchment area, there is a trade-off between the gauging stations in the calibration process regarding simulation of transmission losses, with a need to find a balance between the Steinbach (Essig gauge) and Swist watercourse (Morenhoven and Weilerswist gauges).

5.4. Model Validation

5.4.1. Split-Sample Validation

The 32 parameter sets that produce behavioral model performance for both streamflow and NO_3^--N in-stream loads in the calibration period from 2012 to 2016 have been validated in extended model runs that add the years 2017 and 2018. Crucially, this validation period includes exceptionally dry conditions not present in the relatively wet calibration period, especially in the year 2018 (see Figure 2.4 and Table 2.2 in Chapter 2). In this respect, the present study thus aims at a *differential split-sample* strategy for model validation, as advocated by Klemeš (1986). Figure 5.7 gives an overview of the model performances in both the calibration and validation period across the 32 behavioral parameter sets. Interestingly, there is no clear difference in the objective function medians between calibration and validation with respect to both streamflow and NO_3^--N loads (see the two rightmost panels in the bottom row of Figure 5.7). Therefore, in terms of the objective functions, the validation period does not yield a systematic decline in model performance as typically reported (Gassman et al. 2007, Gassman et al. 2014), despite the fact that there is a deliberate discrepancy in environmental conditions between the calibration and validation period. However, there is a clear increase in variability around the medians, meaning that there are both substantially better and worse model performances in the validation period with regard to streamflow as well as NO_3^--N loads. In total, there remain 11 of the original 32 parameter sets that continue to produce behavioral model performance (regarding objective functions based on KGE) for both streamflow and NO_3^--N loads in the validation period.

A closer look at the various other metrics included in Figure 5.7 gives further insight into the differences between the calibration and validation periods. Considering overall streamflow simulation across all gauges, the alternative objective functions based on R^2 and NSE substantially improve from calibration to validation with more or less unchanged variability. In contrast, the alternative objective function that is based on PBIAS systematically increases from values centered around 0 to a mostly positive range of values (see panels in the second column from left). Taken together, these performance ratings indicate an improvement in capturing the streamflow peaks and overall dynamics in the validation period, at the expense of introducing considerable streamflow overprediction (with both developments likely being balanced out in the practically unchanged median of the combined KGE measure). Looking at the model performance ratings at the individual stream gauges (see the three leftmost columns in Figure 5.7), it is evident that the increase in the overall PBIAS rating can solely be attributed to streamflow overprediction in Essig. Here, the PBIAS scores, which already indicate a mod-

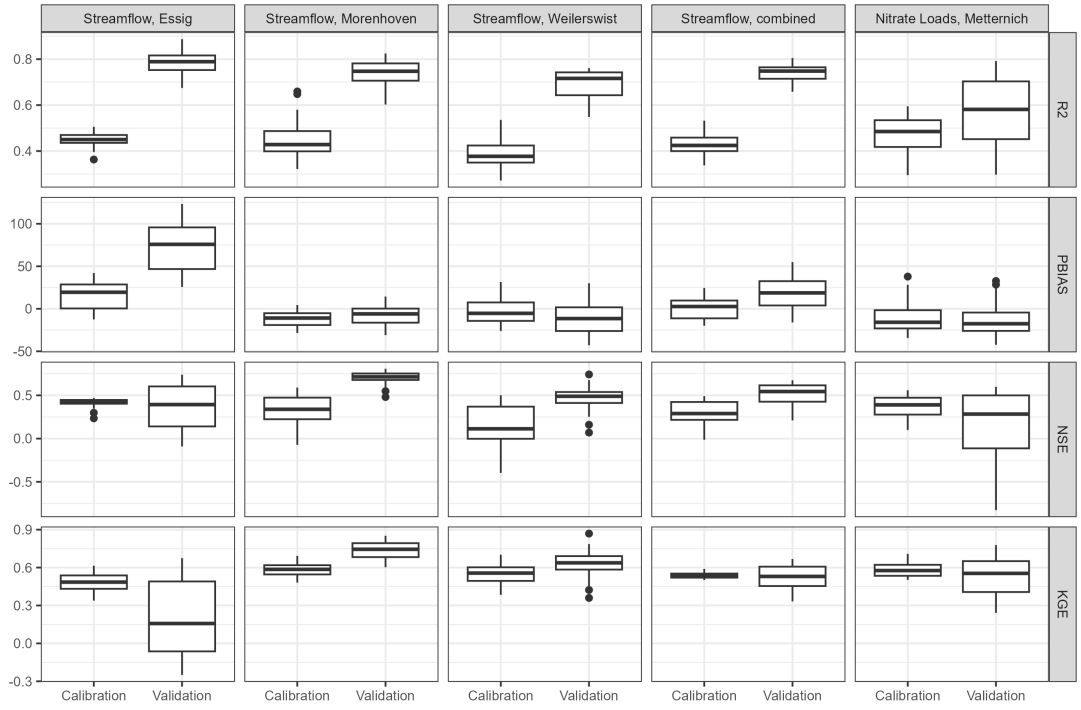


Figure 5.7.: Achieved performance measures for the $n = 32$ calibrated model runs in the calibration (2012 - 2016) and validation period (2017 - 2018). Each row shows the performance results with regard to a different metric (R^2, PBIAS, NSE or KGE). Each column shows the performance results with regard to a different model output ("Streamflow, All" denotes the mean value of the respective performance metric across the three gauging stations in Essig, Morenhoven and Weilerswist). The two rightmost panels in the bottom row represent the results for the objective functions applied for model calibration and validation. The boxplot whisker length is limited to one and half of the interquartile range.

erate upward bias in the calibration period, become excessive in the validation period, with a median value of approximately 75 % and a maximum that exceeds 100 %. This likely reflects extended stretches of low flow at the Essig gauging station in the exceptionally dry year of 2018, which the model fails to capture possibly due to trade-off effects with the Swist gauges in Morenhoven and Weilerswist already discussed at the end of Section 5.3.

Considering NO_3^--N loads in Metternich, the PBIAS metric indicates an equal amount of underestimation in the calibration and validation period, with median PBIAS scores of approximately −16 % and −18 %, respectively. The vast majority of the 32 behavioral model runs yield PBIAS scores that are negative. The systematic underestimation of NO_3^--N in-stream loads is a deficiency already seen in the results of the uncalibrated model examined in Chapter 4. Apparently, model calibration was unable to identify parameter sets that shift the nitrogen regime of the Swist catchment towards higher NO_3^--N exports at the mouth of the river. As

already discussed, the most likely cause for the low nitrate loads simulated by SWAT is an underestimation of applied N fertilizer amounts, although other explanations are also possible. The systematic error in simulated NO_3^--N loads is a crucial issue regarding the aims of this thesis and will therefore be reviewed in greater detail in the upcoming Chapter 6. Looking at the remaining performance metrics, there is an overall increase of R^2 scores for nitrate simulation in the validation period, hinting at better reproduction of the NO_3^--N in-stream dynamics in the drier simulation years. This increase in R^2 scores is however less pronounced than for streamflow. At the same time, NSE scores deteriorate for the validation periods, with minimum values falling well below the critical benchmark at 0, indicating a tendency towards impaired simulation of nitrate load peaks.

Table 5.2 lists the KGE-based objective function results in the calibration and validation period for all 32 behavioral model runs. Additionally, the table specifies the nitrate PBIAS scores for both periods to assess the simulated long-term nitrogen balance in the Swist catchment, which is particularly relevant for quantifying non-point source nitrogen pollution in the following Chapter 6. The 11 model runs that surpass the behavioral threshold for $KGE_{Streamflow}$ and $KGE_{Nitrate}$ in both simulation periods are listed separately at the top of Table 5.2. Most of the other model runs stay well below the behavioral thresholds in the validation period for at least one of the two objective functions. The sole exception is model run 5275, which fails only narrowly in achieving the required validation result for streamflow while surpassing the threshold for NO_3^--N loads and reaching excellent nitrate PBIAS scores both during calibration and validation. Figure 5.8 visualizes the performance results listed in Table 5.2 by showing the behavioral domain of the objective space in the calibration period (also see Figure 5.3) as well as validation period. In addition, the figure identifies the Pareto-optimal solutions in the sampled model runs for both time intervals. Model run 2868 is particularly interesting, being in the Pareto-optimal set during calibration as well as validation, with favorable results for $PBIAS_{Nitrate}$ in both time periods and by far the best rating for $KGE_{Nitrate}$ during validation. For calibration, model run 1434 gives the highest $KGE_{Nitrate}$ result and seems to offer an excellent Pareto-optimal solution with only marginal trade-off between streamflow and nitrate simulation performance. However, keep in mind that the in-sample solutions presented in Figure 5.8 are presumably too sparse to adequately approximate the population Pareto fronts in both simulation periods. For instance, there may be many unexplored parameter sets that are Pareto-optimal and outperform model run 1434 in simulating nitrate in the calibration period (compare with model run 2868 during validation). In other words, Pareto-optimal solutions in the sample are so sparse that they may only give an incomplete picture of the real Pareto front, and the apparently good trade-off properties of model run 1434 during calibration may therefore be misleading.

Table 5.2.: Calibrated model performances in the calibration period (2012 - 2016) and validation period (2017 - 2018). The list is restricted to the 32 model runs that surpass the behavioral threshold of 0.5 for both objective functions $KGE_{Streamflow}$ and $KGE_{Nitrate}$ during calibration. The 11 model runs that additionally surpass both behavioral thresholds in the validation period are listed separately at the top of the table. Both sub-lists are ordered by model run number. $PBIAS_{Nitrate}$ is included as an additional performance metric to assess simulation of the overall nitrogen balance in the Swist catchment in the calibration and validation period.

	Calibration			Validation		
Run	$KGE_{Streamflow}$	$KGE_{Nitrate}$	$PBIAS_{Nitrate}$	$KGE_{Streamflow}$	$KGE_{Nitrate}$	$PBIAS_{Nitrate}$
349	0.540	0.646	−20.6	0.573	0.645	−25.8
1054	0.535	0.571	−27.4	0.611	0.623	−34.4
1434	0.583	0.708	−15.8	0.606	0.574	−20.3
2868	0.583	0.663	−5.3	0.511	0.778	−6.6
4648	0.575	0.564	−29.9	0.533	0.556	−40.5
5221	0.519	0.656	11.3	0.602	0.712	5.8
5502	0.530	0.700	−3.2	0.570	0.552	−8.8
6388	0.554	0.503	−18.9	0.537	0.709	−19.2
7739	0.534	0.529	−34.4	0.567	0.512	−42.3
8638	0.552	0.647	−20.7	0.599	0.680	−27.7
8957	0.573	0.509	−29.3	0.526	0.630	−24.9
228	0.514	0.605	20.4	0.361	0.470	24.4
1227	0.506	0.558	−27.4	0.395	0.621	−34.7
1248	0.526	0.548	−24.5	0.417	0.710	−23.9
1696	0.523	0.550	−6.1	0.665	0.366	−4.4
1994	0.551	0.537	37.9	0.487	0.413	32.7
3878	0.502	0.610	25.4	0.333	0.587	21.1
3883	0.532	0.516	−15.9	0.461	0.517	−11.4
5212	0.557	0.607	−19.3	0.668	0.461	−23.3
5275	0.524	0.683	3.9	0.493	0.687	−4.4
5382	0.506	0.525	−32.2	0.623	0.366	−37.9
5626	0.547	0.507	−32.2	0.635	0.364	−40.4
6842	0.549	0.607	−22.6	0.477	0.668	−26.5
7038	0.548	0.517	−13.3	0.464	0.494	−15.4
7265	0.501	0.581	−14.4	0.370	0.242	−10.6
8006	0.509	0.525	28.2	0.618	0.347	28.4
8294	0.515	0.541	3.0	0.451	0.282	6.3
8503	0.521	0.614	−16.1	0.362	0.711	−17.8
8504	0.523	0.569	−20.9	0.455	0.456	−23.6
8550	0.535	0.670	−4.9	0.621	0.383	−17.4
8774	0.589	0.601	7.1	0.625	0.386	6.3
9485	0.569	0.610	−10.8	0.449	0.637	−15.7

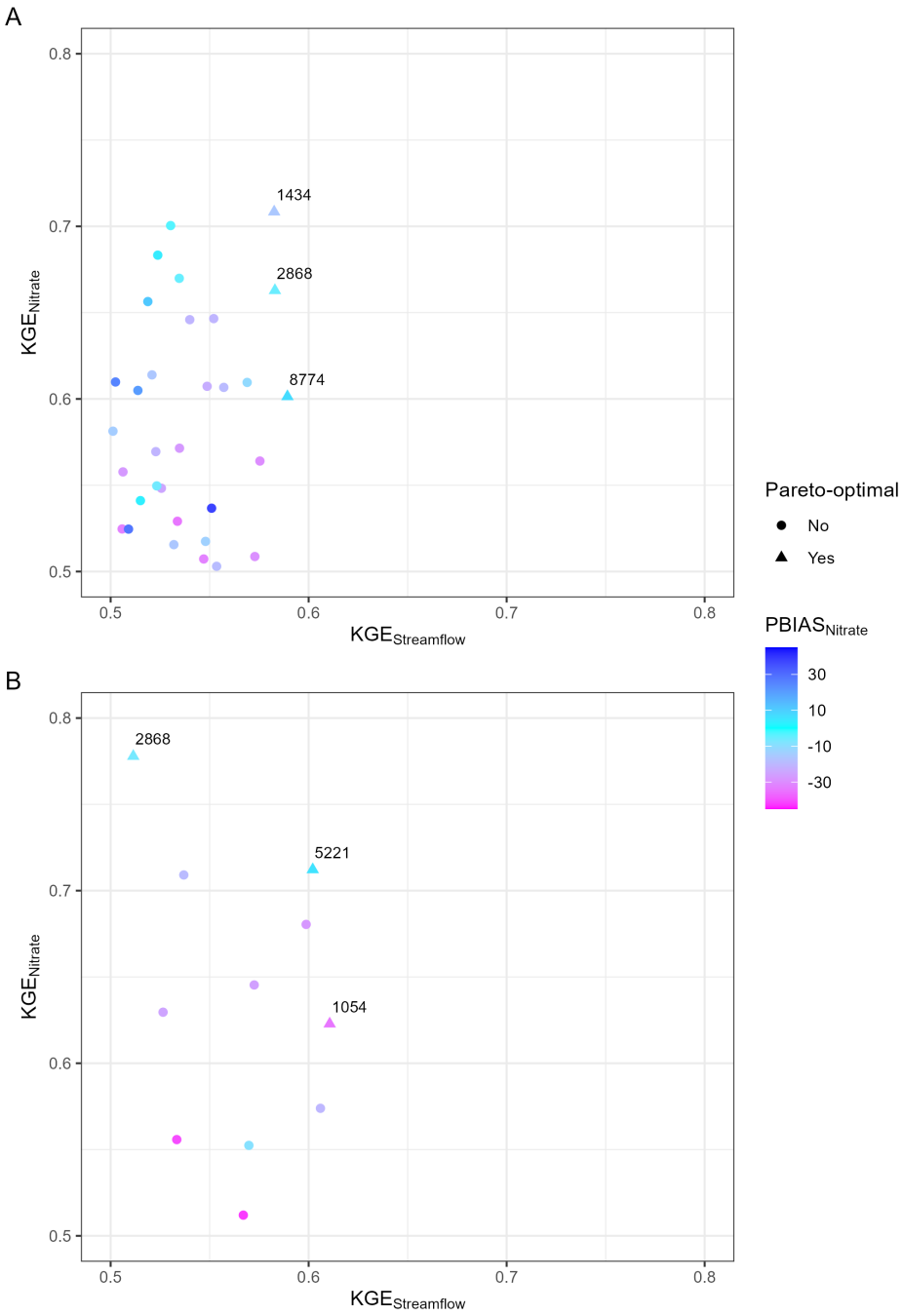


Figure 5.8.: Scatter plots of the calibrated model performances during the calibration (2012 - 2016) and validation period (2017 - 2018). The scatter plots are limited to the behavioral domain of the applied objective space for calibration (A) and validation (B). The color of the points indicates the PBIAS result for simulated NO_3^--N loads in Metternich. The shape of the points signifies Pareto-optimality among the sampled model runs. Notice that Panel A represents a detail of the objective space shown in Figure 5.3. As validation was restricted to the 32 calibrated model runs in Panel A, there may be further model runs among the original 10 000 that surpass both behavioral thresholds during validation but are not included in Panel B.

Figures 5.9 to 5.12 show the observed and simulated time series of streamflow at the Essig, Morenhoven and Weilerswist gauging stations, and of in-stream NO_3^--N loads at the Metternich water quality monitoring station in the calibration and validation period. The figures depict the model results of the 32 calibrated model runs as uncertainty bands from the minimum to the maximum value and from the 10th to the 90th percentile for each simulation day. The time series depictions confirm many of the conclusions already drawn from the calculated performance metrics. Considering the hydrographs at the three gauging stations, the simulated minimum and maximum values manage to bracket the observed data for most of the study period. Exceptions occur most frequently in Essig, where the simulated uncertainty band repeatedly is above the observed values, particularly during low flow periods. This general pattern of low flow *overprediction* in Essig is contrasted by a pattern of low flow *underprediction* at the gauging station in Weilerswist, where the simulated uncertainty band manages to bracket observed low flows during most of the study period, but fails to do so *symmetrically*. A sizable share of model runs even approach streamflow cessation during certain time intervals in Weilerswist, which is not supported by the gauge records. In Morenhoven, the simulation results show a tendency to underpredict low flows as well, but to a clearly lesser extent than in Weilerswist. In this respect, note that base flow in Morenhoven is supported by the continuous wastewater contribution from the upstream WWTP in Flerzheim. The basic patterns of low flow over- and underprediction at the individual gauges are reflected in the PBIAS scores depicted in Figure 5.7, demonstrating the impact of errors in low flow simulation on the general water balance. The divergent model behavior in simulating low flows at the Steinbach and Swist watercourses likely is a consequence of trade-offs in calibrating water transmission losses from the main channel (see Section 5.3).

With regard to streamflow peaks, the models overall perform well in reproducing their occurrence and timing. However, there are some peaks that are notably overpredicted by individual model runs, as indicated by upward spikes of the dotted line marking the maximum simulation results without a correspondingly high peak in the observed hydrograph. Overprediction of streamflow peaks primarily occurs during the calibration period in Morenhoven and Weilerswist. In the validation period, observed streamflow is less dynamic and streamflow peaks are smaller than in the calibration period, especially during the particularly dry year of 2018. Accordingly, overprediction of streamflow peaks is less pronounced in the validation period. In fact, the improvement of performance criteria during validation that emphasize correlation, timing and capturing of peak discharges (R^2, NSE) may be an effect of drought and reduced discharge dynamics in this period, effectively shrinking the room for error in capturing these dynamics during validation. Conversely, reduced observed discharges in the validation period may increase variability and error sensitivity in relative performance measures that assess the

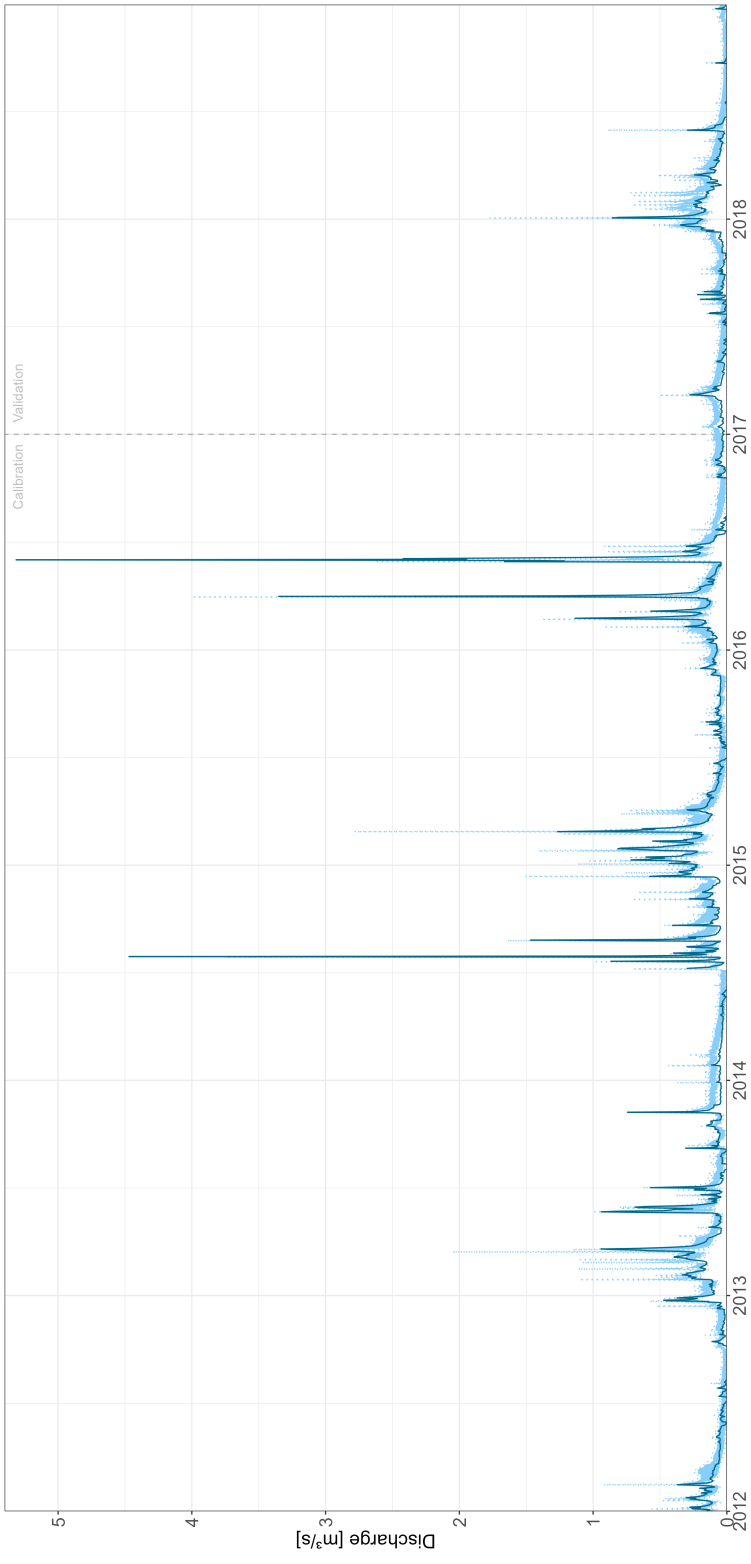


Figure 5.9.: Observed and simulated time series of daily streamflow in Essig. The solid blue line shows observed streamflow. Results of the 32 simulation runs selected during model calibration are depicted in aggregated form. The two dotted light blue lines indicate minimum and maximum simulated streamflow for each day, while the light blue ribbon frames simulated streamflow between the 10th and 90th percentile.

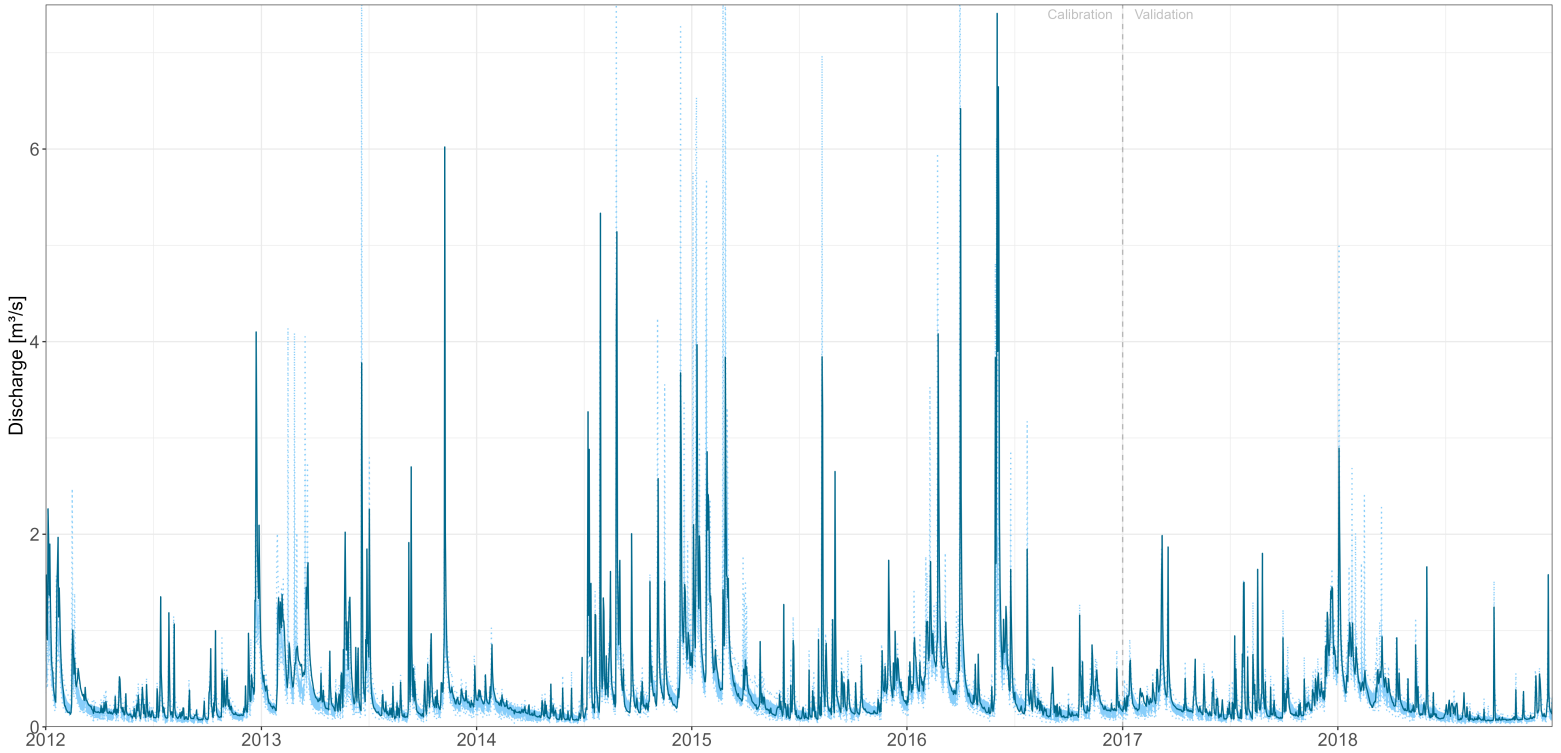


Figure 5.10.: Observed and simulated time series of daily streamflow in Morenhoven. The solid blue line shows observed streamflow. Results of the 32 simulation runs selected during model calibration are depicted in aggregated form. The two dotted light blue lines indicate minimum and maximum simulated streamflow for each day, while the light blue ribbon frames simulated streamflow between the 10th and 90th percentile. For reasons of scale, some simulated peaks are cut off.

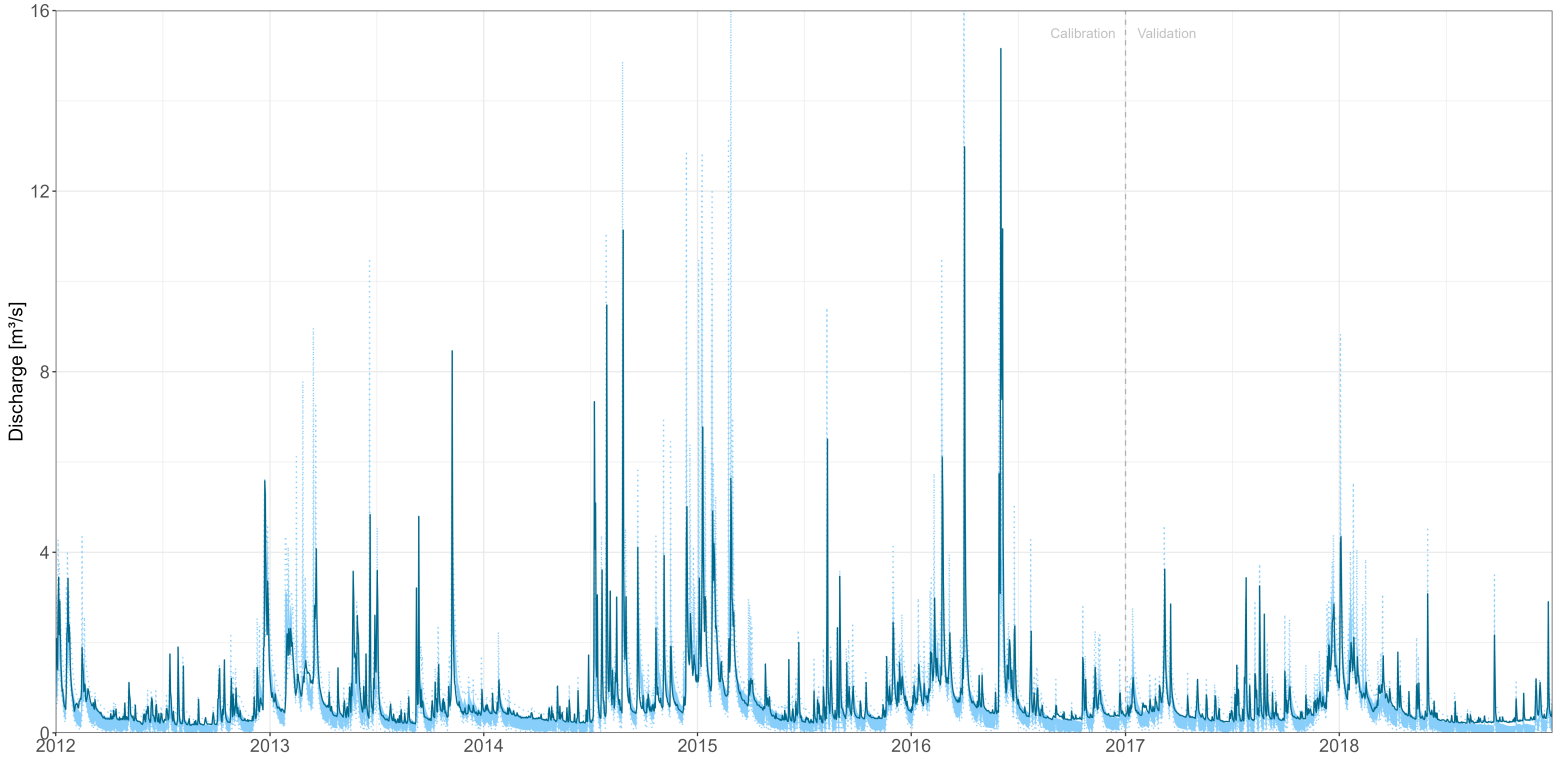


Figure 5.11.: Observed and simulated time series of daily streamflow in Weilerswist. The solid blue line shows observed streamflow. Results of the 32 simulation runs selected during model calibration are depicted in aggregated form. The two dotted light blue lines indicate minimum and maximum simulated streamflow for each day, while the light blue ribbon frames simulated streamflow between the 10th and 90th percentile. For reasons of scale, two simulated peaks are cut off.

overall water balance, like PBIAS (Gupta et al. 1999, Moriasi et al. 2007) and the bias component of KGE (Santos et al. 2018, Clark et al. 2021). This effect may play a role in the extreme PBIAS deterioration for streamflow in the validation period at the Essig gauge, where the general water balance is more dominated by base flow than in Morenhoven and Weilerswist. Comparison of the Essig hydrographs between calibration and validation in Figure 5.9 does not suggest as large an increase in absolute streamflow overprediction as implied by the relative performance measure PBIAS in Figure 5.7.

Looking at the NO_3^--N load data at the Metternich water quality monitoring station in Figure 5.12, the models largely succeed in capturing the overall dynamics in the observed time series both in the calibration and validation period. This may be seen as an expected outcome given the good representation of streamflow dynamics in the simulation results (go back to Figures 5.9 to 5.11 for confirmation) and the typically high correlation between discharge and transported NO_3^--N loads (Oeurng et al. 2010, Pellerin et al. 2014, Bauwe et al. 2020). At the same time, individual model runs introduce additional peaks to the simulated time series not present in the same magnitude in the observed data. While this kind of error is also present in simulated streamflow, it is more pronounced and more frequent in the NO_3^--N load data, both during calibration and validation. Furthermore, the nitrate time series shows that the overall load deficit in the simulations as seen in the mostly negative PBIAS scores can largely be attributed to underprediction of NO_3^--N base loads, with a large portion of the simulated uncertainty band undercutting the observed values in-between the high emission events. This stands in contrast to underprediction of NO_3^--N loads in the uncalibrated model (see Chapter 4), which originates mostly from a failure in reproducing the peaks in the observed load time series while the base loads during low flow are captured better (data not shown).

Generally, the time series plots confirm that there is no systematic decline in model performance between calibration and validation, both with respect to streamflow and NO_3^--N loads. Although not perfect, the model fits to the observed data are not fundamentally different in the calibration and validation period, neither in terms of their strengths nor regarding their weaknesses. Against the background of the sustained drought in the last simulation year, it can thus be stated that the applied SWAT model has been successfully validated in a differential split-sample test regarding climatic conditions. Although no definite prove on its own, this outcome provides confidence that the calibrated parameters represent the Swist catchment characteristics as realistically as possible given the applied model structure and input data. Errors in the model structure and input data are the most likely cause for the remaining deficiencies in the simulation results, effectively posing a ceiling for the calibrated model performance. An example of such an error is the lack of spatial differentiation in the parameter governing

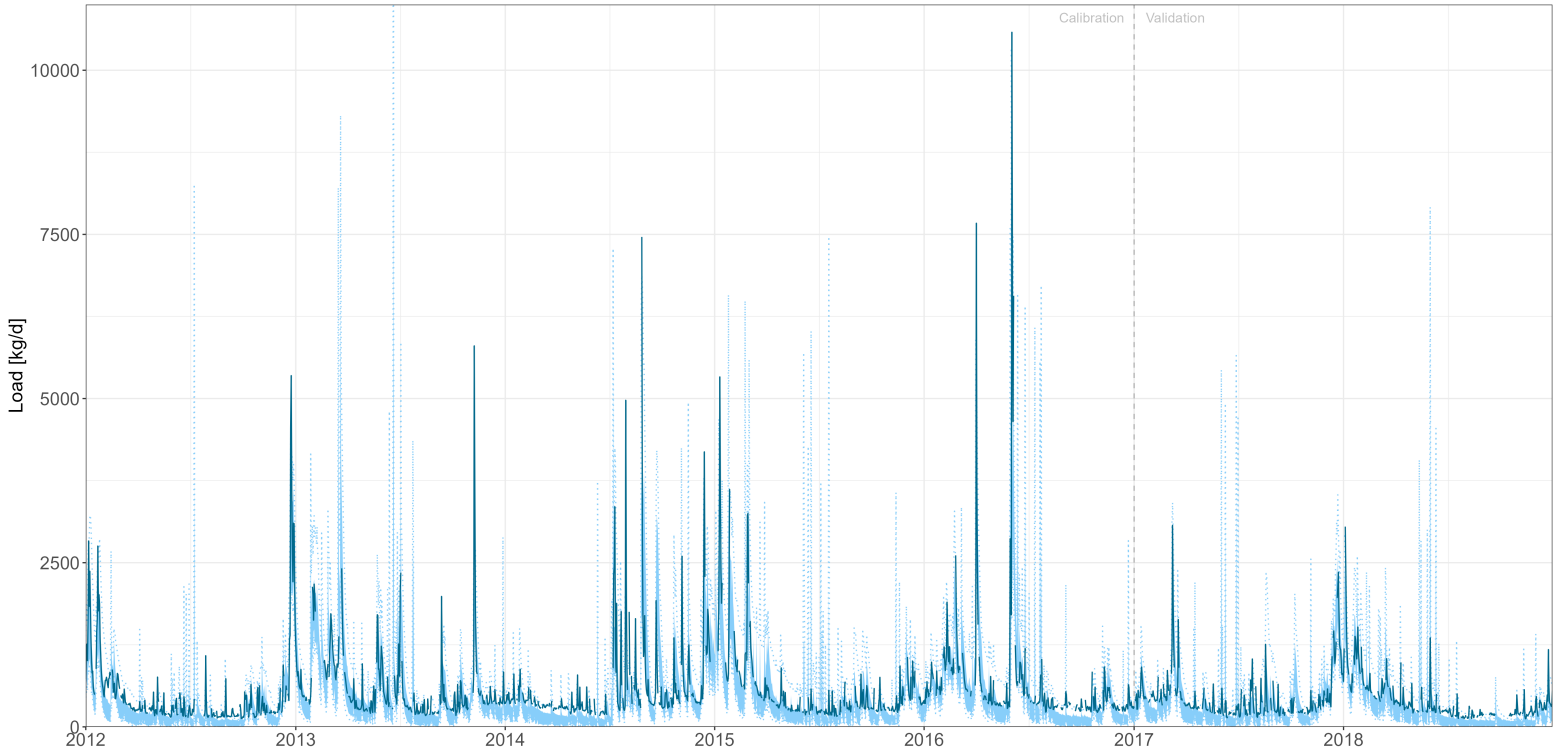


Figure 5.12.: Observed and simulated time series of daily NO_3^--N loads in Metternich. The solid blue line shows observed loads. Results of the 32 simulation runs selected during model calibration are depicted in aggregated form. The two dotted light blue lines indicate minimum and maximum simulated loads for each day, while the light blue ribbon frames all simulated loads between the 10th and 90th percentile. For reasons of scale, one simulated peak is cut off.

transmission losses from the SWAT main channel (CH_K2.rte), which makes it difficult to adequately represent low flow periods at all considered stream gauges simultaneously. In order to gain further understanding of the calibrated simulation results, Subsection 5.4.2 will compare the simulation results to additional monitoring data on in-stream N concentrations.

5.4.2. Comparison with In-Stream N Concentrations

The comparison of the simulation results with data on in-stream N concentrations serves two main purposes: First, it aims to complement split-sample validation of the SWAT models with additional data not previously used in either calibration or validation. The N concentration data originates from the Erftverband routine water quality monitoring network in the Swist catchment (see Figure 2.2 in Chapter 2). Hence, it can be used to examine the simulated water quality along the entire Swist watercourse (and some of its tributaries), rather than exlusively at the Metternich monitoring station near the catchment outlet, effectively testing the model in space instead of only time. Second, the comparison makes it possible to extend model validation towards forms of reactive nitrogen other than NO_3^--N . The Erftverband routine monitoring data encompasses several different fractions of reactive N. The most important of these additional nitrogen fractions is organic N, which most likely constitutes a meaningful part of the nitrogen balance in the catchment, but so far has not been considered in the calibration and validation process. However, note that the Erftverband monitoring data is mostly based on grab samples on a few days per year, with additional qualified grab samples[7] and 24h composite samples taken for research projects. In contrast, the simulated values represent average concentrations for each simulation day. This means that the measured and simulated concentrations are *incommensurate* (Beven 2012, p. 18, also see Subsection 1.2.2) and do not allow a meaningful one-to-one comparison or calculation of standard performance measures. Instead, measured and simulated values are compared in terms of their medians and maxima (i.e. their distributional characteristics).

Table 5.3 lists the median and maximum values of the measured and simulated N concentrations at all considered monitoring points. In total, the monitoring data comprises approximately 300 water samples taken in the entire Swist catchment during the study period (2012 to 2018). The measurements show by far the lowest burden of reactive N in the upstream region of Ersdorfer Bach, a tributary that enters the Swist in the town of Meckenheim. The monitoring point at Ersdorfer Bach represents "pristine" conditions, being situated at the outlet of a small headwater catchment that is dominated by forest with no urban or agricultural

[7] A qualified grab sample is a composite sample consisting of at least 5 grab samples taken with a maximum time gap of 2 min inside a time frame of not more than 2 h.

influences. For the Swist main watercourse, measured N concentrations are lowest in Meckenheim and Flerzheim, although already one order of magnitude higher than at Ersdorfer Bach. Both monitoring points are influenced by intensive agriculture (particularly orchards) in the southeast of the catchment, while still located upstream of the first (and largest) continuous wastewater discharger along the Swist watercourse: WWTP Flerzheim. From there, measured N concentrations rapidly increase until the monitoring point in Miel, which is additionally influenced by wastewater from WWTPs Rheinbach and Miel. Between Miel and Weilerswist, measured N concentrations gradually decrease again, despite further wastewater inflow from WWTP Heimerzheim. The highest measured N burden of all monitoring points is present in the Schiessbach, a tributary that is influenced by intensive cropland farming and enters the Swist between Miel and Heimerzheim. According to WFD monitoring results for the Swist river basin district, the Schiessbach is one of the water bodies that failed to meet the quality requirements for NO_3^- from 2009 to 2011 (State Ministry for Climate Protection, Environment, Agriculture, Nature Conservation and Consumer Protection NRW 2015). However, it met the requirements in the time span from 2015 to 2018 (State Ministry for Environment, Agriculture, Nature Conservation and Consumer Protection NRW 2021).

As expected, the measured data shows NO_3^--N to be the most abundant form of reactive nitrogen in the surface waters at all monitoring points, followed by organic N. In contrast, measured concentrations of NH_4^+-N and NO_2^--N are negligible in comparison: In most cases, the medians and even maxima for both compounds cannot be stated in Table 5.3, as too many concentrations are below the analytical limit of quantification (LOQ). From a purely qualitative point of view, the simulation results replicate this basic pattern, with NO_3^--N exhibiting by far the highest median and maximum concentrations, organic N taking second place, whereas concentrations of NH_4^+-N and NO_2^--N are much lower, not exceeding 1 mg/l at all reference points. In quantitative terms, however, there are some meaningful differences between the measured and simulated data: While median NO_3^--N concentrations by and large agree well between both data sets (especially at the middle and lower reaches of the Swist main watercourse as well as Schiessbach), the models generally overpredict the maximum concentrations of NO_3^--N by several orders of magnitude. Overestimation of the NO_3^--N median concentration in the Ersdorfer Bach by approximately one order of magnitude can be attributed to a spatial mismatch between the monitoring point and the respective subbasin outlet, with the latter being situated further downstream and receiving additional nutrient inputs from agricultural and residential areas. Looking at organic N, the models tend to undercut the medians and exceed the maxima of the measured concentrations, but manage to stay in the same order of magnitude, which overall is a reassuring outcome given that organic N remained unconsidered during model calibration. Here, the monitoring data may actually underestimate peak

Table 5.3.: Measured and simulated concentrations of reactive nitrogen compounds at the Erftverband routine water quality monitoring points in the Swist river network (medians and maxima). The monitoring points are roughly ordered from headwaters to Swist mouth. Measured concentrations mostly stem from grab samples - in some cases qualified grab samples or 24h composite samples - taken during the simulation period (2012-2018). Medians of the measured concentrations are omitted if more than 50 % of the *n* data points are below the limit of quantification (LOQ). Correspondingly, the maxima of the measured concentrations are omitted if 100 % of the *n* data points are below the LOQ. The declared LOQs for NH_4^+-N , NO_2^--N, NO_3^--N and organic N are 0.10 mg/l, 0.10 mg/l, 0.20 mg/l and 1.00 mg/l, respectively. The simulated values are average concentrations for each simulation day, obtained by dividing simulated daily loads by simulated daily streamflow. Medians and maxima of the simulated data represent the simulation results of all 32 behavioral model runs.

		Measured (2012-2018)			Simulated (2012-2018)	
Monitoring Point	Parameter	n	Median [mg/l]	Maximum [mg/l]	Median [mg/l]	Maximum [mg/l]
Ersdorfer Bach	NH_4^+-N	13	-	-	0.00	0.15
Ersdorfer Bach	NO_2^--N	13	-	-	0.00	0.00
Ersdorfer Bach	NO_3^--N	13	0.60	1.3	4.00	105.98
Ersdorfer Bach	Organic N	13	-	2.8	0.00	13.27
Swist Meckenheim	NH_4^+-N	40	-	1.0	0.01	0.35
Swist Meckenheim	NO_2^--N	40	-	-	0.00	0.04
Swist Meckenheim	NO_3^--N	40	5.28	9.3	6.76	3146.45
Swist Meckenheim	Organic N	40	-	2.7	0.00	12.86
Swist Flerzheim	NH_4^+-N	49	-	1.0	0.01	1.47
Swist Flerzheim	NO_2^--N	49	-	-	0.00	0.07
Swist Flerzheim	NO_3^--N	49	4.69	8.0	8.23	7571.88
Swist Flerzheim	Organic N	40	-	2.9	0.00	34.29
Swist Morenhoven	NH_4^+-N	31	-	1.0	0.04	0.59
Swist Morenhoven	NO_2^--N	31	-	-	0.01	0.10
Swist Morenhoven	NO_3^--N	31	5.90	10.0	8.06	272.76
Swist Morenhoven	Organic N	22	1.20	2.4	0.69	8.59
Swist Miel	NH_4^+-N	40	-	1.0	0.08	0.48
Swist Miel	NO_2^--N	40	-	-	0.02	0.16
Swist Miel	NO_3^--N	40	7.16	10.2	7.86	193.65
Swist Miel	Organic N	40	1.10	2.7	0.90	8.29
Schiessbach	NH_4^+-N	9	-	-	0.01	0.52
Schiessbach	NO_2^--N	9	-	-	0.00	0.05
Schiessbach	NO_3^--N	9	8.00	27.4	7.44	1114.06
Schiessbach	Organic N	8	1.40	1.8	0.00	10.20
Swist Heimerzheim	NH_4^+-N	40	-	1.0	0.08	0.63
Swist Heimerzheim	NO_2^--N	40	-	-	0.01	0.21
Swist Heimerzheim	NO_3^--N	40	6.80	10.0	7.21	242.32
Swist Heimerzheim	Organic N	40	1.45	6.4	0.59	8.35
Swist Metternich	NH_4^+-N	16	-	0.4	0.09	0.79
Swist Metternich	NO_2^--N	16	-	-	0.02	0.29
Swist Metternich	NO_3^--N	16	6.50	9.4	7.24	252.27
Swist Metternich	Organic N	15	1.20	2.1	0.54	8.12
Swist Weilerswist	NH_4^+-N	59	-	1.2	0.09	0.80
Swist Weilerswist	NO_2^--N	59	-	-	0.02	0.34
Swist Weilerswist	NO_3^--N	59	6.40	10.0	7.17	254.46
Swist Weilerswist	Organic N	59	1.20	3.0	0.48	7.40

concentrations of organic N, as these typically occur during high flows and are thus most likely missed by grab sampling on only a few days per year.

Note that the simulated median and maximum concentrations in Table 5.3 represent all 32 calibrated model runs at once. In order to compare the simulated N concentrations across the different models, Figure 5.13 shows the concentration distributions at the monitoring point in Weilerswist during the study period for each individual model run as boxplots. Weilerswist was chosen because it is the monitoring point closest to the Swist mouth, integrating discharge and nitrogen emissions from almost the entire Swist catchment area. In addition, it is the monitoring point with the highest number of examined water samples in the study period, with median and maximum values of the measured concentrations indicated in Figure 5.13 as red and blue lines, respectively. Considering NO_3^--N , it is obvious that there is considerable variability in the simulated maximum concentrations between the various model runs.[8] While all model runs overpredict the maximum values to a certain extent, only a minority reach NO_3^--N concentrations of 100 mg/l and more. Other model runs manage to better control the maximum values and produce upward tails in the simulated distributions that are relatively short with no or only a few moderate outliers. An example is model run 2868, which yields favorable results in many other aspects as well, as already discussed in Subsection 5.4.1. The patterns of higher and lower simulated maxima between the different model runs in Weilerswist generally carry over to most of the other monitoring points as well (data not shown).

When interpreting the simulated concentrations, keep in mind that N concentrations are not a SWAT output variable and remained unconsidered in the model calibration process. Instead, the model was calibrated on streamflow and in-stream NO_3^--N loads, and the simulated concentration values are derived from the streamflow and N load results. Consequently, errors in these two model outputs may interact and introduce even larger errors in the simulated N concentration values. In this respect, remember that common errors in the simulations include a tendency to underestimate discharge during low flow periods along the Swist main watercourse as well as massive overestimation of some NO_3^--N load peaks near the catchment outlet in Metternich (see Subsection 5.4.1). Extreme overprediction in the simulated NO_3^--N concentrations may thus result from co-occurrence of these errors on the same simulation day. In order to verify this, Figure 5.14 shows the simulated time series of streamflow at the Weilerswist gauge and NO_3^--N in-stream loads at the Metternich water quality station for two example model runs against the corresponding observed time series in 2016 and 2017 (i.e.

[8] NO_3^--N is chosen here for detailed discussion as it is the most important N compound and clearly dominates the overall N burden in the Swist river network. However, differences in the maximum concentrations between the model runs are also roughly mirrored in the distributions for the other N fractions, as seen in Figure 5.13.

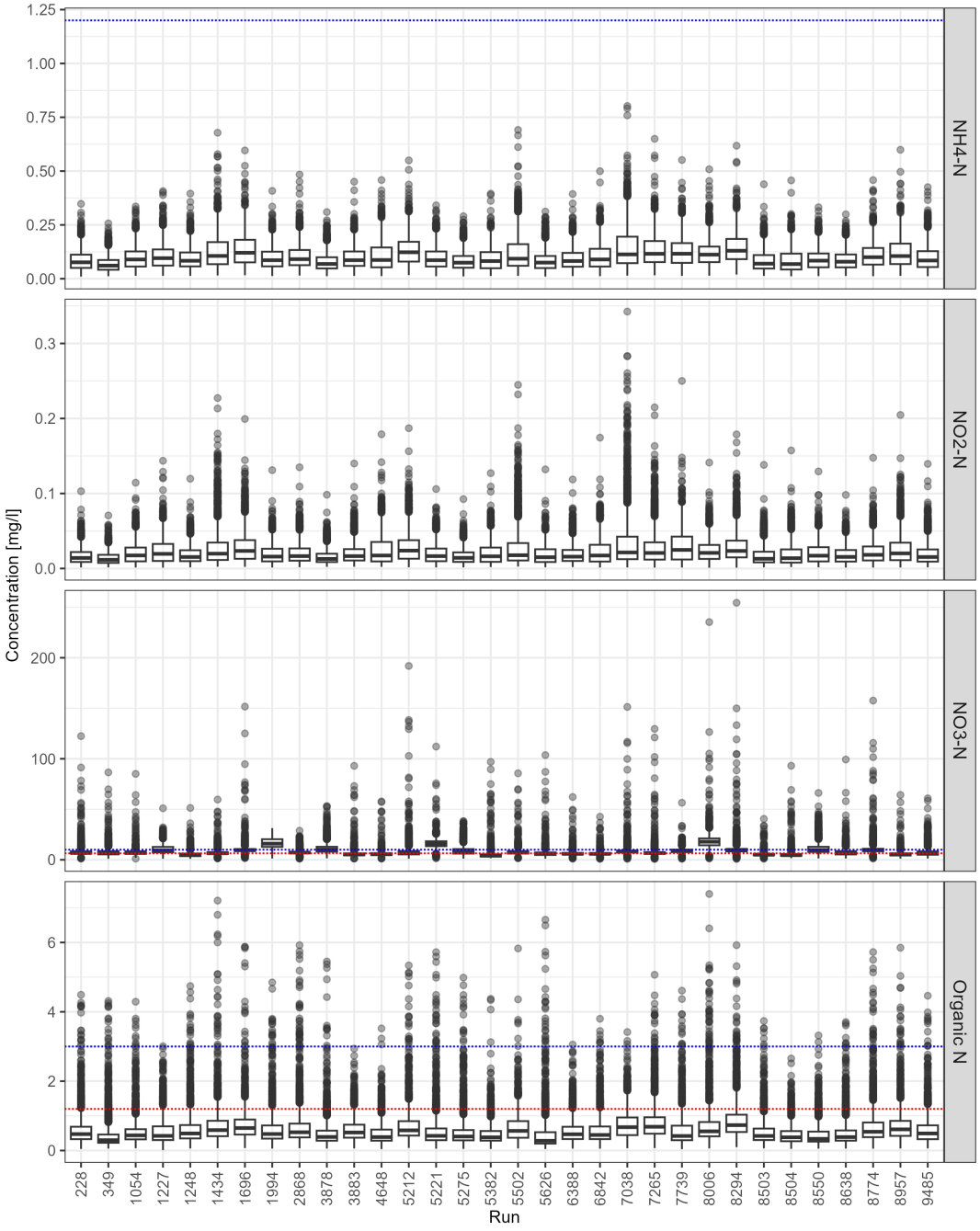


Figure 5.13.: Simulated N concentrations in Weilerswist near the Swist mouth. The boxplots show the distribution of daily simulated N concentrations for each of the 32 calibrated model runs during the study period (2012-2018). Hence, the number of data points behind each boxplot is the number of days in the simulation period (2557). The concentrations are given for each of the four reactive N fractions in surface waters. For quick reference, medians and maxima of the measured concentrations at the corresponding Erftverband monitoring point in Weilerswist are indicated as red and blue dashed lines, respectively (if available for each fraction of reactive N, see Table 5.3).

the last year of the calibration period and the first year of the validation period). The two example model runs (2868 and 5502) were selected as they both surpass the behavioral thresholds for the KGE-based objective functions during calibration and validation as well as yield favorable nitrate PBIAS scores over the entire study period (Table 5.2). However, they give completely opposite results regarding peak concentrations of NO_3^--N , with the predicted maximum concentration being approximately three times higher for model run 5502 than for model run 2868 in Weilerswist (see Figure 5.13).

Looking at Figure 5.14 A, it becomes clear that model run 5502 is prone to low flow underprediction at the Weilerswist gauging station, whereas model run 2868 captures low flows excellently (albeit with a greater tendency to overpredict the hydrograph peaks). At the same time, Figure 5.14 B reveals that model run 5502 immensely overpredicts several NO_3^--N load peaks that model run 2868 mostly captures quite well. For model run 5502, the maximum NO_3^--N concentration in Weilerswist is about 86 mg/l and it occurs end of June 2017, when a longer period of consistently underestimated streamflow coincides with an enormously overestimated load peak. This outcome basically confirms the above-stated notion that concentration outliers are mostly a consequence of simultaneous errors in streamflow and NO_3^--N load simulation. As a final remark on Figure 5.14, notice that model run 2868 not only avoids the errors leading to overpredicted peak concentrations, but also captures NO_3^--N base loads generally well. In contrast, model run 5502 undercuts the observed loads during dry periods throughout 2016 and 2017, providing a good example for the marked underprediction of NO_3^--N base loads that is present in many of the calibrated models examined in this chapter.

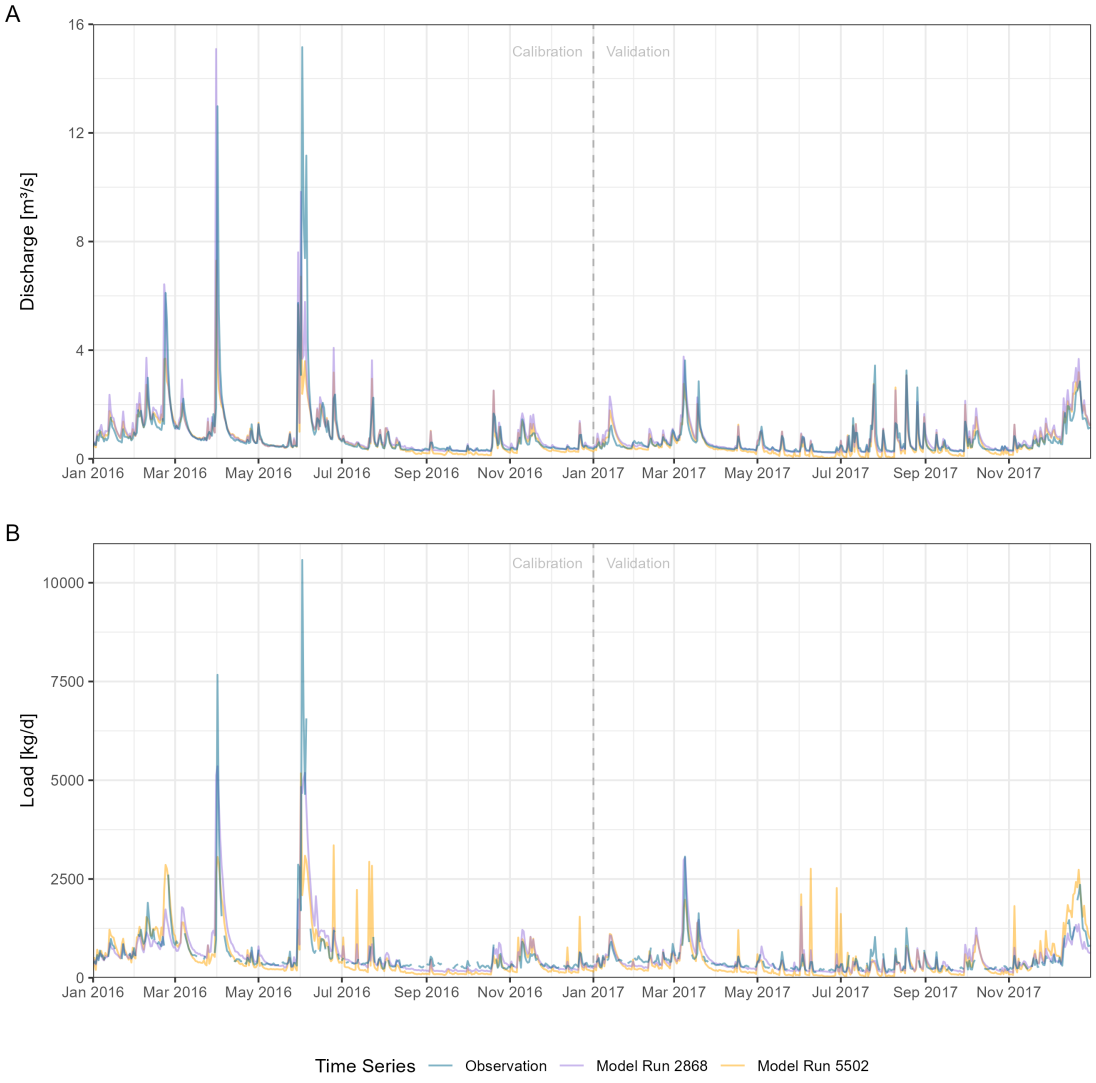


Figure 5.14.: Details from selected simulated time series of daily streamflow in Weilerswist (A) and daily NO_3^--N loads in Metternich (B). The selected simulated time series originate from the behavioral model runs 2868 (violet) and 5502 (yellow). The corresponding observed time series are depicted in blue. The excerpt from the overall time series encompasses the last year of the calibration period (2016) and the first year of the validation period (2017).

5.5. Concluding Discussion

The results of MC-based SWAT model calibration and validation for the Swist catchment presented in this chapter clearly demonstrate the implications of equifinality in the modeling process: On the one hand, several behavioral parameter sets manage to approximate the observed calibration data within reasonably narrow uncertainty bands. From 10 000 model evaluations, a total of 11 parameter sets were identified as behavioral in both the calibration and validation period (2012-2018). On the other hand, although multi-objective calibration helped to better constrain the model parameters compared to a single-objective setting (see Figure 5.5), most calibrated parameters still vary over relatively wide ranges. In a nutshell, different model parameter sets produce similar model outputs. The diversity in the 11 behavioral parameter sets entails a range of N pollution scenarios for the Swist catchment that are consistent with the applied model structure as well as input and calibration data. The different scenarios may be seen as a representation of uncertainty in simulated N emissions and will be examined at length in the forthcoming Chapter 6. Of course, the selected parameter sets represent only a random fraction of an infinite number of behavioral solutions in the parameter space. In addition, they are dependent on many subjective decisions even within an automated MC-based calibration framework as applied in this thesis. These decisions include (among others) the choice of parameter ranges, calibration data sets, objective functions and behavioral thresolds. They come on top of the many subjective decisions already taken during model setup (see Chapter 4).

The selected parameter sets have been classified as behavioral solely in terms of the KGE-based objective funtions assessing streamflow and NO_3^--N in-stream loads. Despite similarly good KGE scores, they differ in a range of aspects, first and foremost their success in capturing the long-term N balance of the Swist catchment. This is quantified by the PBIAS metric for NO_3^--N in-stream loads in Metternich: While a small number of behavioral model runs produce PBIAS scores that are reasonably close to the ideal value of 0, the majority of them are distinctly negative, clearly showing a bias towards underestimating N exports from the Swist catchment. Thinking further, this also suggests underestimation of N emissions to surface waters (although other explanations are plausible as well, like excessive simulated in-stream retention). Given the large number of behavioral model runs that exhibit negative bias in nitrate simulation, calibration as conducted in this study seems mostly unable to identify parameter values that eliminates this bias, suggesting structural model errors or data errors as its most probable cause (e.g. representation of fertilizer inputs, see Chapter 4). The small number of parameter sets that yield good PBIAS results for NO_3^--N loads may then be right for the wrong reasons on the Swist N balance. Again, Chapter 6 will expand on this and give a more detailed analysis of the simulated N emission scenarios simulated by the 11 behavioral model runs.

Apart from the general N balance in the catchment, the behavioral model runs furthermore differ in their ability to reproduce the *concentrations* (as opposed to loads) of NO_3^--N and other forms of reactive nitrogen at various points along the Swist stream network. As worked out in this chapter, this is due to simultaneous errors in streamflow and nitrogen simulation exacerbating each other when looking at the resulting N concentrations. Differences in capturing measured concentrations are thus a consequence in different performances regarding simulation of streamflow and transported N loads, and may furthermore be read as varying model capacity to adequately correlate the dynamics of both quantities (in terms of N mobilization and dilution in response to the catchment hydrology). There are many more potential criteria that can be used to distinguish the 11 selected parameter sets from each other, for example their capability to balance out the trade-offs between the two applied objective functions (i.e. Pareto-optimality between streamflow and NO_3^--N load simulation, see Figure 5.8). In spite of their potential shortcomings and the subjectivity associated with their selection, the 11 selected model runs are the "*best guess*" in this thesis for representing the N emissions at the Swist in the study period as well as estimating the associated (parameter-based) model uncertainty. Consequently, they will serve as the basis for examining non-point source N pollution in the upcoming thesis chapter.

6. Model-Based Nitrogen Source Apportionment

6.1. Introductory Remarks

This chapter examines the nitrogen emission scenarios that correspond to the simulation results of the 11 SWAT model runs validated in Chapter 5. As argued there, these model runs (along with their associated parameter sets) represent the best approximation of actual N pollution in the Swist catchment during the study period (2012-2018), given the applied model setup and calibration strategy. Contrary to model calibration and validation, the analysis in this chapter is no longer based on in-stream simulation results, but on the HRU and subbasin level results for the SWAT land phase prior to channel routing. Hence, the focus of the present chapter are the N loads *entering* the Swist river network. The ultimate goal of the analysis is to quantify overall emission loads for reactive nitrogen in the Swist catchment and to assess the contribution of different emission sources, including a breakdown of point and non-point emission loads and an identification of potential critical source areas (CSAs). In principle, this corresponds to what is commonly referred to as *source apportionment* (Bøgestrand et al. 2005). The resulting emission load estimates are compared to independent data to check for their plausibility. With regard to load contributions from non-point emission sources, there are two reference data sets available for N emissions specifically in the Swist river basin:

1. The first reference data set is based on *edge-of-field* (EOF) monitoring data collected during several Erftverband research projects in the Swist catchment from 2005 to 2012 (Kistemann et al. 2007, Mertens et al. 2012). The monitoring data consists of N concentrations measured in surface runoff, interflow and tile flow from individual fields and hillslopes, covering all major natural land uses in the catchment (cropland, forests, grassland and orchards). The sampling devices applied for collecting surface runoff and interflow are described in Christoffels et al. (2016). The map in Figure 2.2 (see Chapter 2) gives an overview of the various EOF monitoring locations along the Swist and its tributaries during past Erftverband research activities (note that the reference data is based

on water samples from only part of these locations). Using additional data and estimates on point source emissions, Mertens et al. (2012) upscaled the monitoring data to establish a broad surface water emission inventory for total nitrogen in the Swist catchment from 2005 to 2009. For this purpose, catchment-wide emission loads for total nitrogen via surface runoff and interflow were estimated from the medians of the measured N concentrations at all EOF monitoring locations, a general water balance for the Swist river basin as well as the area fractions of different land uses within the basin area. Nitrogen emission loads from groundwater are quantified assuming equivalence between baseflow and groundwater recharge, estimating the average groundwater recharge rate across the entire Swist catchment to be roughly 90 l/s (Kistemann et al. 2007). As catchment-wide tile flow estimates were unavailable, note that the emission inventory does not include N loads from tile drains, offering no reference value for this particular pollution pathway.

2. The second reference data set is based on modeled raster data of diffuse total N emissions, covering the entire German federal state of North Rhine-Westphalia (NRW) at a resolution of 1 ha and representing the years from 2014 to 2016. The raster data has been generated in the GROWA+NRW 2021 project using the various sub-models of the static AGRUM model suite briefly described in Subsection 1.2.1. For each raster cell, this entailed establishing agricultural N budgets with RAUMIS (Gömann et al. 2021), assessing the long-term water balance with mGROWA (Herrmann & Wendland 2021) and estimating the fate of the mobile reactive N fraction in the soil and groundwater compartment with DENUZ and WEKU, respectively (Wendland et al. 2021a). These steps are informed by an identification of areas with agricultural tile drains (Tetzlaff 2021), an evaluation of denitrification potential in soils and aquifers (Kunkel et al. 2021) as well as an estimation of N residence times in different environmental subsurface compartments (Wendland et al. 2021b). The corresponding model output comprises - among other data - distributed average N emissions via surface runoff, erosion, interflow, tile flow and groundwater flow that could be aggregated on the level of SWAT subbasins in the Swist catchment. The result is a georeferenced data set that makes it possible to not only compare the catchment-wide emission loads to surface waters between both modeling approaches (SWAT and the AGRUM models), but also the spatial distribution of emission loads within the catchment boundaries.

The comparison of the SWAT ensemble results with the two reference data sets (the monitoring-based Erftverband emission inventory as well as the georeferenced GROWA+NRW 2021 model data) is used to evaluate the relative strengths and weaknesses of the different approaches in examining N pollution in the Swist case study. In this process, parameter-based uncertainty in

the SWAT emission estimates is assessed by intercomparison of the 11 validated model runs, whereas structural uncertainty is considered by comparing the overall SWAT results with the alternative model results from the GROWA+NRW 2021 project. As hydrology plays a key role for surface water pollution, Section 6.2 first assesses the overall water balance of the Swist catchment simulated by the 11 validated SWAT models, discussing the major implications for the interpretation of the simulated N emission loads. Thereafter, Section 6.3 presents the corresponding nitrogen regimes of the Swist catchment, focusing first on the simulated soil N balances (Subsection 6.3.1) before subsequently examining non-point source emission loads (Subsection 6.3.2) along with their seasonal variability (Subsection 6.3.3), point source emission loads (Subsection 6.3.4) and critical source areas in the catchment (Subsection 6.3.5). Lastly, Section 6.4 gives a joint discussion of the results presented in this chapter.

6.2. Water Balance

Table 6.1 gives an overview of the different water balance components simulated across the 11 selected model runs, representing land phase hydrology in the Swist catchment from 2012 to 2018. Allocating the point-scale precipitation records to individual subbasins by distance, SWAT estimates average areal precipitation for the Swist catchment to be 631.5 mm/a in the study period. In the same time span, the model runs predict the average water yield from the HRUs (i.e. "effective" precipitation) to be 99.7 mm/a, which roughly amounts to 16 % of overall areal precipitation. Across all model runs, simulated water yield fractions roughly vary between 13 % and 22 % of precipitation. This range of values is systematically higher than the approximately 13 % reaching the catchment outlet according to the long-term Weilerswist gauge records (see Chapter 2), but seems plausible given that the water balance components in Table 6.1 refer to the land phase only and do not yet account for transmission losses during channel routing for example along the Steinbach (see Chapter 2). The models estimate actual evapotranspiration in the Swist catchment to be 503.4 mm/a on average, amounting to approximately 80 % of areal precipitation. The remaining difference of 28.4 mm/a - more than 4 % of the catchment precipitation - most likely constitutes water loss in the water balance of the Swist catchment (hydrological conditions in the study period - see Section 2.4 - do not imply significant change in shorter-term water stores like soils). In SWAT, this loss term corresponds either to the fraction of groundwater recharge that is directed to the deep aquifer and is thus permanently removed from the system, or to water that is removed as "consumptive water use" (see the corresponding model parameters in Table 4.2).

Table 6.1.: Simulated water balance in the study period (2012-2018) for the behavioral SWAT model runs (land phase only). The listed values are averages for all simulation years and the entire Swist catchment area. The given percentages denote fractions of average annual precipitation. Average annual precipitation is 631.5 mm/a for all simulation runs. Average annual potential evapotranspiration is 732.2 mm/a for all simulation runs. Total water yield is the sum of surface runoff, interflow, tile flow and groundwater flow. Loss is average annual precipitation minus evapotranspiration and total water yield. As the water balance only encompasses land phase hydrology, transmission losses from the main channel are not considered and thus increase losses (i.e. reduce water yield) at the catchment outlet after channel routing. The sum of total water yield, evapotranspiration and loss gives a closed balance, i.e. equals average annual precipitation (100 %). Minor discrepancies occur due to rounding.

	Water Yield Components [mm/a]				Total Water Yield		Evapotranspiration		Loss	
Model Run	Surface Runoff	Interflow	Tile Flow	Groundwater Flow	[mm/a]	[%]	[mm/a]	[%]	[mm/a]	[%]
349	21.60	78.17	39.73	0.00	139.50	22.09	482.8	76.45	9.20	1.46
1054	21.98	41.72	40.27	0.03	104.00	16.47	501.6	79.43	25.90	4.10
1434	13.30	35.59	31.06	5.41	85.36	13.52	492.3	77.96	53.84	8.53
2868	17.36	34.23	30.54	4.43	86.56	13.71	503.6	79.75	41.34	6.55
4648	12.23	61.29	31.83	2.86	108.21	17.14	497.3	78.75	25.99	4.12
5221	21.87	54.34	31.52	0.01	107.74	17.06	499.8	79.14	23.96	3.79
5502	11.17	35.10	25.76	10.91	82.94	13.13	494.6	78.32	53.96	8.54
6388	14.87	42.53	27.86	0.94	86.20	13.65	521.6	82.60	23.70	3.75
7739	20.62	44.89	29.52	0.84	95.87	15.18	510.2	80.79	25.43	4.03
8638	22.79	53.56	36.10	0.00	112.45	17.81	508.6	80.54	10.45	1.65
8957	17.99	44.38	25.49	0.00	87.86	13.91	524.8	83.10	18.84	2.98
Min	11.17	34.23	25.49	0.00	82.94	13.13	482.8	76.45	9.20	1.46
Median	17.99	44.38	31.06	0.84	95.87	15.18	501.6	79.43	25.43	4.03
Max	22.79	78.17	40.27	10.91	139.50	22.09	524.8	83.10	53.96	8.54
Mean	17.80	47.80	31.79	2.31	99.70	15.79	503.4	79.71	28.42	4.50

The occurrence of underground water loss in the simulations is consistent with the presumed effects of the mining-induced water table draw down on the catchment hydrology formulated in Chapter 2. However, the actual extent of water loss is difficult to verify and depends on reliable estimates of evapotranspiration. There is some independent data that allows for a rough quantification of actual evapotranspiration in the Swist catchment. The Hydrological Atlas of Germany (HAD) (German Federal Ministry for the Environment 2003) specifies annual actual evapotranspiration to be above 500 mm in almost the entire Swist river basin for the reference period from 1961 to 1990.[1] Koch (2004) states that annual actual evapotranspiration for the Swist catchment is 535 mm, without providing specific references. In their application of the mGROWA model in the GROWA+NRW 2021 project, Herrmann & Wendland (2021) calculate annual actual evapotranspiration in NRW for the reference period from 1981 to 2010.[2] Aggregating the mGROWA raster output for the part of the Swist catchment that is located inside NRW yields an estimate of almost exactly 500 mm annual evapotranspiration. In summary, estimates of long-term annual evapotranspiration for the Swist catchment range from approximately 500 to 535 mm. Relating these values to annual precipitation of 600 to 700 mm results in a minimum estimate of slightly more than 70 % and a maximum estimate of slightly less than 90 % of annual precipitation that evapotranspirates. This is in good accordance with simulated annual evapotranspiration across the 11 selected SWAT runs, which varies roughly between 76 % and 83 % of precipitation. The corresponding simulated water loss fractions range from approximately 1.5 to 8.5 % (land phase only) .

Looking at the individual flow components contributing to total water yield in Table 6.1, the generally low values for groundwater flow are particularly noteworthy. On average, modeled groundwater contributions are only at 2.31 mm/a, compared to 47.80 mm/a for interflow - the largest individual flow component - followed by tile flow with 31.79 mm/a and surface runoff with 17.80 mm/a. For almost half of the selected model runs, groundwater flow is practically 0 mm/a, and only one model run exhibits groundwater flow above 10 mm/a. Low or non-existent groundwater flow often coincides with high overall water yields above 100 mm/a, whereas higher groundwater contributions well above 1 mm/a usually fall together with relatively small water yields of less than 90 mm/a (Spearman's rank correlation coefficient between both quantities is −0.75). Here, the model seems to overcompensate missing groundwater contributions by increasing contributions from the other flow components, possibly in an attempt to sustain streamflow even during longer dry periods (as evident from the gauge records from

[1] `https://geoportal.bafg.de/mapapps/resources/apps/HAD/index.html?lang=en`, as retrieved on January 7, 2023.

[2] `https://www.opengeodata.nrw.de/produkte/umwelt_klima/wasser/grundwasser/mgrowa/`, as retrieved on January 7, 2023.

Morenhoven and Weilerswist, streamflow never completely ceases along the Swist main watercourse downstream of WWTP Flerzheim). This compensation seems to come mainly from the interflow component, which exhibits by far the largest variability with an absolute difference between the simulated minimum and maximum value of more than 40 mm/a.

Kistemann et al. (2007) assume overall groundwater recharge in the Swist catchment to equal 90 l/s and use this value as a proxy for baseflow at the catchment outlet in Weilerswist. Subtracting WWTP contributions from annual summer low flows in the Weilerswist gauge records yields a similar baseflow estimate of around 100 l/s (Erftverband personal communication). The only simulated groundwater contribution that roughly corresponds to this flow rate is the maximum value of 10.91 mm/a. All other simulations would fail to reproduce baseflow in the same order of magnitude assuming baseflow to exclusively originate from groundwater contributions. This deficit in groundwater contributions is closely related to the SWAT parameter DEPIMP_BSN.bsn, which defines the depth to an impervious layer in the subsurface. Setting this parameter enables simulation of a perched water table and tile flow in SWAT, while at the same time restricting groundwater recharge with decreasing depth of the impervious zone in the subsurface (Neitsch et al. 2011). For shallow impervious layers, this may lead to a quick depletion of groundwater storage in the shallow aquifer below the threshold value for groundwater flow to occur (see Subection 3.1.2). Indeed, low groundwater flows at or very near 0 mm/a coincide with shallow impervious layers in the simulations, whereas higher groundwater contributions are associated with deeper impervious layers (i.e. higher values for DEPIMP_BSN.bsn, data not shown).

The effect of DEPIMP_BSN.bsn on the simulation results is a plausible explanation for the low sensitivity indices of all groundwater parameters in global sensitivity analysis (see Section 5.2). While the default setup in the uncalibrated model corresponds to virtually unhindered groundwater recharge, parameter variation for sensitivity analysis includes values for DEPIMP_BSN.bsn that prevent groundwater recharge almost entirely, rendering groundwater parameters uninfluential in a significant portion of evaluated model runs. In this respect, note that - for lack of better knowledge about the actual extent of the restrictive subsurface layer in the study area - DEPIMP_BSN.bsn affects the simulated hydrology in the *entire* Swist catchment.[3] Therefore, a possible workaround to this problem would be to vary DEP_IMP.hru - the equivalent of DEPIMP_BSN.bsn on the HRU level - in only a portion of the SWAT subbasins, possibly using the spatial distribution of the known agricultural tile drains (see map in Appendix A) as a proxy for those areas where perched water tables have the most palpable ef-

[3] This rationale is similar to that applied to the hydraulic conductivity of the main channel alluvium (i.e. the SWAT parameter CH_K2.rte, see the statements in Subsection 4.3.2).

fect on soil hydrology. However, the real extent of the restrictive subsurface layer in the Swist catchment remains ultimately uncertain, just as it is unclear in how far the SWAT model structure allows for an accurate representation of the actual runoff generating processes in the Swist catchment. Together with the HRU approach for spatial disaggregation as well as the empirical nature of the SCS curve number method and its misapplication for simulating infiltration and surface runoff in SWAT (Garen & Moore 2005), this reaffirms the basic limitations of SWAT in discerning the individual contributions of different non-point source pollution sources already pointed out in Subsection 3.1.3.

Finally, it should be noted that the selected SWAT model runs in Table 6.1 fail to reproduce water storage in the Steinbach reservoir, with all of them gravely underestimating observed water volumes held in the reservoir throughout the study period[4] (data not shown). Such misrepresentation of the reservoir water balance is not necessarily unexpected, as observed streamflow records from the Kirchheim gauge were fed directly into SWAT to simulate reservoir outflows, to a large degree decoupling the upstream area from model calibration (see Subsection 4.2.1). If simulated streamflow in the reservoir catchment is too low (possibly due to missing groundwater contributions in the majority of model runs), inflows to the reservoir will fail to adequately compensate for the predefined reservoir outflows as well as evaporation losses. Regardless of the exact causes, there is considerable potential for errors in the simulated reservoir storage to affect in-stream nitrogen transport and processing, as reservoirs may act as important N sinks within surface water catchments (Harrison et al. 2009, Schmadel et al. 2018). In this particular case, however, two factors make it unlikely for the reservoir to significantly influence in-stream nitrogen in the Swist catchment. First, the catchment area of the Steinbach reservoir is approximately 14.5 km^2, which amounts to only 5 % of the entire Swist catchment area. Second, the Steinbach reservoir catchment is almost entirely covered by forest and thus mostly unaffected by nitrogen inputs from urban or agricultural sources. Therefore, it is similar to the catchment of the Erftverband water quality monitoring point at the Ersdorfer Bach, which is representative of "pristine" conditions in the Swist catchment and is characterized by only very low levels of in-stream reactive nitrogen (see the measured N concentrations in Table 5.3, note that the considerably higher simulated concentrations refer to a location further downstream already influenced by agricultural and residential areas).

[4] Observed water levels and corresponding water volumes in the Steinbach reservoir were available for most of the study period, albeit with a long gap from late October 2013 to mid-August 2015.

6.3. Nitrogen Balance and Emissions

6.3.1. Soil Nitrogen Balance

Table 6.2 gives an overview of the simulated land phase nitrogen balance across the 11 selected SWAT model runs, representing the entire Swist catchment area from 2012 to 2018. Note that the presented data is not restricted to agricultural land uses. Furthermore, it does not account for in-stream N retention in the Swist river network, which is simulated as part of the SWAT routing phase. Instead, the land phase comprises all soils and vegetation in the catchment, and the nitrogen balance includes all additions of reactive N to these compartments, as well as all removals of reactive N from these compartments. Processes that add reactive N include fertilization, biological N fixation and atmospheric deposition. Processes that remove reactive N encompass crop harvesting, denitrification, ammonia volatilization, leaching of nitrate to the groundwater as well as lateral nitrogen transport to adjacent surface waters. As expected for a study area heavily influenced by intensive crop cultivation, the dominant N input process is fertilization, with a catchment-wide average of 90.9 kg N/(ha a) added to the soil-vegetation system with fertilizer.[5] Inputs from biological N fixation amount to only 3.1 kg N/(ha a) on average. As predefined by the applied ArcSWAT default settings, atmospheric N deposition is inactive in the simulations and therefore omitted from Table 6.2. Potential consequences of missing atmospheric deposition are discussed further below in this subsection.

According to the simulations, the most important loss process for reactive nitrogen from the soil-vegetation system is harvesting from agricultural land[6] which averages 67.8 kg N/(ha a) over all model runs and the entire catchment area. Denitrification constitutes the second most important loss process with 37.0 kg N/(ha a) on average, followed by lateral transport to surface waters (8.1 kg N/(ha a)) and NO_3^--N leaching to groundwater (2.4 kg N/(ha a)). As N fertilizer is added exclusively as NO_3^--N in the applied model setup (see Chapter 4), ammonia volatilization does not occur in any model run and thus remains unaccounted for as a nitrogen loss process in the simulations. Subtracting aggregated N losses from aggregated N gains to the soil-vegetation system yields an average, catchment-wide N budget of −21.2 kg N/(ha a). Although N budgets show a considerable degree of variation between the simulations, negative values are a stable feature of all calibrated model runs listed in Table 6.2, indicating an overall reduction of N stocks in the Swist catchment soils and plant biomass.

[5] Although agricultural fertilizer quantities are specified in the applied management schedules (see Appendix C) and thus should remain unaffected during calibration, there is some variability in fertilizer inputs between model runs originating from auto-fertilization of green spaces in urban/residential areas.

[6] Including some non-agricultural N losses due to mowing of green spaces in urban/residential areas.

Table 6.2.: Simulated nitrogen balance in the study period (2012-2018) for the behavioral SWAT model runs (land phase only). The listed values are averages for all simulation years and the entire Swist catchment area. The land phase N balance comprises all soils as well as vegetation, but excludes groundwater (in contrast to the water balance in Table 6.1). The N budget is the difference between all nitrogen gains and losses in the soil-vegetation system and thus represents the simulated change in storage during the study period, with minor discrepancies due to rounding. Atmospheric N deposition and ammonia volatilization have been inactive in the simulations and are therefore omitted from the gain and loss processes, respectively. Fertilizer inputs vary between model runs due to differences in SWAT auto-fertilization of green spaces in urban areas. N cycling involves processes that neither add nor remove nitrogen from the soil-vegetation system. For comparison, the bottom row shows the simulated nitrogen balance for the uncalibrated default model setup.

	N Gains [kg/(ha a)]		N Cycling [kg/(ha a)]		N Losses [kg/(ha a)]				
Model Run	Fertilizer	Biol. Fixation	Plant Uptake	Mineralization	Harvest	Denitrification	Leaching	Lateral Loss	Budget [kg/(ha a)]
349	95.4	3.0	108.2	84.4	59.1	52.0	0.0	11.4	−24.0
1054	93.3	2.8	115.6	90.0	62.9	52.2	0.4	8.7	−28.1
1434	93.0	3.1	105.9	61.1	60.7	32.7	3.9	6.0	−7.2
2868	94.1	3.5	107.6	76.6	59.9	47.6	4.6	6.1	−20.5
4648	89.5	3.2	126.2	92.8	70.9	40.2	1.2	6.4	−25.9
5221	89.0	3.2	128.0	99.7	71.9	28.7	4.1	15.9	−28.3
5502	88.8	2.7	139.4	108.3	78.6	32.8	7.9	6.5	−34.3
6388	90.1	3.3	119.3	72.0	67.6	29.7	1.2	5.3	−10.3
7739	88.1	3.0	129.6	95.1	73.0	30.3	3.1	8.3	−23.6
8638	92.5	2.9	115.9	83.5	63.7	42.4	0.1	9.3	−20.0
8957	86.1	3.1	136.5	94.9	77.4	18.2	0.4	4.7	−11.4
Min	86.1	2.7	105.9	61.1	59.1	18.2	0.0	4.7	−34.3
Median	90.1	3.1	119.3	90.0	67.6	32.8	1.2	6.5	−23.6
Max	95.4	3.5	139.4	108.3	78.6	52.2	7.9	15.9	−7.2
Mean	90.9	3.1	121.1	87.1	67.8	37.0	2.4	8.1	−21.2
Uncalibrated	85.8	3.5	125.1	60.0	73.2	6.9	3.4	1.2	4.8

Since many N loss processes (e.g. denitrification) are notoriously hard to quantify and typically associated with large uncertainties, mass balance studies for environmental nitrogen often do not explicitly account for them (Oenema et al. 2003). Instead, they focus on N losses with crop harvesting and possibly ammonia volatilization during fertilization, and compare these with the combined inputs through fertilization, biological N fixation and atmospheric deposition, usually yielding relatively large *N surpluses* (see for example Batool et al. (2022) for a recent evaluation of N surpluses in Europe from 1850 to 2019). These surpluses represent the overall potential for surface water N pollution, but may equally be released to the atmosphere via denitrification (Seitzinger et al. 2006) or be retained via long-term storage in soils, the vadose zone and aquifers (Van Meter et al. 2016, Ascott et al. 2017). The N budgets in Table 6.2 explicitly account for most lost processes and thus cannot be compared to N surplus estimates from other studies. When calculating N surpluses instead of budgets from the data in Table 6.2 (ignoring denitrification, leaching and lateral transport of N), the results range from 12 to 39 kg N/(ha a) over all model runs (not included in Table 6.2). Several independent N surplus estimates with a significant spatial and temporal overlap are generally higher than this range of values:

1. For the municipalities in the Swist catchment, Gömann et al. (2021) estimate agricultural N surpluses to range roughly from 30 to 75 kg N/(ha a) between 2014 and 2016, not accounting for atmospheric N deposition.

2. For the Euskirchen district and the Rhein-Sieg-Kreis district, respectively, the Chamber of Agriculture NRW (2021) estimates agricultural N surpluses to be 45 and 36 kg N/(ha a) between 2015 and 2017, as well as 55 and 58 kg N/(ha a) in 2020, again not accounting for atmospheric N deposition.

3. For the entire land area of Germany, Batool et al. (2022) estimate the total N surplus (for both agricultural and non-agricultural soils) to roughly range from 40 to 55 kg N/(ha a) in 2019.

However, when comparing this data to the SWAT simulation results, keep in mind that Gömann et al. (2021) as well as the Chamber of Agriculture NRW (2021) only consider agricultural land in their estimations, whereas the Swist catchment also includes significant portions of forested areas without fertilizer inputs as well as urban areas. Hence, the N surplus estimates from Gömann et al. (2021) and the Chamber of Agriculture NRW (2021) should be expected to be higher than N surpluses simulated by SWAT for the entire Swist catchment. Moreover, the N surplus estimate from Batool et al. (2022) applies to all of Germany. Therefore, it is unlikely to be representative of the land use distribution and management practices in the Swist catchment, making direct comparison with the SWAT simulation results difficult.

Besides the different N gain and loss processes, Table 6.2 furthermore provides some information on the internal N cycling within the soil-vegetation system, in the form of plant uptake of inorganic N as well as mineralization of organic N. As these processes neither add nor remove reactive N from the overall system, they do not contribute to the specified N budgets. However, they give further insight into the internal N fluxes and partitioning. For example, it is evident from the additional information that plants take up more N than supplied via fertilization, with plant uptake and mineralization being highly correlated, as emphasized by a Pearson's correlation coefficient of $r \approx 0.74$. This suggests that plant nitrogen demands in the SWAT simulations are to a large extent met with help of N that is mineralized from soil organic matter. Compared to the estimations compiled by Galloway & Cowling (2002) for crop production, the associated ratio of N plant uptake to fertilization seems too high, as does the nitrogen use efficiency (i.e. the ratio of N harvest removal to fertilization). However, keep in mind again that the data in Table 6.2 includes non-agricultural land uses like forests that are neither fertilized nor harvested, as well as green spaces in urban or residential areas that are (auto-)fertilized and mowed. Consequently, although Table 6.2 may possibly hint at an undersupply of field crops with fertilizer in the SWAT simulations, calculating agricultural uptake ratios and use efficiencies for N fertilizer from that data is potentially misleading.

Remember that PBIAS scores indicate underestimation of NO_3^--N loads at the catchment outlet already for the uncalibrated model setup (Chapter 4). Furthermore, remember that point source N inputs from wastewater treatment plants are fixed, model calibration ignores all parameters affecting in-stream N retention as well as groundwater N (see Chapter 5) and calibrated models mostly shut down groundwater contributions. Consequently, upward correction of NO_3^--N loads in the calibrated models most likely requires adjustment of the soil N balance to increase N emissions to surface waters. In order to assess the direction in which the calibration process changes the simulated soil N balance, the bottom row of Table 6.2 lists the individual parts of the N balance for the uncalibrated default model (see Chapter 4). Indeed, comparison of the lateral N losses in the uncalibrated model with the corresponding average for the calibrated models reveals a more than sixfold increase in diffuse N emissions from the catchment soils to surface waters (from 1.2 to 8.1 kg N/(ha a)), which comes along with a surge in denitrification from 6.9 kg N/(ha a) in the uncalibrated model to an average of 37.0 kg N/(ha a) in the calibrated models. The other two simulated loss processes for nitrogen from the soil-vegetation system - leaching and harvest removal - remain mostly unchanged on average, although leaching is greatly reduced in the majority of calibrated model runs probably due to impeded groundwater recharge. The relatively constant harvest removals are reassuring as they indicate that the crop yields in Table 4.6 remain stable in most calibrated model

runs (albeit with a substantial drop for a subset of model runs when looking at the lower end of simulated harvest removals).

From a purely qualitative point of view, the ranking of the simulated N loss processes is generally plausible in both the uncalibrated model and the calibrated models, with the vast majority of reactive N in agricultural landscapes being removed with crop harvesting, followed by denitrification and exports to ground- and surface waters in variable proportions (Galloway et al. 2004). From a quantitative view, the accuracy of simulated denitrification is particularly difficult to assess given the usually high variability of environmental denitrification rates and the large uncertainties associated with their quantification (Seitzinger et al. 2006). A first point of reference for denitrification is the worldwide estimate for agricultural soils of about 14 kg N/(ha a) mentioned by Seitzinger et al. (2006), a value that is roughly consistent with the uncalibrated model output of 6.9 kg N/(ha a) for the entire Swist catchment (denitrification rates can be expected to be higher in agricultural soils due to greater availability of reactive N from fertilization). More specific to the study area, Kunkel et al. (2021) rate the overall potential for soil-borne denitrification to be "low" to "moderate" in the Swist catchment (classes 2 and 3 in a classification system ranging from 1 - "very low" - to 5 - "very high"), which corresponds to a maximum denitrification rate between 10 and 50 kg N/(ha a). These values mirror the range of denitrification rates simulated by the 11 calibrated model runs quite well, although it is important to stress that they merely represent denitrification *potential* and that actual denitrification is probably lower. In summary, the available reference data in both Seitzinger et al. (2006) and Kunkel et al. (2021) suggests that calibrated denitrification rates in this thesis may be too high, albeit with considerable uncertainty.

As the calibration process needs to raise NO_3^--N loads transported to the catchment outlet, it is striking to see that model calibration leads to an increase in denitrification across all selected model parameterizations, effectively reducing the amount of reactive N available for displacement to surface waters. According to the theoretical SWAT documentation (Neitsch et al. 2011, pp. 194-195), two of the the calibrated model parameters have a direct impact on simulated denitrification: The threshold soil water content for denitrification SDNCO.bsn and soil organic carbon content SOL_CBN.sol. In addition, other soil parameters like SOL_K.sol affect denitrification indirectly by controlling water residence times in the soil profile (Neitsch et al. 2011, pp. 151-152, 162-164). Looking at the posterior distributions in Figure 5.5, it is evident that calibration of these parameters works in opposite directions regarding denitrification: On the one hand, nitrate calibration clearly tends to move SDNCO.bsn up from the default value of 1.1, while also increasing SOL_K.sol to reduce water residence times in the soil profile. Taken on their own, both changes should lead to decreased denitrification. On the

other hand, however, streamflow calibration appears to necessitate lower values for SOL_K.sol (i.e. longer water residence times in the soil profile), to a large extent erasing the effects of nitrate calibration in the combined posterior distribution, which is centered around the default value of 0. When considering SOL_CBN.sol, nitrate calibration favors higher soil organic carbon contents, presumably to increase the amount of organic N in the catchment soils, while at the same time facilitating denitrification. Furthermore, low values of DEPIMP_IMP.bsn may help to raise soil water levels above the denitrification threshold (SDNCO.bsn) by impeding percolation and creating perched water tables throughout the Swist catchment.

In summary, the calibrated SWAT models produce much higher N losses due to denitrification and surface water emissions than the uncalibrated model. At the same time, they largely maintain agricultural crop yields and the associated harvest N removals. Taken together, both requires higher supplies of soil reactive N. However, as neither fertilization nor biological N fixation rise enough to compensate for the additional N losses, model calibration instead leads to a marked increase in mineralization of soil organic N from 60.0 kg N/(ha a) to an average of 87.1 kg N/(ha a) (see Table 6.2). This increase in mineralization is the reason why overall N budgets turn from slightly positive in the uncalibrated model to distinctly negative in the calibrated models, corresponding to an overall depletion in soil organic N stocks. While the ultimate fate of N surpluses is difficult to investigate, there is evidence that they may actually contribute to N accumulation in many agricultural soils (Galloway et al. 2004, Van Meter et al. 2016), supporting the uncalibrated model output and contradicting the calibrated model results. Hence, while the calibrated models succeed in elevating soil-borne N emissions to surface waters, the uncalibrated model presumably generates more realistic estimates for soil N accumulation and denitrification. In conclusion, there seems to be a general shortage of soil reactive N in the simulations, with both the uncalibrated model and the calibrated models failing in different ways to realistically partition the available reactive N and simulate a plausible nitrogen balance for the Swist catchment.

There are several options to resolve the apparent shortage of reactive N in the SWAT simulations. First, as detailed in Chapter 4, the amounts of applied N fertilizer in the models are probably too low to adequately represent agricultural management practices in the Swist catchment during the study period. In order to assess the effect of higher fertilization on the simulation results, the 11 calibrated SWAT models have been rerun with a general 20 % increase of fertilization rates across all agricultural land uses in the Swist catchment (cropland, grassland and orchards). Figure 6.1 illustrates the corresponding impact on various performance metrics assessing NO_3^--N loads at the catchment outlet in both the calibration and validation period. Indeed, the additional N fertilizer leads to an elevated distribution of PBIAS scores that

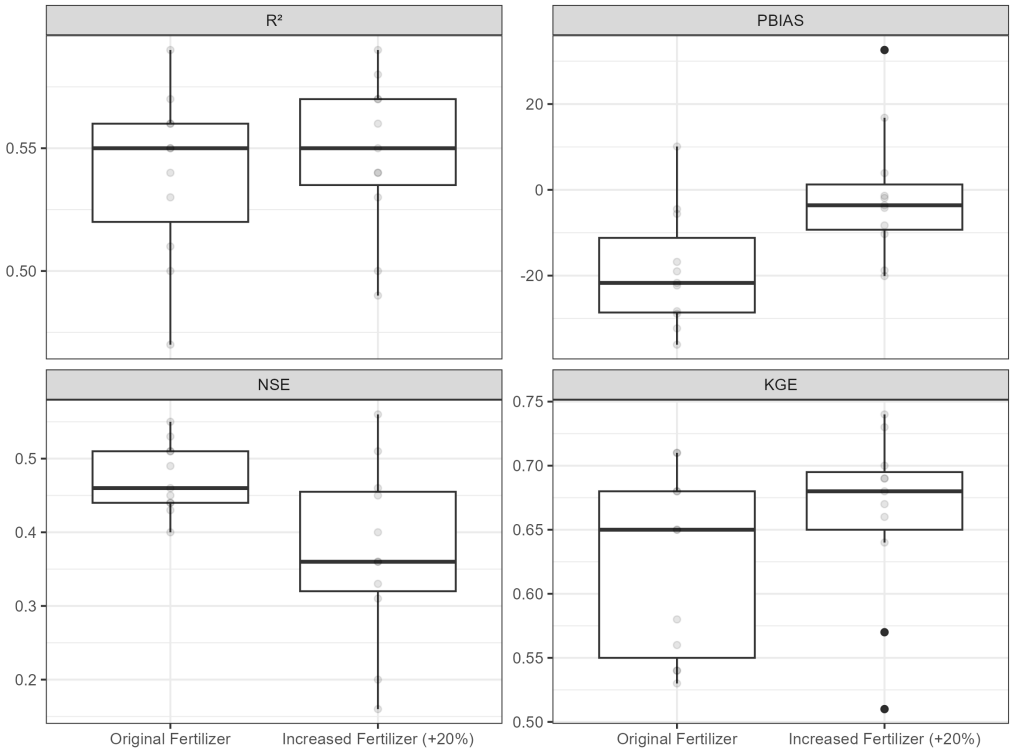


Figure 6.1.: Distributions of various performance measures evaluating nitrate simulation with and without additional N fertilizer in the Swist catchment during the study period (2012-2018). The performance measures are calculated based on in-stream NO_3^--N loads near the catchment outlet in Metternich. Their distributions include the $n = 11$ calibrated (and successfully validated) SWAT models with and without a 20 % increase in fertilization rates for agricultural land uses (cropland, grassland, orchards). For the boxplots, the length of the whiskers is limited to one and a half of the interquartile range. In addition to the boxplots, the individual model performances are plotted as transparent points.

is centered around a value near 0, effectively eliminating much of the systematic error present in the original simulations. Interestingly, the added amounts of fertilizer also improve R^2 and KGE, whereas NSE scores deteriorate, most likely due to misrepresentation of NO_3^--N load peaks (keep in mind that model calibration is based on the original fertilizer quantities listed in Table 4.4). In general, these results demonstrate that low fertilization is a reasonable explanation for underestimated NO_3^--N loads in the simulations.

As an alternative to an adjustment of fertilization rates, inputs of reactive N to the catchment soils may also be raised by activating atmospheric deposition in the SWAT models. According to the GROWA+NRW 2021 data, atmospheric deposition roughly contributes an additional 13.5 kg N/(ha a) to the Swist catchment area on average. For the entire state area of North Rhine-Westphalia, the corresponding average value equals approximately 15 kg N/(ha a)

(Wendland et al. 2021a). Although considerably lower than N inputs via fertilization, these numbers show that atmospheric deposition likely constitutes a meaningful part of the N balance in the study area and could therefore explain part of the apparent nitrogen deficit in the behavioral SWAT models. However, keep in mind that the impact of missing atmospheric N deposition might to a certain extent be offset by omission of ammonia volatilization in the models. Hence, in order to explicitly account for the combined effect of both processes in the simulations, the possible future inclusion of atmospheric deposition in the SWAT models should ideally come along with the likewise inclusion of ammonia volatilization as a loss process, necessitating an estimation (or educated guess) of NH_4^+-N shares in the fertilizers applied to the Swist catchment area.

Instead of raising nitrogen inputs via fertilization and/or atmospheric deposition, supply with reactive N can also be increased by lowering nitrogen losses for example via denitrification. A possible starting point for this is the DEPIMP_BSN.bsn parameter, which defines the depth to an impervious layer in the subsurface needed to simulate perched water tables and tile flow. As suggested in Section 6.2, use of the DEP_IMP.hru (instead of DEPIMP_BSN.bsn) parameter would allow to restrict the impervious layer to subbasins with prevalent agricultural tile drainage, increase catchment-wide groundwater recharge and reconnect surface waters to the shallow aquifer in those calibrated SWAT models with low values of DEPIMP_BSN.bsn. Simultaneously, the same approach may also help to better drain the soil profile in the rest of the subbasins and thus reduce soil-borne denitrification in a large part of the catchment area. In addition, reconnecting surface waters with the shallow aquifer entails reconnecting them to the pool of groundwater N initialized at simulation start (see the initial NO_3^--N concentrations in the shallow aquifer specified in Table 4.2), which in turn may alleviate the in-stream load deficits of NO_3^--N during low flows present in most calibrated SWAT models (see the uncertainty band of simulated NO_3^--N loads in Figure 5.12).

Finally, in-stream NO_3^--N loads at the catchment outlets may also be adjusted without any changes to land phase nitrogen cycling and transport, by instead modifying model parameters that affect in-stream nitrogen retention during the SWAT routing phase. As stated in Chapter 5, QUAL2E parameters for simulation of in-stream water quality and retention remained unconsidered for model calibration. Despite this, in-stream N retention varies noticeably between the different calibrated model runs. According to the *SWAT check* software for model error screening (White et al. 2014), simulated overall retention in the Swist river network is between 5.9 % and 30.2 % of transported N across all model runs, with a median of 14.3 % and a mean of 16.1 %. This variability can largely be attributed to differences in N emissions between the model runs, with a Pearson's correlation coefficient of $r = 0.81$ between lateral N losses (see

Table 6.2) and in-stream N retention rates. The remaining variability possibly originates from different spatiotemporal patterns of N inputs to surface waters. The simulated range of in-stream N retention generally agrees well with literature data, although the median and mean value may be too low in comparison: The model-based estimates of Seitzinger et al. (2002) correspond to in-stream retention of approximately 35 % of annual nitrogen inputs for a drained area equaling that of the Swist catchment (290 km^2), while the monitoring-based maximum estimate of Schwientek & Selle (2016) for the Steinlach catchment (140 km^2) in southern Germany is 27 % of annual nitrogen inputs. Note that all in-stream N retention in SWAT can be ultimately attributed to sedimentation and burial of organic N (Neitsch et al. 2011), although in reality denitrification is the only permanent loss process for reactive N and probably substantial in rivers and streams (Seitzinger et al. 2002, Seitzinger et al. 2006).

6.3.2. Non-Point Source Nitrogen Emissions

Figure 6.2 shows the simulated N emissions from non-point pollution sources in the Swist catchment during the study period (2012-2018). Emissions are depicted for each simulation year and pollution pathway[7] as a separate bar chart, with the bars visualizing the annual nitrogen loads for the 11 behavioral SWAT model runs. The bar charts in the bottom row and rightmost column show total annual emission loads as well as pathway-specific average loads over all simulation years, respectively, with the panel in the bottom right corner aggregating the total emissions over the entire study period. For each panel, the dotted horizontal lines in black indicate average emission loads across all behavioral model runs. In the bottom right panel, this corresponds to an overall estimate for non-point source N emissions of approximately 8.1 kg N/(ha a). Due to the mostly subdued influence of groundwater flow in the SWAT models, this value is practically equal to the mean estimate for lateral N losses from the soil profile in Table 6.2. As it represents the entire Swist catchment area and is not restricted to - for example - agricultural areas, it is difficult to compare this overall estimate to the results of other catchment-scale N pollution studies with divergent land use compositions and management practices. However, the general order of magnitude is completely plausible, with for instance Malagó et al. (2017) deriving a SWAT estimate of about 6.1 kg N/(ha a) for total diffuse nitrogen emissions in the Danube catchment from 1995 to 2009.

The emission loads vary considerably from year to year: As evident from the bottom row of Figure 6.2, total load averages across all model runs range from approximately 5 kg N/(ha a) in 2018 to more than 10 kg N/(ha a) in 2013. This interannual variability of N emissions appears

[7] Nitrogen emissions via surface runoff encompass both dissolved NO_3^--N and eroded organic N.

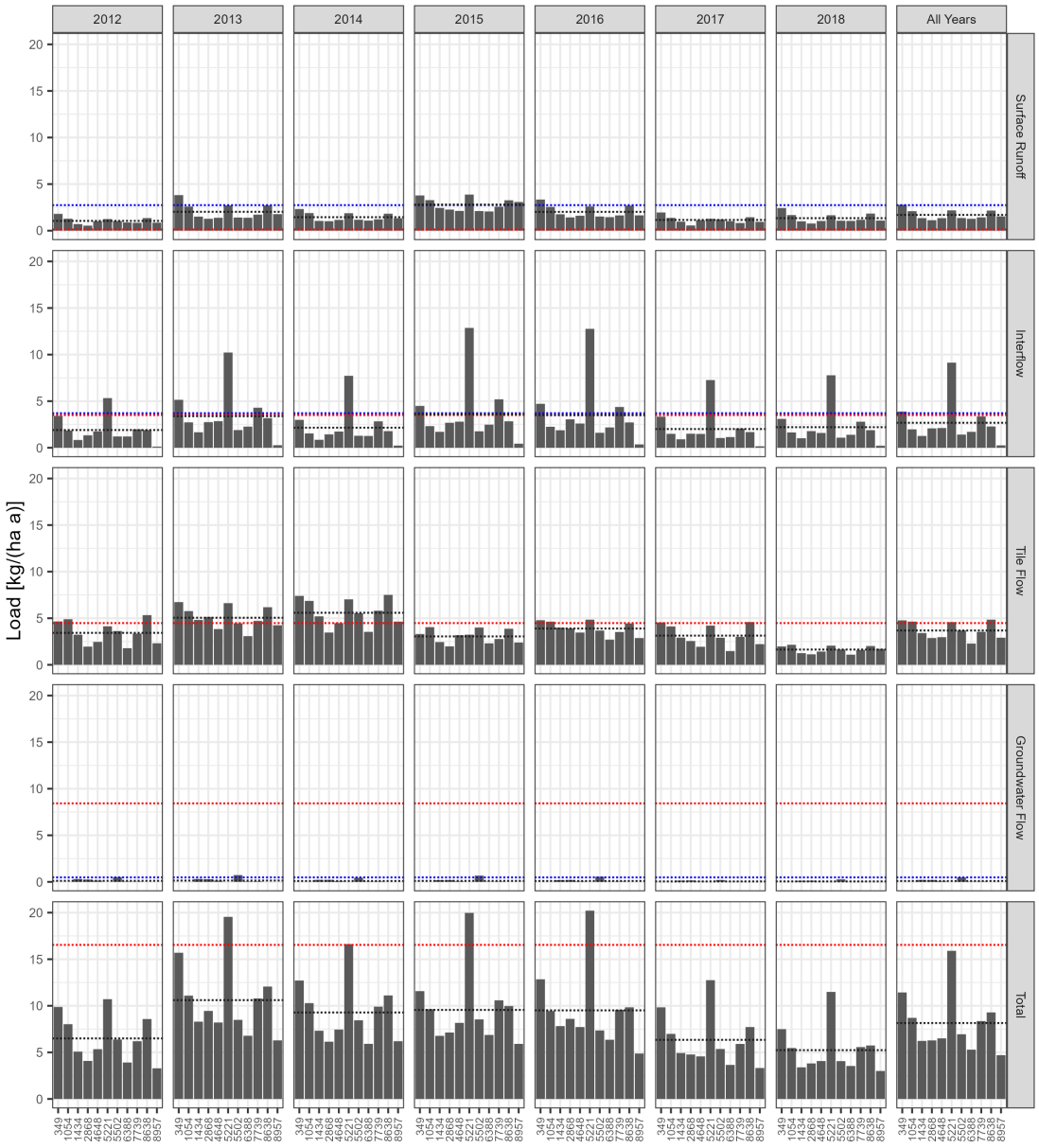


Figure 6.2.: Simulated non-point source N emission loads in the Swist catchment during the study period (2012-2018) for the behavioral SWAT model runs. The bars indicate the average simulated load of the individual model runs for the respective simulation year and emission pathway. N emissions via surface runoff encompass dissolved NO_3^--N as well as organic N transported with eroded soil. The dotted black lines indicate the average emission loads across all model runs. The dotted blue and red lines indicate the reference loads obtained from upscaled edge-of-field (EOF) monitoring data (Mertens et al. 2012), as well as modeling results from the GROWA+NRW 2021 project (Wendland et al. 2021a) aggregated for the Swist catchment, respectively. As the former reference data set lacks information on N emission loads via tile flow, calculation of a total load is omitted to discourage comparison with modeled total loads.

to be related to the general wetness conditions in the Swist catchment: The years from 2013 to 2016 all are wetter than average according to the annual precipitation data and gauge records presented in Section 2.4 (see Figure 2.4 and Table 2.2, respectively). At the same time, these simulation years unanimously produce high emission loads around 10 kg N/(ha a). In contrast, annual emission estimates for the years 2012, 2017 and 2018 - all of them drier than average - remain distinctively below 10 kg N/(ha a), with the exceptionally dry year of 2018 coinciding with the lowest simulated N emissions. A possible explanation for this apparent relationship between N pollution and wetness conditions are differences in soil-borne N mineralization and possibly nitrification. In the SWAT equations, both processes are directly dependent on soil water content (Neitsch et al. 2011, pp. 187-194), modeling the general influence of water availability on microbial activity. In the calibrated SWAT models, this effect may be particularly pronounced due to the prominent role of mineralization in supplying the catchment soils with inorganic N compounds that are mobile and plant-available (see Subsection 6.3.1).

Table 6.3 provides a summary of the same data visualized in Figure 6.2 averaged over the entire seven-year simulation period. It thus serves to complement Figure 6.2 by explicitly listing the emission load values, breaking down the contribution of the individual emission pathways to the total N loads and quantifying the spread between the different simulation results. Regarding the individual emission pathways, both Table 6.3 and the rightmost column in Figure 6.2 clearly demonstrate the lack of groundwater contribution to the overall N pollution in the simulations, with emission loads from groundwater averaging merely 0.1 kg N/(ha a) across all model runs. According to the simulations, the most important non-point source emission pathway is tile flow with an average emission load of 3.7 kg N/(ha a), followed by interflow (2.7 kg N/(ha a)) and surface runoff (1.7 kg N/(ha a)). The dominant role of tile flow in the overall emission loads is all the more remarkable because tile drainage is present in only part of the agricultural area of the Swist catchment, whereas surface runoff and interflow may occur across the entire catchment. This highlights the substantial degree to which agricultural tile drains may affect catchment-scale nutrient balances.

Considering the spread between the lowest and highest N emission estimates, both Figure 6.2 and Table 6.3 reveal substantial differences between the 11 behavioral SWAT models. For example, simulated total emission loads range from 4.7 kg N/(ha a) to 15.9 kg N/(ha a) for model runs 8957 and 5221, respectively, amounting to a more than threefold difference between the minimum and maximum estimate. As the total emission load simulated by model run 5221 is extremely high and mainly driven by an extraordinary amount of N contributions from interflow, this value may be treated as an outlier and excluded from the analysis. However, even then, the second highest simulation result of 11.4 kg N/(ha a) for model run 349 still exceeds

Table 6.3.: Nitrogen emission loads in the study period (2012-2018) for the behavioral SWAT simulation runs. The listed values are averages for all simulation years and the entire Swist catchment area. N emissions via surface runoff encompass dissolved NO_3^--N as well as organic N transported with eroded soil. The stated total loads are the sum of the pathway specific loads, with minor discrepancies due to rounding. The bottom rows show reference loads obtained from upscaled edge-of-field (EOF) monitoring data (Mertens et al. 2012) as well as modeling results from the GROWA+NRW 2021 project (Wendland et al. 2021a) aggregated for the Swist catchment. As the former reference data set lacks information on N emission loads via tile flow, calculation of a total load is omitted to discourage comparison with modeled total loads.

	Nitrogen Emission Loads [kg/(ha a)]				
Model Run	Surface Runoff	Interflow	Tile Flow	Groundwater Flow	Total Load
349	2.7	3.9	4.8	0.0	11.4
1054	2.1	2.0	4.7	0.0	8.7
1434	1.3	1.3	3.4	0.2	6.2
2868	1.1	2.1	2.9	0.2	6.3
4648	1.3	2.1	3.0	0.1	6.5
5221	2.2	9.1	4.6	0.0	15.9
5502	1.4	1.4	3.7	0.5	6.9
6388	1.3	1.7	2.3	0.0	5.3
7739	1.4	3.4	3.5	0.0	8.4
8638	2.2	2.3	4.8	0.0	9.3
8957	1.5	0.3	2.9	0.0	4.7
Min	1.1	0.3	2.3	0.0	4.7
Median	1.4	2.1	3.5	0.0	6.9
Max	2.7	9.1	4.8	0.5	15.9
Mean	1.7	2.7	3.7	0.1	8.1
EOF Monitoring	2.7	3.7	-	0.5	-
GROWA+NRW 2021	0.1	3.5	4.5	8.4	16.5

the minimum by a factor of well above 2. Looking at the individual emission pathways, the variability between the models is even more pronounced for N emissions with interflow: Here, the minimum value is 0.3 kg N/(ha a) and the maximum value is 9.1 kg N/(ha a) for exactly the same model runs as above (8957 and 5221, respectively), identifying emissions via this pathway to be the most important source of intermodel variability. This has major implication for the proportional makeup of the total loads, with interflow accounting for more than half of all N inputs in the highest emission scenario but less than a tenth in the lowest emission scenario. Although not as distinctive as for interflow, emissions via surface runoff and tile flow are also quite varied, with a more than twofold difference between minimum and maximum for both pathways. As stated above, groundwater emissions are generally negligible and practically non-existent for the majority of model runs, with model run 5502 being the most notable exception at 0.5 kg N/(ha a).

Since the differences in the 11 behavioral model runs originate exclusively from parameter calibration, intermodel variability is an indicator of parameter-based model uncertainty. As evident from the water balance components presented in Table 6.1, this kind of uncertainty already affects the simulated catchment hydrology to a notable degree. For example, the simulated total water yields increase by almost 70 % between the minimum and maximum value (82.94 mm/a and 139.5 mm/a, respectively). For the corresponding N loads transported with these water fluxes, parameter-based uncertainty is even greater, with the aforementioned increase of more than a factor of 3 between the simulated minimum and maximum emission loads (see Table 6.3). As discussed in Chapter 5, the behavioral SWAT model runs differ in their ability to reproduce long-term average NO_3^--N loads at the catchment outlet and capture measured in-stream N concentrations along the Swist and its tributaries. Model run 2868 is the only one that excels in both of these aspects, while also yielding Pareto-optimal solutions for streamflow and nitrate simulation during both calibration and validation (see Figure 5.8). Consequently, the total emission load of 6.3 kg N/(ha a) simulated by this model may be seen as a more reliable estimate for overall non-point source N pollution in the Swist catchment than the mean emission load over all behavioral model runs (8.1 kg N/(ha a)). The result for model run 2868 is also nearer to the median estimate of 6.9 kg N/(ha a) (corresponding to model run 5502), which is less affected by the potential outlier from model run 5221 than the mean estimate. However, keep in mind that errors in the model input data and model structure may introduce bias and inaccuracies that interfere with parameter calibration. Thus, there is no guarantee that the predictions of model run 2868 really are superior to those of the remaining SWAT models.

In summary, despite the wealth of available monitoring data in the Swist catchment and the considerable effort devoted to model setup and parameter calibration, the uncertainty in the SWAT emission load estimates for nitrogen is considerable. Assuming a fixed number of influential parameters and associated parameter ranges, the most promising option to reduce parameter-based uncertainty is the inclusion of additional objectives to the calibration and/or validation process. As discussed in Section 5.3, model calibration probably requires at least three instead of only two objectives to adequately constrain the 18 involved model parameters. An obvious candidate is the PBIAS metric for nitrate simulation ($PBIAS_{Nitrate}$) to place additional emphasis on systematic error in the simulated NO_3^--N loads. Applying a third behavioral threshold for $PBIAS_{Nitrate}$ of ±10 % while otherwise retaining the same calibration scheme would result in a much stricter filter for behavioral model performances and thus a substantial reduction in the number of behavioral simulations (see Table 5.2 for confirmation). A smaller sample of behavioral model parameterizations implies a less reliable assessment of model uncertainty. Possible solutions for this problem include increasing the overall number

of calibration runs from the outset as well as adjusting the model setup to enhance model performances (see the upcoming Chapter 7 for a list of potential improvements to the model setup).

The blue and red dotted lines in Figure 6.2 indicate the non-point source N emission loads obtained from two reference data sets (also see Section 6.1): First, load estimates from upscaled Erftverband *edge-of-field* (EOF) monitoring data to create an emission inventory for the entire Swist catchment (Mertens et al. 2012). Second, distributed estimates for emission loads in the Swist catchment from the NRW-wide application of the AGRUM model suite in the GROWA+NRW 2021 project (Wendland et al. 2021a). Keep in mind that these reference values are not necessarily representative for the entire SWAT study period from 2012 to 2018: The EOF monitoring data behind the Erftverband emission inventory reflect conditions between 2005 and 2009, whereas the GROWA+NRW 2021 modeling results are based on agricultural N surplus estimates for the years 2014 to 2016 (Gömann et al. 2021). In the bottom rows of Table 6.3, the pathway-specific and the total N emission loads for both reference data sets are stated explicitly. As the EOF monitoring-based emission inventory does not include agricultural tile drainage, the respective tile flow emissions and the resulting total emissions are omitted.

Excluding the unknown reference N load from tile flow, the data from the emission inventory of Mertens et al. (2012) largely confirms the general ranking of the emission pathways obtained from the SWAT simulations: Almost negligible N contributions from groundwater are exceeded by considerably higher N emission loads from surface runoff and interflow. This is a clear hint that groundwater N emissions indeed only play a minor role in the Swist catchment due to the large-scale groundwater extraction in the region, although the calibrated SWAT models may overestimate this effect by excessively impeding groundwater recharge (see Section 6.2 and Subsection 6.3.1). From a more quantitative point of view, the reference values and SWAT simulation results also agree well on the general order of magnitude in N emissions. Taken together, the N emission load estimates for surface runoff, interflow and groundwater flow by Mertens et al. (2012) amount to 6.9 kg N/(ha a), which remains below the average total N emission load from SWAT (8.1 kg N/(ha a)) but is plausible given that the latter additionally comprises N emissions via tile flow. When comparing pathway-specific emission loads in Table 6.3, however, it becomes clear that the SWAT results consistently undercut the reference data. Thinking further, this suggests that tile flow emission estimates obtained from EOF monitoring *must* contradict the simulation results one way or another: On the one hand, if they were high enough to match the SWAT predictions for tile flow, the emission load sum across all pathways would surpass the corresponding model average of 8.1 kg N/(ha a). On the other hand, if EOF

monitoring revealed markedly lower tile flow emissions than predicted by SWAT, this would imply a different ranking and relative importance of the various emission pathways.

Regarding the second reference data set that is based on the GROWA+NRW 2021 modeling results (Wendland et al. 2021a), the most striking aspect is the total non-point source emission load for the Swist catchment of 16.5 kg N/(ha a) - a more than twofold difference compared to the average value from the SWAT simulations. This surge in N input loads can almost completely be attributed to vast emissions from groundwater: With 8.4 kg N/(ha a), groundwater contributes more than half of the total emission load in the GROWA+NRW 2021 data set. This result stands in stark contrast to the almost non-existent groundwater N emissions in the SWAT simulations and also the Erftverband emission inventory of Mertens et al. (2012). It is furthermore implausible considering the widespread disconnect in the Swist catchment between surface waters and groundwater (see Section 2.3). For the GROWA+NRW 2021, groundwater flow directions and the receiving surface water bodies were identified using modeled groundwater levels for the whole of NRW (Wendland et al. 2021b), most likely yielding incorrect results for the Swist catchment. Validation of the GROWA+NRW 2021 emission loads for nitrogen confirms this: While the combination of non-point source N emission loads from the RAUMIS-mGROWA-WEKU-DENUZ model chain with point source N emission loads and in-stream N retention from the MONERIS model agrees well with observed river nitrogen loads in most of NRW, validation fails for the Erft catchment. The study authors explain this with the effects of lignite mining and groundwater extraction in the Lower Rhine Embayment (Wendland et al. 2021a). As a consequence, it is plausible to assume that the groundwater N emission loads in the GROWA+NRW 2021 data are an extreme overestimate.

Subtracting groundwater emissions from total emissions in the GROWA+NRW 2021 data produces a remaining surface water burden of 8.1 kg N/(ha a), which corresponds exactly to the average SWAT simulation result. However, this remaining N load originates almost exclusively from tile flow at 4.5 kg N/(ha a) and interflow at 3.5 kg N/(ha a), with only negligible contributions from surface runoff at 0.1 kg N/(ha a). This result contradicts the considerably higher N emissions from surface runoff predicted by the behavioral SWAT models and most likely reflects the major restrictions imposed on this emission pathway in the GROWA+NRW 2021 data (Wendland et al. 2021a). These restrictions include for example a requirement for hydrologic connectivity: If the terrain does not provide a direct topographic link between potential source areas and nearby water bodies, N emissions via surface runoff or erosion are ruled out for these parts of the catchment. This is a crucial difference to SWAT, which makes use of the HRU approach to represent spatial heterogeneity and thus enables N emissions via surface runoff from the complete area of each modeled subbasin, disregarding the effects of small-scale topography

on surface water pollution (see Subsection 3.1.3). While the Erftverband emission inventory from Mertens et al. (2012) provides an even higher N emission load estimate for surface runoff than the SWAT simulations, it may be similarly hampered given that the inventory is based on *edge-of-field* monitoring data (collected in direct proximity of the receiving streams) upscaled to the *entire* Swist catchment area.

6.3.3. Seasonality of Non-Point Source Nitrogen Emissions

A major benefit of dynamic models like SWAT (as opposed to static models like the AGRUM models applied in the GROWA+NRW 2021 project) are their ability to simulate temporal trends and seasonality. Figure 6.3 gives an impression of the seasonality in the simulated non-point source N emissions by visualizing the distribution of monthly average N loads across all behavioral SWAT models as boxplots - both pathway-specific and in total. The means across all SWAT models depicted as blue diamonds in Figure 6.3 are also explicitly listed in Table 6.4. In general, the simulated N emissions are highest in winter and lowest in summer, with the cross-model means of monthly average total N emission loads reaching a maximum of 39.3 g N/(ha d) in January and a minimum of 11.8 g N/(ha d) in August. This seasonality is most pronounced in emissions via interflow and groundwater, although groundwater emissions are

Table 6.4.: Monthly means of simulated non-point source N emission loads in the Swist catchment during the study period (2012-2018) across all 11 behavioral model runs. The total loads are the summed pathway specific loads, with minor discrepancies due to rounding. The pathway-specific annual means in the bottom row are equal to the cross-model means stated in Table 6.3 (again with minor discrepancies due to rounding). The monthly values correspond to the blue diamonds in Figure 6.3.

	Monthly Nitrogen Emission Loads [g/(ha d)]				
Month	Surface Runoff	Interflow	Tile Flow	Groundwater Flow	Total Load
Jan	7.3	11.3	20.5	3.2×10^{-1}	39.3
Feb	7.1	12.0	15.7	3.9×10^{-1}	35.2
Mar	6.5	11.8	11.7	4.3×10^{-1}	30.5
Apr	2.0	10.8	3.8	4.0×10^{-1}	17.1
May	3.8	7.7	2.2	3.1×10^{-1}	14.1
Jun	11.5	6.3	7.4	2.2×10^{-1}	25.4
Jul	6.8	5.1	5.0	1.5×10^{-1}	17.1
Aug	1.3	3.9	6.5	1.3×10^{-1}	11.8
Sep	0.6	3.6	9.9	1.4×10^{-1}	14.4
Oct	0.5	3.6	8.7	1.8×10^{-1}	13.0
Nov	3.7	4.7	10.8	2.0×10^{-1}	19.3
Dec	4.6	7.4	19.1	2.3×10^{-1}	31.3
Mean	4.6	7.4	10.1	2.6×10^{-1}	22.4

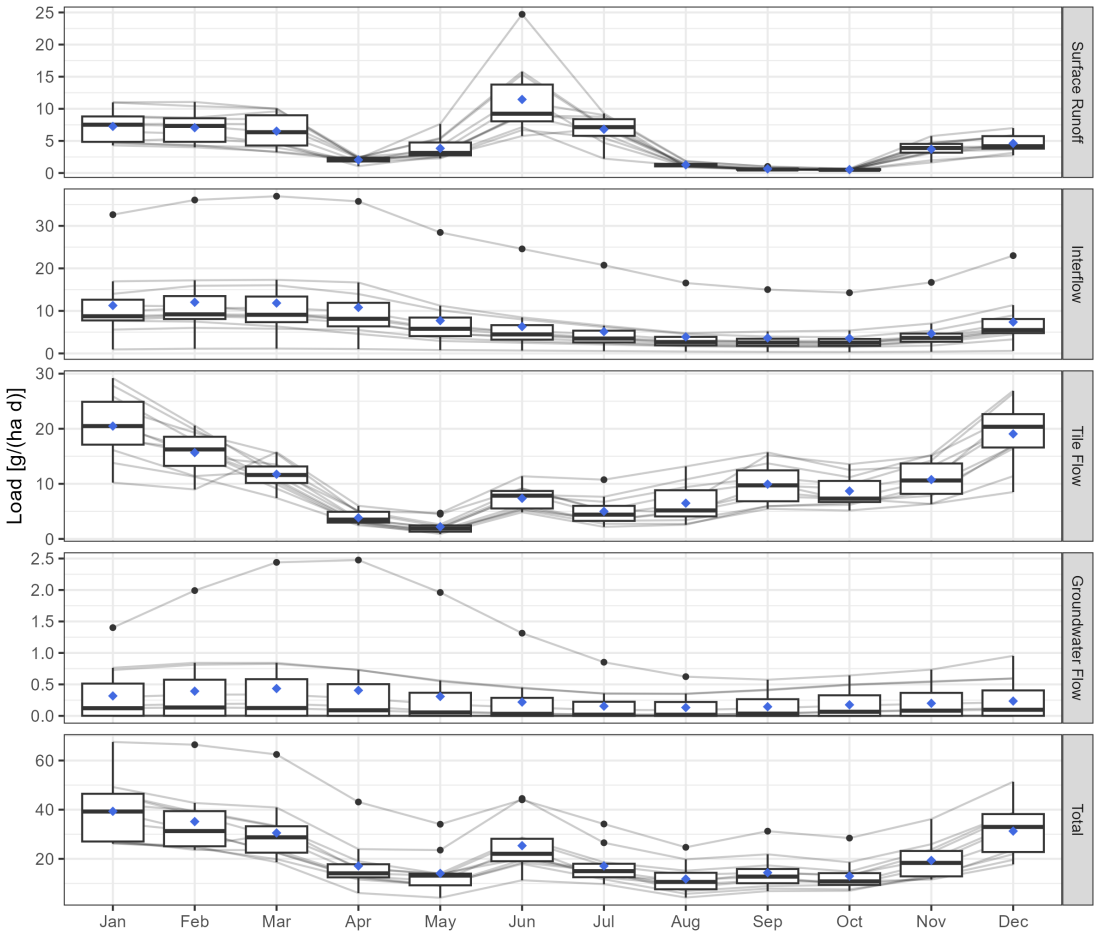


Figure 6.3.: Seasonality of non-point source N emission loads in the Swist catchment during the study period (2012-2018) as simulated by the 11 behavioral model runs. The boxplots illustrate the distribution of the monthly averages from all model runs. The length of the whiskers is limited to one and half of the interquartile range. The blue diamonds indicate the mean across all model runs and correspond to the values listed in Table 6.4. The shaded lines in the background are meant as a help to better track the seasonality in the individual model runs. Please note the different scales of the y-axes, which serve to better resolve the seasonal differences for each emission pathway.

too minuscule to meaningfully shape the annual course of the total emissions. Markedly increased non-point source N emissions during winter is a phenomenon typically observed in monitoring campaigns (Bauwe et al. 2020, Kothe et al. 2021) and can be attributed to higher NO_3^--N availability due to plant dormancy as well as elevated water yields.

Interestingly, the monthly averages of total emission loads show a temporary peak in June before dropping to their lowest values in August. Although not as high as total emissions in winter and early spring, this peak is a distinctive interruption of otherwise relatively low N emissions during summer. It originates mainly from emissions with surface runoff, which exhibit a clear maximum in June. When looking at the cross-model means in Table 6.4, surface runoff contributes almost half of all N emissions in June - namely 11.4 of 25.4 g N/(ha d). At 7.4 g N/(ha d), the second largest component are emissions from tile flow that also increase in June but remain well below their winter counterparts. The resulting overall increase of summer N emissions can most likely be attributed to (heavy) rainfall events co-occuring with elevated levels of soil reactive nitrogen due to fertilization. Regarding the spread of monthly average emissions between the different behavioral SWAT models, the applied convention for the boxplot whiskers in Figure 6.3 identifies some marked outliers in the simulated emission data: As already noted in Subsection 6.3.2, model run 5221 produces extremely high N emissions via interflow, which Figure 6.3 reveals to be systematically shifted upwards over the entire course of the year. The same can be said for emissions from groundwater in model run 5502. The highest N emissions from surface runoff in June stem from model run 349, which also produces the maximum overall emission value for this pathway (as well as tile flow, see Table 6.3).

6.3.4. Point Source Nitrogen Emissions

Apart from non-point source N emissions, the SWAT models also include point source N emissions from wastewater treatment plants (WWTP). The WWTP emission loads are calculated from daily observed time series of wastewater discharge volumes and N concentrations in the WWTP effluents (see Subsection 4.2.2 for a detailed description of the N concentration data as well as the linear regression approach used to fill time series gaps). The resulting WWTP emission loads are fed directly into SWAT and do not differ between the 11 behavioral model variants; they are listed in Table 6.5. In total, N emissions from the four active WWTPs in the catchment amount to nearly 80 t N/a released into the Swist river network during the study period. The individual N loads from each WWTP are roughly proportional to the annual wastewater quantities as stated in Table 2.1 for 2018, with WWTP Flerzheim emitting about twice as much reactive N as WWTP Rheinbach, followed by the considerably smaller WWTP Miel and WWTP Heimerzheim. The interannual variability in WWTP emissions mirrors that in the

Table 6.5.: Annual emissions of total N from the four WWTPs in the Swist catchment during the study period (2012-2018). The emission loads are calculated from observed wastewater discharge volumes and effluent N concentrations as detailed in Subsection 4.2.2. WWTP N emissions are the same in each of the 11 behavioral SWAT models.

	Annual Nitrogen Emissions from WWTPs [t/a]				
Year	Flerzheim	Rheinbach	Miel	Heimerzheim	All
2012	43.97	20.67	7.58	5.51	77.73
2013	46.98	21.58	7.79	5.41	81.75
2014	43.81	21.93	7.13	4.45	77.32
2015	48.38	22.72	8.24	4.77	84.10
2016	47.52	23.71	8.43	4.67	84.34
2017	44.08	22.21	8.10	4.96	79.35
2018	37.91	19.74	7.25	4.38	69.29
Annual Average	44.66	21.79	7.79	4.88	79.13

non-point source emissions (see Subsection 6.3.2), with the relatively wet years of 2015 and 2016 exhibiting the highest total loads and the exceptionally dry year of 2018 producing the lowest total load. Note however that the relative differences between the annual WWTP loads are much less pronounced than for their non-point source counterparts.

Figure 6.4 shows the simulated overall WWTP emission loads in relation to the corresponding non-point source emission loads examined in Subsection 6.3.2. On average, the WWTPs contribute 26.8 % to the total simulated N emissions in the Swist catchment between 2012 and 2018, with the WWTP load fraction varying from 14.7 to 36.7 % across the different SWAT models. As the WWTP emissions are fixed for all models, this variability originates exclusively from the differences in the simulated non-point source N emission loads. Just like the spread in the simulated non-point source emissions, the spread in the WWTP load fractions is thus a consequence of parameter-based model uncertainty. In general, the SWAT-based estimates of WWTP contributions to the overall Swist N burden is in good agreement with independent literature data: From their monitoring-based emission inventory roughly representing the years 2005 to 2009, Mertens et al. (2012) estimate the N emission load fraction from WWTPs to be 23 % in the Swist catchment. For NRW between 2014 and 2016, the modeling results from GROWA+NRW 2021 suggest the WWTP emission load fraction to be approximately 16 %, with the regional estimate for the entire Erft catchment being slightly lower (Wendland et al. 2021a). In a long-term continuation of source apportionment studies such as Fuchs et al. (2010), the German Federal Environment Agency (UBA) publishes surface water pollution data derived from the MoRE modeling system (Fuchs et al. 2017).[8] For the reference period from 2012 to 2016, MoRE esti-

[8] `https://www.umweltbundesamt.de/daten/wasser/fliessgewaesser/eintraege-von-naehr-schadstoffen-in-die`, as retrieved on May 18, 2024.

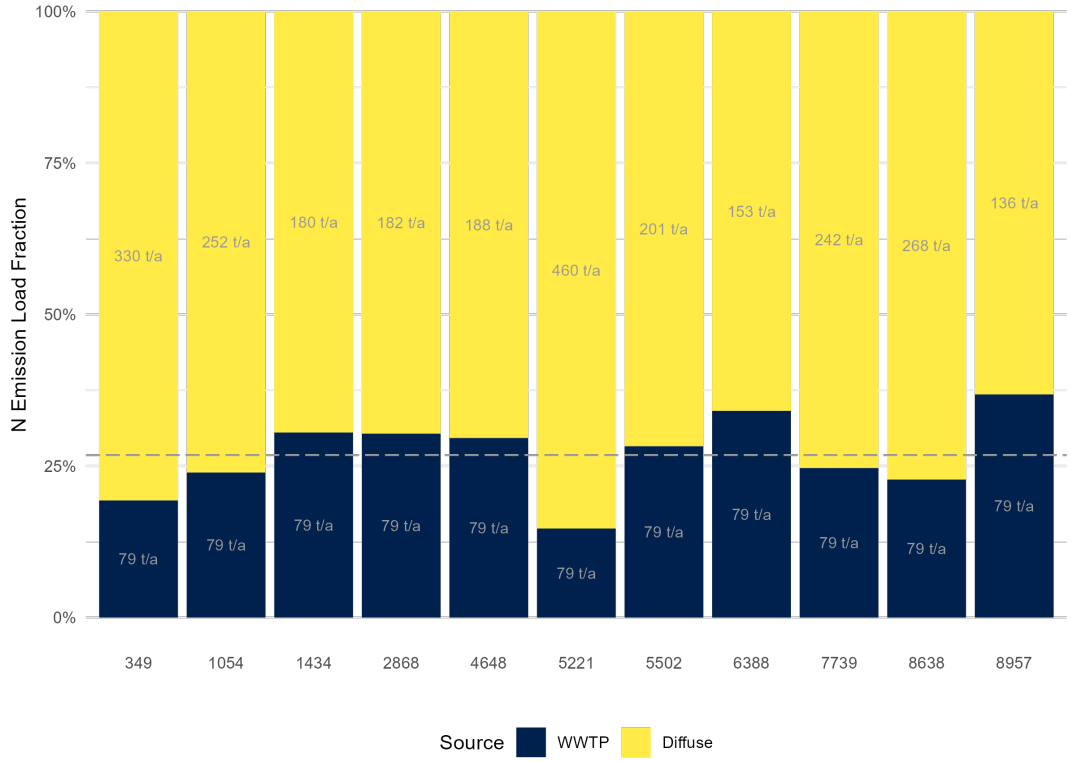


Figure 6.4.: Fractions of the total N emission loads from WWTPs and non-point (diffuse) sources during the study period (2012-2018) for the 11 behavioral model runs. The small labels state the absolute N emissions from WWTPs and diffuse sources, with the former being independent of the calibrated model parameters and thus constant across all model runs (also see Table 6.5). The dashed horizontal line indicates the average WWTP load share across all model runs (26.8 %).

mates the contribution of WWTPs to overall N pollution of German surface waters to be about 17 %. The noticeably higher WWTP contributions for the Swist catchment in both the SWAT results and the Mertens et al. (2012) data are plausible given the generally high wastewater load in the Swist (see Section 2.3).

When interpreting the above-stated numbers, be aware that the SWAT simulations omit further types of point sources that are mostly considered in the cited literature data. These include discharges from combined sewer overflow (CSO) events, stormwater discharges from separate sewer outlets (SSO) as well as direct discharges from industry or trade. In their monitoring-based emission inventory for the Swist catchment between 2005 and 2009, Mertens et al. (2012) estimate CSO and SSO contributions to overall N pollution in the Swist catchment to be 4 % and 2 %, respectively. The authors do not consider direct discharges from industry or trade in their study. The GROWA+NRW 2021 data yields a combined contribution from CSO and SSO

to overall N emissions of 8 % for all of NRW between 2014 and 2016, with direct discharges from industry or trade being negligible (less than 0.1 %) (Wendland et al. 2021a). For all of Germany, the MoRE results rates the combined contribution from CSO and SSO to be 5 % of overall N pollution between 2012 and 2016, while estimating direct industrial contributions in the same time period to be about 1.5 %. In summary, all the reference data sets agree that the remaining emissions from point source types not considered in the SWAT simulations amount to a value of less than 10 % of overall N pollution. This in turn suggests that the simulated WWTP N load fractions that are depicted in Figure 6.4 are slight overestimates, but that they probably reproduce the general order of magnitude of WWTP nitrogen emissions.[9] Inclusion of point sources other than WWTPs would nevertheless be an appealing option for refinement of the SWAT models applied in this thesis, but is likely going to be challenging due to data constraints (see Section 6.4).

6.3.5. Critical Source Areas

The spatial variability of non-point source N emissions as simulated by SWAT can be evaluated either at the HRU level (Ghebremichael et al. 2010, Winchell et al. 2015, Lee et al. 2018, among others) or at the subbasin level (Tripathi et al. 2003, Niraula et al. 2013, Evenson et al. 2021, among others). This thesis makes use of the second option, acknowledging several drawbacks of the HRU-level approach: First, HRUs represent area fractions that - while being more or less uniform in terms of land use, soils and slope - cannot be meaningfully identified with continuous and specifically managed stretches of land in most SWAT studies, including the present thesis. Second, even if special measures are taken to better relate HRUs to well-defined landscape units like individual cropland fields, there is usually not enough information on management practices in each landscape unit available to notably improve the simulation results, as demonstrated for example by Apostel et al. (2021). Third, as SWAT directly transfers runoff and pollutants from each HRU to the stream network, basically neglecting the spatiotemporal variability of runoff generation as well as pollutant mobilization and transport in the landscape, it is highly doubtful that SWAT is able to accurately recreate differences in non-point source pollution patterns at the HRU level. Therefore, resorting to subbasins avoids purporting a level of spatial precision in the simulation results SWAT is probably unable to provide.

[9] On another note, the omission of certain point sources in the SWAT models may imply bias in calibrating the model against observed in-stream N loads in Metternich (see Chapter 5). However, keep in mind that emissions from the omitted point sources are mostly restricted to high rainfall events. Furthermore, remember that the model is calibrated solely on NO_3^--N loads, whereas the presumably highest unconsidered N emissions from CSO events mainly comprise NH_4^+-N and organic N (the predominant N forms in untreated wastewater). Therefore, omission of point sources other than WWTPs is considered unlikely to meaningfully affect the model calibration process with regard to N emissions.

Figure 6.5 gives an overview of the simulated non-point source N emissions in the different SWAT subbasins of the Swist catchment during the entire study period, averaged over all behavioral SWAT model runs. Mirroring the presentation of the catchment-wide emission loads in Figure 6.2, the individual maps show the subbasin emission loads for each simulation year and pollution pathway (with the exception of groundwater flow, which contributes so little N that the spatial differences cannot be visibly resolved on the same scale against the other pathways). The bottom row and rightmost column depict the sum of emission loads over all pathways and the pathway-specific average emission loads over all simulation years, respectively. The panel in the bottom right corner aggregates the subbasin N emission loads for all pathways (including groundwater flow) and the entire study period. According to the various maps in Figure 6.5, the simulated non-point source N emissions primarily seem to concentrate in two separate parts of the Swist catchment: First, the Schiessbach headwater region in the west of the study area, and second, the region along the upper and middle reaches of the Swist main watercourse around Meckenheim in the southeast of the catchment. Both are located in areas with intensive cropland and/or orchard agriculture. In contrast, forested areas - especially the Eifel foothills in the south and southwest of the catchment - exhibit generally low N emission loads.

Although the subbasins show a notable degree of interannual variation in overall emissions (see bottom row of Figure 6.5), elevated N emission loads along the Schiessbach headwaters as well as the upper and middle Swist main watercourse are a recurring result in all simulation years. The same can be said when comparing simulated total emissions for the entire study period between the behavioral SWAT model runs (data not shown). Hence, both regions are the most obvious candidates for nitrogen critical source areas (CSA) in the Swist catchment. Looking at the contributions of the individual emission pathways in Figure 6.5, it becomes clear that high overall input regions for nitrogen predominantly co-occur with high tile flow emissions: Pearson's correlation coefficient between subbasin tile emission loads and the total subbasin loads is $r = 0.74$, with the corresponding correlation coefficients for the other emission pathways all remaining below 0.5. This underscores the important role of tile drains as pathways for nitrogen pollution in the SWAT simulations. Hence, high N availability due to fertilization and tile drainage in agricultural areas seem to be determining factors for the occurrence of critical source areas in the Swist catchment according to the simulation results.

Figure 6.6 compares the SWAT overall N emission map (see the bottom right panel of Figure 6.5) to the corresponding emission map for the GROWA+NRW 2021 raster data, aggregated for each SWAT subbasin. Since N emissions from groundwater in the GROWA+NRW 2021 data are excessive for the Swist catchment (see Subsection 6.3.2) and would override any meaning-

Figure 6.5.: Simulated non-point source N emission loads for the SWAT subbasins of the Swist catchment during the study period (2012-2018), averaged over all behavioral SWAT models. As N emissions from groundwater are minuscule for every simulation year, they are not depicted. However, they are included in the total N emission loads (bottom row).

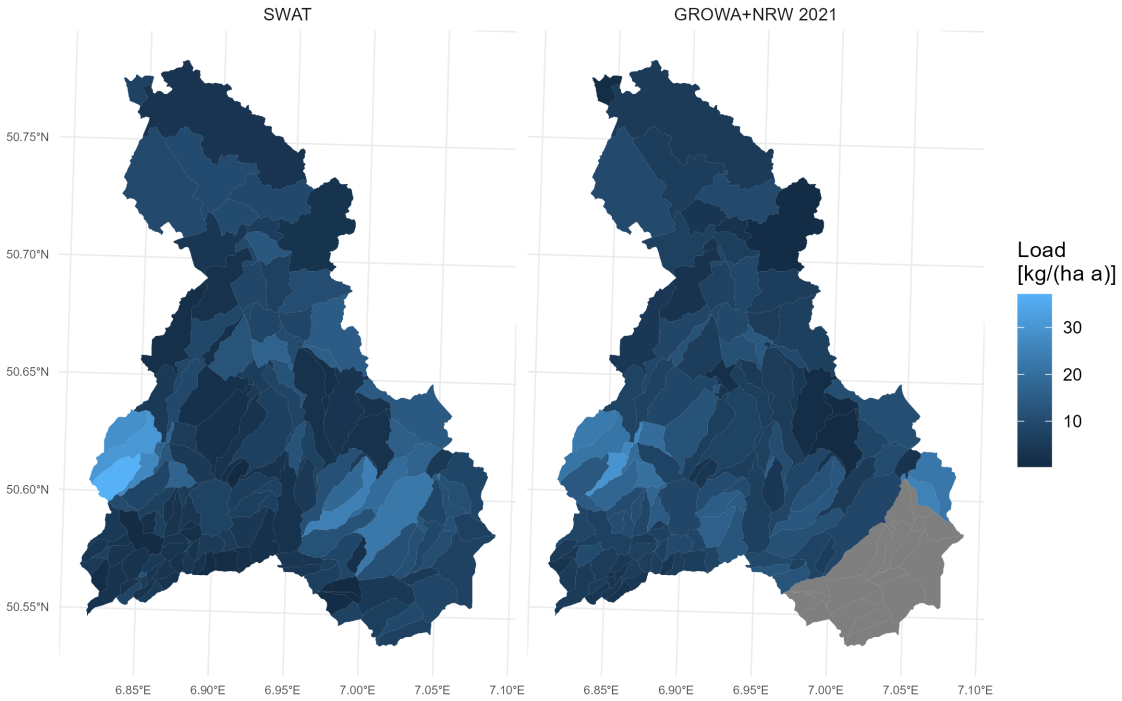


Figure 6.6.: Overall non-point source N emission loads for the SWAT subbasins in the Swist catchment from the behavioral SWAT models (2012-2018) and GROWA+NRW 2021 raster data (2014-2016). The map on the left hand-side depicts subbasin emissions according to the SWAT simulations and corresponds exactly to the map in the bottom right panel of Figure 6.5, albeit with a rescaled color-coding for the emission loads. The map on the right-hand side depicts subbasin emissions that result from aggregating N loads transported to surface waters from the GROWA+NRW 2021 raster data, excluding groundwater emissions. As the GROWA+NRW 2021 raster data does not extend beyond North Rhine-Westphalia, it provides no information for the SWAT subbasins located in Rhineland-Palatinate (i.e. the grey area in the southeast of the catchment).

ful differences between the individual subbasins, data aggregation excludes the groundwater pathway, only considering N emissions from surface runoff, erosion, interflow and tile flow. As evident from the higher contrast in the colors on the left-hand side of Figure 6.6, the non-point source N emission loads simulated by SWAT clearly exhibit a higher degree of spatial heterogeneity than the GROWA+NRW 2021 results. Both data sets also differ in the exact location of many high-input areas, especially when looking at the region along the upper to middle reaches of the Swist in the southeastern part of the catchment (excluding the subbasins located in Rhineland-Palatinate, for which there are no GROWA+NRW 2021 results). Apart from these differences, both SWAT and the AGRUM models behind the GROWA+NRW 2021 data agree on the high N inputs near the Schiessbach headwaters, where tile-drained cropland areas co-

incide with high simulated water yields (data not shown), providing additional evidence that this region may be a critical source area for nitrogen in the Swist catchment.

In general, the results presented in Figure 6.6 are consistent with other studies that compare CSA predictions from different models: Although there often is a certain amount of overlap in CSAs, models tend to disagree on a large part of CSA locations, indicating a high degree of cross-model uncertainty when evaluating the spatial variability of simulated pollutant emission loads. For example, Evenson et al. (2021) found that for their two different multi-model ensembles (including various SWAT models) of the Maumee catchment in the Great Lakes region of the US, only 16 % and 27 % of CSA candidates were identified as such by more than one model. Of course, areas for which different models do agree on elevated pollutant loads - such as the Schiessbach headwaters in the present study - may *vice versa* be flagged as potential CSAs with greater confidence. For the Schiessbach headwaters, this is further substantiated by the Erftverband water quality monitoring data in Table 5.3, showing that observed concentrations of NO_3^--N and organic N tend to be higher in the Schiessbach than the Swist main watercourse right before their confluence. In addition, the Schiessbach is one of the water bodies in the Swist catchment that has failed to meet quality requirements for NO_3^- according to WFD monitoring conducted by the NRW state authorities between 2009 and 2011 (State Ministry for Climate Protection, Environment, Agriculture, Nature Conservation and Consumer Protection NRW 2015).

Figure 6.7 shows the cumulative load curves for non-point source N emissions according to the different data sets. The more the load curves deviate from the identity line between catchment area fraction on the x-axis and the corresponding load contribution to the catchment-wide N emissions on the y-axis, the more pronounced are the differences in the non-point source emission loads from the individual subbasins. According to the cumulative load curve for the SWAT results averaged over all 11 behavioral model runs, the leading 10 % (50 %) of the Swist catchment area contribute 29 % (79 %) of all non-point source N emissions. Taking into account parameter-based model uncertainty (see the light black load curves for the individual SWAT model runs in Figure 6.7), the corresponding load contributions vary between approximately 24 % (72 %) and 35 % (87 %) of all N emissions. The load contributions for the GROWA+NRW 2021 data set correspond almost exactly to the lower end of this uncertainty range for the SWAT simulations: In large parts, the light blue load curve for the GROWA+NRW 2021 data in Figure 6.7 is practically identical to the load curve for the outlier model run 5221 that defines the minimum values of the SWAT uncertainty interval given above. Although contributing by far the highest absolute non-point source N emission loads of all behavioral SWAT model runs (see Table 6.3), the extraordinarily high interflow emission

loads override much of the spatial heterogeneity across the Swist catchment area for model run 5221 (data not shown). In general, all load curves presented in Figure 6.7 are comparable to other N load curves in the literature, for example from Niraula et al. (2012).

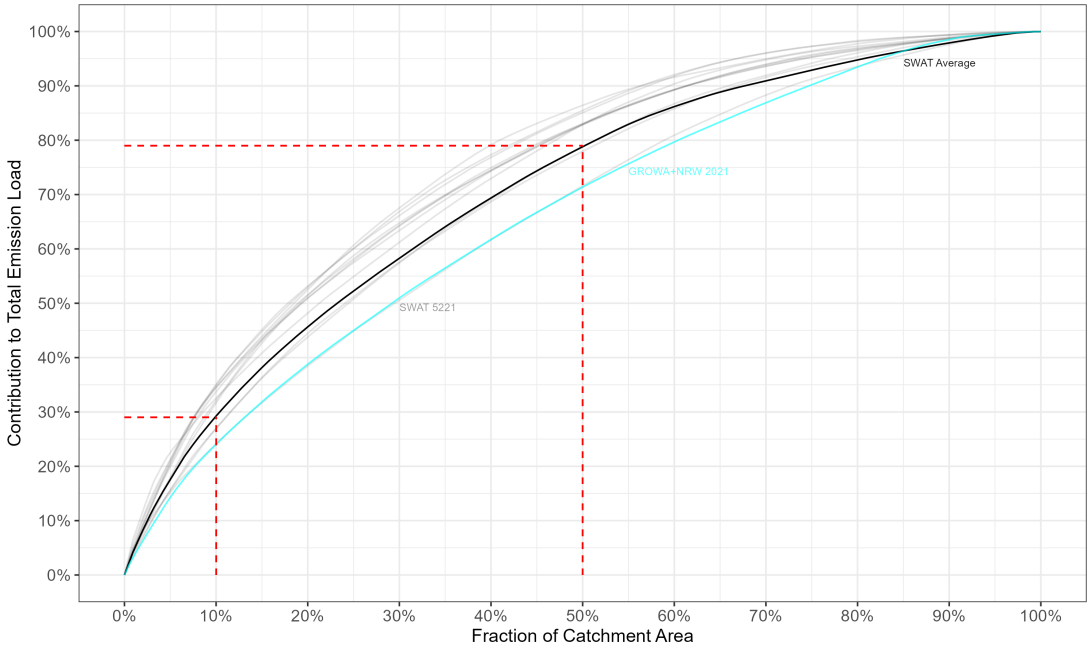


Figure 6.7.: Cumulative nitrogen emission load curves for the Swist catchment during the study period (2012-2018) as simulated by the 11 behavioral SWAT models. The curves are based on non-point source N emission loads from the individual subbasins. The light black lines represent the cumulative load curves for each individual SWAT model run. The bold black line represents the cumulative load curve that results from averaging subbasin emission loads over all SWAT model runs (see the map on the left hand-side of Figure 6.6). The light blue line indicates the corresponding cumulative load curve for the GROWA+NRW 2021 data excluding groundwater emissions (note that it partly overlaps with the load curve for the SWAT outlier model run 5221). Keep in mind that the GROWA+NRW 2021 data represents a shorter time period (2014-2016) and does not include the model subbasins located in Rhineland-Palatinate (see the map on the right hand-side of Figure 6.6).

6.4. Concluding Discussion

Taken together, the behavioral SWAT model results presented in this chapter suggest a plausible range of non-point source N emissions in the Swist catchment somewhere between 4.7 and 11.4 kg N/(ha a) during the study period (2012-2018), corresponding to approximately 136 and 330 t/a for the entire catchment area. A single SWAT model that simulated a considerably higher emission load of 15.9 kg N/(ha a) (460 t/a) can probably be classified as an outlier. The average simulated non-point source emission load across all behavioral SWAT model runs amounts to 8.1 kg N/(ha a), corresponding to 235 t/a for the entire study area. The median, which is less affected by the presumptive outlier value mentioned above, is 6.9 kg N/(ha a) (200 t/a). When comparing these simulation results to literature data, note that they represent total N emission loads for the entire Swist catchment and are not restricted to agricultural areas. Comparison of the simulation results to other studies can probably be facilitated by HRU-level evaluation of the SWAT emission loads to distinguish between different land uses and possibly even field crops. However, this was beyond the primary scope of this thesis.

Since most of the behavioral SWAT models underestimate in-stream NO_3^--N loads near the catchment outlet (see PBIAS scores for nitrate in Table 5.2), this may translate to biased N emissions in the simulations, suggesting that the above-stated range of emission loads as a whole may also be too low. This is not implausible given the potential underestimation of agricultural fertilization and overestimation of soil-borne denitrification in the SWAT models, as discussed in Subsection 6.3.1. However, it is unclear how these uncertainties interact with possible model errors in the SWAT routing phase *after* nitrogen has entered the stream network, especially considering the wide spread in the simulated in-stream N retention rates (see Subsection 6.3.1). This is probably best exemplified by SWAT model run 2868, which is one of the few behavioral runs that manages to produce nearly unbiased in-stream NO_3^--N loads although the corresponding catchment-wide emission load of 6.3 kg N/(ha a) (182 t/ha) is in the lower half of values simulated by all included SWAT models.

The uncertainty in the simulation results is quite large given that the upper bound of the range of plausible SWAT emission estimates is more than twice as high as the corresponding lower bound and more than three times as high when also considering the presumed outlier result of 15.9 kg N/(ha a) (for model run 5221). As brought up in Subsection 6.3.2, it may be possible to make a difference between the individual SWAT models based on their performance during calibration and validation and possibly further criteria, thereby establishing a ranking of emission load estimates. Such a ranking of different behavioral models for the purpose of uncertainty assessment is a feature of the GLUE method, for which Beven & Binley (1992) suggest to use

rescaled objective function scores as likelihood weights to derive a distribution function for the simulation results. However, performance advantages of any one model (such as SWAT model 2868) are always conditional on the model setup, which has not been part of the calibration process in this thesis. The performance advantages may vanish or even be reversed in a more rigorous calibration/validation design involving more and other kinds of observed data (or, for that matter, other objective functions to rate model performance, see Figure 5.4). Hence, the results of all behavioral SWAT models are treated as equally plausible in this thesis, with the possible exception of the outlier model run 5221. As discussed in Subsection 6.3.2, the most promising way forward to reduce the uncertainty in the simulation results of this study is the refinement of the model setup and the inclusion of additional calibration objectives to better constrain the model parameters.

In this respect, keep in mind that the simulated range of N emission loads exclusively represents parameter-based uncertainty. In particular, it does not include the effects of structural model uncertainty, which in many cases may be the most dominant source of modeling uncertainty (van Griensven & Meixner 2006). Consequently, the spread in viable N emission estimates may be even greater when also considering the results of additional models. Indeed, the AGRUM models applied in the GROWA+NRW 2021 project predict a non-point source N emission load for the Swist catchment that is clearly above the upper end of the parameter-based SWAT uncertainty interval, with a more than twofold difference to the average SWAT result of 8.1 kg N/(ha a) (see Table 6.3). However, this difference can mainly be attributed to excessive N contributions from groundwater in the GROWA+NRW 2021 results, most likely due to an insufficient consideration of the disconnect between groundwater and surface waters in large parts of the Swist catchment and thus a structural error in the operational setup of the AGRUM model chain employed in North Rhine-Westphalia (see Subsection 6.3.2). When omitting groundwater-borne N emissions to surface waters in the GROWA+NRW 2021 data, the remaining N load from non-point sources corresponds exactly to the average SWAT estimate.

In general, the parameter-based SWAT uncertainty interval of 4.7 and 11.4 kg N/(ha a) is considered to be a realistic appraisal of non-point source N emissions in the Swist catchment during the study period. This assessment is supported by…

1. …the overall consistency of the SWAT simulation results with the (corrected) non-point source N load estimates from the two reference data sets for the Swist catchment.

2. …the plausible N emission characteristics provided by the SWAT simulations, particularly regarding the seasonal and interannual variability of non-point source N loads as well as their spatial heterogeneity.

3. ...the good agreement of the *relative* N contributions from WWTPs with independent literature data (Fuchs et al. 2010, Mertens et al. 2012, Wendland et al. 2021a) as well as UBA estimates for all of Germany.

Due to the empirical nature of the SCS curve number technique and the inability of the HRU approach to represent pollutant mobilization and transport in the landscape from source areas to water bodies (see Subsection 3.1.3), it is not entire clear whether the SWAT simulations accurately capture the relative N contributions from the individual non-point source emission pathways. Notwithstanding this, the diminished role of the groundwater pathway in the simulations is in line with the hydrogeological features of the Swist catchment. Similarly, the high simulated tile flow N loads are not unexpected, given the high mobility of NO_3^--N in soils and the widespread agricultural tile drains in the Swist catchment. Tile drains probably play a key role for the occurence of CSAs in the catchment according to the SWAT results, although spatial variability of non-point source N pollution may also be affected by other aspects of agricultural management which are however considered uniform across the entire catchment area in the applied SWAT model setup (see Subsection 4.2.3). For a refinement of the source apportionment study presented in this chapter, it would be desirable to explicitly add event-specific point sources such as seperate sewer outlets (SSO) and combined sewer overflows (CSO) to the employed SWAT model. The abundance of potential discharge sites (more than 50 only along the Swist main watercourse according to Erftverband information) and lack of available data make this a challenging task, although approaches like those developed by Christoffels (2008), Mailhot et al. (2015) or Quaranta et al. (2022) may show a way forward to close data gaps regarding discontinuous point source emissions.

7. General Conclusions and Outlook

7.1. Introductory Remarks

Ecohydrological catchment modeling with SWAT in this thesis revealed a series of novel insights into surface water pollution with nitrogen (N) in the Swist river basin. These include the identification of likely critical source areas in the catchment and the determination of dominant emission pathways for N release to surface waters, both of which may become valuable information when it comes to devising emission control strategies and implementing effective mitigation measures (some Swist tributaries have been repeatedly identified as water bodies with elevated nitrate levels in the past). At the same time, the SWAT ensemble results and reference data sets highlight the considerable uncertainties associated with model-based assessment of emission loads and source apportionment, while also showing opportunities to reduce these uncertainties. In this regard, the present thesis demonstrates the crucial role of available monitoring data and a sound understanding of the local conditions in a given catchment along with its most distinctive features (hydrological and other).

This chapter provides an overview of the main research results in the present thesis and discusses them in an overarching manner. For this purpose, Section 7.2 addresses the individual research questions stated in Chapter 1, focusing first on the *similarities* in the different N emission estimates to identify the most substantiated findings on diffuse N pollution in the Swist catchment during the study period (Subsection 7.2.1). Thereafter, the *discrepancies* in the modeling results are discussed to analyze the associated uncertainties (Subsection 7.2.2), before the various approaches for source apportionment brought up in this thesis (*edge-of-field* monitoring, static modeling with AGRUM and dynamic modeling with SWAT) are compared to find the *origins* of the discrepancies in the results and reach a general conclusion about the relative strengths and weaknesses of the different approaches. This also entails an assessment of their suitability and efficiency in deriving reliable estimates of diffuse nitrogen pollution (Subsection 7.2.3). In closing of this thesis, Section 7.3 gives an outlook on open research needs and perspectives as well as possible improvements for a future continuation of model-based source apportionment with SWAT regarding N pollution in the Swist catchment.

7.2. General Conclusions

7.2.1. Nitrogen Pollution in the Swist Catchment

Research Question 1

How high are the (average) annual non-point source nitrogen emission loads in the Swist catchment during the study period, and what are the contributions of the different emission pathways?

On average, the 11 behavioral SWAT simulations evaluated in this thesis predict a non-point source N emission load of 8.1 kg/(ha a) (or 235 t/a) for the Swist catchment from 2012 to 2018. The corresponding median, which is less affected by a potential outlier result, is 6.9 kg N/(ha a) (or 200 t N/a). Although these results are associated with a considerable degree of uncertainty (see Subsection 7.2.2), their general order of magnitude is consistent with the two available reference data sets: An emission inventory for the years 2005 to 2009 derived from *edge-of-field* monitoring data (Kistemann et al. 2007, Mertens et al. 2012) and the static AGRUM modeling results from the GROWA+NRW 2021 project representing the time span from 2014 to 2016 (Wendland et al. 2021a). The order of magnitude of the SWAT estimates is also roughly consistent with the outcomes of further studies on diffuse N pollution in other surface water catchments, although the results are often difficult to compare to one another due to differences in land use distribution and agricultural management between the investigated catchments as well as differing choices in study design like spatiotemporal scale or evaluated N parameters (e.g. NO_3^--N versus total N). Looking at the temporal variability of diffuse N emissions, the SWAT simulation results show pronounced but generally plausible interannual and seasonal differences that are in line with prior expectations.

Considering only the average simulation results, the 11 SWAT models identify tile flow as the most important diffuse emission pathway for nitrogen in the Swist catchment (3.7 kg N/(ha a)), followed by interflow (2.7 kg N/(ha a)) and surface runoff (1.7 kg N/(ha a)), although this ranking is different for some of the individual model runs. The models unanimously agree however on the diminished role of groundwater flow, occupying the last place among the different non-point source emission pathways for nitrogen in every behavioral SWAT model run evaluated for this thesis (the average groundwater contribution across all model runs is very close to 0 kg N/(ha a)). This is consistent with the large-scale groundwater extraction in the Rhenish lignite mining region and the resulting disconnect between groundwater and many surface waters in the study area (see Chapter 2), effectively shutting down groundwater flow as a nitrogen emission pathway particularly in the lower parts of the catchment where most agricultural cropland with intensive fertilizer use is situated (see the map in Figure 2.2). The combination of widespread tile drains and impeded groundwater flow seems to be a special characteristic

of the Swist catchment (and probably also the surrounding river catchments) with regard to N pollution, shifting the individual load contributions from slow (groundwater flow) to fast (tile flow) emission pathways. This in turn reduces the residence times of reactive N in underground environmental compartments and decreases the potential for denitrification to alleviate the overall N burden on nearby surface waters (Seitzinger et al. 2006).

The prominent role of tile flow as an N emission pathway in the SWAT simulations is completely plausible given the high mobility of NO_3^--N in soils due to anion exclusion at the interface between soil solution and soil matrix. Agricultural tile drains are often indispensable to maintain farmland productivity. However, tile drains pose a challenge for the successful implementation of mitigation measures, as they circumvent classical nutrient barriers like riparian buffer zones and filter strips. Of course, further optimization of fertilization practices may help to reduce overall non-point source nutrient emissions, but stricter requirements are already in place since the latest amendments to the German Fertilizer Application Ordinance in 2017 and 2020. Direct treatment of tile flow at or near the drainage outlets for example in reactive ditches using wood chips can be effective in reducing N emissions to surface waters (Pfannerstill et al. 2016, Kothe et al. 2021). However, implementation of such mitigation measures is dependent on local conditions and must probably be put into practice on a site-by-site basis (Kothe et al. 2021). A promising and more generally applicable alternative are controlled tile drainage systems that enable the storage of soil water during dry periods and thus prolong residence times in the soil profile, potentially reducing the exports of nutrients and other pollutants to adjacent surface waters while also enhancing cropland productivity (Sunohara et al. 2016).

Research Question 2

How important are point source nitrogen emissions from wastewater treatment plants in the Swist catchment compared to non-point source nitrogen emissions?

When combining the simulated non-point source N emission loads for the Swist catchment with point source N emission loads from the four wastewater treatment plants (WWTP) in the study area, SWAT predicts a WWTP contribution to overall N pollution of 26.8 % on average (from 2012 to 2018). This value agrees quite well with independent literature data: For the Swist catchment from 2005 to 2009, Mertens et al. (2012) state a WWTP contribution to overall N pollution of 23 %. For the years from 2014 to 2016, Wendland et al. (2021a) estimate the WWTP contribution to overall N pollution in all of North Rhine-Westphalia (NRW) to be approximately 16 %. And finally, official estimates from the German Federal Environment Agency (UBA) derived with the static MoRE modeling system (Fuchs et al. 2017) assume an average WWTP contribution to overall N pollution of 17 % for all of Germany from 2012 to 2016. The

elevated WWTP contributions for the Swist catchment (compared to the average estimates for all of NRW and Germany) most likely orginates from the high municipal wastewater load in the Swist (approximately 40 % of annual streamflow at the catchment outlet, Christoffels et al. (2016)). There is probably still potential for lowering the N emission loads from wastewater treatment plants. However, note that the authorative quality requirements for wastewater discharged into the Swist and its tributaries are already comparatively high (Erftverband personal communication) and long-term efforts to improve wastewater treatment have generally led to a steady reduction of WWTP contribution to surface water pollution in Germany over past decades that tended to outpace emission reductions from other source types (Fuchs et al. 2010).

Research Question 3

Which spatial patterns of non-point source nitrogen emissions in the Swist catchment emerge in the modeling results? In particular, do the results indicate the existence of critical source areas for diffuse nitrogen pollution in the Swist catchment, and where are they located?

At the level of the subbasins delineated for the modeling project in this thesis, the SWAT simulation results show a considerable degree of spatial heterogeneity in the Swist catchment (from 2012 to 2018). On average, the leading 10 % of the catchment area contribute 29 % of all non-point source N emissions. Correspondingly, the leading 50 % of the catchment area contribute 79 % of all non-point source N emissions. There is a considerable degree of correlation between total and tile flow emission loads from the individual subbasins, with high emission areas often co-inciding with subbasins dominated by tile-drained cropland or orchards. This underscores the important role of agricultural tile drains as pathways for surface water N pollution in the Swist catchment according to the SWAT simulations. All behavioral SWAT models and the GROWA+NRW 2021 distributed modeling results agree on particularly high non-point source N emission loads near the Schiessbach headwaters, making this region a likely critical source area (CSA) for nitrogen pollution in the Swist catchment. This finding is further substantiated by routine water quality monitoring data from the Schiessbach: Erftverband monitoring results for the study period reveal elevated nitrogen concentrations in the lower reaches of the Schiessbach compared to the Swist main watercourse (Table 5.3). Furthermore, NRW state authorities identified the Schiessbach as one of the water bodies in the Swist catchment that failed to meet the requirements for the good chemical status based on NO_3^--N concentrations observed during the Water Framework Directive (WFD) monitoring cycle from 2009 to 2011 (State Ministry for Climate Protection, Environment, Agriculture, Nature Conservation and Consumer Protection NRW 2015).

7.2.2. Remaining Uncertainties

Research Question 4

How wide is the range of simulated non-point source nitrogen emissions in the Swist catchment between different calibrated SWAT models (i.e. how large is the parameter-based model uncertainty)?

In the study period from 2012 to 2018, the behavioral SWAT models predict a range of total non-point source N emission loads in the Swist catchment between 4.7 and 15.9 kg/(ha a) (or 136 and 460 t/a), amounting to a more than threefold difference between the minimum and maximum simulation result. When excluding the potential outlier model run that produces the value at the upper end of this interval, the second highest result of 11.4 kg N/(ha a) (or 330 t/a) is still more than twice as high as the lowest value, indicating a considerable degree of uncertainty associated with the SWAT simulations. There is also a substantial amount of variability in the pathway-specific emission loads, with a particularly large difference between the simulated minimum and maximum for nitrogen inputs via interflow (0.3 to 9.1 kg N/(ha a)). The spread in the simulated total non-point source emission loads translates directly to variable N contributions from WWTPs that range from 14.7 % to 36.7 % of overall nitrogen pollution. With regard to the spatial variability of non-point source emissions, the N contributions from the leading 10 % of the catchment area to overall non-point source pollution vary between 24 % and 35 % (72 % and 87 % for the leading 50 % of the catchment area). In this respect, the region along the upper and middle Swist reaches near Meckenheim emerges as the only obvious CSA candidate in the catchment apart from the Schiessbach headwaters.

As the differences between the behavioral SWAT models are restricted to the parameter values, all the described variability between the individual simulations can exclusively be attributed to parameter-based model uncertainty. The behavioral parameter sets are the product of a model calibration (and validation) process (see Chapter 5) in which the individual parameter values may compensate for each other as well as deficiencies in the model setup (i.e. model structure and input data) to meet the in-stream calibration objectives. The calibrated SWAT models are therefore subject to equifinality (Beven 2006), making it practically impossible to distinguish them in terms of their ability to accurately reproduce surface water pollution in the Swist catchment: From the perspective of the calibration objectives, they all may be similarly good predictors of diffuse nitrogen emission loads, with little potential to further narrow the above-stated uncertainty interval from 4.7 to 11.4 kg N/(ha a). Furthermore, parameter calibration is conditioned on the model setup as described in Chapter 4, which in several ways seems to promote a shortage of reactive N in the Swist catchment (see the analysis of simulated N balances in Subsection 6.3.1). Accordingly, a majority of calibrated SWAT models tend to un-

derestimate in-stream NO_3^--N loads at the catchment outlet that - if not overly affected by excessive N retention during the SWAT routing phase - may translate to similar underestimation of N emission loads as they enter the Swist river network. Section 7.3 will go into more detail on potential improvements in the model setup that may help to reduce this potential source of bias in the simulation results.

Research Question 5

How much do the simulated non-point source N emissions as well as critical source areas differ between SWAT and AGRUM (i.e. how large is the structural model uncertainty)?

The AGRUM modeling results in the mGROWA+NRW 2021 project (Wendland et al. 2021a) yield an overall estimate of non-point source N emissions for the Swist catchment area in NRW of 16.5 kg/(ha a) (2014-2016). This amounts to a more than twofold difference compared to the average simulation result from the behavioral SWAT models, while also being outside the parameter-based uncertainty interval defined by the lowest and highest SWAT results (see above). The discrepancy between the AGRUM and SWAT results suggests a substantial amount of structural model uncertainty. However, the surplus in N emissions predicted by the AGRUM results can almost completely be attributed to N emissions from groundwater, which constitute more than half of the entire diffuse N surface water burden according to the GROWA+NRW 2021 data. Such high groundwater contributions are implausible given the widespread disconnect between groundwater and surface waters in the Swist catchment due to groundwater extraction near the Rhenish lignite mining pits. Although the GROWA+NRW 2021 results fit observed in-stream nitrogen loads well for most of NRW, validation fails for the Erft catchment most likely due to human interference with the regional hydrogeology (Wendland et al. 2021a), suggesting an error in the operational setup of the AGRUM model chain as employed in the GROWA+NRW 2021 project for the Erft catchment and its surroundings (including the Swist catchment).

Subtracting the groundwater N emissions from the overall non-point source N emissions in the GROWA+NRW 2021 data set produces a remaining N load that corresponds exactly to the average SWAT estimate (8.1 kg N/(ha a)). Hence, the order of magnitude of the SWAT results is generally consistent with the (corrected) GROWA+NRW 2021 data. As the same is true regarding the SWAT results and the monitoring-based emission inventory of Mertens et al. (2012), it is concluded in this thesis that the parameter-based uncertainty interval of 4.7 to 11.4 kg N/(ha a) probably is a realistic appraisal of non-point N pollution in the Swist catchment during the study period (excluding the potential outlier result at 15.9 kg N/(ha a)). Alternatively, the range of likely non-point loads may also be stated as being in the vicinity of *approximately* 5 to 10 kg N/(ha a) to not purport a level of precision that the applied SWAT models are unable to

provide. Looking at the three non-point source emission pathways other than groundwater, the GROWA+NRW 2021 results reproduce the same ranking as the average SWAT results, although with a substantially lower estimate for nitrogen emissions via surface runoff. This is a consequence of the greater restrictions imposed on nutrient emissions via surface runoff in the distributed AGRUM models, among them a requirement for potential source areas to be hydrologically connected to nearby surface water bodies (Wendland et al. 2021a), standing in stark contrast to the HRU approach in SWAT that allows pollutant transfer to the river network from any point in the landscape.

Regarding the spatial variability of the non-point source N emissions in the Swist catchment, the load estimates obtained from the GROWA+NRW 2021 raster data (without N contributions from groundwater) are more uniform between the different subbasins than the corresponding estimates from the behavioral SWAT simulations. The relative nitrogen contributions from the leading 10 % and 50 % of the catchment area in the GROWA+NRW 2021 data correspond almost exactly to the lower end of the parameter-based uncertainty interval from the SWAT simulation results (approximately 25 % and 72 %), indicating a lower degree of spatial heterogeneity than most SWAT model runs. With respect to critical source areas, there is no clear overlap in high emission subbasins between the SWAT and GROWA+NRW 2021 results apart from the Schiessbach headwater region. This confirms the findings of Evenson et al. (2021) on cross-model uncertainty in CSA identification: On the one hand, different models usually tend to disagree on a majority of potential CSA locations. On the other hand, when different models do agree on certain areas with elevated pollutant loads - such as the region along the upper reaches of the Schiessbach - these can subsequently be targeted with all the more confidence in the case of focused monitoring campaigns or the implementation of mitigation measures.

7.2.3. Strengths and Weaknesses of Source Apportionment Approaches

Research Question 6

What are the relative strengths and weaknesses of the different approaches to assess catchment-wide non-point source pollution (edge-of-field monitoring, static AGRUM modeling or dynamic SWAT modeling)? Which level of effort is necessary and feasible to draw reliable conclusions particularly on nitrogen pollution?

Comparing the non-point source emission estimates for nitrogen in the Swist catchment from *edge-of-field* (EOF) monitoring (Mertens et al. 2012), static AGRUM modeling (Wendland et al. 2021a) and dynamic SWAT modeling in this study, it is reassuring to see that all approaches - with some exceptions - agree on the general order of magnitude of the predicted emission

loads and the overall ranking of emission pathways, an outcome not necessarily expected beforehand. However, going beyond these more qualitative aspects, many quantitative results are associated with considerable uncertainties, as exemplified by the relatively wide range of total and pathway-specific load estimates predicted by the various SWAT models as well as some major discrepancies between the average SWAT modeling results, the AGRUM modeling results and the results from EOF monitoring. Most notably, these discrepancies include the excessively high nitrogen emissions from groundwater flow in the AGRUM data, probably originating from insufficient consideration of the mining-induced water table draw down affecting a large portion of the Swist catchment. Here, although successfully validated for the vast majority of higher-order stream catchments in NRW for the GROWA+NRW 2021 project (Wendland et al. 2021a), the AGRUM model chain fails when confronted with exceptional local conditions the model structure is unable to accommodate.

On the one hand, the overestimation of groundwater N contributions in the Swist catchment by the AGRUM models highlights the need for appropriate consideration of local conditions in different (sub-)catchments when assessing surface water pollution. This of course is difficult for a large-scale, "*one-size-fits-all*" approach as applied in the GROWA+NRW 2021 project for the entire state area of NRW, and can be more naturally realized with a catchment-scale model like SWAT. On the other hand, AGRUM is a more parsimonious modeling sytem than the complex and potentially overparameterized SWAT. As argued by Radcliffe et al. (2009) in the context of phosphorus modeling, the use of parsimonious models may be beneficial in many cases as they can be more readily applied and sidestep the typical problems arising from model calibration as well as parameter equifinality. At the same time, the lack of formal parameter calibration in many static models may reduce transparency about persisting uncertainties (originating from the input data, the model structure or scale issues) and suggest an unwarranted level of confidence in the model predictions. Again, quantitative as well as qualitative local knowledge about the study area ("*hard*" and "*soft*" data, Seibert & McDonnell (2002), Arnold et al. (2015)) may be used to better constrain the viable parameter ranges in complex calibrated catchment models like SWAT and validate the simulation results. In this respect, the wealth of data available for the Swist catchment (see Section 2.5) has been extraordinarily helpful for nitrogen source apportionment in this thesis, as has been intercomparison of multiple independent reference data sets on catchment-wide N pollution to verify the most common findings and evaluate the associated uncertainties. Combination of various methods that complement each other may in fact be the most promising approach to draw reliable conclusions on surface water pollution, but will presumably be difficult in many catchments due to data constraints.

In general, the method of choice for catchment-wide assessment of non-point source pollution will depend on the specific research context and motivation. If the goal is to estimate annual emission loads, a static modeling system like AGRUM may suffice, whereas the determination of seasonal or event-specific pollutant loads requires a dynamic model like SWAT. *Edge-of-field* monitoring alone is insufficient for the identification of critical source areas on the catchment scale, but may be helpful or even mandatory for the correct interpretation of distributed or semi-distributed modeling results for this purpose. In order to fulfill legal reporting obligations for large-scale river catchments or state areas (like those imposed by the WFD on the EU member states), a static and relatively comprehensible modeling system such as AGRUM is probably the most appropriate tool. In contrast, planning of effective mitigation measures to reduce pollutant emissions in a specific catchment requires targeted monitoring efforts and/or the application of more complex, dynamic and process-based computer models. Additionally, the pollutant of interest may also play a role in model selection: For example, the semi-distributed HRU approach in SWAT to represent spatial heterogeneity may be better suited to assess nitrogen than phosphorus pollution. Whereas phosphorus mobilization is often dominated by erosive processes in the small-scale topography of a subbasin, diffuse nitrogen pollution tends to be governed by transport of dissolved NO_3^--N in groundwater flow or tile drainage systems, making it less dependent on natural terrain and surface runoff flow paths.

7.3. Outlook

In a future continuation of SWAT modeling in the Swist catchment to assess nitrogen pollution, there are several potential improvements to the SWAT models that could reduce uncertainty in the simulation results and give additional insights into the likely amounts and origins of reactive N entering the Swist river network. Here, model improvement mainly pertains to the model setup (i.e. the model input data and model structure) prior to parameter calibration and pursues two closely related objectives: First, it strives to increase the achievable model performance ratings, allowing the introduction of additional calibration objectives to better constrain the model parameters while at the same time maintaining (or increasing) the number of behavioral SWAT parameter sets available for uncertainty analysis after model calibration and validation. Second, it attempts to eliminate the shortage of reactive N in the models that likely contributes to underestimation of in-stream NO_3^--N loads and promotes mineralization as well as depletion of soil organic N stocks in the model calibration process, correcting the representation of internal catchment processes and counteracting potential bias in the simulated N emissions. In summary, improvement of the SWAT model setup aims to enhance both the

precision and *accuracy* of the simulated uncertainty intervals for non-point source N pollution in the Swist catchment.

Analysis of the SWAT simulation results in this thesis suggests an increase in N fertilizer quantities to be one of the most promising options for change in the model setup to reduce underestimation of in-stream NO_3^--N loads. Although the fertilizer quantities applied in the model (see Table 4.4) capture the order of magnitude of fertilization in the Swist catchment reasonably well (see Subsection 4.3.4), a certain degree of underestimation is not implausible especially since the fertilization rates are based on best practice recommendations by Paffrath et al. (2021) that take into account stricter legal requirements for fertilizer use in Germany not yet active during the study period. In line with that, an experimental increase of N fertilizer quantities by 20 % in the behavioral SWAT models almost completely eliminated the bias in the simulated in-stream NO_3^--N loads (see the top right panel in Figure 6.1), showing a mostly positive impact of increased fertilization rates on model performance. As an alternative, inputs of reactive N to the catchment soils may equally be raised by activating atmospheric deposition in the SWAT models. However, keep in mind that the effects of missing atmospheric N deposition in the simulations might be partially offset by likewise omission of ammonia volatilization, which should consequently also be included in the SWAT models when adding atmospheric deposition as an input process.

Instead of raising nitrogen inputs via fertilization and/or atmospheric deposition, supply with reactive N can also be increased by lowering nitrogen losses for example via denitrification. Despite high uncertainties in the quantification of catchment-scale denitrification, comparison of simulated denitrification losses (see Table 6.2) with literature data suggests that soil-borne denitrification in the SWAT models is presumably too high. A possible starting point to lower denitrification in the model results is the simulation of perched water tables on top of an impervious subsurface layer in SWAT: Restricting the occurrence of perched water tables only to places where they are necessary to simulate tile flow may help to better drain the soils in the rest of the catchment and thus reduce soil-borne denitrification. The same approach may simultaneously stimulate groundwater flow and better connect surface waters to preexisting N pools in the shallow aquifer, potentially alleviating simulated in-stream load deficits of NO_3^--N particularly during base flow conditions. For the stream segments in the lower parts of the catchment without groundwater contact, simulation of base flow and transmission losses would probably benefit from spatial differentiation in the hydraulic conductivities of the channel bed. In a simultaneous calibration scheme as applied in this thesis, conductivities for example may be set independently for the Swist and its tributaries to resolve trade-offs between the different stream gauges.

There are several other possible model improvements that could enhance performance ratings and refine catchment representation, most notably the explicit inclusion of discontinuous point sources for nitrogen emissions like combined sewer overflows (CSO) or separate sewer outlets (SSO) in the model structure. Although N emissions from these sources are generally considered to be of minor importance on an annual time scale (Fuchs et al. 2010, Mertens et al. 2012, Wendland et al. 2021a), they probably still constitute a meaningful part of the entire N burden in most stream catchments with substantial shares of developed residential or commercial areas. Moreover, with a dynamic model like SWAT, it may be possible to examine the short-term effects of single-event N emission loads from these sources on surface water quality, especially with regard to NH_4^+-N and NO_2^--N in basically untreated wastewater from CSO events. However, difficulties arise from the unpredictable nature of emission events and the abundance of potential discharge locations in a given river catchment. Therefore, despite growing awareness for the significance of CSO and SSO discharges regarding surface water quality and ongoing efforts to close persisting data gaps (Christoffels 2008, Mailhot et al. 2015, Quaranta et al. 2022), accounting for these emission sources in catchment-scale models will likely remain challenging in the foreseeable future.

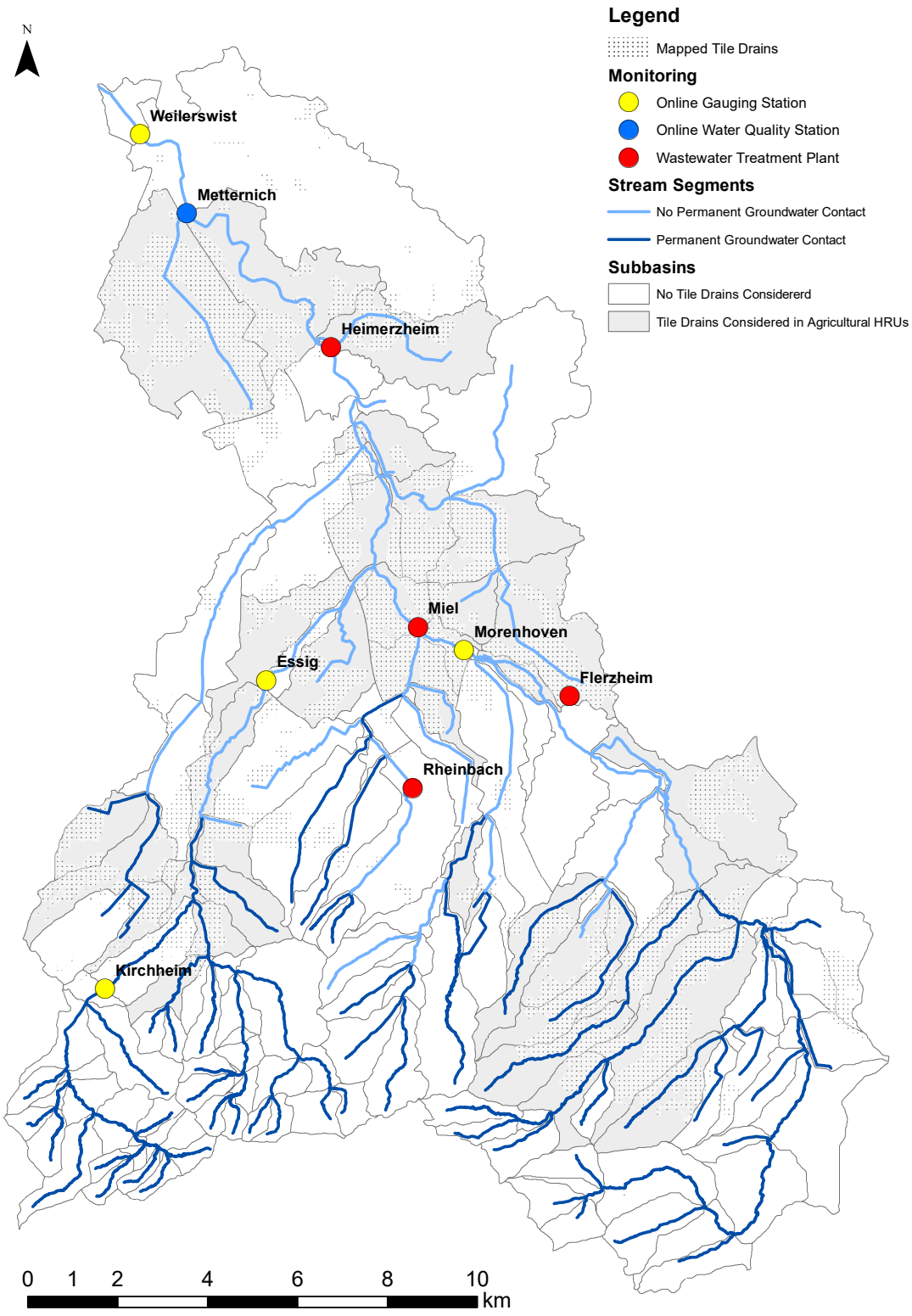
An intriguing research perspective would be to incorporate the effects of the latest amendments to the German Fertilizer Application Ordinance ("Düngeverordnung") into the SWAT models. These amendments impose stronger restrictions for agricultural fertilizer use in Germany and should therefore have led to a noticeable decrease in N fertilization since 2017. In their current state, the SWAT models assume all agricultural management practices to be constant over time, meaning there is no systematic shift in the applied management schedules during the simulation period (see Appendix C for the various schedules used to model crop rotation). In a combined effort to both raise the levels of reactive N in the model (see above) and account for the changed regulations, quantities of N fertilizer specified in the management schedules could be raised until the end of 2016 and be left at their current values afterwards. Extending the simulation period beyond the year 2018 would subsequently allow to examine how the change in fertilizer use affects long-term water quality in the simulations, with special emphasis on the occurrence of *legacy effects* that might cause a long delay in the associated response of in-stream nitrogen levels (Withers et al. 2014, Ascott et al. 2021, Basu et al. 2022). Van Meter et al. (2016) distinguish between a *biogeochemical* legacy of organic N stocks in soils and a *hydrological* legacy of inorganic N traveling through the vadose zone and aquifers. While SWAT should in principle be able to capture biogeochemical N legacies (see Figure 3.2), it is thought to be largely incapable of accounting for hydrological N legacies (Lutz et al. 2022, Golden et al. 2023). However, the latter may be of minor importance in the study area, if the premise is true that

the surface-groundwater disconnect and widespread tile drainage favor fast flow paths to the surface waters of the Swist catchment.

Finally, consideration of alternatives to the spatial model setup of SWAT may prove helpful to explore another facet of model uncertainty (scale dependencies) and build confidence in model-based identification of critical source areas. This includes the number and size of delineated subbasins and HRUs, both of which may have a considerable impact on the SWAT simulation results in particular regarding sediment and nutrients (Jha et al. 2004, Her et al. 2015). A refinement in the number and size of the model subbasins (see Appendix A) may become necessary in the future to accommodate additional monitoring points and discharge sites for discontinuous point sources like CSO and SSO along the network of stream segments. Investigating the effects of different land use and soil data sets on model performance in Chapter 4 resulted in diverging HRU configurations in SWAT (see Table 4.1) and thus already entails a basic uncertainty assessment in this regard. However, a more explicit consideration of HRU threshold values applied during model setup as in Her et al. (2015) may give a more comprehensive picture of the associated error potential. With regard to the inherent shortcomings of the HRU approach regarding pollutant transport in the landscape, transition from SWAT to its successor SWAT+ (Bieger et al. 2017) may be worthwhile for future modeling applications in the Swist catchment. SWAT+ keeps the HRU approach, but introduces *landscape units* (LSU) as an additional level of spatial differentiation in-between subbasins and HRUs enabling transfer of water and pollutants between them (Bieger et al. 2019). Alternatively, SWAT+ also provides an option for a fully distributed grid-based model setup adopted from Rathjens et al. (2015), which may however become computationally prohibitive in larger catchments and/or tied to MC-based model calibration.

A. SWAT Model Setup Map of the Swist River Basin

Figure A.1.: SWAT model setup map of the Swist river basin (see next page). The stream network was subdivided into 153 stream segments, with the same number of delineated subbasins. Stream segments were classified in terms of their groundwater contact by comparing 2018 groundwater levels with streambed elevations (Erftverband personal communication). Agricultural areas with tile drains were mapped by Kothe et al. (2021). The SWAT models simulated tile flow in all subbasins with at least 40 % tile drained agricultural area (depicted in grey).


N
Legend
Mapped Tile Drains
Monitoring
Online Gauging Station
Online Water Quality Station
Wastewater Treatment Plant
Stream Segments
No Permanent Groundwater Contact
Permanent Groundwater Contact
Subbasins
No Tile Drains Considererd
Tile Drains Considered in Agricultural HRUs
Weilerswist
Metternich
Heimerzheim
Miel
Morenhoven
Essig
Flerzheim
Rheinbach
Kirchheim
0 1 2 4 6 8 10
km

B. Description of SWAT Parameters

Table B.1.: Description of SWAT model parameters included in sensitivity analysis (see next page). Parameters are ordered according to their influence on simulation of in-stream nitrate loads (as in Table 5.1 on page 95). The horizontal line separates the influential parameters above from the non-influential parameters below. Parameter descriptions according to the SWAT Input/Output Documentation (Arnold et al. 2013).

Parameter	Description	Unit
TRNSRCH.bsn	Fraction of transmission losses from main channel that enters deep aquifer	-
CH_K2.rte	Effective hydraulic conductivity in main channel alluvium	mm H_2O/h
DEPIMP_BSN.bsn	Depth to impervious layer for modeling perched water tables	mm
ALPHA_BNK.rte	Recession constant for bank storage	-
SDNCO.bsn	Denitrification threshold soil water content	-
ANION_EXCL_BSN.bsn	Fraction of soil porosity from which anions are excluded	-
CN2.mgt	Initial SCS runoff curve number for moisture condition II	-
SOL_AWC.sol	Available water capacity of a specific soil layer	mm H_2O/mm soil
CDN.bsn	Denitrification exponential rate coefficient for the soil	-
NPERCO.bsn	Nitrate percolation coefficient for the soil	-
CNCOEF.bsn	Plant ET curve number coefficient	-
DIS_STREAM.hru	Average distance to the stream	m
SOL_BD.sol	Moist bulk density of a specific soil layer	g/cm^3
EPCO.bsn	Plant water uptake compensation factor	-
SOL_K.sol	Saturated hydraulic conductivity of a specific soil layer	mm H_2O/h
CMN.bsn	Rate factor for humus mineralization of active organic nutrients (N and P).	-
ESCO.bsn	Soil evaporation compensation factor	-
CANMX.hru	Maximum canopy storage	mm H_2O
SOL_CBN.sol	Organic carbon content of a specific soil layer	% (m/m)
TDRAIN.mgt	Time to drain soil to field capacity via tile drains	h
SOL_ALB.sol	Moist soil albedo	-
CH_N2.rte	Manning's roughness coefficient for the main channel	s/m$^{1/3}$
SURLAG.bsn	Surface runoff lag coefficient	-
SMTMP.bsn	Snow melt base temperature	°C
RCHRG_DP.gw	Deep aquifer percolation fraction	-
TIMP.bsn	Snow pack temperature lag factor	-
N_UPDIS.bsn	Plant nitrogen uptake distribution parameter	-
SFTMP.bsn	Snowfall temperature	°C
GWQMN.gw	Threshold storage in the shallow aquifer required for groundwater flow	mm H_2O
SNO50COV.bsn	Fraction of snow volume that corresponds to 50 % snow cover	-
SNOCOVMX.bsn	Minimum snow water content that corresponds to 100 % snow cover	mm H_2O
SMFMN.bsn	Melt factor for snow on December 21	mm H_2O/°C
GW_REVAP.gw	Groundwater "revap" coefficient	-
REVAPMN.gw	Threshold storage in the shallow aquifer for "revap"	mm H_2O
OV_N.hru	Manning's roughness coefficient for overland flow	s/m$^{1/3}$
SMFMX.bsn	Melt factor for snow on June 21	mm H_2O/°C
EVRCH.bsn	Reach evaporation adjustment factor	-
HLIFE_NGW_BSN.bsn	Half-life of nitrate in the shallow aquifer	d
DRAIN_CO_BSN.bsn	Daily drainage coefficient (maximum tile flow)	mm H_2O/d
GW_DELAY.gw	Groundwater flow delay time	day
CH_N1.sub	Manning's roughness coefficient for the tributary channels	s/m$^{1/3}$
RCN.bsn	Concentration of nitrogen in rainfall	mg N/l
ALPHA_BF.gw	Baseflow recession constant for the shallow aquifer	1/d
SHALLST_N.gw	Initial concentration of NO_3^--N in the shallow aquifer	mg N/l
RES_K.res	Hydraulic conductivity of the reservoir bottom	mm H_2O/h

C. Management Schedules and Crop Rotation Cycles

Table C.1.: Canola, wheat, peas, wheat, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 62: Peas, 75: Polish Canola, 76: Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i' for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
3	1	0	3	1			70	0				
4	1	0	3	1			50	0				
7	30	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
8	10	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	21	0	6	2			0					
4	1	0	1	62		0	1800	0	0	0	0	0
8	12	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
8	7	0	6	1			0					
8	25	0	1	76		0	1800	0	0	0	0	0
8	25	0	3	1			50	0				
			0									

Table C.2.: Canola, wheat, oats, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 32: Oats, 75: Polish Canola, 76: Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
3	1	0	3	1			70	0				
4	1	0	3	1			50	0				
7	30	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
8	12	0	6	1			0					
8	30	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	7	0	6	2			0					
3	21	0	1	32		0	1800	0	0	0	0	0
3	21	0	3	1			40	0				
4	21	0	3	1			40	0				
8	4	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
8	7	0	6	1			0					
8	25	0	1	76		0	1800	0	0	0	0	0
8	25	0	3	1			50	0				
			0									

Table C.3.: Canola, wheat, sugar beet, wheat, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 69: Sugar Beet, 75: Polish Canola, 76: Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
3	1	0	3	1			70	0				
4	1	0	3	1			50	0				
7	30	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
8	15	0	6	1			0					
8	29	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
4	8	0	1	69		0	1800	0	0	0	0	0
4	8	0	3	1			60	0				
4	30	0	3	1			50	0				
10	7	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
8	7	0	6	1			0					
8	25	0	1	76		0	1800	0	0	0	0	0
8	25	0	3	1			50	0				
			0									

Table C.4.: Canola, wheat, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat 76: Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
3	1	0	3	1			70	0				
4	1	0	3	1			50	0				
7	30	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
8	7	0	6	1			0					
8	25	0	1	76		0	1800	0	0	0	0	0
8	25	0	3	1			50	0				
			0									

Table C.5.: Canola, wheat, barley, corn, peas.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 19: Corn, 28: Wheat, 62: Peas, 75: Polish Canola, 76: Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
3	1	0	3	1			70	0				
4	1	0	3	1			50	0				
7	30	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
7	31	0	6	1			0					
8	15	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	30	0	6	2			0					
4	25	0	1	19		0	1800	0	0	0	0	0
4	25	0	3	1			70	0				
6	7	0	3	1			80	0				
10	26	0	5				0					
			0									
3	21	0	6	2			0					
4	1	0	1	62		0	1800	0	0	0	0	0
8	12	0	5				0					
8	19	0	6	1			0					
8	25	0	1	76		0	1800	0	0	0	0	0
8	25	0	3	1			50	0				
			0									

Table C.6.: Canola, wheat, wheat, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 76: Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
3	1	0	3	1			70	0				
4	1	0	3	1			50	0				
7	30	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
8	7	0	6	1			0					
8	25	0	1	76		0	1800	0	0	0	0	0
8	25	0	3	1			50	0				
			0									

Table C.7.: Corn silage, wheat, peas, wheat, rye.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 20: Corn Silage, 28: Wheat, 30: Rye, 62: Peas, 75: Polish Canola.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	6	2			0					
4	25	0	1	20		0	1800	0	0	0	0	0
4	25	0	3	1			70	0				
6	7	0	3	1			80	0				
9	20	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
8	10	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	21	0	6	2			0					
4	1	0	1	62		0	1800	0	0	0	0	0
8	12	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	5	0	1	30		0	1800	0	0	0	0	0
			0									
3	15	0	3	1			50	0				
4	15	0	3	1			50	0				
5	15	0	3	1			30	0				
7	25	0	5				0					
8	5	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.8.: Corn silage, wheat, peas, wheat, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 20: Corn Silage, 28: Wheat, 62: Peas, 75: Polish Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	6	2			0					
4	25	0	1	20		0	1800	0	0	0	0	0
4	25	0	3	1			70	0				
6	7	0	3	1			80	0				
9	20	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
8	10	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	21	0	6	2			0					
4	1	0	1	62		0	1800	0	0	0	0	0
8	12	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
7	31	0	6	1			0					
8	15	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.9.: Corn silage, wheat, oats, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 20: Corn Silage, 28: Wheat, 32: Oats, 75: Polish Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	6	2			0					
4	25	0	1	20		0	1800	0	0	0	0	0
4	25	0	3	1			70	0				
6	7	0	3	1			80	0				
9	20	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
8	12	0	6	1			0					
8	30	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	7	0	6	2			0					
3	21	0	1	32		0	1800	0	0	0	0	0
3	21	0	3	1			40	0				
4	21	0	3	1			40	0				
8	4	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
7	31	0	6	1			0					
8	15	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.10.: Corn silage, wheat, rye.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 20: Corn Silage, 28: Wheat, 30: Rye, 75: Polish Canola.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	6	2			0					
4	25	0	1	20		0	1800	0	0	0	0	0
4	25	0	3	1			70	0				
6	7	0	3	1			80	0				
9	20	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	5	0	1	30		0	1800	0	0	0	0	0
			0									
3	15	0	3	1			50	0				
4	15	0	3	1			50	0				
5	15	0	3	1			30	0				
7	25	0	5				0					
8	5	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.11.: Corn silage, wheat, rye, corn, peas.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 19: Corn, 20: Corn Silage, 28: Wheat, 30: Rye, 62: Peas, 75: Polish Canola.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	6	2			0					
4	25	0	1	20		0	1800	0	0	0	0	0
4	25	0	3	1			70	0				
6	7	0	3	1			80	0				
9	20	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	5	0	1	30		0	1800	0	0	0	0	0
			0									
3	15	0	3	1			50	0				
4	15	0	3	1			50	0				
5	15	0	3	1			30	0				
7	25	0	5				0					
8	5	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	30	0	6	2			0					
4	25	0	1	19		0	1800	0	0	0	0	0
4	25	0	3	1			70	0				
6	7	0	3	1			80	0				
10	26	0	5				0					
			0									
3	21	0	6	2			0					
4	1	0	1	62		0	1800	0	0	0	0	0
8	12	0	5				0					
8	21	0	6	1			0					
9	1	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.12.: Corn silage, wheat, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 20: Corn Silage, 28: Wheat, 75: Polish Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	6	2			0					
4	25	0	1	20		0	1800	0	0	0	0	0
4	25	0	3	1			70	0				
6	7	0	3	1			80	0				
9	20	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
7	31	0	6	1			0					
8	15	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.13.: Corn silage, wheat, barley, corn, peas.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 19: Corn, 20: Corn Silage, 28: Wheat, 62: Peas, 75: Polish Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	6	2			0					
4	25	0	1	20		0	1800	0	0	0	0	0
4	25	0	3	1			70	0				
6	7	0	3	1			80	0				
9	20	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
7	31	0	6	1			0					
8	15	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	30	0	6	2			0					
4	25	0	1	19		0	1800	0	0	0	0	0
4	25	0	3	1			70	0				
6	7	0	3	1			80	0				
10	26	0	5				0					
			0									
3	21	0	6	2			0					
4	1	0	1	62		0	1800	0	0	0	0	0
8	12	0	5				0					
8	21	0	6	1			0					
9	1	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.14.: Corn silage, wheat, wheat, rye.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 20: Corn Silage, 28: Wheat, 30: Rye, 75: Polish Canola.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	6	2			0					
4	25	0	1	20		0	1800	0	0	0	0	0
4	25	0	3	1			70	0				
6	7	0	3	1			80	0				
9	20	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	5	0	1	30		0	1800	0	0	0	0	0
			0									
3	15	0	3	1			50	0				
4	15	0	3	1			50	0				
5	15	0	3	1			30	0				
7	25	0	5				0					
8	5	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.15.: Corn silage, wheat, wheat, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 20: Corn Silage, 28: Wheat, 75: Polish Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	6	2			0					
4	25	0	1	20		0	1800	0	0	0	0	0
4	25	0	3	1			70	0				
6	7	0	3	1			80	0				
9	20	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
7	31	0	6	1			0					
8	15	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.16.: Potato, wheat, peas, wheat, rye.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 30: Rye, 62: Peas, 70: Potato, 75: Polish Canola.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	1	70		0	1800	0	0	0	0	0
3	30	0	3	1			70	0				
4	25	0	3	1			60	0				
8	10	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
8	10	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	21	0	6	2			0					
4	1	0	1	62		0	1800	0	0	0	0	0
8	12	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	5	0	1	30		0	1800	0	0	0	0	0
			0									
3	15	0	3	1			50	0				
4	15	0	3	1			50	0				
5	15	0	3	1			30	0				
7	25	0	5				0					
8	5	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.17.: Potato, wheat, peas, wheat, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 62: Peas, 70: Potato, 75: Polish Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	1	70		0	1800	0	0	0	0	0
3	30	0	3	1			70	0				
4	25	0	3	1			60	0				
8	10	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
8	10	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	21	0	6	2			0					
4	1	0	1	62		0	1800	0	0	0	0	0
8	12	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
7	31	0	6	1			0					
8	15	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.18.: Potato, wheat, oats, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 32: Oats, 70: Potato, 75: Polish Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	1	70		0	1800	0	0	0	0	0
3	30	0	3	1			70	0				
4	25	0	3	1			60	0				
8	10	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
8	12	0	6	1			0					
8	30	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	7	0	6	2			0					
3	21	0	1	32		0	1800	0	0	0	0	0
3	21	0	3	1			40	0				
4	21	0	3	1			40	0				
8	4	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
7	31	0	6	1			0					
8	15	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.19.: Potato, wheat, rye.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 30: Rye, 70: Potato, 75: Polish Canola.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	1	70		0	1800	0	0	0	0	0
3	30	0	3	1			70	0				
4	25	0	3	1			60	0				
8	10	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	5	0	1	30		0	1800	0	0	0	0	0
			0									
3	15	0	3	1			50	0				
4	15	0	3	1			50	0				
5	15	0	3	1			30	0				
7	25	0	5				0					
8	5	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.20.: Potato, wheat, rye, corn.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 19: Corn, 28: Wheat, 30: Rye, 70: Potato, 75: Polish Canola.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	1	70		0	1800	0	0	0	0	0
3	30	0	3	1			70	0				
4	25	0	3	1			60	0				
8	10	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	5	0	1	30		0	1800	0	0	0	0	0
			0									
3	15	0	3	1			50	0				
4	15	0	3	1			50	0				
5	15	0	3	1			30	0				
7	25	0	5				0					
8	5	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	30	0	6	2			0					
4	25	0	1	19		0	1800	0	0	0	0	0
4	25	0	3	1			70	0				
6	7	0	3	1			80	0				
10	26	0	5				0					
11	1	0	6	1			0					
			0									

Table C.21.: Potato, wheat, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 70: Potato, 75: Polish Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	1	70		0	1800	0	0	0	0	0
3	30	0	3	1			70	0				
4	25	0	3	1			60	0				
8	10	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
7	31	0	6	1			0					
8	15	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.22.: Potato, wheat, barley, corn.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 19: Corn, 28: Wheat, 70: Potato, 75: Polish Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	1	70		0	1800	0	0	0	0	0
3	30	0	3	1			70	0				
4	25	0	3	1			60	0				
8	10	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
7	31	0	6	1			0					
8	15	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	30	0	6	2			0					
4	25	0	1	19		0	1800	0	0	0	0	0
4	25	0	3	1			70	0				
6	7	0	3	1			80	0				
10	26	0	5				0					
11	1	0	6	1			0					
			0									

Table C.23.: Potato, wheat, wheat, rye.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 30: Rye, 70: Potato, 75: Polish Canola.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	1	70		0	1800	0	0	0	0	0
3	30	0	3	1			70	0				
4	25	0	3	1			60	0				
8	10	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	5	0	1	30		0	1800	0	0	0	0	0
			0									
3	15	0	3	1			50	0				
4	15	0	3	1			50	0				
5	15	0	3	1			30	0				
7	25	0	5				0					
8	5	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.24.: Potato, wheat, wheat, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 70: Potato, 75: Polish Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
3	30	0	1	70		0	1800	0	0	0	0	0
3	30	0	3	1			70	0				
4	25	0	3	1			60	0				
8	10	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
7	31	0	6	1			0					
8	15	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.25.: Sugar beet, wheat, peas, wheat, rye.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 30: Rye, 62: Peas 69: Sugar Beet, 75: Polish Canola.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
4	8	0	1	69		0	1800	0	0	0	0	0
4	8	0	3	1			60	0				
4	30	0	3	1			50	0				
10	7	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
8	10	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	21	0	6	2			0					
4	1	0	1	62		0	1800	0	0	0	0	0
8	12	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	5	0	1	30		0	1800	0	0	0	0	0
			0									
3	15	0	3	1			50	0				
4	15	0	3	1			50	0				
5	15	0	3	1			30	0				
7	25	0	5				0					
8	5	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.26.: Sugar beet, wheat, peas, wheat, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 62: Peas 69: Sugar Beet, 75: Polish Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
4	8	0	1	69		0	1800	0	0	0	0	0
4	8	0	3	1			60	0				
4	30	0	3	1			50	0				
10	7	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
8	10	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	21	0	6	2			0					
4	1	0	1	62		0	1800	0	0	0	0	0
8	12	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
7	31	0	6	1			0					
8	15	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.27.: Sugar beet, wheat, oats, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 32: Oats 69: Sugar Beet, 75: Polish Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
4	8	0	1	69		0	1800	0	0	0	0	0
4	8	0	3	1			60	0				
4	30	0	3	1			50	0				
10	7	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
8	12	0	6	1			0					
8	30	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	7	0	6	2			0					
3	21	0	1	32		0	1800	0	0	0	0	0
3	21	0	3	1			40	0				
4	21	0	3	1			40	0				
8	4	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
7	31	0	6	1			0					
8	15	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.28.: Sugar beet, wheat, rye.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 30: Rye, 69: Sugar Beet, 75: Polish Canola.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
4	8	0	1	69		0	1800	0	0	0	0	0
4	8	0	3	1			60	0				
4	30	0	3	1			50	0				
10	7	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	5	0	1	30		0	1800	0	0	0	0	0
			0									
3	15	0	3	1			50	0				
4	15	0	3	1			50	0				
5	15	0	3	1			30	0				
7	25	0	5				0					
8	5	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.29.: Sugar beet, wheat, rye, corn, peas.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 19: Corn, 28: Wheat, 30: Rye, 62: Peas, 69: Sugar Beet, 75: Polish Canola.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
4	8	0	1	69		0	1800	0	0	0	0	0
4	8	0	3	1			60	0				
4	30	0	3	1			50	0				
10	7	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	5	0	1	30		0	1800	0	0	0	0	0
			0									
3	15	0	3	1			50	0				
4	15	0	3	1			50	0				
5	15	0	3	1			30	0				
7	25	0	5				0					
8	5	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	30	0	6	2			0					
4	25	0	1	19		0	1800	0	0	0	0	0
4	25	0	3	1			70	0				
6	7	0	3	1			80	0				
10	26	0	5				0					
			0									
3	21	0	6	2			0					
4	1	0	1	62		0	1800	0	0	0	0	0
8	12	0	5				0					
8	21	0	6	1			0					
9	1	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.30.: Sugar beet, wheat, sugar beet, wheat, rye.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 30: Rye, 69: Sugar Beet, 75: Polish Canola.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
4	8	0	1	69		0	1800	0	0	0	0	0
4	8	0	3	1			60	0				
4	30	0	3	1			50	0				
10	7	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
8	15	0	6	1			0					
8	29	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
4	8	0	1	69		0	1800	0	0	0	0	0
4	8	0	3	1			60	0				
4	30	0	3	1			50	0				
10	7	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	5	0	1	30		0	1800	0	0	0	0	0
			0									
3	15	0	3	1			50	0				
4	15	0	3	1			50	0				
5	15	0	3	1			30	0				
7	25	0	5				0					
8	5	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.31.: Sugar beet, wheat, sugar beet, wheat, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 69: Sugar Beet, 75: Polish Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
4	8	0	1	69		0	1800	0	0	0	0	0
4	8	0	3	1			60	0				
4	30	0	3	1			50	0				
10	7	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
8	15	0	6	1			0					
8	29	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
4	8	0	1	69		0	1800	0	0	0	0	0
4	8	0	3	1			60	0				
4	30	0	3	1			50	0				
10	7	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
7	31	0	6	1			0					
8	15	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.32.: Sugar beet, wheat, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 69: Sugar Beet, 75: Polish Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
4	8	0	1	69		0	1800	0	0	0	0	0
4	8	0	3	1			60	0				
4	30	0	3	1			50	0				
10	7	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
7	31	0	6	1			0					
8	15	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.33.: Sugar beet, wheat, barley, corn, peas.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 19: Corn, 28: Wheat, 62: Peas, 69: Sugar Beet, 75: Polish Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
4	8	0	1	69		0	1800	0	0	0	0	0
4	8	0	3	1			60	0				
4	30	0	3	1			50	0				
10	7	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
7	31	0	6	1			0					
8	15	0	1	75		0	1800	0	0	0	0	0
			0									
1	1	0	8									
3	30	0	6	2			0					
4	25	0	1	19		0	1800	0	0	0	0	0
4	25	0	3	1			70	0				
6	7	0	3	1			80	0				
10	26	0	5				0					
			0									
3	21	0	6	2			0					
4	1	0	1	62		0	1800	0	0	0	0	0
8	12	0	5				0					
8	21	0	6	1			0					
9	1	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.34.: Sugar beet, wheat, wheat, rye.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 30: Rye, 69: Sugar Beet, 75: Polish Canola.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
4	8	0	1	69		0	1800	0	0	0	0	0
4	8	0	3	1			60	0				
4	30	0	3	1			50	0				
10	7	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	5	0	1	30		0	1800	0	0	0	0	0
			0									
3	15	0	3	1			50	0				
4	15	0	3	1			50	0				
5	15	0	3	1			30	0				
7	25	0	5				0					
8	5	0	6	1			0					
8	20	0	1	75		0	1800	0	0	0	0	0
			0									

Table C.35.: Sugar beet, wheat, wheat, barley.
Operations: 1: Plant, 3: Fertilize, 5: Harvest, 6: Plow, 8: Kill, 0: End of year flag.
Field Crops: 28: Wheat, 69: Sugar Beet, 75: Polish Canola, 99: Barley.
Each table row represents a management operation (column "op") carried out on the specified day and month. Quantities of N fertilizer in kg NO_3^--N per ha are given in column "mgt4i" for fertilizing operations. Please refer to Arnold et al. (2013, pp. 257-296) for further information.

mon	day	HU	op	mgt1i	mgt2i	mgt3i	mgt4i	mgt5i	mgt6i	mgt7i	mgt8i	mgt9i
1	1	0	8									
4	8	0	1	69		0	1800	0	0	0	0	0
4	8	0	3	1			60	0				
4	30	0	3	1			50	0				
10	7	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
10	14	0	6	1			0					
10	26	0	1	28		0	1800	0	0	0	0	0
			0									
3	20	0	3	1			60	0				
4	20	0	3	1			60	0				
5	20	0	3	1			60	0				
7	31	0	5				0					
9	15	0	6	1			0					
10	3	0	1	99		0	1800	0	0	0	0	0
			0									
3	7	0	3	1			50	0				
4	7	0	3	1			50	0				
5	7	0	3	1			40	0				
7	14	0	5				0					
7	31	0	6	1			0					
8	15	0	1	75		0	1800	0	0	0	0	0
			0									

Bibliography

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), 1340–1352.

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. C., Vaghefi, S. & Srinivasan, R. (2017), '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.

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.

Abbott, M., Bathurst, J. C., Cunge, J. A., O'Connell, P. E. & Rasmussen, J. (1986*a*), 'European Hydrological System - Système Hydrologique Européen , "SHE", 1: History and Philosophy of a Physically-Based, Distributed Modelling System', *Journal of Hydrology* **87**, 45–59.

Abbott, M., Bathurst, J., Cunge, J., O'Connell, P. & Rasmussen, J. (1986*b*), 'An Introduction to the European Hydrological System - Système Hydrologique Européen , "SHE", 2: Structure of a Physically-Based, Distributed Modelling System', *Journal of Hydrology* **87**, 61–77.

Ad-hoc AG Boden (2005), *Bodenkundliche Kartieranleitung. 5. verbesserte und erweiterte Auflage*, Schweizerbart Science Publishers, Hannover.

AG Bodenkunde (1982), *Bodenkundliche Kartieranleitung, 3. Auflage*, Hannover.

Allen, R. G., Pereira, L. S., Raes, D. & Smith, M. (1998), *Crop evaporation - Guidelines for computing crop water requirements, FAO Irrigation and drainage paper 56*, FAO, Rome.
URL: *https://www.fao.org/3/x0490e/x0490e00.htm, as retrieved on June 30, 2024*

Apostel, A., Kalcic, M., Dagnew, A., Evenson, G. R., Kast, J., King, K., Martin, J., Muenich, R. L. & Scavia, D. (2021), 'Simulating internal watershed processes using multiple SWAT models', *Science of the Total Environment* **759**, 143920.

Arnold, J. G., Allen, P. M., Volk, M., Williams, J. R. & Bosch, D. D. (2010), 'Assessment of Different Representations of Spatial Variability on SWAT Model Performance', *Transactions of the ASABE* **53**(5), 1433–1443.

Arnold, J. G. & Fohrer, N. (2005), 'SWAT2000: Current capabilities and research opportunities in applied watershed modelling', *Hydrological Processes* **19**(3), 563–572.

Arnold, J. G., Kiniry, J. R., Srinivasan, R., Williams, J. R., Haney, E. B. & Neitsch, S. L. (2013), 'Soil & Water Assessment Tool Input/Output Documentation'.

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R., van Griensven, A., Van Liew, M. W., Kannan, N. & Jha, M. (2012), 'SWAT: Model Use, Calibration, and Validation', *Transactions of the ASABE* **55**(4), 1491–1508.

Arnold, J. G., Srinivasan, R., Muttiah, R. S. & Williams, J. R. (1998), 'Large area hydrologic modeling and assessment part I: Model development', *Journal of the American Water Resources Association* **34**(1), 73–89.

Arnold, J. G., Youssef, M. A., Yen, H., White, M. J., Sheshukov, A. Y., Sadeghi, A. M., Moriasi, D. N., Steiner, J. L., Amatya, D. M., Skaggs, R. W., Haney, E. B., Jeong, J., Arabi, M. & Gowda, P. H. (2015), 'Hydrological Processes and Model Representation: Impact of Soft Data on Calibration', *Transactions of the ASABE* **58**(6), 1637–1660.

Ascott, M. J., Gooddy, D. C., Fenton, O., Vero, S., Ward, R. S., Basu, N. B., Worrall, F., Van Meter, K. J. & Surridge, B. W. (2021), 'The need to integrate legacy nitrogen storage dynamics and time lags into policy and practice', *Science of the Total Environment* **781**, 146698.

Ascott, M. J., Gooddy, D. C., Wang, L., Stuart, M. E., Lewis, M. A., Ward, R. S. & Binley, A. M. (2017), 'Global patterns of nitrate storage in the vadose zone', *Nature Communications* **8**(1), 1416.

Bachmann, J., Horn, R. & Peth, S. (2014), *Einführung in die Bodenphysik*, 4 edn, Schweizerbart Science Publishers, Stuttgart.

Barber, C., Lamontagne, J. R. & Vogel, R. M. (2020), 'Improved estimators of correlation and R2 for skewed hydrologic data', *Hydrological Sciences Journal* **65**(1), 87–101.

Basu, N. B., Van Meter, K. J., Byrnes, D. K., Van Cappellen, P., Brouwer, R., Jacobsen, B. H., Jarsjö, J., Rudolph, D. L., Cunha, M. C., Nelson, N., Bhattacharya, R., Destouni, G. & Olsen, S. B. (2022), 'Managing nitrogen legacies to accelerate water quality improvement', *Nature Geoscience* **15**(2), 97–105.

Batool, M., Sarrazin, F. J., Attinger, S., Basu, N. B., Van Meter, K. J. & Kumar, R. (2022), 'Long-term annual soil nitrogen surplus across Europe (1850–2019)', *Scientific Data* **9**(1), 1–22.

Bauwe, A., Kahle, P. & Lennartz, B. (2019), 'Evaluating the SWAT model to predict streamflow, nitrate loadings and crop yields in a small agricultural catchment', *Advances in Geosciences* **48**, 1–9.

Bauwe, A., Kahle, P., Tiemeyer, B. & Lennartz, B. (2020), 'Hydrology is the key factor for nitrogen export from tile-drained catchments under consistent land-management', *Environmental Research Letters* **15**(9), 094050.

Beven, K. (1989), 'Changing ideas in hydrology — The case of physically-based models', *Journal of Hydrology* **105**(1-2), 157–172.

Beven, K. (2000), 'Uniqueness of place and process representations in hydrological modelling', *Hydrology and Earth System Sciences* **4**(2), 203–213.

Beven, K. (2001), 'How far can we go in distributed hydrological modelling?', *Hydrology and Earth System Sciences* **5**(1), 1–12.

Beven, K. (2006), 'A Manifesto for the Equifinality Thesis', *Journal of hydrology* **320**(1-2), 18–36.

Beven, K. (2012), *Rainfall-Runoff Modelling - The Primer*, 2nd edn, Wiley-Blackwell, Lancaster University, UK.

Beven, K. (2016), 'Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication', *Hydrological Sciences Journal* **61**(9), 1652–1665.

Beven, K. & Binley, A. (1992), 'The future of distributed models: Model calibration and uncertainty prediction', *Hydrological Processes* **6**(3), 279–298.

Beven, K. & Binley, A. (2014), 'GLUE: 20 years on', *Hydrological Processes* **28**(24), 5897–5918.

Bieger, K., Arnold, J. G., Rathjens, H., White, M. J., Bosch, D. D. & Allen, P. M. (2019), 'Representing the Connectivity of Upland Areas to Floodplains and Streams in SWAT+', *Journal of the American Water Resources Association* **55**(3), 578–590.

Bieger, K., Arnold, J. G., Rathjens, H., White, M. J., Bosch, D. D., Allen, P. M., Volk, M. & Srinivasan, R. (2017), 'Introduction to SWAT+, A Completely Restructured Version of the Soil and Water Assessment Tool', *JAWRA Journal of the American Water Resources Association* **53**(1), 115–130.

Bijay-Singh & Craswell, E. (2021), 'Fertilizers and nitrate pollution of surface and ground water: an increasingly pervasive global problem', *SN Applied Sciences* **3**(4), 1–24.

Blöschl, G., Bierkens, M. F., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J. J., Savenije, H. H., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B., Viglione, A., Aksoy, H., Allen, S. T., Amin, A., Andréassian, V., Arheimer, B., Aryal, S. K., Baker, V., Bardsley, E., Barendrecht, M. H., Bartosova, A., Batelaan, O., Berghuijs, W. R., Beven, K., Blume, T., Bogaard, T., Borges de Amorim, P., Böttcher, M. E., Boulet, G., Breinl, K., Brilly, M., Brocca, L., Buytaert, W., Castellarin, A., Castelletti, A., Chen, X., Chen, Y., Chen, Y., Chifflard, P., Claps, P., Clark, M. P., Collins, A. L., Croke, B., Dathe, A., David, P. C., de Barros, F. P., de Rooij, G., Di Baldassarre, G., Driscoll, J. M., Duethmann, D., Dwivedi, R., Eris, E., Farmer, W. H., Feiccabrino, J., Ferguson, G., Ferrari, E., Ferraris, S., Fersch, B., Finger, D., Foglia, L., Fowler, K., Gartsman, B., Gascoin, S., Gaume, E., Gelfan, A., Geris, J., Gharari, S., Gleeson, T., Glendell, M., Gonzalez Bevacqua, A., González-Dugo, M. P., Grimaldi, S., Gupta, A. B., Guse, B., Han, D., Hannah, D., Harpold, A., Haun, S., Heal, K., Helfricht, K., Herrnegger, M., Hipsey, M., Hlaváčiková, H., Hohmann, C., Holko, L., Hopkinson, C., Hrachowitz, M., Illangasekare, T. H., Inam, A., Innocente, C., Istanbulluoglu, E., Jarihani, B., Kalantari, Z., Kalvans, A., Khanal, S., Khatami, S., Kiesel, J., Kirkby, M., Knoben, W., Kochanek, K., Kohnová, S., Kolechkina, A., Krause, S., Kreamer, D., Kreibich, H., Kunstmann, H., Lange, H., Liberato, M. L., Lindquist, E., Link, T., Liu, J., Loucks, D. P., Luce, C., Mahé, G., Makarieva, O., Malard, J., Mashtayeva, S., Maskey, S., Mas-Pla, J., Mavrova-Guirguinova, M., Mazzoleni, M., Mernild, S., Misstear, B. D., Montanari, A., Müller-Thomy, H., Nabizadeh, A., Nardi, F., Neale, C., Nesterova, N., Nurtaev, B., Odongo, V. O., Panda, S., Pande, S., Pang, Z., Papacharalampous, G., Perrin, C., Pfister, L., Pimentel, R., Polo, M. J., Post, D., Prieto Sierra, C., Ramos, M. H., Renner, M., Reynolds, J. E., Ridolfi, E., Rigon, R., Riva, M., Robertson, D. E., Rosso, R., Roy, T., Sá, J. H., Salvadori, G., Sandells, M., Schaefli, B., Schumann, A., Scolobig, A., Seibert, J., Servat, E., Shafiei, M., Sharma, A., Sidibe, M., Sidle, R. C., Skaugen, T., Smith, H., Spiessl, S. M., Stein, L., Steinsland, I., Strasser, U., Su, B., Szolgay, J., Tarboton, D., Tauro, F., Thirel, G., Tian, F., Tong, R., Tussupova, K., Tyralis, H., Uijlenhoet, R., van Beek, R., van der Ent, R. J., van der Ploeg, M., Van Loon, A. F., van Meerveld, I., van Nooijen, R., van Oel, P. R., Vidal, J. P., von Freyberg, J., Vorogushyn, S., Wachniew, P., Wade, A. J., Ward, P., Westerberg, I. K., White, C., Wood, E. F., Woods, R., Xu, Z., Yilmaz, K. K. & Zhang, Y. (2019), 'Twenty-three unsolved problems in hydrology (UPH)–a community perspective', *Hydrological Sciences Journal* **64**(10), 1141–1158.

Bøgestrand, J., Kristensen, P., Kronvang, B., Thyssen, N. & Werner, B. (2005), 'Source apportionment of nitrogen and phosphorus inputs into the aquatic environment: A report for the European Environment Agency', EEA Report No7/2005.
URL: *https://www.eea.europa.eu/publications/eea_report_2005_7, as retrieved on June 30, 2024*

Borah, D. & Bera, M. (2003), 'Watershed-scale Hydrologic and Nonpoint-Source Pollution Models: Review of Mathematical Bases', *Transactions of the ASAE* **46**(6), 1553–1566.

Borah, D., Yagow, G., Saleh, A., Barnes, P., Rosenthal, W., Krug, E. & Hauck, L. (2006), 'Sediment and Nutrient Modeling for TMDL Development and Implementation', *Transactions of the ASABE* **49**(4), 967–986.

Boretti, A. & Rosa, L. (2019), 'Reassessing the projections of the World Water Development Report', *npj Clean Water* **2**(1).

Breiman, L. (2001), 'Random Forests', *Machine learning* **45**(1), 5–32.

Brown, L. C. & Barnwell, T. O. (1987), *The Enhanced Stream Water Quality Models QUAL2E and QUAL2E-UNCAS: Documentation and User Manual*, EPA.

Brunsch, A. F., ter Laak, T. L., Rijnaarts, H. H. & Christoffels, E. (2018), 'Pharmaceutical concentration variability at sewage treatment plant outlets dominated by hydrology and other factors', *Environmental Pollution* **235**, 615–624.

Campolongo, F., Cariboni, J. & Saltelli, A. (2007), 'An effective screening design for sensitivity analysis of large models', *Environmental Modelling and Software* **22**(10), 1509–1518.

Campolongo, F., Saltelli, A. & Cariboni, J. (2011), 'From screening to quantitative sensitivity analysis. A unified approach', *Computer Physics Communications* **182**(4), 978–988.

Carnell, R. (2022), 'lhs: Latin Hypercube Samples'.
URL: *https://github.com/bertcarnell/lhs, as retrieved on June 30, 2024*

Carpenter, S. R., Stanley, E. H. & Vander Zanden, M. J. (2011), 'State of the world's freshwater ecosystems: physical, chemical, and biological changes', *Annual review of Environment and Resources* **36**, 75–99.

Chamber of Agriculture NRW (2018), 'Nährstoffbericht NRW 2017'.
URL: *https://www.landwirtschaftskammer.de/landwirtschaft/ackerbau/pdf/naehrstoffbericht-2017.pdf, as retrieved on June 30, 2024*

Chamber of Agriculture NRW (2021), 'Nährstoffbericht NRW 2021'.
URL: *https://www.landwirtschaftskammer.de/landwirtschaft/ackerbau/pdf/naehrstoffbericht-2021.pdf, as retrieved on June 30, 2024*

Christoffels, E. (2008), Monitoring und Modellanwendung - Entwicklung eines Immissionsinventars am Beispiel der Erft, PhD thesis, University of Trier.

Christoffels, E., Brunsch, A. F., Wunderlich-Pfeiffer, J. & Mertens, F. M. (2016), 'Monitoring micropollutants in the Swist river basin', *Water Science and Technology* **74**(10), 2280–2296.

Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J., Tang, G., Gharari, S., Freer, J. E., Whitfield, P. H., Shook, K. R. & Papalexiou, S. M. (2021), 'The Abuse of Popular Performance Metrics in Hydrologic Modeling', *Water Resources Research* **57**(9), 1–16.

Coffey, R., Benham, B., Krometis, L. A., Wolfe, M. L. & Cummins, E. (2014), 'Assessing the Effects of Climate Change on Waterborne Microorganisms: Implications for EU and U.S. Water Policy', *Human and Ecological Risk Assessment* **20**(3), 724–742.

Cronshey, R. (1986), *Urban Hydrology for Small Watersheds, Technical Release No. 55*, US Department for Agriculture, Soil Conservation Service, Engineering Division.

Daggupati, P., Pai, N., Ale, S., Douglas-Mankin, K. R., Zeckoski, R. W., Jeong, J., Parajuli, P. B., Saraswat, D. & Youssef, M. A. (2015), 'A Recommended Calibration and Validation Strategy for Hydrologic and Water Quality Models', *Transactions of the ASABE* **58**(6), 1705–1719.

Daniel, E. B., Camp, J. V., LeBoeuf, E. J., Penrod, J. R., Dobbins, J. P. & Abkowitz, M. D. (2011), 'Watershed Modeling and its Applications: A State-of-the-Art Review', *The Open Hydrology Journal* **5**(1), 26–50.

De Girolamo, A. M. & Lo Porto, A. (2020), 'Source Apportionment of Nutrient Loads to a Mediterranean River and Potential Mitigation Measures', *Water* **12**(2), 577.

de Jong, S. M. & Jetten, V. G. (2007), 'Estimating spatial patterns of rainfall interception from remotely sensed vegetation indices and spectral mixture analysis', *International Journal of Geographical Information Science* **21**(5), 529–545.

Douglas-Mankin, K. R., Srinivasan, R. & Arnold, J. G. (2010), 'Soil and Water Assessment Tool (SWAT) Model: Current Developments and Applications', *Transactions of the ASABE* **53**(5), 1423–1431.

Du, B., Arnold, J. G., Saleh, A. & Jaynes, D. B. (2005), 'Development and Application of SWAT to Landscapes with Tiles and Potholes', *Transactions of the ASAE* **48**(3), 1121–1133.

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z. I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A. H., Soto, D., Stiassny, M. L. & Sullivan, C. A. (2006), 'Freshwater biodiversity: Importance, threats, status and conservation challenges', *Biological Reviews of the Cambridge Philosophical Society* **81**(2), 163–182.

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

Erisman, J. W., Galloway, J. N., Seitzinger, S. P., Bleeker, A., Dise, N. B., Roxana Petrescu, A. M., Leach, A. M. & de Vries, W. (2013), 'Consequences of human modification of the global nitrogen cycle', *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**(1621).

Evenson, G. R., Kalcic, M., Wang, Y.-C., Robertson, D., Scavia, D., Martin, J., Aloysius, N., Apostel, A., Boles, C., Brooker, M., Confesor, R., Dagnew, A. T., Guo, T., Kast, J., Kujawa, H., Muenich, R. L., Murumkar, A. & Redder, T. (2021), 'Uncertainty in critical source area predictions from watershed-scale hydrologic models', *Journal of Environmental Management* **279**, 111506.

Fu, B., Merritt, W. S., Croke, B. F., Weber, T. R. & Jakeman, A. J. (2019), 'A review of catchment-scale water quality and erosion models and a synthesis of future prospects', *Environmental Modelling & Software* **114**, 75–97.

Fuchs, S., Kaiser, M., Kiemle, L., Kittlaus, S., Rothvoß, S., Toshovski, S., Wagner, A., Wander, R., Weber, T. & Ziegler, S. (2017), 'Modeling of Regionalized Emissions (MoRE) into Water Bodies: An Open-Source River Basin Management System', *Water (Switzerland)* **9**(4), 1–13.

Fuchs, S., Scherer, U., Wander, R., Behrendt, H., Venohr, M., Opitz, D., Hillenbrand, T., Marscheider-Weidemann, F. & Götz, T. (2010), 'Calculation of Emissions into Rivers in Germany using the MONERIS Model - Nutrients, Heavy Metals and Polycyclic Aromatic Hydrocarbons', Umweltbundesamt.
URL: *https://www.umweltbundesamt.de/publikationen/calculation-of-emissions-into-rivers-in-germany, as retrieved on June 30, 2024*

Galloway, J. N., Bleeker, A. & Erisman, J. W. (2021), 'The Human Creation and Use of Reactive Nitrogen: A Global and Regional Perspective', *Annual Review of Environment and Resources* **46**, 255–288.

Galloway, J. N. & Cowling, E. B. (2002), 'Reactive nitrogen and the world: 200 Years of change', *Ambio* **31**(2), 64–71.

Galloway, J. N., Dentener, F. J., Capone, D. G., Boyer, E. W., Howarth, R. W., Seitzinger, S. P., Asner, G. P., Cleveland, C. C., Green, P. A., Holland, E. A., Karl, D. M., Michaels, A. F., Porter, J. H., Townsend, A. R. & Vörösmarty, C. J. (2004), 'Nitrogen Cycles: Past, Present, and Future', *Biogeochemistry* **70**(2), 153–226.

Garen, D. C. & Moore, D. S. (2005), 'Curve number hydrology in water quality modeling: Uses, abuses, and future directions', *Journal of the American Water Resources Association* **41**(2), 377–388.

Gassman, P. W., Reyes, M. R., Green, C. H. & Arnold, J. G. (2007), 'The Soil and Water Assessment Tool: Historical Development, Applications, and Future Research Directions', *Transactions of the ASABE* **50**(4), 1211–1250.

Gassman, P. W., Sadeghi, A. M. & Srinivasan, R. (2014), 'Applications of the SWAT Model Special Section: Overview and Insights', *Journal of Environmental Quality* **43**(1), 1–8.

Geobasis NRW (2019), 'Digital Basic-Landscape Model'.
URL: *https://www.bezreg-koeln.nrw.de/geobasis-nrw/produkte-und-dienste/landschaftsmodelle/aktuelle-landschaftsmodelle/digitales-basis, as retrieved on June 30, 2024*

Geological Survey NRW (2015), 'Bodenkarte von Nordrhein-Westfalen 1:50000'.
URL: *https://www.opengeodata.nrw.de/produkte/geologie/boden/BK/ISBK50/, as retrieved on June 30, 2024*

Geological Survey NRW (2019), 'Bodenkarte von Nordrhein-Westfalen 1:50000'.
URL: *https://www.opengeodata.nrw.de/produkte/geologie/boden/BK/ISBK50/, as retrieved on June 30, 2024*

German Federal Agency for Cartography and Geodesy (2012), 'CORINE Land Cover 10ha'.
URL: *https://sg.geodatenzentrum.de/web_public/gdz/dokumentation/deu/clc10.pdf, as retrieved on June 30, 2024*

German Federal Ministry for the Environment (2003), 'Hydrological Atlas of Germany'.
URL: *https://geoportal.bafg.de/mapapps/resources/apps/HAD/index.html?lang=en, as retrieved on June 30, 2024*

Ghebremichael, L. T., Veith, T. L. & Watzin, M. C. (2010), 'Determination of critical source areas for phosphorus loss: Lake champlain basin, vermont', *Transactions of the ASABE* **53**(5), 1595–1604.

Gijsman, A. J., Thornton, P. K. & Hoogenboom, G. (2007), 'Using the WISE database to parameterize soil inputs for crop simulation models', *Computers and Electronics in Agriculture* **56**(2), 85–100.

Gitau, M. W., Veith, T. L. & Gburek, W. J. (2004), 'Farm-level Optimization of BMP Placement for Cost-effective Pollution Reduction', *Transactions of the American Society of Agricultural Engineers* **47**(6), 1923–1931.

Golden, H. E., Evenson, G. R., Christensen, J. R. & Lane, C. R. (2023), 'Advancing Watershed Legacy Nitrogen Modeling to Improve Global Water Quality', *Environmental Science and Technology* **57**(7), 2691–2697.

Gömann, H., Kreins, P., Brandes, E. & Pfingsten, T. (2021), 'Kooperationsprojekt GROWA + NRW 2021 Teil I: Regionalisierte Quantifizierung der landwirtschaftlichen Flächenbilanzüberschüsse in Nordrhein-Westfalen'.
URL: *https://www.lanuv.nrw.de/fileadmin/lanuvpubl/3_fachberichte/30110a.pdf, as retrieved on June 30, 2024*

Grüneberg, E., Ziche, D. & Wellbrock, N. (2014), 'Organic carbon stocks and sequestration rates of forest soils in Germany', *Global Change Biology* **20**(8), 2644–2662.

Gumm, L. P., Bense, V. F., Dennis, P. F., Hiscock, K. M., Cremer, N. & Simon, S. (2016), 'Dissolved noble gases and stable isotopes as tracers of preferential fluid flow along faults in the Lower Rhine Embayment, Germany', *Hydrogeology Journal* **24**(1), 99–108.

Guo, T., Engel, B. A., Shao, G., Arnold, J. G., Srinivasan, R. & Kiniry, J. R. (2019), 'Development and improvement of the simulation of woody bioenergy crops in the Soil and Water Assessment Tool (SWAT)', *Environmental Modelling and Software* **122**, 1–35.

Guo, T., Gitau, M., Merwade, V., Arnold, J. G., Srinivasan, R., Hirschi, M. & Engel, B. (2018), 'Comparison of performance of tile drainage routines in SWAT 2009 and 2012 in an extensively tile-drained watershed in the Midwest', *Hydrology and Earth System Sciences* **22**(1), 89–110.

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.

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', *Journal of Hydrologic Engineering* **4**(2), 135–143.

Guse, B., Kiesel, J., Pfannerstill, M. & Fohrer, N. (2020), 'Assessing parameter identifiability for multiple performance criteria to constrain model parameters', *Hydrological Sciences Journal* **65**(7), 1158–1172.

Guzman, J. A., Shirmohammadi, A., Sadeghi, A. M., Wang, X., Chu, M. L., Jha, M., Parajuli, P. B., Harmel, R., Khare, Y. P. & Hernandez, J. E. (2015), 'Uncertainty Considerations in Calibration and Validation of Hydrologic and Water Quality Models', *Transactions of the ASABE* **58**(6), 1745–1762.

Haas, H., Reaver, N. G., Karki, R., Kalin, L., Srivastava, P., Kaplan, D. A. & Gonzalez-Benecke, C. (2022), 'Improving the representation of forests in hydrological models', *Science of the Total Environment* **812**, 151425.

Harmel, R., Cooper, R., Slade, R., Haney, R. & 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. & 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.

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

Harrison, J. A., Maranger, R. J., Alexander, R. B., Giblin, A. E., Jacinthe, P. A., Mayorga, E., Seitzinger, S. P., Sobota, D. J. & Wollheim, W. M. (2009), 'The regional and global significance of nitrogen removal in lakes and reservoirs', *Biogeochemistry* **93**(1-2), 143–157.

Hellweger, F. L. (1997), 'AGREE - DEM Surface Reconditioning System'.
URL: *https://www.caee.utexas.edu/prof/maidment/GISHYDRO/ferdi/research/agree/agree.html, as retrieved on June 30, 2024*

Helton, J. C. & Davis, F. J. (2003), 'Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems', *Reliability Engineering & System Safety* **81**(1), 23–69.

Her, Y. & Chaubey, I. (2015), 'Impact of the numbers of observations and calibration parameters on equifinality, model performance, and output and parameter uncertainty', *Hydrological Processes* **29**(19), 4220–4237.

Her, Y., Frankenberger, J., Chaubey, I. & Srinivasan, R. (2015), 'Threshold Effects in HRU Definition of the Soil and Water Assessment Tool', *Transactions of the ASABE* **58**(2), 367–378.

Herrera, P. A., Marazuela, M. A. & Hofmann, T. (2022), 'Parameter estimation and uncertainty analysis in hydrological modeling', *Wiley Interdisciplinary Reviews: Water* **9**(1), 1–23.

Herrmann, F. & Wendland, F. (2021), 'Kooperationsprojekt GROWA + NRW 2021 Teil IIa: Modellierung des Wasserhaushalts in Nordrhein-Westfalen mit mGROWA'.
URL: *https://www.lanuv.nrw.de/fileadmin/lanuvpubl/3_fachberichte/30110b.pdf, as retrieved on June 30, 2024*

Hoang, L., van Griensven, A., van der Keur, P., Refsgaard, J. C., Troldborg, L., Nilsson, B. & Mynett, A. (2014), 'Comparison and Evaluation of Model Structures for the Simulation of Pollution Fluxes in a Tile-Drained River Basin', *Journal of Environmental Quality* **43**(1), 86–99.

Holmes, T., Stadnyk, T. A., Kim, S. J. & Asadzadeh, M. (2020), 'Regional Calibration With Isotope Tracers Using a Spatially Distributed Model: A Comparison of Methods', *Water Resources Research* **56**(9), 1–18.

Horn, A. L., Rueda, F. J., Hörmann, G. & Fohrer, N. (2004), 'Implementing river water quality modelling issues in mesoscale watershed models for water policy demands––an overview on current concepts, deficits, and future tasks', *Physics and Chemistry of the Earth, Parts A/B/C* **29**(11-12), 725–737.

Hrachowitz, M., Benettin, P., van Breukelen, B. M., Fovet, O., Howden, N. J., Ruiz, L., van der Velde, Y. & Wade, A. J. (2016), 'Transit times—the link between hydrology and water quality at the catchment scale', *Wiley Interdisciplinary Reviews: Water* **3**(5), 629–657.

Jakeman, A. J. & Hornberger, G. M. (1993), 'How much complexity is warranted in a rainfall-runoff model?', *Water Resources Research* **29**(8), 2637–2649.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013), *An Introduction to Statistical Learning*, Vol. 103 of *Springer Texts in Statistics*, Springer New York, New York, NY.
URL: *https://www.statlearning.com/, as retrieved on June 30, 2024*

Jha, M., Gassman, P. W., Secchi, S., Gu, R. & Arnold, J. G. (2004), 'Effect of Watershed Subdivision on SWAT Flow, Sediment, and Nutrient Preditions', *Journal of the American Water Resources Association* **40**(3), 811–825.

Kern, U., Wendland, F. & Christoffels, E. (2007), Catchment Modeling of Emissions from the Perspective of WFD Implementation, *in* B. Westrich & U. Förstner, eds, 'Sediment Dynamics and Pollutant Mobility in Rivers', Springer, chapter 5.1, pp. 172–185.

Kirchner, J. W. (2003), 'A double paradox in catchment hydrology and geochemistry', *Hydrological processes* **17**(4), 871–874.

Kirchner, J. W. (2006), 'Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology', *Water Resources Research* **42**(3), 1–5.

Kistemann, T., Christoffels, E., Franke, C., Rechenburg, A., Willkomm, M. & Exner, M. (2007), 'Abschlussbericht zum Forschungsprojekt Mikrobielle Belastung der Fließgewässer aus diffusen Eintragspfaden am Beispiel der Swist („ Swist III “)'.
URL: *https://intern.geohealth-centre.de/images/Abschlussberichte/Abschlussbericht_Swist%20III.pdf, as retrieved on June 30, 2024*

Kistemann, T., Koch, C., Herbst, S., Rechenburg, A. & Exner, M. (2001), 'Abschlußbericht zum Forschungsprojekt Untersuchungen zur Mikrobiellen Fließgewässerbelastung durch Kläranlagen'.
URL: *https://intern.geohealth-centre.de/images/Abschlussberichte/Abschlussbericht_SwistI.pdf, as retrieved on June 30, 2024*

Klemeš, V. (1986), 'Operational testing of hydrological simulation models', *Hydrological Sciences Journal* **31**(1), 13–24.

Knoben, W. J., Freer, J. E. & Woods, R. A. (2019), 'Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores', *Hydrology and Earth System Sciences* **23**(10), 4323–4331.

Koch, C. (2004), Vergleichende Untersuchung der Belastung von Fließgewässern mit Cryptosporidium sp. und Giardia lamblia in vier Einzugsgebieten mit unterschiedlicher naturräumlicher Ausstattung und menschlicher Nutzung, PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn.

Koch, S., Bauwe, A. & Lennartz, B. (2013), 'Application of the SWAT Model for a Tile-Drained Lowland Catchment in North-Eastern Germany on Subbasin Scale', *Water Resources Management* **27**(3), 791–805.

Kothe, M., Ahring, A., Jansen, T., Bönsch, D., Mertens, F. M., Brunsch, A. F., Stephanus, H., Knorz, K., Gattke, C. & Rose, U. (2021), 'Identifizierung und Umsetzung von Maßnahmen zur Reduzierung der Stoffeinträge aus landwirtschaftlichen Dränausleitungen in die Oberflächengewässer'.
URL: *https://www.flussgebiete.nrw.de/system/files/atoms/files/agro-diffus_forschungsbericht.pdf, as retrieved on June 30, 2024*

Krause, P., Boyle, D. P. & Bäse, F. (2005), 'Comparison of different efficiency criteria for hydrological model assessment', *Advances in Geosciences* **5**, 89–97.

Krysanova, V. & Arnold, J. G. (2008), 'Advances in ecohydrological modelling with SWAT—a review', *Hydrological Sciences Journal* **53**(5), 939–947.

Krysanova, V. & White, M. (2015), 'Advances in water resources assessment with SWAT—an overview', *Hydrological Sciences Journal* **60**(5), 1–13.

Kunkel, R., Bergmann, S., Eisele, M. & Wendland, F. (2021), 'Kooperationsprojekt GROWA + NRW 2021 Teil IV: Denitrifikation Boden und im Grundwasser Nordrhein-Westfalens'.
URL: *https://www.lanuv.nrw.de/fileadmin/lanuvpubl/3_fachberichte/30110e.pdf, as retrieved on June 30, 2024*

Lam, Q. D., Schmalz, B. & Fohrer, N. (2010), 'Modelling point and diffuse source pollution of nitrate in a rural lowland catchment using the SWAT model', *Agricultural Water Management* **97**(2), 317–325.

Lamontagne, J. R., Barber, C. A. & Vogel, R. M. (2020), 'Improved Estimators of Model Performance Efficiency for Skewed Hydrologic Data', *Water Resources Research* **56**(9), 1–25.

Lee, S., Sadeghi, A. M., McCarty, G. W., Baffaut, C., Lohani, S., Duriancik, L. F., Thompson, A., Yeo, I. Y. & Wallace, C. (2018), 'Assessing the suitability of the Soil Vulnerability Index (SVI) on identifying croplands vulnerable to nitrogen loss using the SWAT model', *Catena* **167**, 1–12.

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

Lenk, S. (2008), Grundwasserbeschaffenheit und Hydrogeochemische Prozesse in Rheinischen Braunkohlenabraumkippen und in deren Abstrom, PhD thesis, Ruhr-Universität Bochum.

Lutz, S. R., Ebeling, P., Musolff, A., Van Nguyen, T., Sarrazin, F. J., Van Meter, K. J., Basu, N. B., Fleckenstein, J. H., Attinger, S. & Kumar, R. (2022), 'Pulling the rabbit out of the hat: Unravelling hidden nitrogen legacies in catchment-scale water quality models', *Hydrological Processes* **36**(10), 1–7.

Mailhot, A., Talbot, G. & Lavallée, B. (2015), 'Relationships between rainfall and Combined Sewer Overflow (CSO) occurrences', *Journal of Hydrology* **523**, 602–609.

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.

McDowell, R. W., Dils, R. M., Collins, A. L., Flahive, K. A., Sharpley, A. N. & Quinn, J. (2016), 'A review of the policies and implementation of practices to decrease water quality impairment by phosphorus in New Zealand, the UK, and the US', *Nutrient Cycling in Agroecosystems* **104**(3), 289–305.

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), 239–245.

Mertens, F. M., Schreiber, C., Brunsch, A. F., Zacharias, N., Knorz, K., Grasser, N., Heinkel, S.-B., Kistemann, T. & Christoffels, E. (2017), 'Abschlussbericht zum Forschungsprojekt Überprüfung innovativer Maßnahmen zur Reduzierung von Spurenstoffen und Mikroorganismen in Fließgewässern (ReSMo)'.
URL: *https://www.lanuv.nrw.de/fileadmin/forschung/abschlussbericht_resmo_endversion_final.pdf, as retrieved on June 30, 2024*

Mertens, F. M., Schreiber, C., Heinkel, S.-B., Kistemann, T. & Christoffels, E. (2012), 'Abschlussbericht zum Forschungsprojekt Überprüfung und Bewertung von Maßnahmen zur Reduzierung der chemisch-physikalischen und hygienisch-mikrobiologischen Belastungen von Fließgewässern am Beispiel der Swist'.
URL: *https://intern.geohealth-centre.de/images/Abschlussberichte/swistiv_abschlussbericht.pdf, as retrieved on June 30, 2024*

Miller, D. M. (1984), 'Reducing Transformation Bias in Curve Fitting', *The American Statistician* **38**(2), 124 – 126.

Mockus, V. (1972), Section 4: Hydrology, *in* 'National Engineering Handbook', Soil Conservation Service - U.S. Department Of Agriculture.

Møller, A. B., Beucher, A., Iversen, B. V. & Greve, M. H. (2018), 'Predicting artificially drained areas by means of a selective model ensemble', *Geoderma* **320**, 30–42.

Monteith, J. (1965), 'Evaporation and Environment', *Symposia of the Society for Experimental Biology* **19**, 205–234.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. & 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., Gitau, M. W., Pai, N. & Daggupati, P. (2015), 'Hydrologic and water quality models: Performance measures and evaluation criteria', *Transactions of the ASABE* **58**(6), 1763–1785.

Moriasi, D. N., Rossi, C. G., Arnold, J. G. & Tomer, M. D. (2012), 'Evaluating hydrology of the Soil and Water Assessment Tool (SWAT) with new tile drain equations', *Journal of Soil and Water Conservation* **67**(6), 513–524.

Morris, M. D. (1991), 'Factorial Sampling Plans for Preliminary Computational Experiments', *Technometrics* **33**(2), 161–174.

Musyoka, F. K., Strauss, P., Zhao, G., Srinivasan, R. & Klik, A. (2021), 'Multi-step calibration approach for SWAT model using soil moisture and crop yields in a small agricultural catchment', *Water (Switzerland)* **13**(16).

Musyoka, F. K., Strauss, P., Zhao, G., Strohmeier, S., Mutua, B. M. & Klik, A. (2023), 'Evaluating the impacts of sustainable land management practices on water quality in an agricultural catchment in Lower Austria using SWAT', *Environmental Monitoring and Assessment* **195**(4).

Näschen, K., Diekkrüger, B., Evers, M., Höllermann, B., Steinbach, S. & Thonfeld, F. (2019), 'The Impact of Land Use/Land Cover Change (LULCC) on Water Resources in a Tropical Catchment in Tanzania under Different Climate Change Scenarios', *Sustainability (Switzerland)* **11**(24).

Nash, E. & Sutcliffe, V. (1970), 'River flow forecasting through conceptual models Part I - A discussion of principles', *Journal of Hydrology* **10**, 282–290.

Neitsch, S. L., Arnold, J. G., Kiniry, J. & Williams, J. R. (2011), 'Soil and Water Assessment Tool Theoretical Documentation - Version 2009, Technical Report no 406'.

Niraula, R., Kalin, L., Srivastava, P. & Anderson, C. J. (2013), 'Identifying Critical Source Areas of Nonpoint Source Pollution With SWAT and GWLF', *Ecological Modelling* **268**, 123–133.

Niraula, R., Kalin, L., Wang, R. & Srivastava, P. (2012), 'Determining Nutrient and Sediment Critical Source Areas with SWAT: Effect of Lumped Calibration', *Transactions of the ASABE* **55**(1), 137–147.

Oenema, O., Kros, H. & De Vries, W. (2003), 'Approaches and uncertainties in nutrient budgets: Implications for nutrient management and environmental policies', *European Journal of Agronomy* **20**(1-2), 3–16.

Oeurng, C., Sauvage, S. & Sánchez-Pérez, J. M. (2010), 'Temporal variability of nitrate transport through hydrological response during flood events within a large agricultural catchment in south-west France', *Science of the Total Environment* **409**(1), 140–149.

Paerl, H. W., Scott, J. T., McCarthy, M. J., Newell, S. E., Gardner, W. S., Havens, K. E., Hoffman, D. K., Wilhelm, S. W. & Wurtsbaugh, W. A. (2016), 'It Takes Two to Tango: When and Where Dual Nutrient (N & P) Reductions Are Needed to Protect Lakes and Downstream Ecosystems', *Environmental Science and Technology* **50**(20), 10805–10813.

Paffrath, P., Schauhoff-Tholen, V. & Röhling, D. (2021), *Ratgeber Pflanzenbau und Pflanzenschutz*, Landwirtschaftskammer Nordrhein-Westfalen.

Pandey, A., Himanshu, S. K., Mishra, S. K. & Singh, V. P. (2016), 'Physically based soil erosion and sediment yield models revisited', *Catena* **147**, 595–620.

Pappenberger, F. & Beven, K. (2006), 'Ignorance is bliss: Or seven reasons not to use uncertainty analysis', *Water Resources Research* **42**(5), 1–8.

Pellerin, B. A., Bergamaschi, B. A., Gilliom, R. J., Crawford, C. G., Saraceno, J., Frederick, C. P., Downing, B. D. & Murphy, J. C. (2014), 'Mississippi river nitrate loads from high frequency sensor measurements and regression-based load estimation', *Environmental Science and Technology* **48**(21), 12612–12619.

Pfannerstill, M., Kühling, I., Hugenschmidt, C., Trepel, M. & Fohrer, N. (2016), 'Reactive ditches: A simple approach to implement denitrifying wood chip bioreactors to reduce nitrate exports into aquatic ecosystems?', *Environmental Earth Sciences* **75**(13).

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B. & Wagener, T. (2016), 'Sensitivity analysis of environmental models: A systematic review with practical workflow', *Environmental Modelling & Software* **79**, 214–232.

Pianosi, F., Sarrazin, F. J. & Wagener, T. (2015), 'A Matlab toolbox for Global Sensitivity Analysis', *Environmental Modelling and Software* **70**, 80–85.

Poeplau, C., Jacobs, A., Don, A., Vos, C., Schneider, F., Wittnebel, M., Tiemeyer, B., Heidkamp, A., Prietz, R. & Flessa, H. (2020), 'Stocks of organic carbon in German agricultural soils—Key results of the first comprehensive inventory', *Journal of Plant Nutrition and Soil Science* **183**(6), 665–681.

Poméon, T., Diekkrüger, B., Springer, A., Kusche, J. & Eicker, A. (2018), 'Multi-objective validation of SWAT for sparsely-gauged West African river basins - A remote sensing approach', *Water (Switzerland)* **10**(4).

Pronzato, L. & Müller, W. G. (2012), 'Design of computer experiments: Space filling and beyond', *Statistics and Computing* **22**(3), 681–701.

Quaranta, E., Fuchs, S., Jan Liefting, H., Schellart, A. & Pistocchi, A. (2022), 'A hydrological model to estimate pollution from combined sewer overflows at the regional scale: Application to Europe', *Journal of Hydrology: Regional Studies* **41**, 101080.

R Core Team (2022), 'R: A Language and Environment for Statistical Computing'.
URL: *https://www.r-project.org/, as retrieved on June 30, 2024*

Radcliffe, D., Freer, J. & Schoumans, O. (2009), 'Diffuse Phosphorus Models in the United States and Europe: Their Usages, Scales, and Uncertainties', *Journal of Environmental Quality* **38**(5), 1956–1967.

Rajabi, M. M. & Ataie-Ashtiani, B. (2014), 'Sampling efficiency in Monte Carlo based uncertainty propagation strategies: Application in seawater intrusion simulations', *Advances in Water Resources* **67**, 46–64.

Rajabi, M. M., Ataie-Ashtiani, B. & Janssen, H. (2015), 'Efficiency enhancement of optimized Latin hypercube sampling strategies: Application to Monte Carlo uncertainty analysis and meta-modeling', *Advances in Water Resources* **76**, 127–139.

Rathjens, H., Oppelt, N., Bosch, D. D., Arnold, J. G. & Volk, M. (2015), 'Development of a grid-based version of the SWAT landscape model', *Hydrological Processes* **29**(6), 900–914.

Reckhow, K. H. (1994), 'Water quality simulation modeling and uncertainty analysis for risk assessment and decision making', *Ecological Modelling* **72**(1-2), 1–20.

Refsgaard, J. C. & Storm, B. (1995), MIKE SHE, *in* V. P. Singh, ed., 'Computer Models of Watershed Hydrology', Water Resources Publications, Colorado, pp. 809–847.

Refsgaard, J. C., Storm, B. & Clausen, T. (2010), 'Système Hydrologique Européen (SHE): Review and perspectives after 30 years development in distributed physically-based hydrological modelling', *Hydrology Research* **41**(5), 355–377.

Reid, A. J., Carlson, A. K., Creed, I. F., Eliason, E. J., Gell, P. A., Johnson, P. T., Kidd, K. A., MacCormack, T. J., Olden, J. D., Ormerod, S. J., Smol, J. P., Taylor, W. W., Tockner, K., Vermaire, J. C., Dudgeon, D. & Cooke, S. J. (2019), 'Emerging threats and persistent conservation challenges for freshwater biodiversity', *Biological Reviews* **94**(3), 849–873.

Rissman, A. R. & Carpenter, S. R. (2015), 'Progress on nonpoint pollution: Barriers & opportunities', *Daedalus* **144**(3), 35–47.

Ritter, A. & Muñoz-Carpena, R. (2013), 'Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments', *Journal of Hydrology* **480**, 33–45.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. & Tarantola, S. (2008), *Global sensitivity analysis: The primer*, Wiley.

Salvetti, R., Azzellino, A. & Vismara, R. (2006), 'Diffuse source apportionment of the Po river eutrophying load to the Adriatic sea: Assessment of Lombardy contribution to Po river nutrient load apportionment by means of an integrated modelling approach', *Chemosphere* **65**(11), 2168–2177.

Santos, L., Thirel, G. & Perrin, C. (2018), 'Technical note: Pitfalls in using log-transformed flows within the KGE criterion', *Hydrology and Earth System Sciences* **22**(8), 4583–4591.

Sarrazin, F. J., Pianosi, F. & Wagener, T. (2016), 'Global Sensitivity Analysis of environmental models: Convergence and validation', *Environmental Modelling and Software* **79**, 135–152.

Schaffhauser, T., Lange, S., Tuo, Y. & Disse, M. (2023), 'Shifted discharge and drier soils: Hydrological projections for a Central Asian catchment', *Journal of Hydrology: Regional Studies* **46**, 101338.

Schmadel, N. M., Harvey, J. W., Alexander, R. B., Schwarz, G. E., Moore, R. B., Eng, K., Gomez-Velez, J. D., Boyer, E. W. & Scott, D. (2018), 'Thresholds of lake and reservoir connectivity in river networks control nitrogen removal', *Nature Communications* **9**(1), 2779.

Schoumans, O. F., Silgram, M., Groenendijk, P., Bouraoui, F., Andersen, H. E., Kronvang, B., Behrendt, H., Arheimer, B., Johnsson, H., Panagopoulos, Y., Mimikou, M., Lo Porto, A., Reisser, H., Le Gall, G., Barr, A. & Anthony, S. G. (2009), 'Description of nine nutrient loss models: Capabilities and suitability based on their characteristics', *Journal of Environmental Monitoring* **11**(3), 506–514.

Schoups, G., van de Giesen, N. C. & Savenije, H. H. G. (2008), 'Model complexity control for hydrologic prediction', *Water Resources Research* **44**(12), 1–14.

Schreiber, C. (2011), Einträge, Vorkommen, Verbreitung und gesundheitliche Bedeutung antibiotikaresistenter Bakterien in Abwasser und Gewässern - Ein sozial-ökologischer Beitrag zur Geographischen Gesundheitsforschung, PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn.

Schreiber, C., Rechenburg, A., Koch, C., Christoffels, E., Claßen, T., Willkomm, M., Mertens, F. M., Brunsch, A. F., Herbst, S., Rind, E. & Kistemann, T. (2016), 'Two decades of system-based hygienic–microbiological research in Swist river catchment (Germany)', *Environmental Earth Sciences* **75**, 1393.

Schürz, C. (2019), 'SWATplusR: Running SWAT2012 and SWAT+ Projects in R'.
URL: *https://github.com/chrisschuerz/SWATplusR, as retrieved on June 30, 2024*

Schürz, C., Strauch, M., Mehdi, B. & Schulz, K. (2017), SWATfarmR : A simple rule-based scheduling of SWAT management operations, *in* 'Proceedings of the 2017 International SWAT Conference', Warsaw, Poland, pp. 97–98.

Schwientek, M. & Selle, B. (2016), 'Quantifying in-stream retention of nitrate at catchment scales using a practical mass balance approach', *Environmental Monitoring and Assessment* **188**(2), 1–12.

Seibert, J. & McDonnell, J. J. (2002), 'On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration', *Water Resources Research* **38**(11), 23–1–23–14.

Seitzinger, S. P., Harrison, J. A., Böhlke, J. K., Bouwman, A. F., Lowrance, R., Peterson, B., Tobias, C. & Van Drecht, G. (2006), 'Denitrification across landscapes and waterscapes: A synthesis', *Ecological Applications* **16**(6), 2064–2090.

Seitzinger, S. P., Styles, R. V., Boyer, E. W., Alexander, R. B., Billen, G., Howarth, R. W., Mayer, B. & Van Breemen, N. (2002), Nitrogen retention in rivers: model development and application to watersheds in the northeastern U.S.A., *in* 'The Nitrogen Cycle at Regional to Global Scales', Springer Netherlands, Dordrecht, pp. 199–237.

Shin, M. J., Guillaume, J. H., Croke, B. F. & Jakeman, A. J. (2013), 'Addressing ten questions about conceptual rainfall-runoff models with global sensitivity analyses in R', *Journal of Hydrology* **503**, 135–152.

Shirmohammadi, A., Chaubey, I., Harmel, R., Bosch, D. D., Muñoz-Carpena, R., Dharmasri, C., Sexton, A. M., Arabi, M., Wolfe, M. L., Frankenberger, J., Graff, C. & Sohrabi, T. M. (2006), 'Uncertainty in TMDL models', *Transactions of the ASABE* **49**(4), 1033–1049.

Sprenger, M., Llorens, P., Gallart, F., Benettin, P., Allen, S. T., Area, E. S., Berkeley, L. & Science, E. (2022), 'Precipitation fate and transport in a Mediterranean catchment through models calibrated on plant and stream water isotope data', **26**(15), 4093–4107.

Srinivasan, M. S., Gérard-Marchant, P., Veith, T. L., Gburek, W. J. & Steenhuis, T. S. (2005), 'Watershed Scale Modeling of Critical Source Areas of Runoff Generation and Phosphorus Transport', *Journal of the American Water Resources Association* **41**(2), 361–377.

Srinivasan, R., Zhang, X. & Arnold, J. G. (2010), 'SWAT Ungauged: Hydrological Budget and Crop Yield Predictions in the Upper Mississippi River Basin', *Transactions of the ASABE* **53**(5), 1533–1546.

State Ministry for Climate Protection, Environment, Agriculture, Nature Conservation and Consumer Protection NRW (2015), 'Steckbriefe der Planungseinheiten in den nordrhein-westfälischen Anteilen von Rhein, Weser, Ems und Maas - Oberflächengewässer und Grundwasser im Teileinzugsgebiet Rhein/Erft NRW - Bewirtschaftungszeitraum 2016-2021'.
URL: *https://www.flussgebiete.nrw.de/system/files/atoms/files/pe-stb_2016-2021_erft_final.pdf, as retrieved on June 30, 2024*

State Ministry for Environment, Agriculture, Nature Conservation and Consumer Protection NRW (2021), 'Steckbriefe der Planungseinheiten in den nordrhein-westfälischen Anteilen von Rhein, Weser, Ems und Maas - Oberflächengewässer und Grundwasser im Teileinzugsgebiet Rhein/Erft NRW - Bewirtschaftungszeitraum 2022-2027'.
URL: *https://www.flussgebiete.nrw.de/system/files/atoms/files/pe-steckbriefe_erftnrw_2022-2027.pdf, as retrieved on June 30, 2024*

State Office for Geology and Mining RLP (2019), 'Bodenflächendaten 1:50000, Reference Number 16/2019'.

State Office for Information and Technology NRW (2016), 'Agricultural Structure Survey'.
URL: *https://www.landesdatenbank.nrw.de/link/statistikTabellen/41141#abreadcrumb, as retrieved on June 30, 2024*

State Office for Land Surveying and Geoinformation RLP (2001), 'Digital Basic-Landscape Model'.

State Office for Land Surveying and Geoinformation RLP (2006), 'Digital Elevation Model (DEM10)'.

State Office for Land Surveying NRW (2002), 'Digital Elevation Model (DEM10)'.

State Office for Statistics RLP (2016), 'Agricultural Structure Survey'.

Strauch, M., Schweppe, R. & Schürz, C. (2016), 'TopHRU: Threshold optimization for HRUs in SWAT (v1.2)'.
URL: *https://doi.org/10.5281/zenodo.154379, as retrived on June 30, 2024*

Styczen, M., Thorsen, M. & Refsgaard, A. (1999), Non-point pollution modelling at different scales and resolution, based on MIKE SHE, *in* 'Proceedings of the 3rd DHI Software Conference', Helsingør, Denmark, pp. 1–25.

Sunohara, M. D., Gottschall, N., Craiovan, E., Wilkes, G., Topp, E., Frey, S. K. & Lapen, D. R. (2016), 'Controlling tile drainage during the growing season in Eastern Canada to reduce nitrogen, phosphorus, and bacteria loading to surface water', *Agricultural Water Management* **178**(3), 159–170.

Tetzlaff, B. (2021), 'Kooperationsprojekt GROWA + NRW 2021 Teil IIb: Ausweisung potenziell dränierter Flächen unter landwirtschaftlicher Nutzung in Nordrhein-Westfalen'.
URL: *https://www.lanuv.nrw.de/fileadmin/lanuvpubl/3_fachberichte/30110c.pdf, as retrieved on June 30, 2024*

Tetzlaff, B., Kuhr, P. & Wendland, F. (2009), 'A new method for creating maps of artificially drained areas in large river basins based on aerial photographs and geodata', *Irrigation and Drainage* **58**(5), 569–585.

Tripathi, M. P., Panda, R. K. & Raghuwanshi, N. S. (2003), 'Identification and Prioritisation of Critical Subwatersheds for Soil Conservation Management using the SWAT Model', *Biosystems Engineering* **85**(3), 365–379.

Tuppad, P., Douglas-Mankin, K. R., Lee, T., Srinivasan, R. & Arnold, J. G. (2011), 'Soil and Water Assessment Tool (SWAT) Hydrologic/Water Quality Model: Extended Capability and Wider Adoption', *Asabe* **54**(5), 1677–1684.

van Griensven, A. & Meixner, T. (2006), 'Methods to Quantify and Identify the Sources of Uncertainty for River Basin Water Quality Models', *Water Science and Technology* **53**(1), 51–59.

Van Meter, K. J., Basu, N. B., Veenstra, J. J. & Burras, C. L. (2016), 'The nitrogen legacy: Emerging evidence of nitrogen accumulation in anthropogenic landscapes', *Environmental Research Letters* **11**(3).

Venohr, M., Hirt, U., Hofmann, J., Opitz, D., Gericke, A., Wetzig, A., Natho, S., Neumann, F., Hürdler, J., Matranga, M., Mahnkopf, J., Gadegast, M. & Behrendt, H. (2011), 'Modelling of Nutrient Emissions in River Systems - MONERIS - Methods and Background', *International Review of Hydrobiology* **96**(5), 435–483.

Walker, D., Baumgartner, D., Gerba, C. & Fitzsimmons, K. (2019), Surface Water Pollution, *in* 'Environmental and Pollution Science', 3 edn, Elsevier Inc., pp. 261–292.

Walker, W., Harremoës, P., Rotmans, J., van der Sluijs, J., van Asselt, M., Janssen, P. & Krayer von Krauss, M. (2003), 'Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support', *Integrated Assessment* **4**(1), 5–17.

Wellbrock, N., Grüneberg, E., Riedel, T. & Polley, H. (2017), 'Carbon stocks in tree biomass and soils of German forests', *Central European Forestry Journal* **63**(2-3), 105–112.

Wendland, F., Bergmann, S., Eisele, M., Elbers, J., Gömann, H., Kreins, P. & Kunkel, R. (2021a), 'Kooperationsprojekt GROWA + NRW 2021 Teil V: Stickstoffeintrag ins Grundwasser und die Oberflächengewässer Nordrhein-Westfalens (2014-2016)'.
URL: *https://www.lanuv.nrw.de/fileadmin/lanuvpubl/3_fachberichte/30110f.pdf, as retrieved on June 30, 2024*

Wendland, F., Bergmann, S., Eisele, M., Herrmann, F., Kunkel, R., Linder, B., Schöbel, T. & Schuster, H. (2021b), 'Kooperationsprojekt GROWA + NRW 2021 Teil III: Modellierung des Sickerwassers in der ungesättigten Zone und der Fließzeiten des Grundwassers in Nordrhein-Westfalen'.
URL: *https://www.lanuv.nrw.de/fileadmin/lanuvpubl/3_fachberichte/30110d.pdf, as retrieved on June 30, 2024*

White, M. J., Harmel, R. D., Arnold, J. G. & Williams, J. R. (2014), 'SWAT Check: A Screening Tool to Assist Users in the Identification of Potential Model Application Problems', *Journal of Environmental Quality* **43**(1), 208–214.

White, M. J., Storm, D. E., Busteed, P. R., Stoodley, S. H. & Phillips, S. J. (2009), 'Evaluating Nonpoint Source Critical Source Area Contributions at the Watershed Scale', *Journal of Environmental Quality* **38**(4), 1654–1663.

Whittaker, G., Confesor, R., Di Luzio, M. & Arnold, J. G. (2010), 'Detection of overparameterization and overfitting in an automatic calibration of SWAT', *Transactions of the ASABE* **53**(5), 1487–1499.

Williams, J. R. (1975), Sediment-yield prediction with universal equation using runoff energy factor, *in* 'Present and Prospective Technology for Predicting Sediment Yield and Sources: Proceedings of the Sediment-Yield Workshop, USDA Sedimentation Laboratory, Oxford, Miss., Nov. 28-30, 1972', Vol. 40, Agricultural Research Service, US Department of Agriculture, pp. 244–252.

Winchell, M. F., Folle, S., Meals, D., Moore, J., Srinivasan, R. & Howe, E. A. (2015), 'Using SWAT for sub-field identification of phosphorus critical source areas in a saturation excess runoff region', *Hydrological Sciences Journal* **60**(5), 1–19.

Withers, P., Neal, C., Jarvie, H. & Doody, D. (2014), 'Agriculture and Eutrophication: Where Do We Go from Here?', *Sustainability* **6**(9), 5853–5875.

Wright, M. N., Wager, S. & Probst, P. (2022), 'ranger: A Fast Implementation of Random Forests'.
URL: *https://cran.r-project.org/web/packages/ranger/index.html, as retrieved on June 30, 2024*

Yang, Q., Almendinger, J. E., Zhang, X., Huang, M., Chen, X., Leng, G., Zhou, Y., Zhao, K., Asrar, G. R., Srinivasan, R. & Li, X. (2018), 'Enhancing SWAT simulation of forest ecosystems for water resource assessment: A case study in the St. Croix River basin', *Ecological Engineering* **120**, 422–431.

Yang, Q. & Zhang, X. (2016), 'Improving SWAT for simulating water and carbon fluxes of forest ecosystems', *Science of The Total Environment* **569-570**, 1478–1488.

Yang, Q., Zhang, X., Almendinger, J. E., Huang, M., Leng, G., Zhou, Y., Zhao, K., Asrar, G. R., Li, X. & Qiu, J. (2019), 'Improving the SWAT forest module for enhancing water resource projections: A case study in the St. Croix River basin', *Hydrological Processes* **33**(5), 864–875.

Yen, H., White, M. J., Jeong, J., Arabi, M. & Arnold, J. G. (2015), 'Evaluation of Alternative Surface Runoff Accounting Procedures Using the SWAT Model', *International Journal of Agricultural and Biological Engineering* **8**(3), 54–68.

Yuan, Y., Khare, Y., Wang, X., Parajuli, P. B., Kisekka, I. & Finsterle, S. (2015), 'Hydrologic and Water Quality Models: Sensitivity', *Transactions of the ASABE* **58**(6), 1721–1744.

Zambrano-Bigiarini, M. (2020), 'hydroGOF'.
URL: *https://github.com/hzambran/hydroGOF, as retrieved on June 30, 2024*

Zhong, F., Jiang, S., Van Dijk, A. I., Ren, L., Schellekens, J. & Miralles, D. G. (2022), 'Revisiting large-scale interception patterns constrained by a synthesis of global experimental data', *Hydrology and Earth System Sciences* **26**(21), 5647–5667.

Zinnbauer, M., Eysholdt, M., Henseler, M., Herrmann, F., Kreins, P., Kunkel, R., Nguyen, H., Tetzlaff, B., Venohr, M., Wolters, T. & Wendland, F. (2023), 'Quantifizierung aktueller und zukünftiger Nährstoffeinträge und Handlungsbedarfe für ein deutschlandweites Nährstoffmanagement – AGRUM-DE, Thünen Report 108'.
URL: *https://literatur.thuenen.de/digbib_extern/dn066338.pdf, as retrieved on June 30, 2024*