

**SCHRIFTENREIHE  
HYDROLOGIE/WASSERWIRTSCHAFT**

**Entwicklung eines selbstkalibrierenden  
Niederschlag-Abfluss-Modells auf Basis  
der geomorphologischen Einheitsganglinie  
und Methoden des Machine Learning**

von Henning Oppel



Die vorliegende Arbeit wurde von der Fakultät  
für Bau- und Umweltingenieurwissenschaften als Dissertation angenommen.

Doktorarbeit eingereicht am: 13.03.2019  
Tag der mündlichen Prüfung: 06.05.2019

Berichter:

Prof. Dr. rer. nat. habil. Andreas Schumann, Ruhr-Universität Bochum  
Prof. Ph. D. Thorsten Wagener, University of Bristol

Copyright: Lehrstuhl für Hydrologie, Wasserwirtschaft und Umwelttechnik  
Ruhr-Universität Bochum, 2019  
Universitätsstraße 150, 44801 Bochum  
Tel. +49 (0) 234 32-24693, Fax. -14153  
ISSN 0949-5975

Herausgeber: Prof. Dr. rer. nat. habil. Andreas Schumann

## Kurzfassung

Niederschlags-Abfluss (N-A) Modelle nutzen die deterministische Beziehung zwischen Niederschlag und Abfluss zur Vorhersage von Hochwassern, wobei die Parameter des Modells mittels gemessener Abflussganglinien kalibriert werden. In Einzugsgebieten ohne Abflussmessungen können klassische N-A Modelle nicht angewendet werden. Durch die fehlenden Daten kann keine Kalibrierung der Modelle durchgeführt werden und daher keine operationelle Hochwasservorhersage oder Berechnung von Bemessungsganglinien durchgeführt werden. Modelle, die eine geomorphologische Einheitsganglinie (GIUH) verwenden, bieten die Möglichkeit die Modellstruktur an räumliche Muster eines Einzugsgebiets anzupassen und somit wichtige Informationen über die Abflussgenese zu berücksichtigen. Die wenigen freien Parameter des Modells sowie das Einbeziehen von Gebietsstrukturen sind die wichtigsten Vorteile dieser Modelle gegenüber anderen, konzeptionellen Modellen. Der Nachteil von GIUH-Modellen ist die benötigte ereignisweise Parametrisierung. Da bislang keine Methode zur Verfügung steht diese Parameter zu schätzen, beschränkte sich die Anwendung der GIUH-Modelle bislang auf die Analyse vergangener Ereignisse.

In dieser Arbeit wurde ein Ensemble aus *Machine Learning* (ML) Algorithmen eingesetzt, um aus klimatischen und meteorologischen Randbedingungen sowie der Vorbedingungen des Einzugsgebietes ereignisbasiert die nötigen Parameter zu schätzen. Acht Algorithmen wurden verwendet und in einer *Leave-one-out* Studie in drei Untersuchungsgebieten im Süd-Osten Deutschlands getestet. Die Parameterschätzungen der ML-Algorithmen wurden an ein weiterentwickeltes GIUH-Modell übergeben, mit dessen Hilfe 2-dimensionale Niederschlagsfelder in Prognosen von Abflussganglinien transformiert wurden.

Die Modellgüte des GIUH-Modells und der ML-Algorithmen wurden separat untersucht. Es konnte gezeigt werden, dass das GIUH-Modell in der Lage ist verschiedene Einzugsgebietsformen, Verteilungen von Fließwegen sowie unterschiedliche Charakteristika der Fließwege zu berücksichtigen. Darüber hinaus wurde eine Variation der Fließgeschwindigkeit abhängig vom Fließweg implementiert. Diese Erweiterung erwies sich in den Fallstudien als sinnvolle Ergänzung. Zudem konnte gezeigt werden, dass die Parameter des Modells direkt aus Ganglinienanalysen gewonnen werden können. Eine Kalibrierung der Parameter blieb aber zu bevorzugen. Die Randbedingungen der ML-Anwendung wurden modellgestützt ausgewählt. Ermittelt wurden die besten Prädiktoren sowie eine optimale Segmentierung der Eingangsdaten. Zur Bildung eines Regionalmodells wurden alle verfügbaren Daten eines Untersuchungsgebiets zur Anpassung der ML-Algorithmen verwendet. Validiert wurden die Modelle mit zurückgehaltenen Daten eines einzelnen Teilgebietes. Der mittlere relative Fehler lag bei 20% für die Prognose der Entwässerungsgeschwindigkeit und 40% für das Abflussvolumen. Der Fehler konnte durch eine selektive Datenkomposition der Lernmenge weiter reduziert werden. Dabei wurde nur eine begrenzte Anzahl der ähnlichsten Gebiete zur Anpassung der Modelle verwendet.

Anschließend wurden die Komponenten gemeinsam angewendet. Die mittleren Gütekomponenten der Vorhersagen, die zeitliche Übereinstimmung, Volumen und Varianz der Ganglinien berücksichtigten, lagen nahe am jeweiligen Optimum. Die Varianz der Ergebnisse wies aber darauf hin, dass die Algorithmen nur als Ensemble zu verwenden sind, da die einzelnen ML-Algorithmen nicht in der Lage waren die vollständige Komplexität hydrologischer Prozesse abzubilden. Nur im Zusammenwirken der Algorithmen konnten gute Ergebnisse erreicht werden. Die Modellergebnisse wurden mit einem HBV-Modell verglichen, dessen Parameter in unbeobachteten Gebieten regionalisiert wurden. Die Ergebnisse des mit ML kombinierten GIUH Modells waren besser in unbeobachteten und gleichwertig in den beobachteten Einzugsgebieten.

Abschließend wurde die Möglichkeit getestet über die erlernten Zusammenhänge der Algorithmen Annahmen über natürliche Prozesse abzuleiten. Aus den Ergebnissen der Arbeit ließ sich ein Zusammenhang zwischen einem hohen Füllen-Scheitel Verhältnis des Ereignisses und den Niederschlagscharakteristika herausstellen. Geringe Verhältniswerte wiesen eine Verknüpfung zu Anfangszuständen der Einzugsgebiete auf.

## Abstract

In data scarce regions common hydrological models cannot be applied. Due to the missing data for calibration conceptual rainfall-runoff models cannot be parametrized and, hence, not be used for operational predictions or definition of design hydrographs. Geomorphological instantaneous unit hydrographs (GIUH) offer the unique possibility to adapt model structure to catchment structure and thereby increase model accuracy in data scarce regions. The parsimony as well as the incorporation of catchment structures is a valuable advantage for prediction in ungauged basins. The drawback of GIUH-models is the required parametrization for each individual event. Hence, applications of GIUH-models have been limited to scientific reanalysis of past rainfall-runoff events.

In this study an ensemble of machine learning (ML) algorithms was applied for the estimation of the required parameters in ungauged basins. Indicators of meteorological forcing and initial catchments states were used as predictors for the estimation of drainage velocity and runoff coefficient. Eight algorithms were applied and their performance has been evaluated in a leave-one-out study in three major catchments in South-East Germany. Predictions provided by the algorithms were given to an improved GIUH-model to transform 2-dimensional precipitation data into an ensemble prediction of hydrographs in ungauged basins.

The performance of the improved GIUH-model and the ML-Algorithms were evaluated separately. The GIUH-structure proved to be as flexible as demanded. In a synthetic case study it was able to incorporate different catchment shapes, flowpath distributions and characteristics into the shape of predicted hydrographs. A variation of drainage velocity by flowpath was implemented and improved simulation results. Moreover, a parametrization directly from rainfall-runoff event analysis seemed possible, yet calibrated parameters led to better performance. The setup of the ML-module has been evaluated with respect to the predictors and data segmentation by model approach. In a subsequent regional application, data from all available gauges were used to train the algorithms. Withheld data was used to imitate a prediction in ungauged basin. The models showed an average relative error for drainage velocity of 20% and 40% for runoff volume. The error were lower afterwards by selective data composition, considering only a limited number of similar catchments for model training.

The combination of both model components were tested subsequently. The mean efficiencies considering hydrograph timing, volume and variance were close to optimum value. Yet the model worked only in ensemble mode, because a single ML-algorithms proved not to be capable of imitating the full range of hydrological complexity. A comparison to a regionalized HBV-model showed superior results for the GIUH-ML model in ungauged catchments and equal results for gauged catchments. Finally, the possibility of deriving assumptions about hydrological processes from trained ML-dependencies has been discussed. For the performed case studies an assumption about changing dependencies of driving factors and the resulting ratio of flood volume and peak was derived.

## **Vorwort**

War die Hydrologie in der Vergangenheit eher durch eine Knappheit an Daten geprägt, so haben neue automatisierte Datenerfassungssysteme die zeitliche Auflösung der Abflusswerte wesentlich erhöht. Statt mittleren Tageswerten stehen nunmehr Wasserstände bzw. Abflüsse im Abstand von 5 Minuten digital zur Verfügung. Daraus und aus den nunmehr digital verfügbaren Zeitreihen von einigen Jahrzehnten Länge ergeben sich neue Möglichkeiten der Datenanalyse. Die Informatik im Bereich der Künstlichen Intelligenz bietet hierzu in Form des Machine Learnings geeignete Hilfsmittel an, um aus der Analyse von „Big Data“ neue Erkenntnisse zu gewinnen. Herr Oppel hat sich in seiner Arbeit intensiv mit diesen Möglichkeiten der Datenanalyse befasst, in dem er die Zusammenhänge zwischen Niederschlag und Abfluss bei Hochwasser durch Auswertung einer Vielzahl von Ereignissen mit Hilfe des Machine Learnings charakterisierte. In der Vergangenheit wurde oftmals versucht, derartige Zusammenhänge durch die Auswertung einer möglichst großen Zahl von abgelaufenen Hochwasserereignissen zu ermitteln und in Form von nichtlinearen statistischen Beziehungen, meist in graphischer Form als Koaxialdiagramme, für die Abschätzung möglicher Reaktionen unbeobachteter Gebiete zu nutzen. Die Problematik dieser Ansätze bestand neben der begrenzten Datenverfügbarkeit in der großen Variabilität der hydrologischen Prozesse, die die Allgemeingültigkeit derartiger Beziehungen in Frage stellte. Herr Oppel hatte sich das Ziel gesetzt, mit modernen Verfahren einen vergleichbaren Ansatz zu entwickeln, der in Einzugsgebieten ohne Abflussmessungen angewendet werden kann. Er wählte hierzu als Grundmodell die geomorphologische Impulsantwort (GIUH), da diese die Möglichkeit bietet, die Modellstruktur mit Hilfe einer GIS-basierten Analyse des Gewässernetzes an die Struktur des jeweiligen Einzugsgebiets anzupassen. Somit wurden zunächst die Abflusskonzentrationsbedingungen analytisch definiert, wodurch sich die Zahl der freien Parameter verringerte. Für die Anwendung war es jedoch notwendig, zwei weitere ereignisabhängig stark variierende Parameter, die Entwässerungsgeschwindigkeit und das Abflussvolumen, zu ermitteln. Hierzu nutzte Herr Oppel Verfahren des Machine Learning (ML), um unter Beachtung der hydrologischen Anfangsbedingungen und der klimatischen und meteorologischen Randbedingungen zunächst die Hochwasserereignisse zu kategorisieren und dann die Parameter in Abhängigkeit vom Ereignistyp zu schätzen. In seiner Studie in drei Flussgebieten in Süddeutschland wendete er vergleichend acht Machine Learning Algorithmen an. Der somit entwickelte und parametrisierte Modellansatz ist seinem Anwendungsergebnissen mit denen eines herkömmlichen konzeptionellen Modells vergleichbar und belegt die volle Funktionalität der gewählten Methodik. Herr Oppel hat mit seiner durch die Anwendung von Verfahren des Machine Learning originäre Vorgehensweise sein Ziel, einen neuen modellbasierten Regionalisierungsansatz zu entwickeln, erreicht. Mit dieser Arbeit wurde ein wichtiger innovativer Beitrag zur Fortschreibung hydrologischer Verfahren unter Anwendung neuer Ansätze der Informatik vorgelegt, der hoffentlich eine breitere Anwendung finden wird.

Prof. Andreas Schumann

## **Danksagung**

Diese Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Hydrologie, Wasserwirtschaft und Umwelttechnik der Ruhr-Universität Bochum.

Mein besonderer Dank gilt an dieser Stelle Herrn Prof. Dr. rer. nat. Andreas Schumann für die Betreuung und Ermöglichung dieser Arbeit.

Herrn Prof. Ph. D. Thorsten Wagener danke ich sehr herzlich für die Übernahme des Korreferats.

Benjamin Mewes, Christoph Tyralla, Svenja Fischer und Gordon Horn möchte ich für die wertvollen Diskussionen danken, die diese Arbeit sehr bereicherten. Ihnen und allen weiteren aktuellen und ehemaligen Kollegen am Lehrstuhl möchte ich für ihre Unterstützung danken.

Zu guter Letzt danke ich meiner Familie für ihre Hilfe und Unterstützung ohne die diese Arbeit nicht möglich gewesen wäre.

# Inhaltsverzeichnis

---

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Niederschlag-Abfluss-Modellierung in unbeobachteten Einzugsgebieten</b>	<b>5</b>
2.1	Methoden.....	5
2.1.1	Regionalisierung .....	6
2.1.2	Physikalisch begründete Parametrisierung .....	8
2.1.3	Regionale Strategien .....	8
2.2	Konzeptionelle Niederschlags-Abfluss Modelle.....	9
2.3	Datengetriebene Modelle .....	10
2.4	Zusammenfassung und Ansatzpunkte.....	13
<b>3</b>	<b>Grundlagen und Datenaufbereitung</b>	<b>15</b>
3.1	Untersuchungsgebiete .....	15
3.1.1	Oberer Main.....	17
3.1.2	Regen .....	19
3.1.3	Iller.....	20
3.1.4	Sonstige Gebiete .....	22
3.2	Ereignisindikatoren .....	23
3.2.1	Zielgrößen.....	23
3.2.2	Klimatische Randbedingungen .....	25
3.2.3	Anfangszustände.....	25
3.3	Prozessorientierte Klassifikation.....	28
3.3.1	Klassifikation mittels Fuzzy-Entscheidungsbaum.....	29
3.3.2	Klassifikation mittels Füllen-Scheitel Verhältnis .....	32
3.4	Güteparameter .....	33
<b>4</b>	<b>Modellentwicklung</b>	<b>35</b>
4.1	Zielsetzung .....	36
4.2	Darstellung der Geomorphologie .....	37
4.3	Abflussdynamik .....	41

4.4	Modellsynthese .....	44
4.4.1	Niederschlag .....	45
4.4.2	Hangabfluss .....	47
4.4.3	Gewässerabfluss .....	51
4.4.4	Dynamikmodul .....	53
4.4.4.1	Multiple lineare Regression .....	55
4.4.4.2	Machine Learning Algorithmen .....	55
4.4.4.3	Ensemble Machine Learning Schätzer .....	58
4.5	Zusammenfassung Modellentwicklung .....	59
<b>5</b>	<b>Analyse Abflussmodell</b> .....	<b>62</b>
5.1	Funktionsweise Modellstruktur .....	63
5.2	Sensitivität Parametrisierung .....	67
5.3	Intra-Ereignis Variation analytischer Dynamikparameter .....	70
5.3.1	Räumliche Variation .....	70
5.3.2	Zeitliche Variation .....	73
5.4	Fehleranalyse und Anpassung .....	75
5.4.1	Fehleranalyse Teilgebiete .....	76
5.4.2	Anpassung der Modellparameter .....	78
5.4.3	Ereignisanalyse .....	82
5.5	Zusammenfassung Abflussmodell .....	84
<b>6</b>	<b>Analyse Dynamikmodul</b> .....	<b>86</b>
6.1	Vorhersage Klassifikation & Selektion .....	87
6.2	Prozessorientierte Unterscheidung .....	90
6.3	Prädiktoren .....	91
6.3.1	Kombinationen .....	91
6.3.2	Anwendung der Kombinationen .....	93
6.4	Algorithmen .....	95
6.5	Zusammenfassung Dynamikmodul .....	100

<b>7</b>	<b>Datenkomposition</b>	<b>103</b>
7.1	Einfluss der Lernmenge auf Regionalisierung .....	104
7.2	Metriken zur Schätzung von Ähnlichkeit.....	106
7.2.1	Centroid .....	106
7.2.2	Top-Kriging .....	107
7.2.3	DFF-Ranking .....	108
7.3	Selektive Datenkomposition .....	109
7.4	Alternative Methoden.....	114
<b>8</b>	<b>Ergebnisse der Fallstudien</b>	<b>116</b>
8.1	Einfluss Modellstruktur.....	119
8.2	Ensemble Performance.....	120
8.3	Güte individueller Prognosen.....	126
8.4	Vergleich Kalibrierungsergebnisse .....	128
8.5	Einfluss Fehlklassifikationen .....	130
8.6	Abhängigkeit Modellgüte zu Ereignisgröße .....	131
8.7	Ereignischarakteristika .....	135
<b>9</b>	<b>Weiterführende Analysen</b>	<b>139</b>
9.1	Komparative Modellstudie.....	139
9.2	Prozessabbildung der ML-Algorithmen.....	143
<b>10</b>	<b>Zusammenfassung</b>	<b>149</b>
<b>11</b>	<b>Fazit und Ausblick</b>	<b>154</b>
<b>12</b>	<b>Literaturverzeichnis</b>	<b>157</b>
	<b>Abbildungsverzeichnis</b>	<b>169</b>
	<b>Tabellenverzeichnis</b>	<b>175</b>
	<b>Abkürzungs- und Symbolverzeichnis</b>	<b>176</b>
	<b>Anhänge</b>	<b>179</b>

# 1 Einleitung

---

Die Variabilität des Abfluss in natürlichen Flussläufen stellt für die Ingenieurwissenschaften eine duale Herausforderung dar. Zum einen erfüllen Flüsse eine Versorgungsleistung, indem sie beispielsweise Trinkwasser bereitstellen oder als Verkehrsweg genutzt werden können. Zum anderen stellen sie eine Gefahr für die Gesellschaft dar. Dabei wiegen Niedrigwasserereignisse, bei denen die Versorgungsleistung nicht mehr erfüllt werden kann, aus volkswirtschaftlicher Sicht genauso schwer wie Hochwasserereignisse, die Schäden an Infrastruktur und Gebäuden verursachen. Extreme Hochwasserereignisse erzeugten im Zeitraum von 1961 – 2000 deutschlandweit einen jährlichen Schaden von ca. 465 Mio. € pro Jahr (Hattermann et al., 2014). Hochwasserereignisse stellen zudem eine Gefahr für die Anwohner von Gewässerläufen und deren Umland dar. Daher kommt der Senkung von Hochwasserrisiken eine besondere Rolle zu.

Das Hochwasserrisiko definiert sich über die Wahrscheinlichkeit von extremen Ereignissen und den zu erwartenden Schäden bei Eintritt der Ereignisse. Vor allem die Häufigkeit von Starkregenereignissen und den daraus resultierenden Hochwasserereignissen nimmt nach Studien von Murawski, Zimmer & Merz (2016) und Uhlemann, Thieken & Merz (2010) in Deutschland zu. Dem damit verbundenen Anstieg des Hochwasserrisikos kann und muss durch Hochwasserschutzmaßnahmen entgegengewirkt werden. Dies umfasst bauliche Vorsorge, bspw. durch den Bau von Poldern, Entlastungsgerinnen oder Deichen, aber auch Bau- und Verhaltensvorsorge, die auf die Reduzierung von Hochwasserschäden abzielen.

Für alle Maßnahmen ist es jedoch notwendig die Ausmaße eines Hochwasserereignisses vor dessen Eintreten abzuschätzen. Sei es zur Bestimmung von Hochwassermarken für die Dimensionierung von Hochwasserschutzanlagen, oder die operationelle Vorhersage eines Ereignisses, um die Bevölkerung zu warnen und eventuell Evakuierungsmaßnahmen einzuleiten. Für jede dieser Aufgaben werden Niederschlags-Abfluss (N-A) Modelle verwendet, die in einer Simulation über einen kurzen Zeitraum die Transformation von Starkregen in Gewässerabfluss modellieren und somit Vorhersagen oder Bemessungsganglinien produzieren.

Es existiert keine einheitliche hydrologische Theorie mit der zweifelsfrei die Genese von Abfluss aus Niederschlag beschrieben werden kann (Sivapalan, 2005), daher gibt eine Vielzahl von N-A Modellen, die auf den verschiedensten Theorien bzw. Beobachtungen basieren. Konzeptionelle Modelle basieren dabei häufig auf abstrahierten Prozessgleichungen, die sich aus, meist kleinräumigen, Beobachtungen hydrologischer Prozesse der Mikroskala (d.h. wirksam auf einer Fläche  $\leq 1\text{km}^2$ ) ergaben. Zwar entsprechen die Gleichungen den beobachteten Prozessen, aber sie gelten nur auf der jeweiligen Prozessskale. Angewendet auf ein

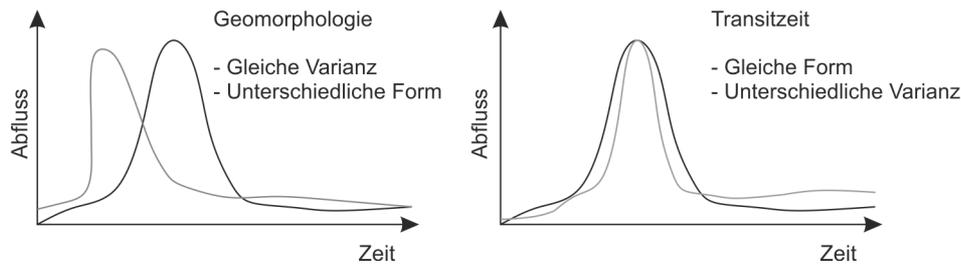
natürliches Einzugsgebiet in der Mesoskala (2 - 2000 km<sup>2</sup>) können aber keine homogenen Bedingungen mehr angenommen werden und somit ist die Validität der Prozessgleichungen fraglich. Zwar sind die Prozesse weiterhin existent, auf der Skala des Einzugsgebiets vermischen sie sich aber mit anderen Prozessen oder werden überlagert.

Der Übertrag der Prozessgleichungen auf die räumlich (und auch zeitlich) höhere Mesoskala ist erforderlich, da dies die wasserwirtschaftliche Beobachtungsskala ist (Blöschl & Sivapalan, 1995). Pegel, die den Abfluss an einem Gewässer aufzeichnen, registrieren das Integral aller abflussbildenden Prozesse des oberliegenden Einzugsgebietes, das selten eine Fläche < 1 km<sup>2</sup> umfasst. Diese Diskrepanz zwischen Beobachtungs- und Prozessskala macht die Kalibrierung von N-A Modellen erforderlich. Mit Hilfe von Aufzeichnungen vergangener N-A Ereignisse werden die Parameter eines Modells angepasst, um mit Hilfe dieser Parameter zukünftige Ereignisse simulieren zu können. Dieses Vorgehen ist nicht nur auf konzeptionelle N-A Modelle beschränkt sondern auch für systemanalytische Ansätze vonnöten. Systemanalytische Ansätze verwenden keinerlei Prozessannahmen sondern stellen eine rein empirische Beziehung zwischen Niederschlag und Abfluss her.

Für die ingenieurhydrologische Anwendung ist diese Form der Modellanpassung nur bedingt praktikabel. Um kalibrierte Parameter für zukünftige Ereignisse anwenden zu können, muss das System Einzugsgebiet stationär sein, ein Zustand der durch anthropogene Einflüsse oder veränderte klimatische Bedingungen nur selten gilt (Wagener et al., 2010). Einige systemanalytische Ansätze benötigen darüber hinaus ereignisbezogene Parameter, d.h. ihre Parameter werden nicht als konstant angenommen sondern müssen für jede Anwendung neu geschätzt werden. Häufig sind Modellaussagen für Orte und Regionen erforderlich, in denen keine oder nur unzureichende Daten vorliegen. Diese unbeobachteten Gebiete stellen eine besondere Herausforderung dar. Modelle können hier nicht angepasst werden, daher müssen sie über regionale Ansätze parametrisiert werden.

Neben der Anwendung zu Zwecken von wasserwirtschaftlichen Planungen können und werden N-A Modelle aber auch angewendet, um Erkenntnisse über die hydrologischen Prozesse zu gewinnen. In diesem Zusammenhang können sich konzeptionelle Modelle als kontraproduktiv erweisen, da die ihnen zugrunde liegenden Prozessgleichungen die Identifikation von Prozessen der Mesoskala versperren (Klemes, 1983). Um Einblick in die Prozesse auf der Ebene von Einzugsgebieten zu erlangen, müssen also so wenig Prozessannahmen wie möglich getroffen werden und so viele Beobachtungen wie möglich in die Struktur des Modells einfließen.

Geomorphologische Einheitsganglinien (GIUH) Modelle bieten dafür eine Möglichkeit. Bei diesem Modellkonzept werden Einzugsgebiete als funktionale Einheiten betrachtet, d.h. ein Einzugsgebiet definiert als Ganzes die Systemreaktion auf einen eingehenden Niederschlagsimpuls. Die Form der Systemreaktion (auch Einheitsganglinie genannt) definiert sich über die Verteilung von Fließwegen, d.h. die Geomorphologie innerhalb eines Einzugsgebietes. Die Geschwindigkeit des Abflusses bzw. die Transitzeit des Niederschlagswassers definiert die Streuung der Ganglinie. Dieser Unterschied ist in Abbildung 1-1 verdeutlicht.



**Abbildung 1-1: Einfluss (schematisch) der Geomorphologie und Transitzeit auf GIUH-Ganglinien**

Das grundlegende Problem aller systemanalytischen Ansätze gilt aber auch für den GIUH: Eine ereignisweise Parametrisierung des Modells ist erforderlich (Rodríguez-Iturbe & Valdés, 1979; Beven, 1989). Ohne eine Möglichkeit die Transitzeit bzw. die Entwässerungsgeschwindigkeit zu schätzen, können GIUH-Modelle nur zur Reanalyse vergangener Ereignisse verwendet werden. Eine weitere Kritik an gängigen GIUH-Modellen ist die einheitliche Verteilung der Entwässerungsgeschwindigkeit über die Fließwege und auch innerhalb der Fließwege (Blöschl & Sivapalan, 1995). Zwar bestehen in dieser Hinsicht erste Ansätze zu einer differenzierten Betrachtung (Rigon et al., 2016), aufgrund der ungelösten Problematik der Geschwindigkeitsschätzung finden sie bislang jedoch wenig Anwendung.

Um die Vorteile von GIUH-Modellen für praktische Anwendungen nutzen zu können, müssen diese Hürden überwunden werden. In dieser Arbeit soll die Möglichkeit getestet werden mit maschinellem Lernen (*Machine Learning*, ML) die benötigte ereignisweise Schätzung der Entwässerungsgeschwindigkeit für den GIUH durchzuführen. Ein gekoppeltes GIUH-ML Modell soll entwickelt werden, dass sich vor allem für die Anwendung in unbeobachteten Gebieten eignet. Dies erfordert nicht nur den Einsatz von ML sondern auch einen neu konzipierten Aufbau eines GIUH-Modells. Eine adaptive Grundstruktur des Modells soll entwickelt werden, um ein Maximum von verfügbaren Gebietsinformationen und Randbedingungen der N-A-Simulation (bspw. 2-dimensionale Niederschlagsdaten) zu verwenden.

Warum aber gerade ML zur Erstellung der Prognose? In diesem Fall ermöglicht ML eine probabilistische Vorhersage der benötigten Parameter ohne a-priori Prozessannahmen und Gleichungen zu entwickeln. Das Modell ist somit in der Lage sich im Anwendungsfall selbsttätig zu kalibrieren. Die Algorithmen wurden in dieser Arbeit eingesetzt, um einen Zusammenhang aus meteorologischen Bedingungen und Anfangszuständen der Gebiete zur Entwässerungsgeschwindigkeit und Abflussbildung zu entwickeln. Datengrundlage bildeten dabei N-A Beobachtungen verschiedener Einzugsgebiete im Süd-Osten Deutschlands.

Mit der Kombination aus GIUH-Modell und ML-Algorithmen zur Selbstkalibrierung können beide Ziele einer hydrologischen Simulation erfüllt werden. Zum einen konnte eine Ganglinienprognose für wasserwirtschaftliche Zwecke, auch in unbeobachteten Gebieten, erstellt werden. Zum anderen ermöglicht der Hypothesen-freie Ansatz der ML-Algorithmen die Entwicklung von Prozess-Hypothesen, basierend auf den Modellergebnissen. Mittels des maschinellen Lernens können so Theorien über Prozesse der Mesoskala entwickelt werden.

In Kapitel 2 wird zunächst eine Übersicht über den Stand der Forschung im Bezug zur Vorhersage in unbeobachteten Gebiete gegeben. Der Fokus liegt dabei auf den Methoden und Modellen. Dabei wird auch auf die Verwendung von datengetriebenen Modellen, zu denen ML gehört, im Kontext hydrologischer Anwendungen eingegangen.

Anschließend werden die Untersuchungsgebiete des oberen Mains, Regen und Iller im Kapitel 3 vorgestellt. Aus diesen Gebieten stammen die Daten, die in dieser Arbeit verwendet wurden, um das Modell und die Kopplung zu den Algorithmen zu entwickeln und zu validieren. Die dafür verwendeten Gütekriterien werden ebenfalls in diesem Kapitel eingeführt.

Die Kopplung der Modellkomponenten GIUH und ML wird in Kapitel 4 vorgestellt. Zunächst werden die Ziele des Modells genauer definiert und die grundlegende Funktion zur Erfassung der Geomorphologie vorgestellt. Es folgt eine detaillierte Darstellung der einzelnen Modellkomponenten von Niederschlag, Hang- und Gewässerabfluss. Abschließend wird das Dynamikmodul des Modells vorgestellt, das die Schätzung der Ereignisparameter mittels ML beinhaltet.

Für beide Komponenten musste getrennt voneinander geprüft werden, ob sie in ihrer Umsetzung den definierten Zielen entsprachen. Die GIUH-Struktur sollte Variationen von Gebietsform und Fließwegverteilungen berücksichtigen und die Fließgeschwindigkeit sollte die Varianz der Ganglinien beeinflussen. Darüber hinaus war die Parametrisierung des Modells zu diskutieren, denn es war fraglich, ob die zu schätzenden Parameter auf Ganglinienanalysen beruhen konnten oder ob sie an die Modellstruktur angepasst werden mussten. Die Analysen sowie ihre Ergebnisse werden in Kapitel 5 vorgestellt.

Die nachfolgenden Kapitel 6 und 7 befassen dann mit den ML-Algorithmen. In Kapitel 6 werden die Analysen dargestellt, mit denen die Eignung der ML-Algorithmen zur Vorhersage der Parameter mit einer regionalen Modellstrategie überprüft wurde. Zudem wird die Auswahl der Prädiktoren sowie eine optimale Sortierung der Eingangsdaten in diesem Kapitel diskutiert. Kapitel 7 befasst sich dann weiter mit den Daten und wie eine selektive Auswahl der Daten die Vorhersage der ML-Algorithmen verbessern kann.

Beide Komponenten wurden dann gemeinsam in einer Fallstudie angewendet. Die Ergebnisse und Diskussionen werden in Kapitel 8 vorgestellt. Die Ergebnisse wurden in weiterführenden Analysen mit einem konzeptionellen Modell verglichen. Kapitel 9 zeigt diese Ergebnisse sowie eine Analyse der Arbeitsweise der ML-Algorithmen, um daraus eine Hypothese über Prozesse zu entwickeln.

In den folgenden Kapiteln 10 und 11 werden die Ergebnisse dieser Arbeit zusammengefasst, ein Fazit gezogen und ein Ausblick auf die Entwicklungspotentiale sowie weitere laufende Arbeiten auf diesem Themenfeld gegeben.

## 2 Niederschlag-Abfluss-Modellierung in unbeobachteten Einzugsgebieten

---

Der Anwendung von Niederschlags-Abfluss (N-A) Modellen in Einzugsgebieten ohne Pegelaufzeichnungen wurde in der wissenschaftlichen Literatur viel Aufmerksamkeit gewidmet. Nicht erst seit der Dekade der „*Predictions in ungauged basins (PUB)*“ (Sivapalan et al., 2003), die von der von der internationalen Assoziation der hydrologischen Wissenschaft initiiert wurde, stellt die Problematik der N-A-Simulation in unbeobachteten Gebiete ein zentrales Problem der Hydrologie dar. Schon frühe Arbeiten, wie von Klemes (1983), Dooge (1986) oder Beven (1989), befassen sich mit dem Problem der fehlenden Allgemeingültigkeit hydrologischer Modelle und deren Annahmen über Prozesse. Die dadurch mangelnde Übertragbarkeit von Modellen auf unbeobachtete Gebiete bedingt die Problematik der Vorhersage von Abflussganglinien in diesen Gebieten. Eines der Ziele der PUB-Initiative war daher das Verständnis von hydrologischen Prozessen verschiedener Skalen zu verbessern, um damit den Grundstein für eine einheitliche hydrologische Theorie zu legen (Sivapalan et al., 2003; Sivapalan, 2005). Da dieses Ziel nur auf die lange Sicht zu erreichen war bzw. ist, wurde auch verstärkt nach Wegen gesucht mit bestehendem Prozessverständnis und Modellsätzen zu adäquaten Vorhersagen in unbeobachteten Gebieten zu gelangen.

Zum Abschluss der PUB-Initiative wurden von Blöschl et al. (2013), Hrachowitz et al. (2013) und Singh, Mishra & Jain (2014) umfassende Zusammenfassungen zur PUB-Methodik und Forschungsergebnissen veröffentlicht. Während Blöschl et al. (2013) einen Überblick über die Methoden zur Vorhersage verschiedenster Zielgrößen in unbeobachteten Gebieten geben, wurden von Hrachowitz et al. (2013) der wissenschaftliche Fortschritt während der Dekade zusammengefasst. Singh, Mishra & Jain (2014) verfassten eine ergänzende Übersicht zur Anwendung von Einheitsganglinien-Verfahren im Kontext der Anwendung in unbeobachteten Gebieten. Ausgehend von den Ergebnissen der PUB-Dekade soll in diesem Abschnitt der Stand der Forschung zum Thema N-A-Modellierung in unbeobachteten Gebieten dargestellt werden.

### 2.1 Methoden

Es kann zwischen drei grundlegenden Herangehensweisen bei einer N-A Modellierung in unbeobachteten Gebieten differenziert werden: Regionalisierung, regionale Modellierung und die physikalische Begründung von Parametern (Wagener, Wheeler & Gupta, 2004; Hrachowitz et al., 2013). Für die unterschiedlichen Zielgrößen eignen sich unterschiedliche

Methoden und dabei auch unterschiedliche Herangehensweisen. Im Folgenden werden die Methoden grundlegend vorgestellt und Anwendungsbeispiele gegeben.

### 2.1.1 Regionalisierung

Regionalisierung beschreibt den Prozess der räumlichen Übertragung einer ortsspezifischen Beobachtung auf einen unbeobachteten Punkt. Die Übertragung findet dabei unter Zuhilfenahme einer Übertragungsfunktion statt. Die Übertragungsfunktion ist der fundamentale Teil der Regionalisierung. Sie kann prinzipiell auf einem vermuteten Zusammenhang zwischen Zielvariable und Einzugsgebietenkennwerten, auf geographischer Nähe oder auf dem Prinzip homogener Gruppen beruhen.

Prinzipiell liegen all diesen Ansätzen vermutete Zusammenhänge zugrunde, Regressionsmodelle stellen aber einen direkten funktionellen Zusammenhang zwischen kalibrierten Modellparametern und Kennwerten am Ort der Parameteranpassung her. Über die aufgestellte Regressionsgleichung können die Parameter anschließend übertragen werden. Soulsby, Tetzlaff & Hrachowitz (2010) stellen in ihrer Arbeit einen funktionalen Zusammenhang zwischen verschiedenen Bodenkennwerten und Verweilzeiten von Wasser innerhalb der von ihnen betrachteten Gebiete her. Steinschneider, Yang & Brown (2015) stellen dagegen topographische Kennzahlen in Zusammenhang mit den Modellparametern eines kalibrierten konzeptionellen Modells. In einer vergleichenden Studie wendeten Brunner et al. (2018) verschiedene Strategien zur Erstellung von Regressionsgleichungen an, um den Zusammenhang zwischen Topographie und Geomorphologie mit Parametern von Einheitsganglinien zu beschreiben. So wie Rice et al. (2015) verwenden sie außerdem einen Machine Learning Regressor, um eine funktionale Beziehung automatisiert herstellen zu lassen (vgl. Kapitel 2.3.).

Alternativ wird ein auch eine Übertragung von Parametern getestet, Brunner et al. (2018) verwenden als Grundlage für den Übertrag eine Gebietsklassifikation. Bei diesem Ansatz werden Gebiete ermittelt, die als hydrologisch ähnlich angesehen werden. Anschließend wird der Mittelwert der Parameter aller ähnlichen Gebiete bestimmt und in dem unbeobachteten Gebiete verwendet. Als Metrik zur Beurteilung von Ähnlichkeit dienen in dieser Studie Ansätze, die auf der minimalen Distanz zwischen verschiedenen Kennwerten der Gebiete basieren, unter anderem Größe, Lage, Bodenkennwerte, Gefälle etc. Die Definition von homogenen Gebieten beruht dabei auf der Vorstellung von funktionalen Einheiten, wie den von Winter (2001) eingeführten hydrologischen Landschaften oder dem Konzept von Hydrotopen (Flügel, 1995). Hydrologische Landschaften beziehen sich dabei auf die Gesamtheit eines Einzugsgebietes, während beim Konzept der Hydrotopen (engl. *hydrological response units*, HRU) ein Einzugsgebiet aus einer Vielzahl von interagierenden HRU besteht. In einer von Casper, Gronz & Gemmar (2015) vorgestellten Studie wurden bspw. rasterbasiert Landnutzung, Gefälle und Bodenkennwerte miteinander verschnitten und mit hydrologischen Prozessen verknüpft. Ähnlich gehen Gharari et al. (2011) vor, die über Höhe eines Punktes

zum nächsten Gewässer (Nobre et al., 2011) ein Einzugsgebiet in funktionale Einheiten unterteilen. Ooppel & Schumann (2017) schlagen eine fließstreckenorientierte Betrachtung von Einzugsgebieten vor sowie eine Unterteilung basierend auf der Minimierung von Boden- und Topographiekennwerten. Die Ähnlichkeit von Gebieten wurde basierend auf der Lage zum Entwässerungspunkt definiert. Eine der bekanntesten Klassifikationen von Einzugsgebieten ist die HOST-Klassifikation von Einzugsgebieten in Großbritannien von Boorman, Hollis & Lilly (1995), die unter anderem von Dunn & Lilly (2001) zur Übertragung von Modellparametern oder von Yadav, Wagener & Gupta (2007) zur Übertragung von hydrologischen Signaturen zwischen den klassifizierten Gebieten angewendet wurde. Bárdossy (2007) begründete die von ihm durchgeführte räumliche Übertragung auf der Ähnlichkeit der von ihm betrachteten Gebiete des Rheins. Die verwendete Ähnlichkeitsmetrik basiert dabei auf einer Hauptkomponentenanalyse verschiedener Gebietskennwerte. Auf einer wesentlich größeren Skala arbeiteten Beck et al. (2016). Basierend auf einer global angewendeten Ähnlichkeitsmetrik, die vor allem klimatische Bedingungen berücksichtigte, wurden Parameter des HBV-Modells (Lindström et al., 1997) auf globale Einzugsgebiete übertragen. Masih et al. (2010) regionalisierten ebenfalls Parameter eines HBV-Modells, aber innerhalb eines meso-skaligen Einzugsgebietes. In der von ihnen durchgeführten Studie verwenden sie verschiedene Übertragungsfunktionen, basierend auf Ähnlichkeit und auf räumlicher Nähe. Parameter, die über die räumliche Nähe abgeleitet wurden, zeigen in dieser Arbeit die besten Modellergebnisse. Dies deckt sich nicht mit anderen Studien wie Steinschneider, Yang & Brown (2015).

Übertragungsfunktionen, die auf räumlicher Nähe basieren, verwenden meist einfache Ansätze wie die Distanz der Gebietsschwerpunkte (Masih et al., 2010; Patil & Stieglitz, 2015; Zhang et al., 2018), oder die Verwendung von Unter-/Oberlieger Beziehungen (Yao et al., 2014). Skøien, Merz & Blöschl (2006) entwickelten einen Ansatz, in dem räumliche Nähe, die Unter-/Oberlieger Beziehung sowie die Varianz der Zielgröße in die Parameter der Übertragungsfunktion einbezogen werden können.

Aus diesen Studien sowie aus den ausführlichen Darstellungen von Hrachowitz et al. (2013) lässt sich zusammenfassend keine einheitliche Empfehlung zur Regionalisierung von Modellparametern geben. Die Auswahl der Methodik sollte abhängig vom Anwendungsgebiet, Zielvariable und Datenverfügbarkeit gewählt werden. Eine Mehrzahl der Studien zeigen aber, dass Topographie, Geomorphologie und regionale klimatische Bedingungen mit hydrologischen Kennwerten verknüpft sind. Rice et al. (2015) weisen zudem darauf hin, dass diese Kennwerte mit Veränderungen im Abflussregime verknüpft sind. Chen et al. (2015) zeigten außerdem in einer Messkampagne, dass die Ähnlichkeit von Ganglinienformen durch räumliche Nähe erklärt werden kann, alle weiteren Faktoren aber durch andere Kennwerte erklärt werden müssen.

### **2.1.2 Physikalisch begründete Parametrisierung**

Eine direkte Bestimmung von Modellparametern mittels Gebietskennwerten, ohne Regionalisierung, wird als physikalisch begründete Parametrisierung bezeichnet. Diese Methode eignet sich vorrangig zur Parametrisierung von physikalisch basierten Modellen wie bspw. SHE (Abbott et al., 1986). Dabei werden stets die einzelnen Teilprozesse der Abflussgenese parametrisiert. Welche Parameter über welche Beziehung bestimmt werden ist abhängig von der Wahl des Modells. Es bleibt aber zu beachten, dass diese Parametrisierung nie fehlerfrei sein kann. Die Heterogenität natürlicher Einzugsgebiete erlaubt bislang keinen direkten Übertrag, selbst auf voll-verteilte physikalische Modelle (Minns & Hall, 2005).

Dennoch können Parameter, die über Pedo-Transferfunktionen bestimmt wurden, eine solide erste Schätzung darstellen. Weitgehend etabliert sind Ansätze mit denen die Parameter einzelner Prozessgleichungen bestimmt werden können. Als Beispiele können der Ansatz von Rawls, Brakensiek & Saxton, (1982) zur Ermittlung der Parameter für das Green-Ampt Verfahren (Herber Green & Ampt, 1911), der Curve Number Ansatz des SCS-Verfahrens (Mishra & Singh, 2013) oder die Parametrisierung des Muskingum-Verfahrens aus Gewässerkennzahlen (Cunge, 1969; Todini, 2007) genannt werden. Ihnen allen liegt die Annahme der Homogenität des Bodens, Landnutzung oder der Gewässerstrecke zugrunde, die im Falle von meso-skaligen Einzugsgebieten in den wenigsten Fällen gegeben ist.

Hughes & Kapangaziwiri (2007) stellen in ihrer Studie eine Methode zur Parametrisierung eines vollständigen Modells zur Simulation von Grundwasserabflüssen vor. In ihrer Studie stellen sie einen Bezug zwischen der Zusammensetzung des Bodens und den Modellparametern her. Zum Vergleich wurden die Parameter ebenfalls kalibriert. Die von ihnen ermittelten Parameter wichen zwar voneinander ab, die Ergebnisse der Simulationen waren aber vergleichbar. Weitere Beispiele sind Anwendungen von Fang et al. (2010) sowie Grimaldi et al. (2010), die mittels topographischer Informationen die Parameter ihrer Modelle ableiten. Skaugen, Peerebom & Nilsson (2015) stellen ein Modell vor, dessen Parameter über die Verteilung Fließlängen im betrachteten Einzugsgebiet ermittelt werden können. Auch in diesen Studien müssen die abgeleiteten Parameter stets als erste Schätzung interpretiert werden.

### **2.1.3 Regionale Strategien**

Das Ziel einer regionalen Anpassungsstrategie ist einen Parametersatz zu finden, der innerhalb einer definierten Region die mittlere Güte über alle verfügbaren Teilgebiete maximiert. Dafür wird eine Zielfunktion definiert, in die die Anpassung aller Simulationen an ihre jeweiligen Beobachtungen einfließt.

Drouge et al. (2002) wendeten in ihrer Studie mit 16 Einzugsgebieten unter anderem eine regionale Parametrisierung an. Ihre Ergebnisse zeigten, dass eine regionale Parametrisierung im Vergleich zu einer lokalen Parametrisierung eine schlechtere Anpassung in der Kalibrierungsphase aufweist. Im Gegenzug bleibt die Modellgüte in der Validierung aber stabil. Lokale Parametrisierungen zeigten hingegen einen starken Abfall der Modellgüte in der Vali-

dierung. Diese Ergebnisse decken sich mit weiteren Anwendungen von regionalen Strategien. Parajka, Blöschl & Merz (2007) kommen mit einer auf regionale Konsistenz ausgelegten Strategie zum gleichen Ergebnis. Auch Gaborit et al. (2015) sowie Bárdossy, Huang & Wagener (2016) zeigten, dass ein regionaler Parametersatz einen Kompromiss zwischen individueller Anpassung an beobachtete Daten und der Stabilität der Ergebnisse bei räumlicher und zeitlicher Übertragung der Parameter darstellt.

## 2.2 Konzeptionelle Niederschlags-Abfluss Modelle

Konzeptionelle Modelle stellen die am häufigsten verwendete Art von Niederschlags-Abfluss-Modellen im Kontext von unbeobachteten Gebieten dar. Auch während der PUB-Dekade kam ihnen eine zentrale Rolle zu, da sie auf Prozessverständnis aufgebaut sind. Alternative Modellansätze sind statistische Modelle und datengetriebene Modelle (Kap. 2.3).

Jedem konzeptionellen Modell liegt eine funktionale These über den Zusammenhang von natürlichen Gegebenheiten eines Einzugsgebietes mit dem Abfluss zugrunde. Dieser Zusammenhang mag hypothetischer Natur sein, wie im Falle von FLEX-TOPO (Savenije, 2010), oder auf physikalischen Prozessgleichungen der Mikroskala beruhen, wie im Modell SHE (Abbott et al., 1986) oder VIC (Liang et al., 1994).

In den meisten Fällen bestehen konzeptionelle Modelle aus einem, einer Abfolge, oder mehreren parallelen Speichern, mit denen Niederschlag zu Abfluss transformiert wird. Eines der bekanntesten Modelle dieser Art ist das HBV<sub>96</sub>-Modell (Bergström, 1976; Lindström et al., 1997). Die im Modell angeordneten Speicher repräsentieren dabei eine Abfolge verschiedener natürlicher Prozesse wie Interzeption, Infiltration, etc. Das HBV-Modell wurde in einer Vielzahl der zuvor angeführten Studien verwendet (u. a. Bárdossy, 2007; Beck et al., 2016; Masih et al., 2010; Parajka, Blöschl & Merz, 2007) und zeichnet sich durch seine Anpassungsfähigkeit aus. Zudem wird durch die Möglichkeit innerhalb eines Einzugsgebietes zwischen verschiedenen Landnutzungen zu differenzieren eine Anpassung an physikalische Gegebenheiten ermöglicht. Das von Beven & Kirkby (1979) entwickelte TOPMODEL ist wohl das in der Hydrologie bekannteste Modell, da es mit als erstes Modell die Vorteile von Speichermodellen mit topographischen Kennwerten verknüpfte. Weitere speicherbasierte Modelle, die in der Literatur Anwendung finden, sind unter anderem LARSIM (Bremicker, 2000; Casper, Gronz & Gemmar, 2015), GR4 (Perrin, Michel & Andréassian, 2003; Drouge et al., 2002) und MHM (Samaniego, Kumar & Attinger, 2010).

Ein alternativer konzeptioneller Ansatz wird von geomorphologischen Einheitsganglinien Modellen (GIUH) verwendet. Diesen Modellen liegen keine strukturellen Annahmen über Prozesse innerhalb des Gebietes zugrunde. Vielmehr wird das betrachtete Einzugsgebiet als Einheit betrachtet und der Abfluss über die Verteilung von Fließwegen definiert. Die Abflusskonzentration, d.h. die zeitliche Verteilung des Abflusses, beruht allein auf der geomorphologischen Ordnung der Gebiete.

Entwickelt wurde das Konzept von Rodríguez-Iturbe & Valdés (1979) und verallgemeinert von Gupta, Waymire & Wang (1980). Die Darstellung der Geomorphologie beruhte in ihren Ansätzen auf Horton-Kennzahlen und den Strahler Ordnungszahlen innerhalb des betrachteten Gebietes. Mit Hilfe einer angenommenen Fließgeschwindigkeit konnte über die Verteilung der Fließwege eine Einheitsganglinie ermittelt werden. Die Reaktion des Einzugsgebietes auf ein Niederschlagsereignis konnte anschließend mittels dieser Ganglinie unter Berücksichtigung des Super-Positions-Prinzips berechnet werden. Eine signifikante Weiterentwicklung erfuhr das GIUH-Konzept durch die von Kirkby (1976) entwickelte Weite-Funktion, die mit dem Aufkommen von digitalen Geländemodellen eine verbesserte Erfassung der Geomorphologie ermöglichte (Mesa & Mifflin, 1986; Snell & Sivapalan, 1994). Während die vorherigen GIUH-Modelle auf der Schätzung von Transit- und Wartezeiten beruhten, konnte mit den Weite-Funktion-GIUH-Modellen (Rinaldo, Marani & Rigon, 1991; Rinaldo et al., 2006; White et al., 2004) die Berechnung des Abflusses auf die Verteilung von Transitzeiten reduziert werden. Einer der kritischsten Punkte des GIUH-Konzeptes stellt die Schätzung der Entwässerungsgeschwindigkeit dar mit den die Transitzeiten bestimmt werden können (Beven, Wood & Sivapalan, 1988; Gupta & Mesa, 1988; Grimaldi, Petroselli & Nardi, 2011).

Grimaldi, Petroselli & Nardi (2011) zeigen in ihrer Studie, wie ein auf den Parameter der Schwerpunktlaufzeit reduziertes GIUH-Modell auch auf unbeobachtete Gebiete angewendet werden kann. Sie verwenden dabei den von Grimaldi et al. (2010) entwickelten Ansatz um Entwässerungsgeschwindigkeiten zu schätzen. In einer ähnlichen Studie zeigen Kumar et al. (2007) wie aus der Analyse von Niederschlags-Abfluss Ereignissen die Parameter eines GIUH geschätzt werden können. Hallema & Moussa (2014) schlagen zudem eine Unterteilung der Fließstrecken eines Einzugsgebietes nach Hang- und Gewässerfließstrecken vor. Die von ihnen eingeführte Einteilung der Gebiete beruht auf der Zerlegung eines Einzugsgebietes in rechteckige „repräsentative Hangflächen“. Nardi, Annis & Biscarini (2018) zeigen in einer weiteren Studie, wie das von Grimaldi, Petroselli & Nardi (2011) entwickelte Modell mit einer Abflussbildungskomponente verknüpft werden kann. Trotz dieser und anderer Weiterentwicklungen (bspw. Hazenberg et al., 2015; Rigon et al., 2016) besteht für GIUH-Modelle noch immer das Problem, dass die Geschwindigkeit des Wassers für jedes N-A-Ereignis neu geschätzt werden muss.

### 2.3 Datengetriebene Modelle

Um ein konzeptionelles Modell anwenden zu können, werden Daten benötigt, um Parameter an lokale oder regionale Beobachtungen anzupassen. Zwar werden Daten verwendet, trotzdem unterscheiden sich diese Modelle durch ihre fest definierten Prozessannahmen von datengetriebenen Modellen (*data-driven models*, DDM).

Der Hauptaspekt von DDMs ist der Lernprozess. In diesem Prozess werden Verknüpfungen zwischen den Eingangsvariablen und den gesuchten Zielgrößen hergestellt und parametrisiert (Solomatine & Ostfeld, 2008), welche dann in der Anwendung der angelernten Struktur zur Vorhersage verwendet werden können. Maschinelles Lernen (ML) bezeichnet diesen Lernvorgang und kann mit unterschiedlichen Algorithmen realisiert werden (Solomatine & Ostfeld, 2008). Das in jüngster Zeit aufgekommene *Deep Learning* (Goodfellow, Bengio & Courville, 2016) gilt als eine der vielversprechendsten Methoden, um mittels datengetriebener Modelle zu prozessorientierten Erkenntnissen zu gelangen (Shen et al., 2018). Während ML-Algorithmen bereits aufbereitete Daten in Form von Prädiktoren benötigen um einen Zusammenhang zu den Zielvariablen herzustellen werden beim *Deep Learning* die Rohdaten verwendet. Die Extraktion der relevanten Informationen zur Regression wird so vom Algorithmus selbst durchgeführt und muss nicht durch manuelle und damit subjektive Vorauswahl erfolgen. Mount et al. (2016) heben zudem die Eignung von DDMs als ergänzende Modellstrukturen zu physikalisch-basierten Modellen für sozio-hydrologische Anwendungen hervor. Durch die Kombination dieser Modellstrukturen nach ihrer Einschätzung die unübersichtlichen Interaktionen zwischen Gesellschaft und der Ressource Wasser besser abgebildet werden.

Eine der simpelsten Formen von DDMs stellen mechanistische Regressionsmodelle dar, wie sie bspw. von Young (2003), Ratto et al. (2007) und Zhang & Govindaraju (2003) verwendet wurden. In diesen Studien wurden lineare Regressionsmodelle für einen vermuteten Zusammenhang zwischen Niederschlag- und Abfluss verwendet. Eine Besonderheit stellt die Arbeit von Zhang & Govindaraju (2003) dar. Die von ihnen vorgeschlagenen geomorphologischen künstlichen neuronalen Netze stellen einen Bezug zwischen der Geomorphologie des Einzugsgebietes und der Struktur eines künstlichen neuronalen Netzes (ANN) her. Nach ihrer Definition richten sich die Anzahl von Neuronen und die Wichtung ihrer Verknüpfungen zur Zielvariable (dem Abfluss) nach der Anzahl von Fließwegen im Gebiet und deren Pfadwahrscheinlichkeit.

Machine Learning Algorithmen, also ANNs, Support Vector Machines, Regressionsbäume etc. (siehe Alpaydin, 2016), können zudem als Ersatz für konzeptionelle Modelle eingesetzt werden. Alvisi et al. (2006), Kisi, Shiri & Tombul (2013), Piotrowski et al. (2015), Shortridge, Guikema & Zaitchik (2016) oder Heřmanovský et al. (2017) können hier als Beispiele für Studien genannt werden, in denen ML-Algorithmen zur operationellen Vorhersage von Abflusswerten, Wassertemperaturen oder Wasserständen verwendet wurden. Die datengetriebenen Modelle zeigen in diesen Anwendungen im Vergleich zu konzeptionellen Modellen vergleichbare und teilweise bessere Ergebnisse. Heřmanovský et al. (2017) geht zudem auf die Regionalisierung eines lokal angelernten DDMs ein und zeigt, dass es in einem unbeobachteten Gebiet vergleichbare Ergebnisse erzielt wie ein konzeptionelles Modell. Elshorbagy et al. (2010a; 2010b) präsentieren eine ausführliche Übersicht über veröffentlichte Anwendungen von ML-Anwendungen in der Hydrologie. Dazu ergänzend geben Yaseen et al. (2015) und Nourani et al. (2014) weitere Literaturübersichten zu Anwendung

im Bereich operationeller Vorhersage oder der Verwendung von Wavelet-Transformationen im Zusammenspiel mit ML-Algorithmen. Neben der Literaturübersicht präsentieren Elshorbagy et al. (2010a; 2010b) auch eine Studie zur Vorhersage von aktuellen Verdunstungswerten in mehreren Einzugsgebieten. Sie konnten zeigen, dass die Algorithmen auch in dieser Anwendung zu adäquaten Ergebnissen gelangen. Darüber hinaus weisen sie darauf hin, dass unterschiedliche Algorithmen für unterschiedliche Probleme geeignet sind. So zeigten sich ANNs in ihrer Studie eher geeignet für komplexe Probleme, während ein KNN (*k-nearest neighbour*) für lineare Probleme die besten Ergebnisse lieferte. Sie folgern daraus, dass ML-Algorithmen stets als Ensemble verwendet werden sollten.

Neben der Anwendung als operationelles Vorhersagemodell konnten ML-Algorithmen in einer Reihe von Studien zur Ableitung einer regionalen Übertragungsfunktion verwendet werden. Vor allem *Random Forest* (RF) Algorithmen fanden in diesem Zusammenhang vermehrte Anwendung. Als Beispiel sind hier die Studien von Brunner et al. (2018), Zhang et al. (2018) und Addor et al. (2018) zu nennen. In diesen drei Studien wurden mit RFs Beziehungen zwischen Gebietscharakteristiken wie Größe, Landnutzung, etc. und verschiedensten Abflusscharakteristiken hergestellt. Die Ergebnisse liefern hier ein einheitliches Bild: RF-Algorithmen liefern bessere Ergebnisse als klassische lineare Regressionsmethoden. Zhang et al. (2018) stellen zudem fest, dass RF Charakteristiken von Niedrigwasser und Abflussdynamik besser übertragen konnten als ein regionalisiertes konzeptionelles Modell.

Ein weiterer Schwerpunkt liegt auf der Verwendung von ANNs zur Regionalisierung. Hall, Minns & Ashrafuzzaman (2002) verwendeten ein ANN, um über Gebietscharakteristika Hochwasser-Quantile in Großbritannien zu regionalisieren. Ähnliche Studien wurden von Jingyi & Hall (2004) in China und Toth (2016) in Italien durchgeführt.

Elshorbagy et al. (2010a) und Hrachowitz et al. (2013) stimmen in ihren Übersichten überein, dass die Hauptprobleme der datengetriebenen Modellierung eine fehlende Allgemeingültigkeit ihrer Ergebnisse sowie eine schwere Deutbarkeit sind. Da die Modelle sich individuell an lokale Gegebenheiten anpassen, lassen sich aus einer einzelnen Studie mit kleinen Datensätzen nur begrenzt allgemein gültige Regeln ablesen. Die zweite benannte Problematik betrifft die Algorithmen selbst: zwar können mit ihnen adäquate Vorhersagen getroffen werden, die angelernte Struktur lässt aber in alleiniger Betrachtung keine Rückschlüsse auf Prozesse zu. Vor allem die Arbeitsweise von ANNs gilt als gänzlich unverständlich für menschliche Logik (Kelleher, MacNamee & D'Arcy, 2015; Alpaydın, 2008). Elshorbagy et al. (2010a) und Shen et al. (2018) weisen aber auch auf einen Ausweg hin: komparative Studien mit großen Datenmengen und iteratives Vorgehen. Durch wiederholte Lernprozesse in unterschiedlichen Gebieten wird eine Vielzahl von Hypothesen getestet. Durch den Vergleich der Ergebnisse aus unterschiedlichen Regionen und mit verschiedenen Datensätzen kann dann eine Auswahl getroffen werden und bestenfalls Rückschlüsse auf physikalische Prozesse gezogen werden. Somit könnten DDMs einen Beitrag zur Identifizierung bislang unbekannter Verknüpfungen zwischen Gebietskennwerten und Abfluss leisten (Wagener et al., 2010; Shen et al., 2018). Sie lenken zudem den Fokus auf Modellstrukturen, statt auf

Parametrisierungen von bereits bekannten Konzepten (Gupta & Nearing, 2014). Wichtig ist auch die von Mount et al. (2016) hervorgehobene Nützlichkeit von DDMs als Unterstützung von konzeptionellen Modellen. Beispielsweise als Kombination der Modellstrukturen zur Vorhersage einer Zielgröße, bspw. Ratto et al. (2007), oder als Verwendung von ML zur Aufbereitung und Validierung von Eingangsdaten (Mewes, Oppel & Hartmann, 2019; Mewes & Oppel, 2019) für konzeptionelle Modelle.

## 2.4 Zusammenfassung und Ansatzpunkte

In diesem Abschnitt wurden die gängigsten Methoden zur N-A-Modellierung in unbeobachteten Gebieten vorgestellt. Allgemein kann das Vorgehen wie folgt zugefasst werden: Ein Modell wird durch Abflussbeobachtungen von umliegenden Pegeln parametrisiert. Welches Modell oder welche Kalibrierungsstrategie (Regionalisierung von kalibrierten Parametern oder regionale Kalibrierung) gewählt wird, ist dabei nicht nur abhängig vom Ziel der Modellierung sondern in den meisten Fällen auch von den lokalen Gegebenheiten der Einzugsgebiete.

Unabhängig von bestehenden Konzepten wird allgemein empfohlen so viele Informationen über das Zielgebiet wie möglich in die gewählte Modellstruktur einzubeziehen. Die Berücksichtigung von räumlichen Strukturen und Mustern wird allgemein als wichtige Information über die Abflussgenese angesehen. Darüber hinaus soll die Komplexität des Modells so gering wie möglich gehalten werden, da komplexere Modelle einen höheren Kalibrierungsaufwand haben und schlechter auf unbeobachtete Gebiete zu übertragen sind.

Diese allgemeinen Empfehlungen sprechen gegen die Verwendung von gängigen konzeptionellen Modellen in der Regionalisierung, die in der Regel eine Vielzahl von Parametern benötigen und damit viel zu komplex sind. Trotz der Komplexität der Modelle verwenden sie in den meisten Fällen aber nur einen Bruchteil der zur Verfügung stehenden Kennwerte eines Gebietes. In den meisten Fällen dienen sie nur zur Anpassung von Parametern, nicht aber der eigentlichen Modellstruktur. Darüber hinaus bedarf es im Fall einer Regionalisierung der Definition einer Übertragungsfunktion, die weitere subjektive Entscheidungen erfordert.

Als grundsätzlich erstrebenswert muss ein selbstkalibrierendes Modell angesehen werden, das sich komplett, oder zumindest zu einem Großteil, an die zur Verfügung stehenden Daten anpasst. Es wird eine Modellstruktur benötigt, die die Abflussgenese einerseits an die Struktur des betrachteten Einzugsgebiets anpasst und andererseits die beobachtete Abflussdynamik anderer, ähnlicher, Einzugsgebiete mit einbezieht.

Die Geomorphologische Einheitsganglinie bietet diese gesuchte Flexibilität der Modellstruktur im Hinblick auf die räumliche Struktur des Einzugsgebietes. Über die Verteilung von Fließflächen wird die Form der Abflussganglinie bestimmt. Benötigt wird neben dem GIUH eine Methode die Dynamik der umliegenden Gebiete nachzubilden, um das Modell

für unbeobachtete Abflussereignisse anwendbar zu machen. Datengetriebene Modelle, und dabei vor allem die Techniken des *Machine Learning*, eröffnen genau diese Möglichkeit. Als regionale Modellinstanzen, angepasst auf den Zusammenhang zwischen meteorologischem Antrieb sowie Zustand des Einzugsgebiets zur Abflussdynamik, können ML-Algorithmen zur ereignisbezogenen, Echtzeit-Kalibrierung des GIUH-Modells verwendet werden.

## 3 Grundlagen und Datenaufbereitung

### 3.1 Untersuchungsgebiete

Zur Umsetzung der geplanten Analysen war der Aufbau einer umfangreichen und heterogenen Datenbank von Niederschlags-Abfluss (N-A) Ereignissen erforderlich. Heterogen bezieht sich dabei zum einen auf die Genese der Ereignisse und zum anderen auf den Ort ihres Auftretens. So sollte sichergestellt werden, dass das entwickelte Modell ein breites Spektrum natürlicher Prozesse abdecken kann, um in möglichst vielen Regionen zuverlässige Aussagen zu liefern.

Zeitreihen des Abflusses, Niederschlags, Temperatur und Sonnenscheindauer in stündlicher und täglicher zeitlicher Auflösung lagen für diese Arbeit für das gesamte Bundesland Bayern vor. Zur Verfügung gestellt wurden die Daten durch das Bayerische Landesamt für Umwelt (2018). Im Folgenden werden drei Flussgebiete vorgestellt (vgl. Abbildung 3-1), die als Datengrundlage der durchgeführten Analysen dienen.

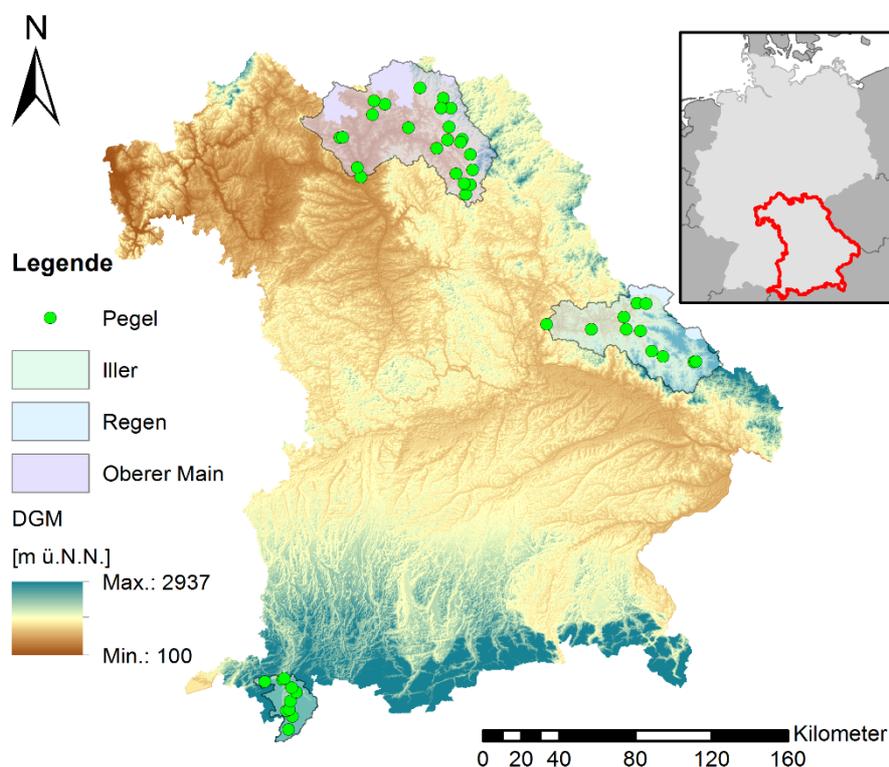


Abbildung 3-1: Übersicht über die ausgewählten Flussgebiete Iller, Regen und Main im Bundesland Bayern

Für alle Gebiete standen Landnutzungsdaten aus *CORINE Land Cover* Daten (Bossard, Feranec & Otahel, 2000) und ein digitales Höhenmodell (DGM) (basierend auf Daten der *Shuttle Radar Topography Mission* (SRTM) (Jarvis et al., 2008)) zur Verfügung. Außerdem wurden mit Hilfe von Pedo-Transferfunktionen (Sponagel, 2005) und der Bodenübersichtskarte Deutschland (Bundesanstalt für Geowissenschaften und Rohstoffe, 2006) Bodenkennwerte (Porenvolumen (GPV), hydr. Leitfähigkeit, Aufbau) ermittelt. Alle Daten lagen in einheitlicher räumlicher Auflösung als Rasterdaten mit einer Zellgröße von  $100 \times 100$  m vor.

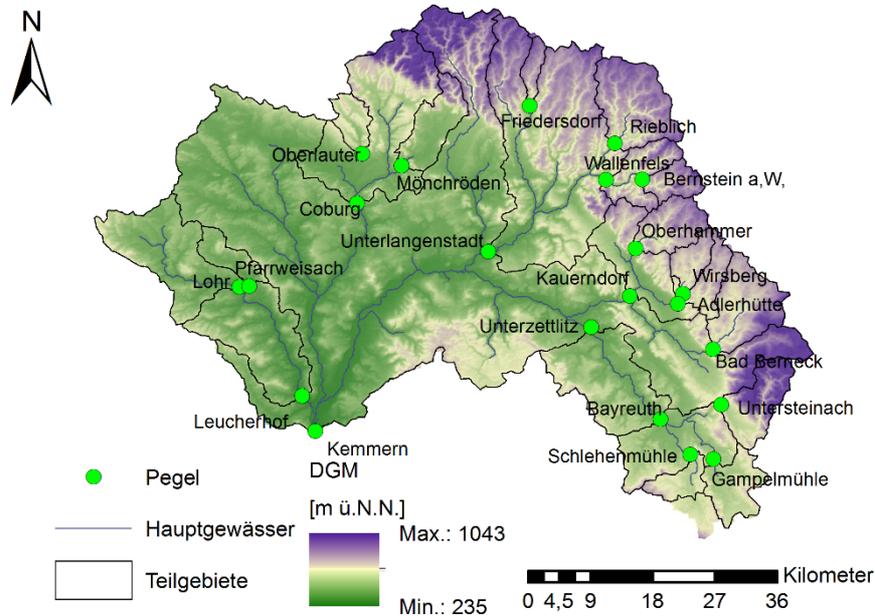
Eine übliche Auswahl von N-A Ereignissen ist die Selektion über Jahreshöchstabflüsse. Wird pro Jahr nur das Ereignis mit dem höchsten Scheitelabfluss gewählt, wird dadurch gewährleistet, dass sich ausschließlich Hochwasserereignisse im Datensatz befinden. Bei geringer Länge der Zeitreihen, aus denen die Ereignisse selektiert werden sollen, führt dies aber zu einer unzureichend umfangreichen Datenbank. Des Weiteren bliebe die Heterogenität der Abflussprozesse im Datensatz anzuzweifeln. In einem alpinen Gebiet, wie das der Iller, in dem kurze intensive Niederschläge im Sommer die höchsten Scheitel verursachen, würde die Selektion über Jahresmaxima u.a. Ereignisse aus Schneeschmelze vernachlässigen, die betreffend des Abflussvolumens deutlich signifikanter als Sommerereignisse sein können. Ein weiteres Problem ist, dass ein Modell, vor allem ein daten-getriebenes Modell, das nur auf Hochwasser angelehrt wurde, immer ein Hochwasser vorhersagen wird. Es ist daher auch erforderlich Ereignisse einzuschließen, in denen ein alarmierender Niederschlag (sonst würde keine Vorhersage benötigt werden) zu keinem Hochwasser geführt hat.

Um diesen Anforderungen zu entsprechen, wurde die Datenbank wie folgt aufgebaut: Zum einen wurden pro Jahr 5 Ereignisse mit dem höchsten Scheitel ausgewählt. So sollte eine ausreichende Anzahl von Hochwasserereignissen in den Datensätzen garantiert werden. Die Anzahl von 5 Ereignissen stellt einen Kompromiss zwischen Anzahl der Ereignisse und Mindestabflusshöhe der Ereignisse dar, bleibt aber eine bewusst erfolgte Entscheidung. Zum anderen wurden 5 Ereignisse pro Jahr selektiert, die auf die höchsten Niederschlagsereignisse folgten, unabhängig von der Abflusshöhe des nachfolgenden Abflussereignisses. Mittels der Methode der Grenzregenpause (Sartor, 1994) wurden die zur Verfügung stehenden kontinuierlichen Reihen in Ereignisse zerlegt und für jedes Ereignis eine mittlere Niederschlagsintensität bestimmt. Wird ein zeitlich zusammenhängender Niederschlagsimpuls für mehr als 36 Stunden unterbrochen, wurde der Impuls als ein Ereignis identifiziert. Außerdem wurde eine minimale Niederschlagsintensität von 0,1 mm/h als Grenzintensität berücksichtigt. Basierend auf diesen Werten wurden pro Jahr bis zu 10 N-A Ereignisse selektiert.

Ereignisbeginn und -ende von Niederschlag und Abfluss wurden manuell für jedes der ausgewählten Ereignisse ermittelt. Die Ereignisse wurden aus den kontinuierlichen Zeitreihen in stündlicher Auflösung extrahiert. Die Anzahl von Ereignissen pro Pegel variiert, da Ereignisse basierend auf Scheitelselektion und Ereignisse aus Niederschlagsselektion teilweise identisch waren. In Summe wurden in allen Gebieten 2.249 Ereignisse separiert (~ 52 pro Pegel) und für jedes Ereignis die zugehörigen Ereignisindikatoren (Kapitel 3.2) ermittelt.

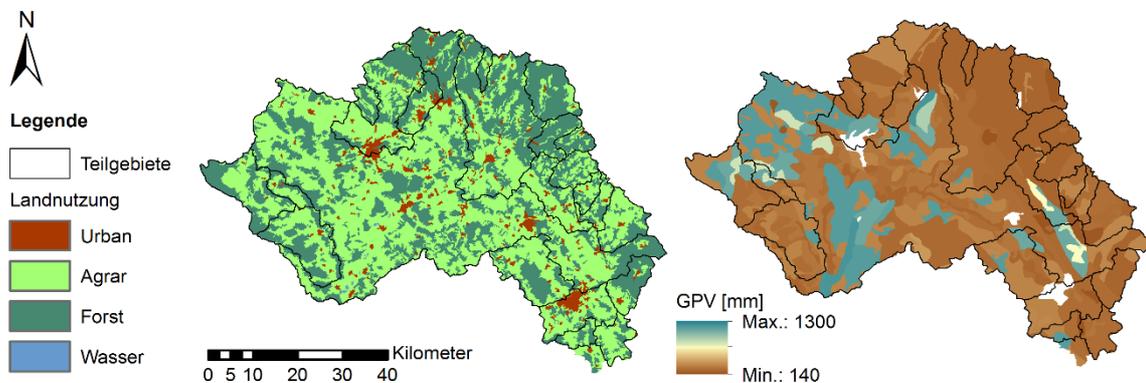
### 3.1.1 Oberer Main

Das am nördlichsten gelegene Flussgebiet, das in dieser Arbeit verwendet wurde, ist das Einzugsgebiet des oberen Mains. Betrachtet wurde das Gebiet im Oberlauf des Pegels Kemmern mit einer Fläche von 4223 km<sup>2</sup>. Innerhalb dieses Gebietes wurden 22 (inkl. Kemmern) Pegel betrachtet, ihre Lage, Teileinzugsgebiete sowie das DGM sind in Abbildung 3-2 dargestellt.



**Abbildung 3-2: DGM, ausgewählte Pegel und Teileinzugsgebiete im oberen Mains**

Der nord-östliche Oberlauf des Einzugsgebietes besteht aus einer Vielzahl kleinerer Einzugsgebiete im Thüringisch-Fränkischen Mittelgebirge. Südlich des Pegels Bayreuth befinden sich weitere Kopfgebiete, die dem Naturraum des Oberpfälzischen-Obermannischen Hügellandes zugeordnet werden. Mittel- und Unterlauf des Mains sowie die Pegel Lohr, Pfarweisach und Leucherhof befinden sich im Fränkischen-Keuper-Lias-land. Die Unterschiede der verschiedenen Naturräume im Gebiet spiegeln sich in Landnutzung und der Beschaffenheit des Bodens (vgl. Abbildung 3-3) wieder.



**Abbildung 3-3: Landnutzung und Gesamtporenvolumen im oberen Mains**

Diese starke Heterogenität innerhalb des Untersuchungsgebietes war einer der Gründe dieses Gebiet in die Untersuchung einzuschließen. Es stehen in vergleichsweise kurzer Distanz mehrere Pegel zur Verfügung, deren Beobachtungen den Abfluss aus unterschiedlichen Naturräumen beschreiben. In diesem Gebiet finden sich auch die meisten anthropogenen Einflüsse aller Untersuchungsgebiete dieser Arbeit. In Abbildung 3-3 sind bspw. die urbanen Regionen der Städte Coburg (Nord-Westen) und Bayreuth (Süd-Osten) zu erkennen. Allerdings befinden sich keinerlei Speicher, Kraftwerke, Kanäle oder größere Seen im Einzugsgebiet, die die Beziehung zwischen Niederschlag und Abfluss verfälschen würden.

Ein weiterer Aspekt bei der Wahl dieses Gebietes ist die Verteilung von Kopfgebieten. Durch die Bogenform des Gewässers (von Ost nach West) und der südlichen Lage innerhalb des Gebietes, befinden sich entlang des Hauptgewässers immer wieder kleinere, beobachtete seitliche Zuflüsse. In verschiedenen Abständen vom Gebietsauslass stehen somit Beobachtungen von Kopfgebieten mit vergleichbaren Größen zur Verfügung (14 von 22 Gebieten sind kleiner als 100 km<sup>2</sup>). Dies erlaubt eine dezidierte Analyse des gemeinsamen Abflusses aus den unterschiedlichen Naturräumen an den Unterliegerpegeln Unterzettlitz, Unterlangenstein und Kemmern.

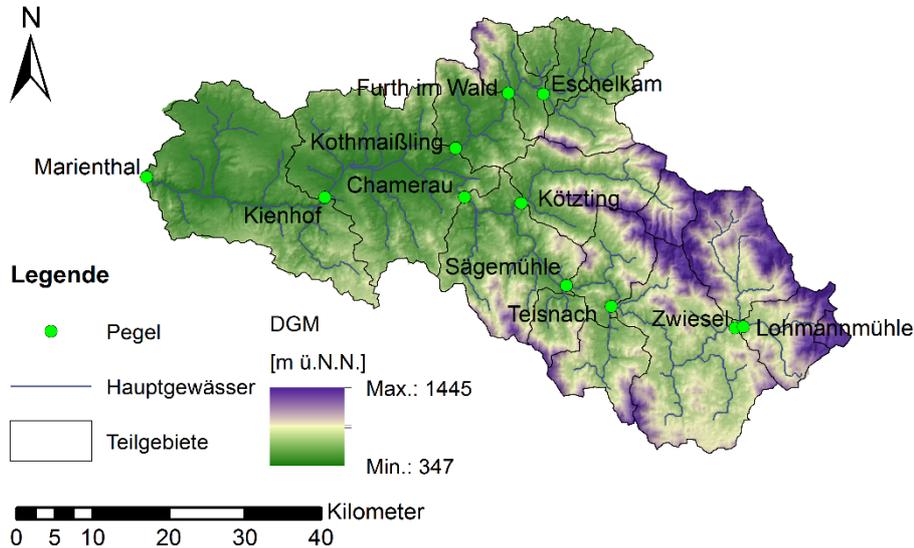
Im Gebiet des oberen Mains wurden im Mittel 48 Ereignisse pro Pegel separiert. Dieser Wert liegt etwas unter denen der anderen Gebiete, was vor allem auf die Länge der verfügbaren stündlichen Zeitreihen zurück zu führen ist (2001 bis 2007). In Tabelle 3-1 ist die Anzahl separierter N-A-Ereignisse pro ausgewähltem Pegel sowie dessen Einzugsgebietsfläche gegeben.

**Tabelle 3-1: Ausgewählte Pegel, Einzugsgebietsgrößen und Anzahl separierter N-A-Ereignisse im EZG des oberen Mains**

Pegel	Fläche [km <sup>2</sup> ]	Anz. NA-E. [-]	Pegel	Fläche [km <sup>2</sup> ]	Anz. NA-E. [-]
Adlerhütte	33.39	37	Mönchröden	70.66	48
Bad Berneck	99.79	43	Oberhammer	64.25	50
Bayreuth	340.28	50	Oberlauter	31.56	45
Bernstein a.W.	35.38	46	Pfarrweisach	36.67	51
Coburg	346.34	50	Rieblich	118.22	49
Friedersdorf	11.14	49	Schlehenmühle	70.95	49
Gampelmühle	62.2	45	Unterlangenstein	713.87	50
Kauerndorf	246.23	51	Untersteinach	73.52	48
Kemmern	4223.84	51	Unterzettlitz	500.35	53
Leucherhof	380.52	52	Wallenfels	96.45	49
Lohr	165.25	45	Wirsberg	76.86	52

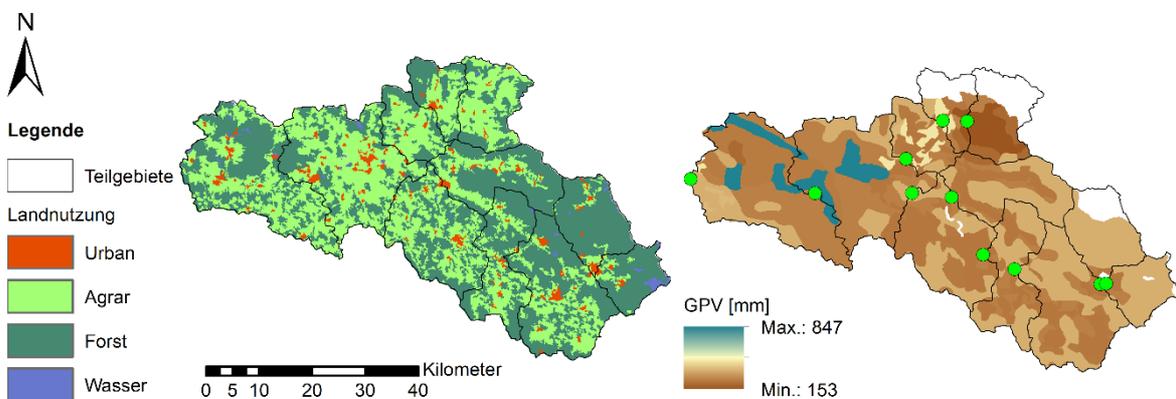
### 3.1.2 Regen

Im Einzugsgebiet des Regens, in dieser Arbeit definiert als das Einzugsgebiet des Pegels Marienthal mit einer Fläche von 2.590 km<sup>2</sup>, wurden 11 Pegel ausgewählt. Die Lage der Pegel und ihre Teileinzugsgebiete sind in Abbildung 3-4 dargestellt.



**Abbildung 3-4: DGM, ausgewählte Pegel und Teileinzugsgebiete im Regen**

Das gesamte Gebiet liegt im Naturraum des bayrischen Waldes und weist eine klare Ost-West Orientierung auf. Der Regen hat in diesem Teil seines Einzugsgebietes nur einen nennenswerten seitlichen Zufluss, den Chamb nördlich des Pegels Kothmaißling. Die Besonderheit dieses Einzugsgebietes und seiner Abflussbeobachtungen besteht in der Verschachtelung der Teileinzugsgebiete. Von Ost nach West ist jedes Gebiet ein Teilgebiet des nächsten Pegels. Dies gilt auch für die Pegel am Chamb: Kothmaißling, Furth im Wald und Eschelkam weisen eine eigene verschachtelte Struktur auf. Einzige Ausnahme ist der Pegel Kötzing, angesiedelt im Mittellauf des Regens. Hervorzuheben ist auch die rund 70 km lange Fließstrecke unterhalb des Zusammenflusses von Chamb und Regen (in der Nähe von Chamerau und Kothmaißling) bis zum Gebietsauslass.



**Abbildung 3-5: Landnutzung und Gesamtporenvolumen im Regen**

Im Vergleich zum Einzugsgebiet des oberen Mains sind im Einzugsgebiet des Regens kaum nennenswerte urbane Nutzungen zu erkennen (vgl. Abbildung 3-5). Auch die Verteilung der Bodenkennwerte, die durch den einheitlichen Naturraum des Gebietes bedingt ist, wirkt deutlich homogener. Aufgrund dieser Differenzen wurde der Regen als zweites Untersuchungsgebiet, das auch ein Repräsentant für Einzugsgebiete in Mittelgebirgsregionen ist, ausgewählt. Ein weiterer Grund war die angesprochene Organisation des Einzugsgebietes. Die Verschachtelung der (Teil-) Einzugsgebiete lies eine unterschiedliche Vermutung zur Nachbarschaftsbeziehung zwischen den Gebieten zu, d.h. eine andere Übertragbarkeit von Informationen als sie im oberen Main vermutet wurde.

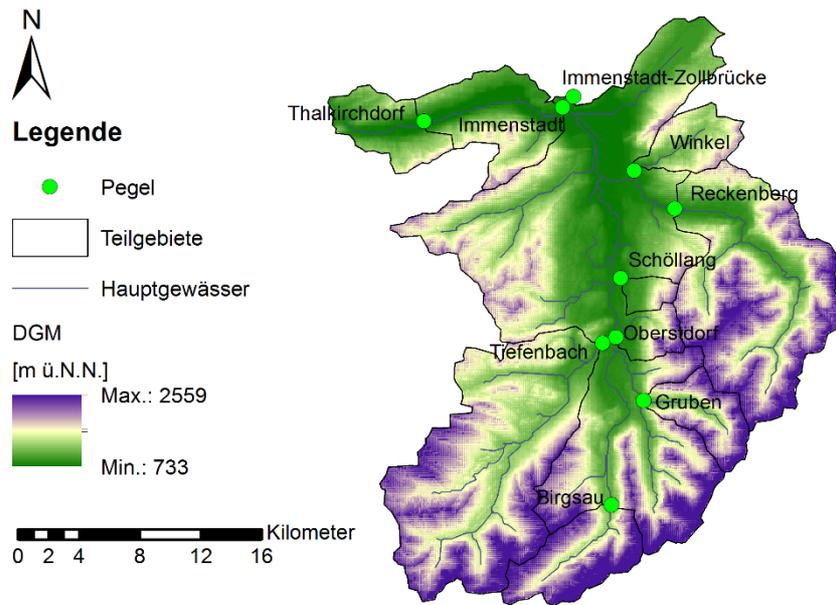
Beobachtungen des Abflusses, Niederschlag, Temperatur und anderer Klimadaten lagen im Zeitraum von 1999 bis 2012 vor. Für die ausgewählten 11 Pegel wurden im Mittel 63 Ereignisse separiert. Eine Auflistung der ausgewählten Pegel, der Fläche ihrer Einzugsgebiete und der Anzahl separierter Ereignisse ist in Tabelle 3-2 gegeben.

**Tabelle 3-2: Ausgewählte Pegel, Einzugsgebietsgrößen und Anzahl separierter N-A-Ereignisse im EZG des Regens**

Pegel	AE [km <sup>2</sup> ]	Anz. NA-E. [-]	Pegel	AE [km <sup>2</sup> ]	Anz. NA-E. [-]
Chamerau	1356.5	68	Lohmannmühle	115.9	66
Eschelkam	178.0	67	Marienthal	2590.4	63
Furth im Wald	276.6	61	Sägemühle	839.3	63
Kienhof	2174.1	67	Teisnach	626.6	67
Kötzing	224.4	63	Zwiesel	293.4	60
Kothmaißling	405.0	58			

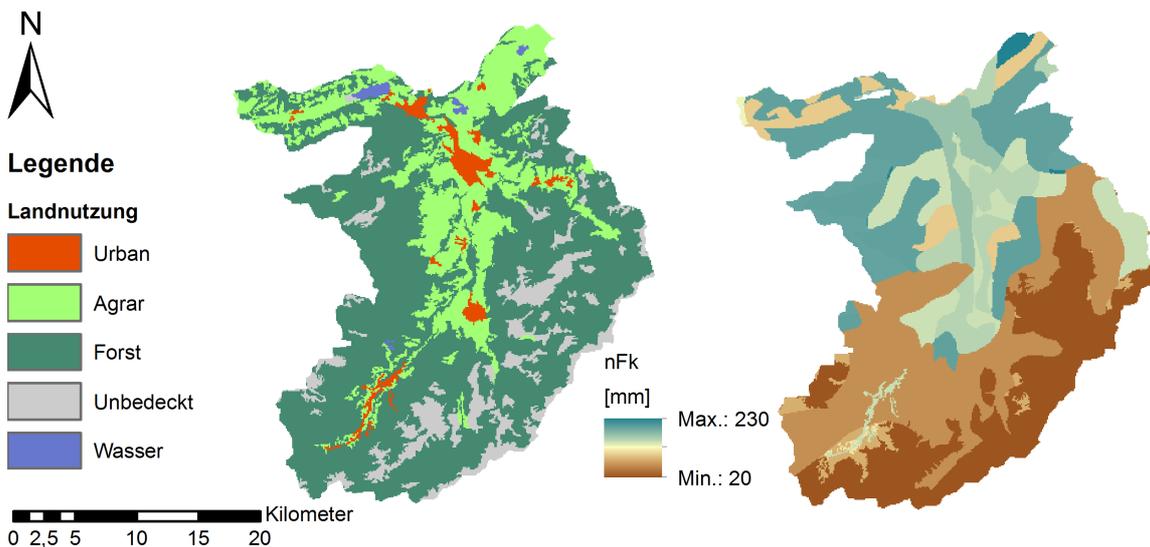
### 3.1.3 Iller

Als letztes Untersuchungsgebiet wurde das alpine Einzugsgebiet der Iller, südlich des Pegels Immenstadt-Zollbrücke, ausgewählt. Mit einer Fläche von 724 km<sup>2</sup> ist dieses Gebiet das kleinste betrachtete Gebiet. In diesem Gebiet wurden 10 Pegel ausgewählt (vgl. Abbildung 3-6).



**Abbildung 3-6: DGM, ausgewählte Pegel und Teileinzugsgebiete der Iller**

Das ausgewählte Einzugsgebiet befindet sich am Rande der nördlichen Kalkalpen. Die Pegel am Gebietsauslass, Immenstadt und Immenstadt-Zollbrücke, liegen bereits im Voralpinen Moor- und Hügelland. Zu erkennen sind drei größere urbane Regionen (Abbildung 3-7) der Städte Immenstadt (Nord), Sonthofen (Nord-Mitte) und Oberstdorf (Süd-Mitte). Im Unterschied zu den übrigen Gebieten ist natürlich die Höhendifferenz innerhalb des Gebietes wesentlich größer, außerdem sind aber auch unbedeckte Flächen auf den Bergkuppen vorhanden. Desweiteren befindet sich im Nord-Westen des Einzugsgebietes ein natürlicher See, der große Alप्see mit einer Fläche von 2,5 km<sup>2</sup> nahe Immenstadt.



**Abbildung 3-7: Landnutzung und nutzbare Feldkapazität der Iller**

Mit seinem alpinen Charakter stellt das Einzugsgebiet der Iller eine Ergänzung des bisher vorgestellten Datensatzes dar. Während die vorherigen Datensätze aus Mittelgebirgsregion-

nen stammen und größere Raumskalen abdecken, wurde mit diesem Gebiet ein alpines kleinräumiges Gebiet in die Entwicklung des Modells einbezogen. Es wurde angenommen, dass die dominanten Prozesse in diesem Gebiet sich signifikant von denen in den anderen Regionen unterscheiden. Daher wurde dieses Gebietes in die Fallstudien dieser Arbeit aufgenommen.

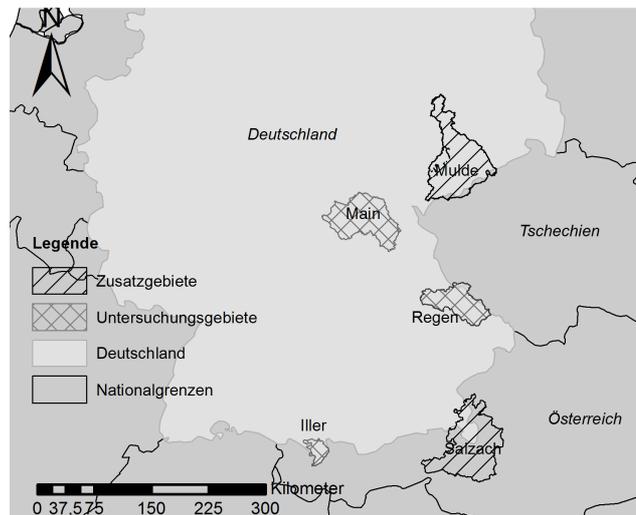
Es wurden 10 Pegel ausgewählt, wobei an den Pegeln Gruben und Winkel aufgrund fehlerhafter und nur kurzer Zeitreihen nur wenige Ereignisse separiert werden konnten. An den restlichen Pegeln konnten im Zeitraum 2001 bis 2012 im Mittel 57 Ereignisse separiert werden.

**Tabelle 3-3: Ausgewählte Pegel, Einzugsgebietsgrößen und Anzahl separierter N-A-Ereignisse im EZG der Iller**

Pegel	AE [km <sup>2</sup> ]	Anz. NA-E. [-]	Pegel	AE [km <sup>2</sup> ]	Anz. NA-E. [-]
Birgsau	35.0	51	Reckenberg	126.7	63
Gruben	24.0	6	Schöllang	2.3	58
Immenstadt	65.8	61	Thalkirchdorf	23.8	57
Immenstadt-Zollbrücke	724.0	61	Tiefenbach	146.7	51
Oberstdorf	76.0	61	Winkel	19.9	14

### 3.1.4 Sonstige Gebiete

Neben den bereits vorgestellten Gebieten wurden zwei weitere Einzugsgebiete im Verlauf der Arbeit verwendet. Zum einen das Einzugsgebiet der Mulde in Sachsen, ein Zufluss der Elbe, und zum anderen das Einzugsgebiet der Salzach in Salzburg (Österreich), der größte Zufluss zum Inn. Die Lage der beiden Gebiete mit Referenz zu den anderen verwendeten Gebieten ist in Abbildung 3-8 dargestellt.



**Abbildung 3-8: Lage und Umriss der Untersuchungs- und Zusatzgebiete**

Die beiden Einzugsgebiete sind größer als die bereits vorgestellten Gebiete (Salzach 5995 km<sup>2</sup> und Mulde 6170 km<sup>2</sup>) und wurden nur zur Analyse der Geomorphologie verwendet. Beide Gebiete sind ebenfalls Vertreter für ein Flussgebiet des Mittelgebirges (Mulde) und

des Alpenraumes (Salzach) und wurden daher als Vergrößerungen der Untersuchungsgebiete verwendet. In Kapitel 4.2 wird die Untersuchung der Geomorphologie auf Grundlage dieser Gebiete vorgestellt.

Für diese Gebiete wurden keine Zeitreihen aufbereitet. Es lagen lediglich DGM, Landnutzung und Bodenkennwerte analog zu den Untersuchungsgebieten vor (Darstellungen in Abbildung 4-1 und Abbildung 4-2).

## 3.2 Ereignisindikatoren

Neben Zeitreihen von Niederschlag und Abfluss, wurden für die selektierten und separierten N-A-Ereignisse verschiedene Indikatoren ermittelt, um die Anfangs- und Randbedingungen, dieser Ereignisse zu beschreiben. Diese Indikatoren wurden verwendet um Charakteristika der Abflussereignisse mit Hilfe von *Machine Learning* (ML) vorherzusagen. In diesem Abschnitt werden die ermittelten Indikatoren vorgestellt. Generell lassen sich die ermittelten Kennzahlen in drei Gruppen einteilen: Zielgrößen, klimatische Randbedingungen des N-A Ereignisses und Anfangszustände des Einzugsgebietes.

Zu beachten ist, dass die ML-Algorithmen in den meisten Fällen normalisierte Daten erfordern. Der Erwartungswert eines Datensatzes (Prädiktoren und Zielvariablen) soll also gleich Null sein, die Varianz gleich eins (Kelleher, MacNamee & D'Arcy, 2015; Pedregosa et al., 2011). Da dies im Falle der meisten Prädiktoren nicht der Fall ist, bspw. sind Niederschlagsdauern stets größer Null, war eine Normalisierung der Daten notwendig. Verwendet wurde eine L2-Norm, da sie im Vergleich zur L1-Norm stabiler und effizienter ist (Han & Kamber, 2010). Dabei wird jede Stichprobe, unabhängig von der Verteilung der einzelnen Variablen, auf eine euklidische Norm von 1 gebracht. Die Werte der Stichprobe liegen nach der Normierung innerhalb einer Einheitskugel.

### 3.2.1 Zielgrößen

Zwei Charakteristiken der separierten N-A Ereignisse wurden als Zielgrößen der ML-Algorithmen ausgewählt (siehe Kapitel 4.1) zum einen die Entwässerungsgeschwindigkeit und zum anderen das Abflussvolumen.

Die Entwässerungsgeschwindigkeit  $v_E$  [m/s] ist die mittlere Fließlänge  $\bar{x}$  [m] (berechnet als Mittelwert aller Fließstrecken bis zum Auslass) im betreffenden Teileinzugsgebiet normalisiert über die Schwerpunktlaufzeit  $t_S$  [s]. Die Schwerpunktlaufzeit ist definiert als Differenz der zeitlichen Schwerpunkte des Niederschlages  $P$  und Abflusses  $Q$ . Die Entwässerungsgeschwindigkeit eines Ereignisses  $v_{E,i}$  wurde wie folgt bestimmt:

$$v_{E;i} = \frac{\bar{x}}{t_s} = \frac{\bar{x}}{\left( \frac{\sum_{t=AQ_i}^{EQ_i} Q_t \cdot t}{\sum_{t=AQ_i}^{EQ_i} Q_t} - \frac{\sum_{t=AP_i}^{EP_i} P_t \cdot t}{\sum_{t=AP_i}^{EP_i} P_t} \right)} \quad [\text{m/s}] \quad 3.1$$

Zur Berechnung der Schwerpunkte wird über die manuell definierten Ereignisgrenzen des Niederschlags (Anfang  $AP$ ; Ende  $EP$ ) und des Abflusses (Anfang  $AQ$ ; Ende  $EQ$ ) summiert. Zu bemerken ist, dass die mittlere Fließstrecke für alle Ereignisse innerhalb eines Gebietes identisch ist und daher ohne Index aufgeführt wird. Allerdings variiert sie mit den betrachteten Untersuchungs- und Teilgebieten. Die Entwässerungsgeschwindigkeit wurde in m/s berechnet.

In Gleichung 3.1 tritt das Abflussvolumen bereits in der Berechnung des Abflussschwerpunktes auf, dennoch soll diese Zielgröße noch einmal separat definiert werden. Das Abflussvolumen eines Ereignisses  $Q_{Vol;i}$  wurde über die Dauer des Ereignisses, abzüglich des Basisabflussvolumens, hier vereinfacht als Abfluss am Beginn des Ereignisses angesetzt, summiert:

$$Q_{Vol;i} = \sum_{t=AQ_i}^{EQ_i} Q_t - Q_{AQ_i} \quad [\text{mm}] \quad 3.2$$

Der Parameter  $v_E$  stellt eine fließlängen-normierte Größe dar und beschreibt die Transformationsgeschwindigkeit von Niederschlag zu Abfluss. Zudem kann  $v_E$  direkt als Modellparameter verwendet werden und wird (vgl. folgende Kapitel 4.4) zur Berechnung der geomorphologischen Einheitsganglinie verwendet.  $Q_{Vol}$  ist hingegen nicht direkt im Modell verwendet. Der Quotient aus geschätztem Abflussvolumen und des Niederschlagsvolumen eines bevorstehenden Ereignisses werden verwendet um einen Abflussbeiwert zu berechnen. Von der direkten Schätzung des Abflussbeiwertes wurde aus zweierlei Gründen abgesehen: mögliche Interpolations- und Datenfehler im Niederschlag und eine doppelte Verwendung des Niederschlagsvolumens. Daten- und Interpolationsfehler waren natürlich auch bei der Schätzung der Entwässerungsgeschwindigkeit zu erwarten. Da  $v_E$  aber über die zeitliche Verteilung des Niederschlags und des Abflusses bestimmt wurde und nicht über das Volumen, das meist stärker von solchen Fehlern betroffen ist, wurde  $v_E$  gemäß Gleichung 3.1 berechnet. Neben dem stärkeren Einfluss von Fehlern bei der Bestimmung des Niederschlagsvolumens würde eben jenes im Vorhersageprozess doppelt verwendet werden. Denn das Niederschlagsvolumen wird im Vorhersagebetrieb des Modells (vgl. Kapitel 6.3) als Prädiktor verwendet und wäre, wenn der Abflussbeiwert prognostiziert werden sollte, gleichzeitig ein Teil der Zielgröße. Um eine Trennung von Prädiktor und Zielvariable zu erhalten, wurde  $Q_{Vol}$  als Zielvariable definiert.

### 3.2.2 Klimatische Randbedingungen

Die Mehrzahl der Indikatoren zur Beschreibung der klimatischen Randbedingungen beschreibt den Ereignisniederschlag, da diesem der größte Einfluss auf das zu erwartende Abflussereignis zugeschrieben wurde. Im Falle der Modellierung von zurückliegenden Ereignissen, wie in dieser Arbeit, können diese Indikatoren aus Niederschlagsaufzeichnungen entnommen werden. In der Prognoseanwendung müssen diese Indikatoren aus einer meteorologischen Vorhersage abgeleitet werden, für eine Bemessung ergeben sich diese Indikatoren aus dem Bemessungsniederschlag (DVWK, 1990; Deutscher Wetterdienst, 2005).

Zunächst wurden für jedes Ereignis grundlegende Niederschlagscharakteristika berechnet: Niederschlagsvolumen  $VP$  [mm], Dauer  $DP$  [h], Mittelwert aller Niederschlagsintensitäten größer Null  $I_0$  [mm/h] und maximale Intensität  $I_{Max}$  [mm/h]. Um den Fehler durch Ausreißer durch mögliche Mess- oder Interpolationsfehler zu reduzieren, wurde die maximale Intensität als Mittelwert der drei höchsten Niederschlagsintensitäten eines Ereignisses berechnet.

Um die zeitliche Verteilung eines Niederschlagsereignisses mit einem einzigen Indikator zu charakterisieren wurde ein weiterer Indikator entwickelt. Zwar liefern Kennwerte wie  $I_0$  und  $I_{Max}$  im Kontext mit  $VP$  und  $DP$  bereits einen Eindruck über die zeitliche Verteilung, dennoch beleuchten sie entweder nur den vollständigen Verlauf ( $I_0$  und  $VP$ ) oder nur eine punktuelle Ungleichförmigkeit ( $I_{Max}$  und  $I_0$ ).

Berechnet wurde für jedes Ereignis das minimale Zeitfenster in dem 50% des Ereignisniederschlags fallen, normiert über die Ereignisdauer. Der Indikator  $D_{50}$  zeigt somit an, wie ungleich der Niederschlag über das Ereignis verteilt ist. Ist  $D_{50}$  gleich 0.5, bedeutet dies, dass der Niederschlag komplett gleichverteilt ist. Je kleiner  $D_{50}$  wird, desto zentrierter traten während des Ereignisses die Niederschläge auf. Ansonsten ist  $D_{50}$  ein komparativer Indikator, der lediglich die Unterschiede der zeitlichen Verteilung von Niederschlagsereignissen verdeutlicht. Daher ist er unabhängig vom gewählten Anteil der Niederschlagssumme, so lange dieser konstant gehalten wird.

Des Weiteren wurde die mittlere Temperatur während des Ereignisses  $T_E$  [°C] und die mittlere Temperatur vor dem Ereignis  $T_{Pre}$  [°C] ermittelt, um das Einsetzen von Schmelzprozessen während der Ereignisse erfassen zu können. Liegt  $T_{Pre}$  unter Null und  $T_E$  über Null ist dies ein Hinweis auf eine abschmelzende Schneedecke während des Ereignisses.  $T_E$  wurde dabei wie die Zielvariablen und Niederschlagscharakteristika über die Ereignisdauer ermittelt,  $T_{Pre}$  als Mittelwert über einen Zeitraum von 5 Tagen vor Ereignisbeginn.

### 3.2.3 Anfangszustände

Neben der Beschreibung des Ereignisses selbst ist der Zustand des Einzugsgebietes zu Beginn des Ereignisses von Bedeutung für die Abflussgenese. Daher wurden mit verschiedenen Indikatoren verschiedene Aspekte von Anfangszuständen des Einzugsgebietes beschrieben.

Als einfachste Einschätzung der Vorfeuchte des Gebietes wurde der Basisabfluss des N-A Ereignisses  $Q_0$  [mm/h], also der Abfluss unmittelbar vor Beginn des Ereignisses  $Q(AQ_i - I)$

für jedes Ereignis ermittelt. Dieser Indikator ermöglicht den Algorithmen zwischen unterschiedlichen Abflussniveaus der Ereignisse zu differenzieren.

Um die Vorfeuchte eines Gebietes zu beschreiben, wurden darüber hinaus der Vorregenindex  $VRI$  [-], als gewichtete Summe des Niederschlages der vorhergehenden 30 Tage sowie der Abflussbeiwert  $\psi_{Pre}$  [-] desselben Zeitraumes bestimmt. Der Vorregenindex  $VRI_i$  eines Ereignisses wurde nach folgender Gleichung bestimmt:

$$VRI_i = \sum_{t=0}^{30 \cdot 24} P_{(AP_i-t)} \cdot \left(1 - \frac{t}{30 \cdot 24}\right) \quad 3.3$$

Der Abflussbeiwert vor dem Ereignis  $\psi_{Pre}$  berechnete sich über den identischen Zeitraum von 30 Tagen mal 24 Stunden und ergibt sich als Quotient aus Summe des Abflusses und des Niederschlages:

$$\psi_{Pre} = \frac{\sum_{t=AP_i-(30 \cdot 24)}^{AP_i} P_t}{\sum_{t=AP_i-(30 \cdot 24)}^{AP_i} Q_t} \quad 3.4$$

Die bisher angeführten Kennwerte wurden allein aus Niederschlag- und Abflusszeitreihen in stündlicher zeitlicher Auflösung berechnet. Die Niederschlagszeitreihen sind Gebietsniederschläge und wurden mittels Thiessen Polygonverfahren und anschließender Mittelung je Teilgebiet, aus Punktdaten berechnet. Zur weiteren Charakterisierung von gespeichertem Wasser im EZG zu Ereignisbeginn weitere Indizes unter Zuhilfenahme weiterer Informationsquellen, wie Verdunstungs-, Temperatur- und Abflusszeitreihen, bestimmt.

Um die im EZG gespeicherte Wassermenge vor und während eines Ereignisses, bzw. die daraus resultierende Differenz, qualitativ beschreiben zu können, wurde eine Wasserhaushaltssimulation auf Tageswertbasis durchgeführt. Über den gesamten Zeitraum, in dem Stundenwerte vorlagen (2001-2007 Main; 1999-2012 Regen, 2001-2012 Iller), wurde mit einem Vorlauf von 5 Jahren folgende Gleichung schrittweise gelöst:

$$S_t = S_{t-1} + P_t - EA_t - Q_{G,t} + \varepsilon_t \quad 3.5$$

Der Speicherterm  $S$  wurde in jedem Zeitschritt  $t$  als Speicherfüllung des vorgehenden Zeitschrittes zuzüglich des Niederschlages  $P$  und eines Fehlerterms  $\varepsilon$  berechnet, abzüglich der aktuellen Verdunstung  $EA$  und des beobachteten Gewässerabflusses  $Q_G$ . Als Startwert wurde eine Füllung von 0 mm angesetzt, da angenommen wurde, dass der Einfluss des Startwertes am Ende der 5 Jahre Vorlauf vernachlässigt werden konnte. Der Gewässerabfluss lag als beobachtete Zeitreihe vor. Zeitreihen des Gebietsniederschlags  $P$  mittels Thiessen-Polygonverfahren berechnet. Die aktuelle Verdunstung  $EA$  wurde mittels des Turc-Wendling Verfahren (DVWK, 1996) und Bestandskoeffizienten nach Allen (1998) aus Zeitreihen von Temperatur- und Sonnenscheindauer sowie geographischen und Landnutzungsparametern der EZGs geschätzt.

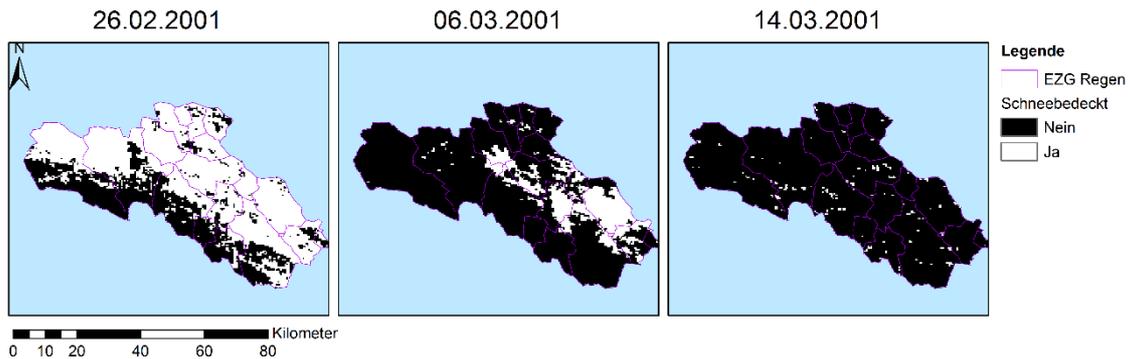
Der Fehlerterm  $\varepsilon$  zur Lösung der Wasserhaushaltsgleichung (Gl.3.5) wurde separat vor mittels eines numerischen Filters geschätzt. Er diente als Ausgleichsterm um Fehler in der Wasserhaushaltsgleichung auszugleichen, die durch Datenfehler oder fehlerbehaftete Schätzungen einzelner Wasserhaushaltskomponenten entstehen können. Anstatt den Wert des Fehlerterms in jedem Zeitschritt zu schätzen oder einen konstanten Wert anzunehmen, wurde je Teileinzugsgebiet eine Zeitreihe für  $\varepsilon$  mit einem numerischen Filter berechnet. Verwendet wurde der Chapman-Filter (Chapman, 1991). Für jeden Zeitschritt  $t$  wurde über den beobachtetem Gewässerabfluss  $Q_G$  und einer Rezessionskonstante  $k$  der Wert des Fehlerterms  $\varepsilon_t$  bestimmt:

$$\varepsilon_t = \frac{k}{2-k} \varepsilon_{t-1} + \frac{1-k}{2-k} Q_{G,t} \quad 3.6$$

Die Rezessionskonstante  $k$  wurde so gewählt, dass die Lösung von Gleichung 3.5 keinen Trend aufwies und über einen langen Simulationszeitraum um  $S_t = 0$  schwankte.

Nach Anwendung des Chapman-Filters, der Berechnung der aktuellen Verdunstung und der abschließenden Lösung von Gl. 3.5 standen für jedes EZG eine simulierte Zeitreihe der gespeicherten Wassermenge  $S$  zur Verfügung. Da die Werte dieser Zeitreihe mit starken Unsicherheiten behaftet waren (Interpolation  $P$  oder Datenunsicherheit  $Q_G$ , um nur einige Fehlerquellen zu nennen) und aufgrund des semi-artifiziellen Ursprungs von  $EA$  und  $\varepsilon$ , mussten die quantitativen Aussagen von  $S$  als unsicher eingestuft werden. Dennoch, der qualitative Verlauf der Speichercurven wurde als ausreichend sicher angesehen, um eine Einschätzung der unterschiedlichen Speicherfüllungen zwischen den Ereignissen zu geben. Mit Hilfe der Zeitreihe von  $S$  wurde für jeden Tag die empirische Unterschreitungswahrscheinlichkeit von  $S$  bestimmt und als Vorfeuchteindikator  $WI$  [-] verwendet.  $WI$  wurde in drei Variationen als Prädiktor verwendet: Als  $WI_E$  am Tag des Scheitels eines betreffenden Ereignisses, als  $WI_{Pre}$  [-] als Mittelwert über die 5 Tage vor Ereignisbeginn und  $\Delta WI$  [-], die mittlere Steigung von  $WI$  über 5 Tage vor dem Ereignis.

Um den ML-Algorithmen die Möglichkeit zu geben die im Gebiet gespeicherte Wassermenge zwischen gefrorenem und flüssigem Wasser zu unterscheiden, wurden die Indikatoren  $SC$  (*Snow cover*) [-] und  $\Delta SC$  [-] in den Datensatz mit aufgenommen. Für jedes Ereignis wurde ermittelt wie viel Prozent das Einzugsgebiet bei Ereignisbeginn mit Schnee bedeckt ist und in welchem Umfang die Schneedecke vor dem Ereignis zurückgegangen ist. In Abbildung 3-9 ist beispielhaft die Veränderung der Schneedecke im Untersuchungsgebiet Regen im Vorfeld eines Hochwasserereignisses, dessen Scheitel am Gebietsauslass am 16.03.2001 erreicht wurde, dargestellt.



**Abbildung 3-9: Veränderung der Schneebedeckung im Untersuchungsgebiet Regen im Vorfeld eines Hochwassers am 16.03.2001 abgeleitet aus MODIS Daten**

Zu erkennen ist, wie die Schneedecke im EZG über 24 Tage im Einzugsgebiet kontinuierlich schrumpft, bis sie zwei Tage vor dem Scheitel des Hochwassers am Gebietsauslass nahezu vollständig verschwunden ist. Der hier dokumentierte Schneeschmelzprozess wird nicht der alleinige Auslöser des Hochwassers gewesen sein, das Abschmelzen des Schnees wird aber eine hohe Anfangsfeuchte bedingt haben.

Die Indikatoren  $SC$  und  $\Delta SC$  wurden aus MODIS Schneebedeckungsdaten (MODIS10A2 Terra) (Hall, Riggs & Solomonson, 2015) abgeleitet, die zeitliche Auflösung dieser Daten beträgt 8 Tage. Daher wurde  $\Delta SC$  aus den Daten der drei letzten vor Ereignisbeginn zur Verfügung stehenden Datensätzen ermittelt. Aus den MODIS Daten wurde eine Zeitreihe abgeleitet, die die prozentuale Bedeckung des jeweiligen Einzugsgebietes angibt. Indikator  $SC$  [-] gibt die prozentuale Schneebedeckung bei Ereignisbeginn an, bzw. dem Zeitpunkt des letzten zur Verfügung stehenden Satellitenbildes. Die gewählten Schneeindikatoren verfügen über keinerlei Information zur Schneemenge, bzw. dem Schnee-Wasser-Äquivalent (SWÄ), sie dienen lediglich als Informationsquelle für die ML-Algorithmen ob vor oder während eines Ereignisses im Teileinzugsgebiet Schnee lag oder nicht. Andere Kennwerte hätten hier ebenfalls verwendet werden können, wie die Temperatur vor Ereignisbeginn, oder eine Simulation des SWÄ. Um die Durchführung einer weiteren Wasserhaushaltssimulation zu vermeiden und um eine weitere Messgröße in diese Studie aufzunehmen, wurde auf MODIS-Daten zurückgegriffen.

### 3.3 Prozessorientierte Klassifikation

Es ist anzunehmen, dass die separierten Ereignisse in ihrer Entstehung sehr heterogen sind, d.h. dass bei unterschiedlichen Ereignissen unterschiedliche hydrologische Prozesse die Abflussgenese dominierten. Um eine Differenzierung der Ereignisse basierend auf den aktiven Prozessen vornehmen zu können, wurden drei unterschiedliche Klassifizierungen der Ereignisse durchgeführt. Dabei wurden zwei unterschiedliche Ansätze zur Unterscheidung von vermuteten aktiven Prozessen angewendet.

Zum einen wurde eine direkte Zuweisung von Abflussprozessen basierend auf Klima- und

Abflussindikatoren durchgeführt. Angelehnt an die Prozessklassifizierung nach Merz & Blöschl (2003) wurde ein Entscheidungsbaum (vergleichbar mit Sikorska, Viviroli & Seibert (2015) und Brunner et al. (2017)) zur Identifizierung des Ereignistyps entwickelt. Dabei wurde zwischen Ereignissen aus langen und kurzen Niederschlägen sowie zwischen Schneeschmelzereignissen und Regen-auf-Schnee Ereignissen unterschieden.

Zum anderen wurde eine indirekte, basierend auf der Analyse von Abflusscharakteristika, Prozessunterscheidung vorgenommen. Dabei wurde für jedes Ereignis das Verhältnis zwischen Abflussvolumen und -Scheitel bestimmt. Bell & Om Kar (1969) wiesen darauf hin, dass eine Variation dieses Verhältnisses auf unterschiedliche aktive Prozesse hinweist. Nach Gaál et al. (2015) können die Verhältniswerte mit den von Merz & Blöschl (2003) eingeführten Prozesstypen in Verbindung gesetzt werden.

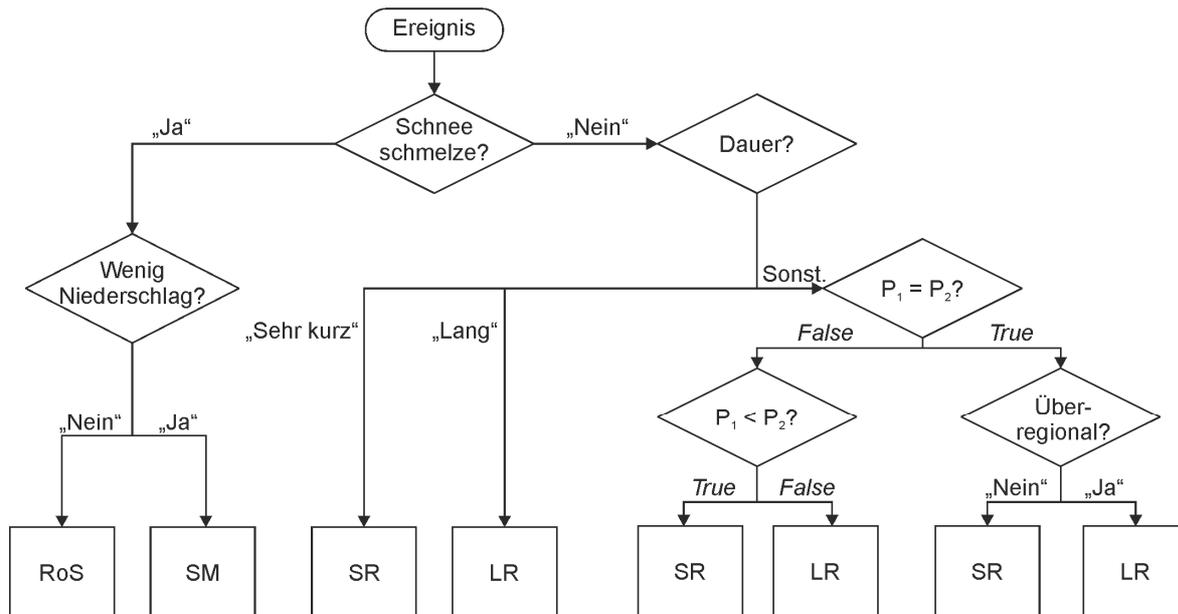
Beide Strategien, die verwendeten Ereignisindizes sowie die Häufigkeiten der identifizierten Ereignistypen, werden im Folgenden näher erläutert.

### 3.3.1 Klassifikation mittels Fuzzy-Entscheidungsbaum

Merz & Blöschl (2003) schlagen eine Unterteilung von Abflussereignissen in die Typen Flut aus Dauerregen, Flut aus kurzem Regen, Sturzflut, Regen auf Schnee und Schneeschmelzereignisse vor. Dazu verwendeten sie Indikatoren der Jahreszeit, Niederschlagsdauer, Niederschlagssumme, Speicherfüllung, Abflussdynamik, räumliche Ausdehnung und einen Schneeeindikator. Während Merz & Blöschl (2003) sich auf fünf Typen beschränkten, führen Sikorska, Viviroli & Seibert (2015) einen weiteren Typ für Ereignisse aus Gletscherschmelzen ein. In der Erweiterung der vorgeschlagenen Methodik von Sikorska, Viviroli & Seibert (2015) werden die Indikatoren zunächst in Fuzzyzahlen umgewandelt und dann in einer Entscheidungsbaumklassifikation verwendet, um Ereignistypen zuzuweisen.

In dieser Arbeit wurden beide Ansätze mit einigen Veränderungen aufgegriffen. Zunächst wurde die Auswahl der Prozesstypen kritisch hinterfragt. In keinem der betrachteten Einzugsgebiete sind Gletscher vorhanden, daher konnte auf den von Sikorska, Viviroli & Seibert (2015) eingeführten Ereignistyp verzichtet werden. Des Weiteren wurde der Typ Sturzflut hier nicht als potentieller Ereignistyp betrachtet. Zum einen ist die zu erwartende Häufigkeit dieser Ereignisse auf der Mesoskala relativ gering und zum anderen ist die Differenzierung zwischen Flut aus kurzem Regen und Sturzflut sehr unsicher. Im Gegensatz zu den übrigen Typen sind sie nicht klar voneinander zu differenzieren, sie werden in der originalen Umsetzung durch Abstufungen einzelner Fuzzyzahlen („Mittel bis hoch“ und „Wenig bis mittel“; „Mittel bis schnell“ und „Schnell“) differenziert.

Die Datensätze sollten demnach in vier Ereignistypen unterteilt werden: Ereignisse aus kurzem Regen (SR) oder aus langem Regen (LR), Schneeschmelzereignisse ohne Regen (SM) oder Regen auf Schnee Ereignisse (RoS). Abbildung 3-10 zeigt den Entscheidungsbaum, mit dem die Ereignistypen identifiziert wurden. Neben den dargestellten Abfragen der Fuzzyzahlen werden zwei Typwahrscheinlichkeiten  $P_1$  und  $P_2$  verwendet, die in den folgenden Gleichungen 3.7 und 3.8 erläutert werden.



**Abbildung 3-10: Entscheidungsbaum zur Identifizierung der Ereignistypen Regen auf Schnee (RoS), Schneeschmelze (SM), kurzes Regenereignis (SR) und langes Regenereignis (LR)**

Am ersten Knoten wird zwischen Ereignissen mit und ohne Schneeeinfluss unterschieden. Grundlage für diese Entscheidung bildet der fuzzyfizierte Indikator  $T_{Pre}$ . Die numerischen Werte des Indikators wurden in die unscharfen Fuzzyzahlen „Ja“ und „Nein“ übertragen und beurteilen ob Schneeschmelze vorliegt oder nicht. Liegt Temperatur unter  $0^{\circ}\text{C}$  vor, ist der Zugehörigkeitswert von „Nein“ größer als der Zugehörigkeitswert von „Ja“ (die Definition der restlichen Grenzwerte der Fuzzyzahlen kann Anhang A, die Fuzzyfunktionen dem digitalen Anhang entnommen werden).

Nach derselben Logik wurde auf der in Abbildung 3-10 linken Verästelung, auf der zweiten Entscheidungsebene nach Niederschlagsvolumen  $VP$  differenziert. Lag der Wert der Fuzzyzahl „wenig“ unter den Werten „mittel“ oder „hoch“ wurde das betreffende Ereignis als *Regen auf Schnee* (RoS) Ereignis klassifiziert. Andernfalls als Ereignis rein aus *Schneeschmelze* SM.

Auf der in Abbildung 3-10 rechten Verästelung wurden die Ereignisse zunächst nach ihrer Dauer  $DP$  differenziert. Handelt es sich um ein langes Niederschlagsereignis kann das Ereignis direkt als *Ereignis aus langem Regen* (LR) klassifiziert werden. Hier hat der Zugehörigkeitswert der Fuzzyzahl „lang“ den höchsten Wert, im Vergleich zu den anderen Fuzzyzahlen der Niederschlagsdauer. Ähnlich eindeutig ist die Zuordnung sollte der Wert der Zahl „sehr kurz“ den höchsten Wert besitzen. Solche Ereignisse werden als *Ereignis aus kurzem Regen* (SR) klassifiziert.

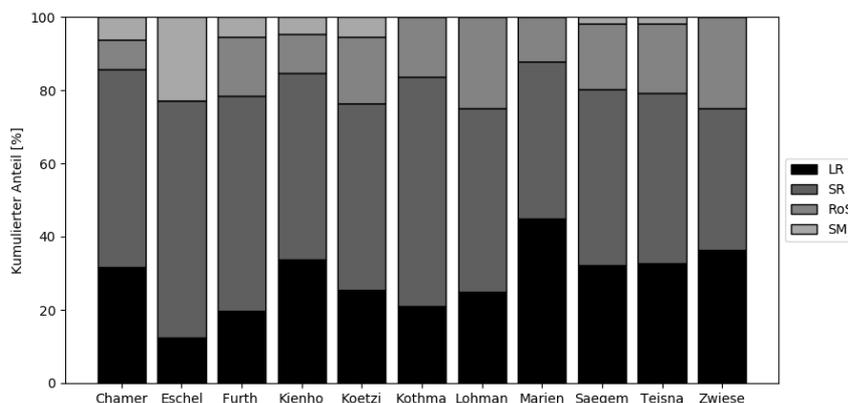
Ist eine Differenzierung aufgrund der Dauer nicht möglich, d.h. Werte der Zahlen „kurz“ und „mittel“ sind höher als „lang“ oder „sehr kurz“, wird auf folgende multiplikativen Typ-Wahrscheinlichkeiten zurückgegriffen:

$$P_1 = \max(D(\text{lang}); D(\text{mittel})) \cdot v(\text{langsam}) \cdot \frac{1}{2}(rk(\text{transreg}) + rk(\text{reg})) \quad 3.7$$

$$P_2 = \max(D(\text{kurz}); D(\text{sehr kurz})) \cdot \max(v(\text{mittel}); v(\text{hoch})) \cdot \frac{1}{2}(rk(\text{lok.}) + rk(\text{reg.})) \quad 3.8$$

Diese Typwahrscheinlichkeiten sind mit Anpassung an die in dieser Arbeit verwendeten Daten der Klassifikation nach (Merz & Blöschl, 2003) entnommen. Beide Typwahrscheinlichkeiten multiplizieren die Werte der fuzzyfizierten Werte von DP,  $v_E$  und der räumlichen Kohärenz<sup>1</sup>. Ist  $P_1$  kleiner als  $P_2$ , wird das Ereignis der SR-Klasse zugerechnet, im anderen Fall LR. Sollten die Typwahrscheinlichkeiten gleich sein (mit einer Fehlertoleranz von 5%) wird eine Entscheidung auf Grundlage der räumlichen Kohärenz getroffen. Überregionale Ereignisse werden dabei als LR klassifiziert, regionale und lokale als SR Ereignisse.

In Abbildung 3-11 sind die Häufigkeiten der Ereignistypen beispielhaft für das Untersuchungsgebiet Regen dargestellt. Da in allen Teilgebieten in etwa dieselbe Anzahl von Ereignissen zur Verfügung (vgl. Tabelle 3-2) stand, wurde die Verteilung der Klassen normiert dargestellt. Zur besseren Übersicht wurden die Gebiete mit ihren ersten 6 Buchstaben abgekürzt. Ersichtlich ist, dass N-A Ereignisse ohne Schneeeinfluss in allen Gebieten den höchsten Anteil haben. Ob die Mehrzahl der Ereignisse aus kurzem oder langem Regen entstand ist dabei lokal variabel. RoS Ereignisse sind in den meisten Gebieten häufiger als reine Schneeschmelzeereignisse. Die hier einzige Ausnahme bildet Eschelkam, in dem alle Schneeeereignisse als durch Schneeschmelzeereignisse klassifiziert wurden. In den ergänzenden Darstellungen zu diesem Kapitel im digitalen Anhang werden die Häufigkeiten in allen Gebieten differenziert dargestellt.



**Abbildung 3-11: Relative Häufigkeiten der Ereignistypen im Untersuchungsgebiet Regen**

<sup>1</sup> Die räumliche Kohärenz (nach Merz & Blöschl, 2003) beschreibt das gleichzeitige Auftreten eines Ereignisses in benachbarten Gebieten. Basierend auf einem Ereignis im betrachteten Gebiet wird der prozentuale Anteil der verfügbaren Gebiete ermittelt, in denen zum gleichen Zeitpunkt ( $\pm 5$  Tage) ein Ereignis auftritt. Da dieser Kennwert in dieser Arbeit nicht weiter verwendet wurde, wird auf Anhang A verwiesen.

An dieser Stelle sollen lediglich die Größenverhältnisse der Häufigkeiten dargestellt werden, um einen Eindruck über die Zusammensetzung des Datensatzes sowie die Funktionalität der Klassifizierungsstrategie zu bekommen. Es soll auch darauf hingewiesen werden, dass diese Klassifikation nicht als die „wahre“ Klassifikation angesehen wird. Sie folgt im Wesentlichen den Angaben und Hinweisen der Literatur und führt zu vergleichbaren Ergebnissen (Nied et al., 2014; Brunner et al., 2017; Keller et al., 2018). Vor allem in den feineren Differenzierungen zwischen LR und SR sowie zwischen RoS und SM sind die Unsicherheiten der Klassenzugehörigkeit groß. Dennoch, die Unterteilung nach Schneebeeinflussung kann als vergleichsweise sicher beurteilt werden und stellt im Hinblick auf aktive Prozesse eine wesentlich wichtigere Unterteilung dar.

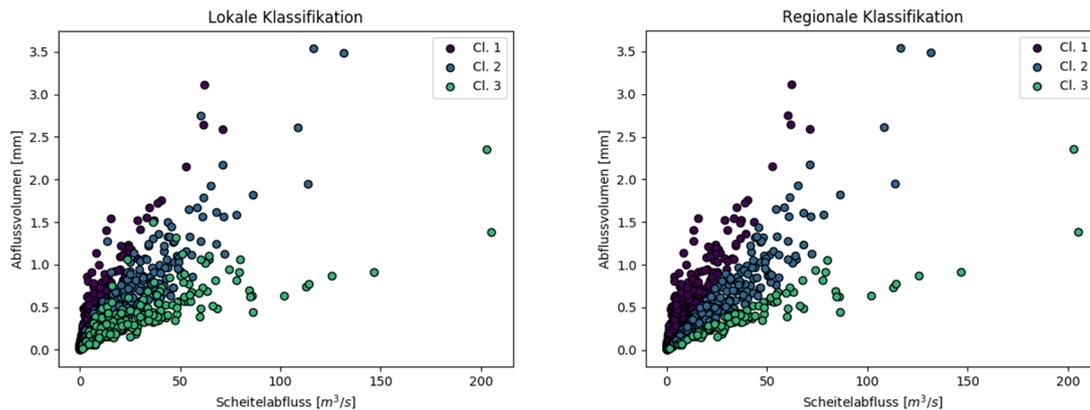
### 3.3.2 Klassifikation mittels Füllen-Scheitel Verhältnis

Wird  $Q_{Vol}$  eines N-A Ereignisses durch den zugehörigen Abflussscheitel  $Q_P$  dividiert, so hat der Quotient die Dimension einer Zeitspanne (Bell & Om Kar, 1969). Diese Zeitspanne ist direkt verknüpft mit der Streuung der zugehörigen Ganglinie und kann daher als zeitliche Varianz der Ganglinie, also ihrer Form, interpretiert werden (Viglione et al., 2010). Da die Form einer Ganglinie durch die aktiven, bzw. dominanten Prozesse definiert wird, kann mittels des Füllen-Scheitel Verhältnisses auf Abflussprozesse rückgeschlossen werden.

Gaál et al. (2012) zeigten, dass zwischen  $Q_{Vol}$  und  $Q_P$  ein linearer Zusammenhang hergestellt werden kann, der für unterschiedliche Ereignistypen unterschiedliche Steigungen aufweist. Dieser Umstand kann dazu genutzt werden zwischen unterschiedlichen Typen zu differenzieren.

In dieser Arbeit wurden zwei Ansätze verfolgt. Zunächst eine Klassifikation auf lokaler Ebene, also eine Differenzierung nach Füllen-Scheitel-Verhältnis, durchgeführt an jedem Pegel. Die Klassifikation wurde mit Hilfe des *k-Means* Algorithmus (Pedregosa et al., 2011) und den  $t_Q$ -Werten (Verhältnis von  $Q_{Vol}$  zu  $Q_P$ ) der Ereignisse durchgeführt. Gesucht wurde nach drei Clustern, da mittels des Füllen-Scheitel Verhältnisses zwischen Ereignissen mit hohem Scheitel im Verhältnis zum Volumen, Ereignissen mit hohem Volumen im Vergleich zum Scheitel und ausgeglichenen Ereignissen differenziert werden kann. Ergebnisse dieser Klassifikation werden nachfolgend als *tq\_lokal* referenziert.

Schon Bell & Om Kar (1969) weisen darauf hin, dass das Füllen-Scheitelverhältnis auch als Gebietscharakteristik genutzt werden kann. Auch Gaál et al. (2012) weisen auf regionale Abhängigkeiten hin. Um diesen Effekt in die Klassifikation einzubeziehen, also Ähnlichkeiten zwischen den Gebieten, wurde eine zweite Klassifikation durchgeführt. Bei diesem Ansatz wurden alle Ereignisse aller Pegel eines Untersuchungsgebietes gemeinsam klassifiziert. Die Ergebnisse dieser Klassifikation werden nachfolgend als *tq\_regional* referenziert. Beide Klassifikationen sind in nachfolgender Abbildung 3-12 dargestellt.



**Abbildung 3-12: Füllen-Scheitel Verhältnisse für das Untersuchungsgebiet oberer Main, lokale (links) und regionale (rechts) Klassifikation**

Zu erkennen ist, dass in der gewählten Darstellung aller Teilgebiete im Untersuchungsgebiet oberer Main, die lokale Klassifikation nicht trennscharf wirkt. In der Betrachtung aller Gebiete ist dies auch der Fall, auf der Ebene jedes eines einzelnen Pegels ist sie aber nicht überlappend. Als Cluster 1 werden die Ereignisse bezeichnet, die im Vergleich zu ihrem Scheitel ein großes Volumen aufweisen, dazu gegensätzlich Ereignisse im Cluster 3 mit hohem Scheitel im Vergleich zum Volumen. Cluster 2 steht für Ereignisse mit einem ausgeglichenen Verhältnis von Fülle und Scheitel.

### 3.4 Güteparameter

Um die Aussagefähigkeit von Modellen zu überprüfen, werden Simulationsergebnisse mit bekannten Größen verglichen. Die Abweichung der Simulation von der bekannten Größe wird dann mit Hilfe eines Gütemaßes beschrieben werden.

Die meisten Gütemaße verwenden die punktuelle Distanz zwischen den individuellen Ordinaten der Simulation  $y$  und der Beobachtung  $x$ . Eines der simpelsten Maße ist der mittlere absolute Fehler (*Mean Absolute Error*, MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad 3.9$$

Da alle Abweichungen im Betrag betrachtet werden, hat der MAE gegenüber dem BIAS, dem Mittelwert aller Abweichungen, den Vorteil, dass sich Über- und Unterschätzungen nicht gegenseitig aufheben. Der MAE ist definiert im Intervall  $[0; \infty]$ , Zielwert ist 0.

Im Vergleich zu anderen Maßen, wie dem mittleren quadratischen Fehler (MSE), die mehrere Aspekte von Anpassungsgüte vereinen (im Falle des MSE wird der BIAS sowie die Fehlervarianz kombiniert), ist der MAE eindeutig interpretierbar (Ramsay & Silverman, 2005). Der MAE wurde in dieser Arbeit vor allem zur Evaluierung von Machine-Learning

Vorhersagen verwendet, da in diesem Falle die Vorhersage jedes einzelnen Wertes der Vorhersage entscheidend ist.

Für die Evaluierung von Simulationen hydrologischer Zeitreihen wird in der Literatur häufig die Nash-Sutcliffe-Effizienz (NSE) verwendet (Nash & Sutcliffe, 1970). Die NSE stellt einen normierten MSE dar, die ermöglicht die Modellgüte zwischen unterschiedlichen Untersuchungsgebieten zu vergleichen. Gupta et al. (2009) konnten zeigen, dass die NSE in mehrere Komponenten zerlegt werden kann: Abbildung der Standardabweichung  $\sigma$ , des Mittelwertes  $\beta$  und der Korrelation  $r$  zwischen Simulation und Beobachtung. In Anlehnung an die Arbeiten von Gupta et al. (2009) wurden folgende Komponenten verwendet:

$$\alpha = \frac{\sigma(y)}{\sigma(x)}, \beta = \frac{\mu(y)}{\mu(x)}, r = \frac{s_{xy}}{s_x \cdot s_y} \quad 3.10$$

Die lineare Korrelation wurde berechnet mittels erwartungstreuer Schätzung der Streuung der Beobachtung  $s_x$ , Simulation  $s_y$  und der Kovarianz zwischen  $x$  und  $y$ ,  $s_{xy}$ . Mit diesen Komponenten konnten die verschiedenen Aspekte der Modellgüte eindeutig beschrieben werden. Mittels  $\alpha$  konnte über die Varianz der Reihen auf die Reaktivität des Modells im Vergleich zur Beobachtung geschlossen werden. Reaktivität steht dabei bspw. für die Abbildung von kurzzeitigen Abflussspitzen o.ä. Parameter  $\beta$  verwendet dem Mittelwert der Reihen. Bei gleicher Länge von Simulation und Beobachtung steht  $\beta$  für das Verhältnis zwischen simuliertem und beobachtetem Volumen. Im Vergleich zu Gupta et al., (2009) wurde auf eine Normierung von  $\sigma(x)$  auf  $\beta$  verzichtet, um eine eindeutige Interpretation zuzulassen. Der lineare Korrelationskoeffizient zeigt in Ergänzung zu den beiden anderen Maßen die zeitliche Übereinstimmung von Simulation und Beobachtung. Zielwerte aller Maße ist 1,  $\alpha$  und  $\beta$  sind definiert in  $[-\infty; +\infty]$ ,  $r$  in  $[-\infty; 1]$ .

Zur Auswahl von Modellen basierend auf dem Vergleich von Zeitreihen sowie zur Kalibrierung von Modellen wurde die Kling-Gupta-Effizienz verwendet, die  $\alpha$ ,  $\beta$  und  $r$  gleichgewichtet verrechnet:

$$KGE = 1 - \sqrt{(1-\alpha)^2 + (1-\beta)^2 + (1-r)^2} \quad 3.11$$

Die KGE ist definiert in  $[-\infty; 1]$ , der Zielwert ist 1 und steht für eine komplette Übereinstimmung zwischen Simulation und Beobachtung.

## 4 Modellentwicklung

---

Modelle sind die formalisierte Darstellung des Systems *Einzugsgebiet* und repräsentieren dessen Aufbau und die Funktionsweise in vereinfachter Form. Bei der Entwicklung oder Auswahl eines Modells ist stets zwischen einer möglichst realistischen Darstellung der Abflussprozesse und dem Grad der Detaillierung im Hinblick auf das Ziel des Modells abzuwägen (Blöschl et al., 2013). Ein weiteres Augenmerk ist auf die räumliche Organisation, d.h. nicht zufallsbedingte räumliche Muster, des betrachteten Einzugsgebietes zu legen, da diese die Funktionsweise des Einzugsgebietes widerspiegeln (Grayson & Blöschl, 2001).

Ein natürliches Einzugsgebiet und der Abfluss am Auslass dieses Gebietes sind ein ko-evolutionäres System. Das bedeutet, das Einzugsgebiet definiert den Fließweg des Wassers und das Wasser definiert durch Erosion die Form des Einzugsgebietes (Kirkby, 2005). Diese gegenseitige Abhängigkeit spielt sich zwar auf unterschiedlichen Zeitskalen ab, ist aber für die Untersuchung der nicht-linearen Natur des Abflussprozesses von entscheidender Bedeutung. Sichtbare räumliche Muster in geologischen Kennwerten und Topologie müssen auf Grund dieser Abhängigkeit wichtige und nutzbare Informationen über den Abflussprozess enthalten (Grayson & Blöschl, 2001). Eines der bekanntesten Konzepte räumliche Muster (d.h. die Geomorphologie eines Einzugsgebietes) mit der Entstehung von Abfluss zu verknüpfen ist die geomorphologische momentane Einheitsganglinie (*Geomorphologic instantaneous unit hydrograph*, GIUH) von Rodríguez-Iturbe & Valdés (1979). Sie befassten sich mit folgender Fragestellung: wie kann eine geordnete Darstellung der Geomorphologie mit der Abflussdynamik verknüpft werden? Ihr Vorschlag stellt eine Einheitsganglinie vor, die als Integral über alle Reisezeiten eines Niederschlagsimpulses durch ein Einzugsgebiet interpretiert werden kann.

Grundsätzlich besteht ein GIUH aus zwei Operatoren: Zum einen der Wahrscheinlichkeit, dass ein Niederschlagsimpuls über einem bestimmten Bereich des Einzugsgebietes niedergeht, und zum anderen aus der zeitabhängigen Wahrscheinlichkeit, dass dieser Impuls an den Gebietsauslass weitergegeben wird. Letzterer Operator wird wiederum aus zwei Komponenten gebildet: einer Übergangsmatrix und einer Wartezeitenmatrix. Die Übergangsmatrix beschreibt die Geomorphologie des Einzugsgebietes mit Hilfe von Strahler-Ordnungszahlen und Horton-Verhältnissen. Im Detail werden die Übergänge von Teilgebieten geringerer Ordnung in unterliegende Teilgebiete mit höherer Ordnung beschrieben. Die Wartezeitenmatrix definierte sich hauptsächlich über die Entwässerungsgeschwindigkeit des Einzugsgebietes. Diese wurde von Rodríguez-Iturbe & Valdés (1979) bereits als variabel definiert und sollte für jedes Ereignis neu bestimmt werden. Dies widerspricht zwar dem

klassischen Konzept einer Einheitsganglinie (Sherman, 1932; Dooge, 1959), ist für die Reproduktion der natürlichen Nicht-Linearität des Abflussprozesses aber entscheidend (Rodríguez-Iturbe, Devoto & Valdés, 1979; Gupta, Waymire & Wang, 1980).

Die Verwendung von Strahler-Ordnung und Horton-Verhältnissen ist unter anderem dem Erscheinungsjahr der Veröffentlichung des Konzeptes zuzurechnen. Da keine flächendeckenden Geodaten zur Verfügung standen, musste vor allem auf Kartenwerke mit Gewässerverläufen zurückgegriffen werden. Die gewählte Repräsentation mittels Horton-Zahlen und Strahler-Ordnung stellte einen einfachen Weg da diese spärlichen Informationen zu verarbeiten (Rigon et al., 2016). Im Laufe der Zeit standen räumliche Daten immer umfangreicher zur Verfügung und es wurden Konzepte gesucht diese geomorphologischen Informationen zu verwenden. Mesa & Mifflin (1986) führten die von Kirkby (1976) entwickelte Weite-Funktion im Rahmen des GIUH ein. Die Weite-Funktion ist eine Beschreibung des Einzugsgebietes basierend auf der Fließstrecke jedes Punktes im Einzugsgebiet zum Gebietsauslass. Snell & Sivapalan (1994) konnten für die von ihnen untersuchten Gebiete zeigen, dass die Weite-Funktion, die auch Flächen-Funktion genannt wird, eine bessere Abbildung der Geomorphologie darstellt als die Verwendung von Horton-Verhältnissen.

Der von Rodríguez-Iturbe & Valdés (1979) und Gupta, Waymire & Wang (1980) gezeichnete Rahmen der Verbindung zwischen Geomorphologie und Abflussdynamik wurde in dieser Arbeit aufgegriffen. Die entwickelte Modellstruktur umfasst Änderungen und Anpassungen des originalen Modells, welche im Folgenden erläutert werden. Dabei wird zunächst auf die geordnete Abbildung der Geomorphologie eingegangen und anschließend auf die Abflussdynamik. Abschließend werden die Modellsynthese und Details des Modells vorgestellt. Zunächst werden aber die Ziele und Rahmenbedingungen dargestellt, an denen sich die Entwicklung des Modells orientierte.

### 4.1 Zielsetzung

Um Modellergebnisse vergleichen zu können und um überhaupt eine funktionierende Modellstruktur zu entwickeln, war es zunächst erforderlich Modellziele und deren Skalen klar zu definieren. Raum- und Zeitskalen der Modellstruktur wurden an diese Definition sowie an die Analyse räumlicher Organisation der Einzugsgebiete angepasst (Blöschl & Sivapalan, 1995). Es wäre nicht zielführend klein-räumige Muster wie bspw. ein urbanes Straßennetz in Einzugsgebieten mit mehreren 1000 km<sup>2</sup> zu berücksichtigen und umgekehrt.

Das allgemeine Ziel dieser Arbeit ist die Bestimmung von Direktabflussganglinien in unbeobachteten Gebieten. Relevant waren demnach alle Prozesse, die Niederschlagswasser unmittelbar in Gewässerabfluss transformieren. Vernachlässigt wurden (zunächst) Verdunstungsprozesse, Hochwasser aus großen Schneeschmelzen oder durch starke Interaktion mit dem Grundwasser sowie Karstprozesse. Die zeitlich-räumliche Skala umfasst nach Blöschl & Sivapalan (1995) Flächen von 1 km<sup>2</sup> bis 10000 km<sup>2</sup> und eine Zeitschrittweite von 1 Mi-

nute bis 1 Monat. Um eine ausreichende Datengrundlage zu gewährleisten, wurde die zeitliche Auflösung des Modells und der Daten auf 1 Stunde definiert und die Länge der betrachteten Zeiträume auf maximal 1 Monat. Die Länge der manuell separierten Abflussergebnisse bleibt in der Regel unter dieser Höchstdauer. Ereignisse, die eine längere Dauer aufweisen, wiesen eine langfristige Änderung des Basisabflusses auf und wurden daher von dieser Studie ausgeschlossen.

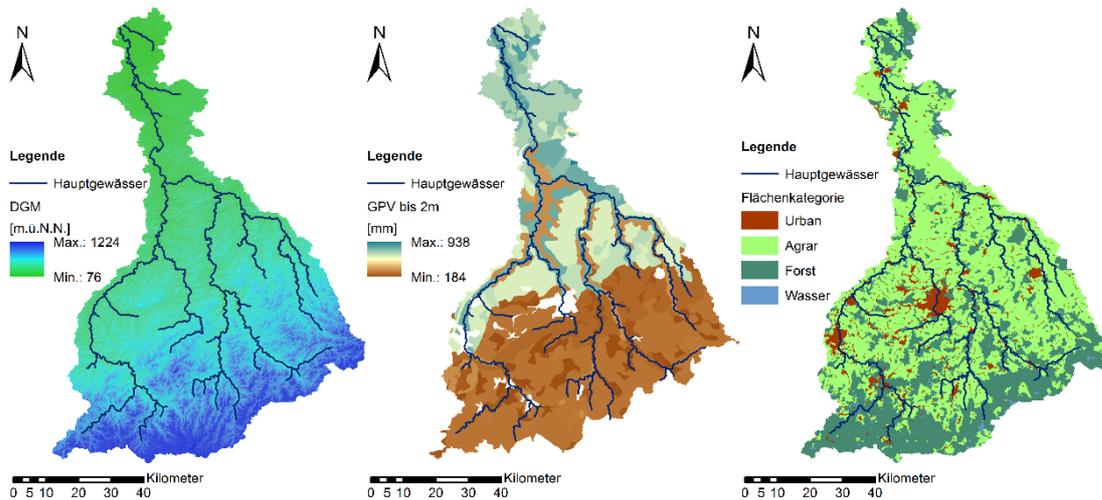
Im Falle von unbeobachteten Gebieten liegen, der Definition nach, keine oder nur unzureichende Daten in dem Gebiet vor. Für dieses Problem werden üblicherweise Regionalmodelle oder regionale Modellierungsansätze angewendet (Wagener, Wheeler & Gupta, 2004; Blöschl et al., 2013). Die beiden Ansätze unterscheiden sich dahingehend, dass ein Regionalmodell eine Modellstruktur ist, die Aussagen für alle Gebiete innerhalb einer Region trifft, und ein regionaler Modellierungsansatz lokale Modelle an beobachtete Gebiete anpasst. Im Anschluss werde diese versucht auf das Zielgebiet zu übertragen. Im hier vorgestellten Modell wurde ein Regionalmodell eingebunden und eine regionale Modellierung wurde beim Vergleich zu bestehenden Ansätzen verwendet.

## **4.2 Darstellung der Geomorphologie**

Eine der grundlegenden Annahmen von Rodríguez-Iturbe & Valdés (1979) ist, dass das System Einzugsgebiet mit der gewählten geordneten Darstellung der Geomorphologie in all seinen möglichen Formen gut beschrieben ist. Alle folgenden Analysen und Zusammenhänge bauen auf der Richtigkeit der gewählten Metrik auf. Die Wahl der geomorphologischen Beschreibung spielt demnach eine zentrale Rolle.

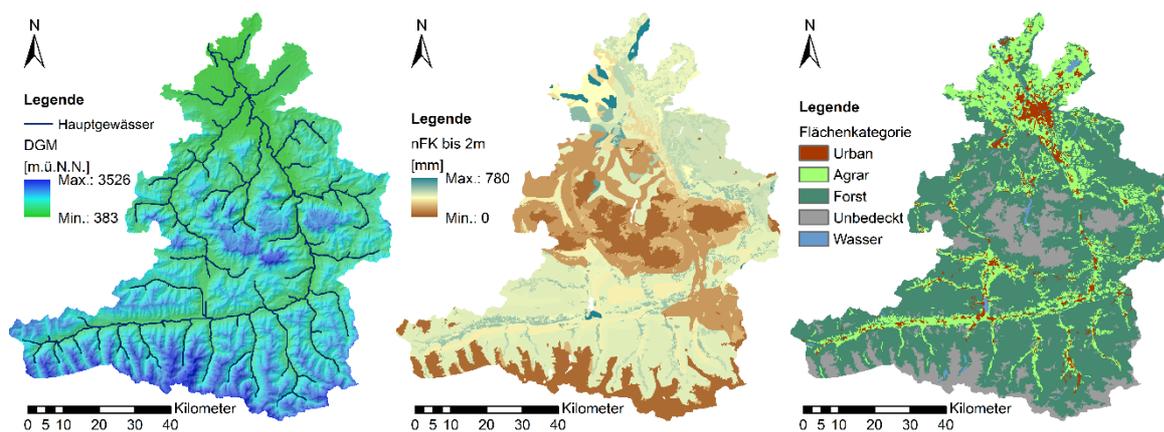
Anhand zweier Beispiele soll gezeigt werden, welche Strukturen und Ordnungsprinzipien gebietsübergreifend die Form eines Einzugsgebiets definieren. Aus diesen Beobachtungen wird abgeleitet, welches die ordnende Größe innerhalb eines Einzugsgebietes ist. Abschließend wird dargestellt wie diese Größe verwendet wurde um räumliche Muster in der Modellstruktur abzubilden.

Zunächst soll das in Abbildung 4-1 dargestellte Einzugsgebiet der Mulde betrachtet werden. Das Einzugsgebiet weist eine klare Süd-Nord Orientierung der Gewässer auf und besteht im Ober- und Mittellauf aus drei größeren parallelen Gewässerarmen. Im Vergleich der räumlichen Strukturen der dargestellten Kennwerte ist zum einen der abrupte Wechsel von geringem Gesamtporenvolumen (GPV) im vornehmlich bewaldeten Süden zu hohem GPV im agrarwirtschaftlich genutzten Norden zu beobachten. Zum anderen sind um die parallelen Flussarme im Mittellauf Flussauen mit geringerem Porenvolumen als die umgebenden Flächen sichtbar. Zu erkennen ist also, wie die Gewässer den fruchtbaren Boden im Laufe der Zeit im Mittel- und Unterlauf abgetragen haben und die vorherrschenden topographischen Verhältnisse die Ausformung der Gewässer- und Landnutzungsstrukturen bedingen.



**Abbildung 4-1: Geokennwerte im Einzugsgebiet der Mulde in Sachsen (v.r.n.l.): Höhe, GPV bis 2m Tiefe und Landnutzung**

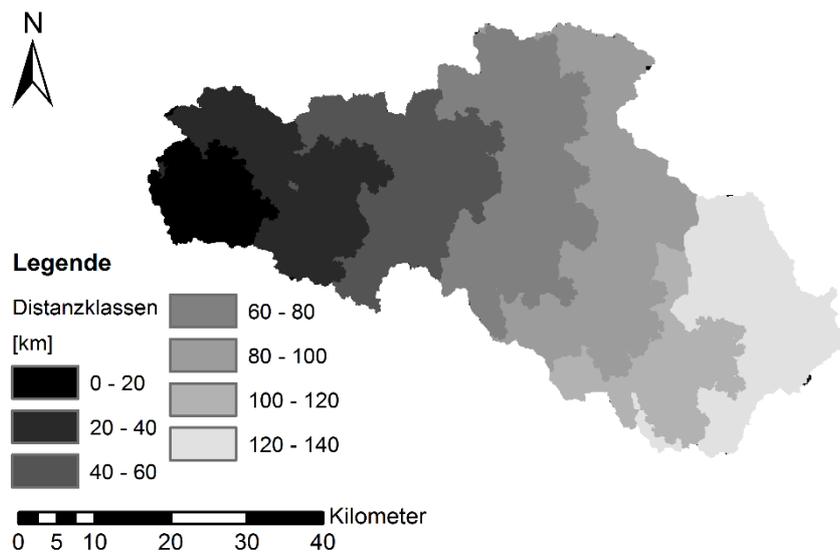
Als zweites Beispiel ist in Abbildung 4-2 das Einzugsgebiet der Salzach dargestellt. Die Topologie dieses Einzugsgebiets unterscheidet sich signifikant vom Einzugsgebiet der Mulde. Zentral im Einzugsgebiet befinden sich die Berchtesgadener Alpen, bestehend aus mehreren Hochplateaus und Gipfeln sowie einigen Seen. Die Salzach, deren Quelle an der süd-westlichen Grenze des Einzugsgebiets liegt, umfließt diese Gebirgsgruppe zunächst im Süden in Richtung Osten und fließt dann östlich davon in Richtung Norden durch die klar zu erkennende urbane Region Salzburg. In den flachen Regionen im Norden des Einzugsgebietes, Nahe des Gebietsauslasses, sind die Porenräume des Bodens wesentlich höher als im Rest des Einzugsgebietes mit Ausnahme der Täler, bzw. den Gewässerauen um die Gewässer. Mit Zunahme des Abstandes vom Gewässer sinkt die nutzbare Feldkapazität des Bodens drastisch bis hin zu unbedecktem Felsboden mit nahezu keinem Speichervermögen. Die Landnutzung folgt dieser Verteilung von Steigung und Porenräumen.



**Abbildung 4-2: Geokennwerte im Einzugsgebiet der Salzach, Österreich (v.r.n.l.): Höhe, nFK bis 2m Tiefe und Landnutzung**

In diesen beiden kurzen Darstellungen geomorphologischer Strukturen in sehr unterschiedlichen Gebieten (hinsichtlich Topographie, Geographische Lage, Geomorphologie) konnte die Organisation des Einzugsgebietes, d.h. nicht zufallsbedingte Strukturen der Oberfläche und des Untergrundes, in beiden Gebieten über die Struktur der Gewässer beschrieben werden. Die Gewässer sowie die Distanz der übrigen Flächen zum nächsten Entwässerungspunkt können als über-regionale Ordnungsgröße angesehen werden. Diese Beobachtungen sprechen für eine Metrik angelehnt an die von Kirkby (1976) eingeführte Weite-Funktion. Erfasst werden sollte ein beliebiger Kennwert, in den hier gezeigten Beispielen Porenraum-Parameter. Da die Weite-Funktion lediglich zur Erfassung der an ein Gewässerobjekt angeschlossenen Fläche dient, wurde die *Distanz-Faktor-Funktion* (DFF) entwickelt. Neben der Fläche kann mit dieser Funktion ein beliebiger Kennwert  $C$  der Fließstrecke zum Gebietsauslass gegenübergestellt (Oppel & Schumann, 2017) werden.

Die Fließstrecke wird in diesem Falle nicht als kontinuierliche Größe erfasst sondern in Klassen unterteilt. Basierend auf der räumlichen Auflösung  $\Delta s$  werden die Fließstrecken innerhalb eines Einzugsgebietes in  $N_S$  Distanzklassen unterteilt. Abbildung 4-3 zeigt beispielhaft das Einzugsgebiet des Pegels Marienthal, das mit einer räumlichen Auflösung von  $\Delta s = 20$  km in sieben Distanzklassen unterteilt wurde.



**Abbildung 4-3: Unterteilung des EZGs Marienthal in Distanzklassen mit  $\Delta s = 20$  km**

Es wurde außerdem zwischen Gewässerfließstrecke bis zum Auslass des Gebietes  $x_S$  und Hangfließstrecke bis zum Gewässer  $x_H$  unterschieden. Dabei haben Gewässer eine Hangfließstrecke von  $x_H = 0$  und Hangflächen wurde das  $x_S$  des nächstgelegenen Gewässers zugeordnet. Die räumliche Auflösung  $\Delta s$  bezieht sich auf die Gewässerfließstrecke und  $\Delta o$  auf die Hangfließstrecke.

Wird nun eine Distanzklasse  $i$  betrachtet, so werden alle Zellen, deren Gewässerfließstrecke  $x_S$  zwischen  $i \cdot \Delta s$  und  $(i+1) \cdot \Delta s$  liegt, dieser Klasse zugerechnet. Diese Zellen lassen sich als Menge  $B$  definieren:

$$B = [i \cdot \Delta s; (i+1) \cdot \Delta s] \quad 4.1$$

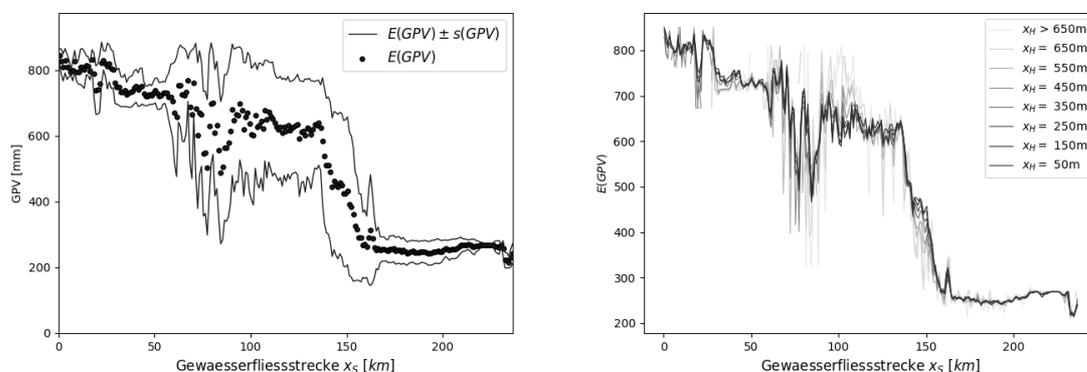
Um nun den Kennwert  $C$  innerhalb von Klasse  $i$  zu beschreiben, wurden der Erwartungswert  $E(C)$  und Standardabweichung  $s(C)$  der Kennwerte der Klasse nach folgenden Gleichungen berechnet:

$$E(C)_i = \frac{1}{W(i \cdot \Delta s)} \sum_{(j|x_{s,j} \in B)} C_j \quad 4.2$$

$$s(C)_i = \sqrt{\frac{1}{W(i \cdot \Delta s)} \sum_{(j|x_{s,j} \in B)} (C_j - E(C)_i)^2} \quad 4.3$$

dabei steht  $W(i \cdot \Delta s)$  für die Anzahl von Zellen innerhalb der Distanzklasse  $i$  und entspricht damit der nicht-normierten Flächenfunktion (Snell & Sivapalan, 1994). In Abbildung 4-4 (links) wurde das GPV im Einzugsgebiet der Mulde mittels der DFF erfasst. Die räumliche Auflösung ist  $\Delta s = 1$  km und  $E(GPV)$  sowie  $s(GPV)$  wurden der Fließstrecke zum Gebietsauslass gegenübergestellt. Auf der rechten Seite von Abbildung 4-4 wurden anstatt  $E(GPV)$  der jeweiligen Distanzklassen GPV-Mittelwerte für  $x_H$ -Klassen dargestellt, mit  $\Delta o = 100$  m. Über den Mittelwert und die Standardabweichung wird der Kennwert auf der gewählten Raumscale erfasst. Dabei bleibt die natürliche Varianz des Kennwertes, die unterhalb der gewählten räumlichen Auflösung liegt, erhalten. Des Weiteren kann bei zusätzlicher Berücksichtigung der Hangflächen die Verteilung des Kennwertes in Abhängigkeit des Fließweges zum Gewässer erfasst und abgebildet werden.

Das Auftreten hoher Varianzen, wie im mittleren Teil des linken Panels in Abbildung 4-4 zu erkennen, weist auf eine notwendige Unterteilung des Einzugsgebietes hin. Oppel & Schumann (2017) stellen eine Methode vor, wie mittels DFF-Analyse Teileinzugsgebiete definiert und zonale Unterteilungen vorgenommen werden können. Die vorgestellte Distanz-Faktor-Funktion sowie die Flächen-Funktion wurden als Grundlage für die Darstellung der geomorphologischen Gegebenheiten im Modell verwendet.



**Abbildung 4-4: Distanz-Faktor-Funktion des gesamten Porenvolumens im EZG Mulde, dargestellt als Mittelwert und Standardabweichung der  $x_s$ -Distanzklasse (links) und Mittelwerte differenziert nach  $x_H$  (rechts)**

### 4.3 Abflussdynamik

Bereits Rodríguez-Iturbe & Valdés (1979) hoben die Wichtigkeit der Entwässerungsgeschwindigkeit  $v_E$  als Dynamikparameter des GIUH hervor. Er muss nach ihrer Darstellung zur Abbildung der nicht-Linearität des Abflussbildungsprozesses als inter-Ereignis Variable angesehen werden. Während eines Ereignisses wurde die mittlere  $v_E$  als stationär angesehen, lediglich eine räumliche Geschwindigkeitsverteilung über die Ordnungszahlen der Teilgebiete war vorgesehen.

Ausgehend von diesen Annahmen konnte eine Datengrundlage für eine empirische Untersuchung von  $v_E$  aufgebaut werden. Die mittlere  $v_E$  konnte demnach als die Schwerpunktlaufzeit eines Ereignisses dividiert durch die mittlere Fließlänge im Gebiet berechnet werden (im Kontext des originalen GIUH wäre es das Horton Längenverhältnis  $R_L$ ). Somit konnte für jedes Ereignis der Datenbasis ein  $v_E$ -Wert geschätzt werden. Ziel des Modells sollte nun sein basierend auf den Ereignisindikatoren, die zu Beginn eines Ereignisses bekannt sind, ein Modell zu entwickeln, um  $v_E$  für jedes Ereignis vorherzusagen, und dieses anschließend mittels eines GIUH in eine Abflussganglinie zu transformieren.

Neben der ereignisweisen Parametrisierung ist ein weiterer Kritikpunkt am GIUH Konzept das Fehlen eines Abflussbildungsmoduls (Rigon et al., 2016). Es stand natürlich eine Vielzahl von Verfahren unterschiedlicher Komplexität zur Verfügung, die vor Berechnung der Abflusskonzentration angewendet werden konnte. Da in unbeobachteten Gebieten a-priori kaum Kenntnisse über Abflussbildungsprozesse vorliegen, wurde die Modellierung des Abflussvolumens so simpel wie möglich gehalten. Modellintern wurde ein Abflussbeiwert-Ansatz verwendet, d.h. der Niederschlag wird vor Berechnung der Abflusskonzentration so korrigiert, dass sein Volumen dem zu erwartenden Abflussvolumen  $Q_{Vol}$  entspricht. Der Abflussbeiwert, d.h. der Koeffizient aus zu erwartendem Abflussvolumen und (erwartetem) Niederschlagsvolumen, ist wie  $v_E$  für jedes Ereignis unterschiedlich und  $Q_{Vol}$  wurde damit zur zweiten Zielvariable der ML-Algorithmen.

Wird ein Modell auf Prozesse angepasst, die im Zielgebiet nicht vorkommen oder nur eine untergeordnete Rolle spielen, kann ein solches Modell keine zuverlässigen Aussagen treffen. Die Übertragbarkeit des Modells auf unbeobachtete Gebiete wäre in einem solchen Fall nicht gegeben (Blöschl et al., 2013). Demnach war die a-priori Einschätzung der hydrologischen Prozesse von ebenso zentraler Bedeutung wie die Wahl der Repräsentation der Geomorphologie. Für die gewählte Raumscale existiert eine Vielzahl von Hypothesen, die den Zusammenhang von Niederschlagscharakteristiken, Vorfeuchte, Temperatur, Schneebedeckung und Gebietsstrukturen mit den Zielvariablen erklären können. Die meisten basieren aber auf klein-räumigen Beobachtungen und wurden in anderen Modellen auf eine höhere Skale übertragen. Ihre Eignung war daher fraglich und ihre Verwendung würde der Entwicklung eines Modells für unbeobachtete Gebiete bei gegebener Datenlage und Zielsetzung zu enge Grenzen setzen.

In dieser Arbeit wurde daher eine von Shen et al. (2018) aufgezeigte neue Herangehensweise zur Erarbeitung einer Hypothese gewählt: Es werden datengetriebene Modelle (DDM), Machine-Learning (ML) und/oder Multiple lineare Regressionen (MLR) angewendet, um die Beziehung zwischen Prädiktoren und Zielvariablen herzustellen. Bei dieser Herangehensweise durchlief die Analyse mehrere Phasen: Zunächst wurde der beschriebene grobe Modellrahmen definiert, Prädiktoren und Zielvariablen gewählt. Anschließend wurde von den ML-Algorithmen eine Vielzahl von Hypothesen entworfen, getestet und verworfen, bis sich die beste aus ihnen herauskristallisierte, für den Algorithmus erkennbar durch die (in der Regel) Minimierung seiner Zielfunktion. Diese wurde mittels vorenthaltener Daten nochmals überprüft. Der nächste Schritt war die Rückübersetzung der „Maschinen“ Theorien in verständliche Prozesstheorien, die dann für weitere Analysen und Verallgemeinerungen verwendet werden konnte.

In den Abschnitten zuvor wurde der grobe Modellrahmen bereits aufgezeigt, Zielvariablen, Anforderungen an Zeit- und Raumscale wurden ebenfalls definiert. Die Auswahl der Prädiktoren konnte entweder modellbasiert oder informationsbasiert durchgeführt werden (bspw. Sharma & Mehrotra, 2014), dabei haben beide Strategien Vor- und Nachteile. Da die Modellstruktur dynamisch ermittelt werden sollte und für eine informationsorientierte Auswahl eine sehr große Datengrundlage erforderlich war, wurde eine modellgestützte Auswahl der Prädiktoren durchgeführt. Im Laufe der Arbeit (Kapitel 6.3) wurden verschiedene Kombinationen von Indikatoren getestet und ausgewertet.

Die Intention dieses alternativen Vorgehens war das Modell zunächst so flexibel wie möglich zu halten. Der fließwegorientierte Aufbau der Modellstruktur ermöglichte bereits die Anpassung an unterschiedliche Modellformen. Eine Festlegung auf starre Konzepte zur Reproduktion des Zusammenhangs zwischen Ereignisindikatoren und Zielvariablen hätte das Modell zu stark eingeschränkt. Außerdem ergab sich durch die Verwendung von DDM Regionalmodellen die Möglichkeit die Hypothesen der Algorithmen in verschiedenen Regionen zu vergleichen und Rückschlüsse auf natürliche Prozesse zu generieren, die unter Umständen bei einer Vorfestlegung übersehen worden wäre.

An dieser Stelle sollen zwei weitere Arbeitshypothesen vorgestellt werden, die im Laufe der Arbeit auf „klassischem“ Wege überprüft wurden, da sie die aufgezeigten Rahmen der DDM-Analyse betreffen.

Zum einen wird eine Parallelstruktur des Dynamikmoduls getestet, die sich durch folgende Hypothese begründet:

Eine Abbildung der Prozesse innerhalb eines Einzugsgebietes erfordert eine prozess-orientierte Differenzierung der Modellstruktur, da eine einheitliche Modellstruktur die Heterogenität der Prozesse nicht nachbilden kann.

Überprüft wurde, ob eine parallele Struktur des Dynamikmoduls bessere Vorhersageleistungen erzielt, als ein einzelnes Modul. Die dahinterstehende Überlegung ist, dass

die Beziehung zwischen Indikatoren und Zielvariablen für unterschiedliche Ereignistypen so stark divergieren, dass sie nicht in einem gemeinsamen Vorhersagesystem abgebildet werden können.

Es wurden daher unterschiedliche Prozessunterscheidungen vorgenommen und getestet. Grundlage zur Überprüfung dieser Hypothese waren die in Abschnitt 3.3 vorgestellten Klassifikationen der Eingangsdaten. Für jede dieser Klassen wurde ein eigenes Dynamikmodul angepasst. Diese Hypothese betrifft zum einen die Zielgröße, da unterschiedliche Modelle angewendet werden und eine zusätzliche Prognose der Klasse im Validierungsfall erforderlich ist. Zum anderen wird die vorhandene Lernmenge der DDM reduziert.

Ein weiterer Vorteil der parallelen Struktur des Dynamikmoduls ist die gegebene Flexibilität gegenüber Systemänderungen. Sollte sich die Häufigkeitsverteilung bestimmter hochwassererzeugender Prozesse verändern, würde sich dies negativ auf ein Modell mit einem einzelnen Dynamikmodul auswirken, da die Stationarität des Systems nicht mehr gegeben wäre (Solomatine & Ostfeld, 2008). Das Modell kann dadurch besser auf sich verändernde Systeme angewendet werden, ein Vorteil der vor allem in Verbindung mit dem GIUH gegeben ist (Wagener et al., 2010). Die parallele Struktur macht die Initialisierung des Modells rechenintensiver, da jede Instanz individuell angepasst werden muss. Zudem steigt die benötigte Datenmenge, da für jeden Cluster ausreichend Daten zu Verfügung stehen müssen.

Die zweite Hypothese baute auf einer der Grundannahmen bei der Übertragung von Modellparametern von beobachteten zu unbeobachteten Gebieten auf (Blöschl et al., 2013):

Der Austausch von Parametern zur Bestimmung der Abflussdynamik ist nur zwischen ähnlichen Gebieten möglich.

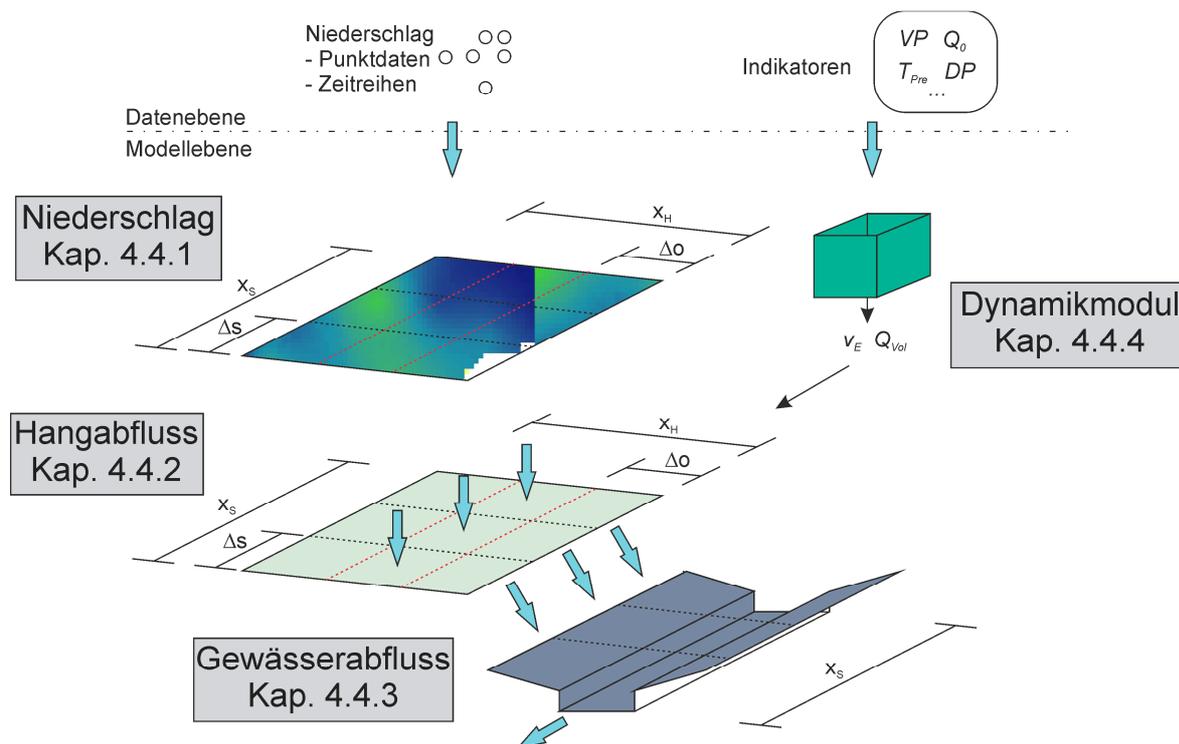
Im Laufe der Arbeit wurden unterschiedliche Metriken angewendet, um hydrologische Ähnlichkeit zu beziffern und im Aufbau der Lerndaten der Regionalmodelle zu berücksichtigen. Ein Vergleich der Metriken untereinander sowie mit Ergebnissen ohne räumliche Beschränkung der Lernmenge dient der Überprüfung dieser Hypothese.

Zusammenfassend wurden innerhalb des Dynamikmoduls zwei Variablen berechnet: Entwässerungsgeschwindigkeit  $v_E$  und Abflussvolumen  $Q_{Vol}$ . Beide Variablen wurden innerhalb des Moduls mittels eines DDMs zu Beginn der Berechnung eines Ereignisses geschätzt und an den GIUH übergeben. Die Auswahl der DDM-Struktur sowie der Prädiktoren werden nachfolgend dargestellt. Die Analyse der Funktionsweise und eine Sensitivitätsanalyse werden im Kapitel 5 dargestellt.

## 4.4 Modellsynthese

In diesem Kapitel werden das entwickelte Modell und seine einzelnen Komponenten vorgestellt. Das Modell kombiniert eine geomorphologische Einheitsganglinie mit einer datengetriebenen Vorhersage der Dynamikparameter  $v_E$  und  $Q_{Vol}$ . Der zuvor aufgezeichnete Rahmen des Modells, wie zuvor definiert, wird in Abbildung 4-5 zusammen mit den Kapitelnummern in denen die jeweiligen Komponenten erläutert werden, dargestellt.

Als Daten zur Berechnung der Direktabflussganglinie eines einzelnen Niederschlagsereignisses stehen Niederschlagszeitreihen von verschiedenen Niederschlagsstationen sowie ein Satz von Indikatoren zur Verfügung.



**Abbildung 4-5: Übersicht über die entwickelte Modellstruktur mit Kapitelnummern**

In jedem Zeitschritt werden die punktuellen Niederschlagsdaten mittels Interpolation in räumliche Informationen umgesetzt und anschließend mittels DFF fließwegorientiert ausgewertet. Dieser Arbeitsschritt wird in Kapitel 4.4.1 dargestellt. Der räumlich und zeitlich differenzierte Niederschlag wird anschließend auf das Hangmodell gegeben und mittels GIUH an das Gewässer übergeben. Die Umwandlung von Niederschlag zu Oberflächenabfluss wird in Kapitel 4.4.2 erläutert. Abschließend erfolgt die Berechnung des Abflusses im Gewässer, Kapitel 4.4.3, an dessen Ende die Direktabflussganglinie ausgegeben wird. Die benötigten Parameter für den GIUH sowie der Abflussbeiwert des Ereignisses werden vom Dynamikmodul, Kapitel 4.4.4, geschätzt. Das Modul verwendet dazu die Indikatoren und greift nicht auf interne Systemzustände zurück.

Das Modell besitzt 3 Parameter, die räumliche und zeitliche Diskretisierung des Modells definieren: zeitliche Auflösung  $\Delta t$ , räumliche Auflösung der Gewässerfließstrecke  $\Delta s$  und

die räumliche Auflösung der Hangfließstrecke  $\Delta o$ . Im Folgenden werden diese Parameter als Strukturparameter bezeichnet. Die zeitliche Auflösung  $\Delta t$  wurde angelehnt an die verfügbaren Zeitreihendaten mit 1 h gewählt. Die räumliche Auflösung der Rasterdaten (bspw. DGM) wurde als maßgebend für die Wahl von  $\Delta o = 100$  m angesehen. Die Auflösung der Gewässerfließstrecke  $\Delta s$  wurde zwischen 1 km und 10 km variiert. Der Einfluss dieses Strukturparameters wird in Kapitel 5.2 diskutiert.

#### 4.4.1 Niederschlag

Für ein Niederschlagsereignis lagen stündliche Intensitäten für eine Dauer von  $N_D$  Stunden an  $N_{Stat}$  Messpunkten vor. In der Iller standen Messungen an 11 Punkten, Main 26 und im Regen 18 Punkten zur Verfügung. Jeder dieser Messpunkte war einer geographischen Position  $[x, y]$  zugeordnet. Für ein betrachtetes Einzugsgebiet wurde zunächst ein Zuordnungsraster erstellt. Zuordnung  $Z$  des Punkt  $j$  ist die Station  $S$  mit geringsten euklidischen Abstand:

$$Z_j = \arg \min_{1 \leq s \leq N_{Stat}} \left( \sqrt{(x_j - x_s)^2 + (y_j - y_s)^2} \right) \quad 4.4$$

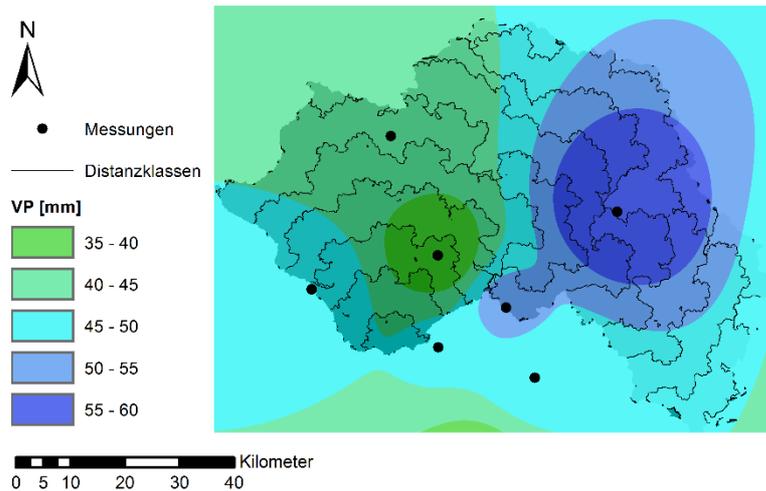
Die Zuteilung folgte dem Thiessen-Verfahren und nimmt an, dass die nächstliegende Messung über die meisten Informationen über einen Interpolationspunkt verfügt (Thiessen, 1911) und daher zeitlich invariant ist. Ein weiterer Vorteil dieses Verfahrens ist, dass im Gegensatz zu Verfahren, in denen eine Mittelung verschiedener Messpunkte vorgenommen wird, die Varianz der Messwerte im interpolierten Niederschlagsfeld erhalten bleibt.

Anschließend wurde für jeden Zeitschritt  $t$  ein Niederschlagsfeld interpoliert. Jedem Punkt  $j$  des Einzugsgebietes wurde dabei der Niederschlagswert  $P(t)$  der zugeordneten Station  $Z_j$  zum Zeitpunkt  $t$  zugeordnet:

$$P_j(t) = P_{Z_j}(t) \quad 4.5$$

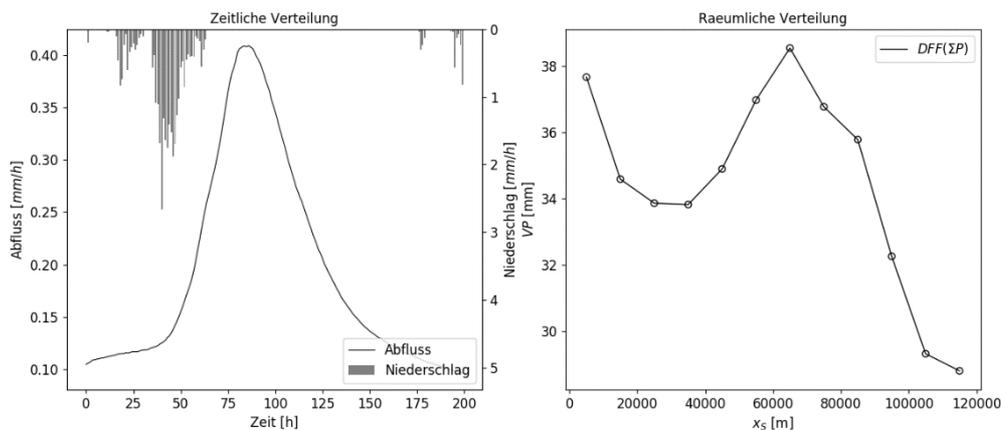
Die räumliche Interpolation der Niederschlagswerte wurde dann für jeden Zeitschritt mittels DFF auf die Fließstecken im Einzugsgebiet (Gleichung 4.1 und 4.2) bezogen. Dadurch stand in jeder Distanzklasse und jedem Zeitschritt ein Niederschlagswert zur Verfügung. Gemäß Gleichung 4.2 ist dieser Wert, der Mittelwert aller Niederschlagswerte, zugehörig zur betrachteten Distanzklasse.

Zur Anschauung dieses Vorgehens wurde ein Niederschlagsereignis im Februar 2002 im Einzugsgebiet des oberen Mains mittels DFF analysiert. Die räumliche Verteilung des Niederschlags wird in Abbildung 4-6 dargestellt. Die punktuellen Niederschlagsmessungen wurden zur Veranschaulichung mittels Inverse-Distanz-Verfahren (E=2) interpoliert.



**Abbildung 4-6: Interpolierte Niederschlagsverteilung Ereignis 14.02.2002 im oberen Main**

Zusätzlich zur räumlichen Verteilung der Niederschlagssumme sind in Abbildung 4-6 die Distanzklassen der DFF-Funktion mit  $\Delta s = 10$  km dargestellt. Innerhalb jeder Distanzklasse wurde der dargestellte Niederschlag gemäß Gleichung 4.2 ausgewertet. In folgender Abbildung 4-7 sind im linken Panel die zeitliche Verteilung des Abflussereignisses sowie des Gebietsniederschlags darstellt. Im rechten Panel die mit der DFF erfasste räumliche Verteilung des Niederschlags.



**Abbildung 4-7: Umwandlung Niederschlagszeitreihe in räumlich differenzierte Reihe, Beispiel Niederschlagsereignis Februar 2002 im Flussgebiet oberer Main**

Das Abflussereignis am Pegel Kemmern wurde durch einen zeitlich kompakten Niederschlag ausgelöst, dessen Niederschlag innerhalb von rund 50 Stunden fiel. Die räumliche Verteilung dieses Ereignisses zeigt eine zweigeteilte Belastung im Westen und mittleren Osten des Flussgebiets. Zu erkennen ist, dass die DFF diese zweigeteilte Belastung erfassen konnte. Nahe dem Gebietsauslass sowie in mittlerer Entfernung zum Gebietsauslass sind die stärksten Niederschlagssummen erfasst worden.

Dieses Beispiel zeigt, dass ein Teil der räumlichen Varianz durch die Verwendung der Distanzklassen erhalten werden kann. In diesem Beispiel beträgt die räumliche Varianz des Niederschlags in Abbildung 4-6  $16,8 \text{ mm}^2$ , die verbleibende Varianz der Erwartungswerte des

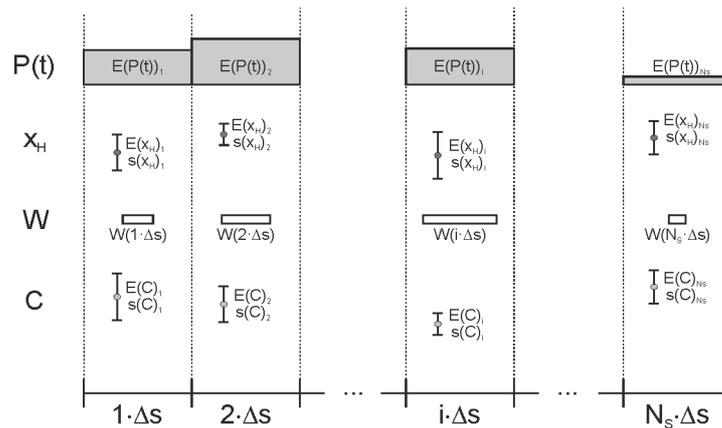
Niederschlags beträgt  $8,7 \text{ mm}^2$ . Durch die Mittelung aller Werte innerhalb einer Distanzklasse halbiert sich die Varianz. Zudem wurde über Teile des Gebiets gemittelt, die zwar zur gleichen Distanzklasse gehören, aber sehr weit auseinander liegen können. Dennoch zeigt die Analyse, dass ein Teil der Niederschlagsvarianz berücksichtigt werden kann. Ein klarer Vorteil gegenüber der Verwendung von Gebietsniederschlag.

#### 4.4.2 Hangabfluss

Abfluss von Flächen eines Einzugsgebietes, die keine Gewässerfläche sind, wurde mittels des Hangabfluss-Moduls simuliert. Innerhalb einer Distanzklasse wurden alle Hangflächen gemeinsam betrachtet.

Generell gibt es innerhalb des Modells zwei Bewegungsrichtungen: vertikal und horizontal. Es ist zu betonen, dass es sich dabei um konzeptionelle Richtungen innerhalb des Modells handelt, die nicht die tatsächliche Bewegungsrichtung des Wassers im EZG beschreiben. Die Bewegungsrichtungen beziehen sich nur auf die Weitergabe innerhalb der Modellstruktur. Vertikale Bewegung beschreibt die Weitergabe von Wasser innerhalb einer Distanzklasse. Dies betrifft die Weitergabe von Niederschlag auf das Hangmodul und Hangabfluss zum Gewässer. Die horizontale Bewegung beschreibt die Weitergabe von Wasser zwischen den Distanzklassen, also den Gewässerabfluss hin zum Gebietsauslass. Beide Richtungen sind unidirektional, d.h. die Weitergabe erfolgt stets von „oben nach unten“ in vertikaler Richtung und von oberliegenden Distanzklassen zu unterliegenden Distanzklassen. Die Hangflächen eines Einzugsgebietes wurden als parallele Objekte behandelt, in denen Wasserströme ausschließlich vertikal auftreten.

Eine Hangfläche der Distanzklasse  $i$  wurde durch fünf grundlegenden Parameter beschrieben, vier davon zeitlich invariant: der Erwartungswert der Hangfließlänge  $E(x_H)_i$ , Standardabweichung  $s(x_H)_i$  und Schiefe  $z(x_H)_i$  und die Fläche  $W$ . Ermittelt wurden diese Werte über die DFF( $x_H$ ) sowie dem Wert der Flächenfunktion  $W(i \cdot \Delta s)$ . Mit diesen Werten kann der GIUH des Hanges bestimmt werden und mittels Flächenfunktion kann die absolute Menge des Wassers am Übergang in das Gewässernetz bestimmt werden. Der fünfte Parameter ist der Erwartungswert der zeitlich variable Niederschlagsintensität  $E(P(t))_i$ . Neben diesen Parametern können weitere Parameter zur Charakterisierung der Hangflächen verwendet werden, bspw. um eine räumliche Variation von  $v_E$  anhand von Bodenkennwerten zu realisieren (vgl. Kapitel 5.3). Abbildung 4-8 zeigt den schematischen Aufbau der Hangmodelle eines Einzugsgebietes, das in  $N_S$  Distanzklassen unterteilt wurde. Jeder Hang verfügt zu jedem Zeitpunkt über einen individuellen Niederschlagsimpuls  $E(P(t))_i$ , eine individuelle Beschreibung der Fließstrecken innerhalb der Distanzklasse bis zum Gewässer (Erwartungswert, Standardabweichung und Schiefe von  $x_H$ ) sowie über die Weite  $W$  des Hanges. Zusätzlich verfügen die dargestellten Hänge über eine weitere Fließweginformation  $C$ . Dies könnte bspw. das Gefälle entlang der Fließwege sein, oder das Porenvolumen. Über die DFF werden Mittelwert, Standardabweichung und Schiefe der ausgewählten Kennwerte ermittelt und im Hangmodell hinterlegt.



**Abbildung 4-8: Konzeptionelle Darstellung der Hangflächen eines Einzugsgebietes mit  $N_S$  Distanzklassen**

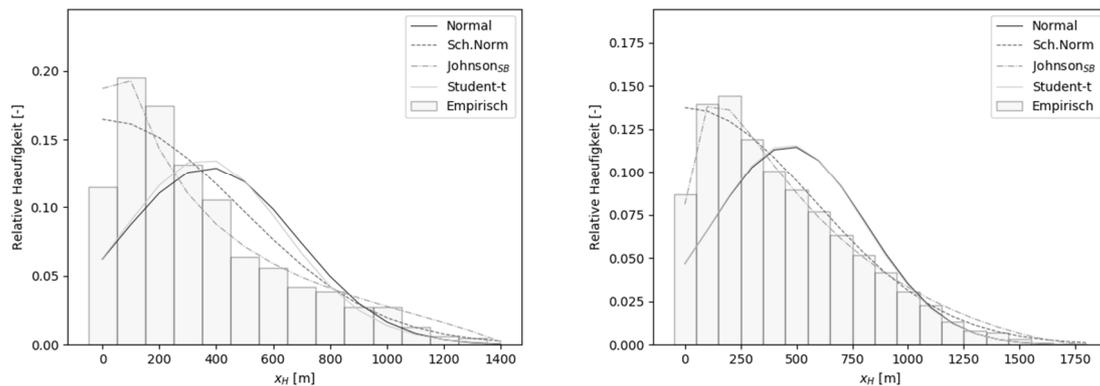
Zu Beginn der Berechnung einer Ganglinie stehen dem Hangmodell eine Schätzung der Entwässerungsgeschwindigkeit  $v_E$  und des Abflussvolumens  $Q_{Vol}$  zur Verfügung (vgl. Kapitel 4.4.4). Diese Dynamikparameter sind für die Vorhersage eines Ereignisses konstant. Zudem ist das zu erwartende Niederschlagsvolumen  $VP$  (vgl. Kap. 3.2) zu Beginn der Simulation bekannt.

Da die Volumina des Niederschlags und des Abflusses zu Beginn des Ereignisses bekannt sind, wurde die Berechnung der Abflussbildung mit einem Abflussbeiwert-Ansatz realisiert. Der effektive Niederschlag zum Zeitpunkt  $t$  in jeder Distanzklasse  $P_{eff}(t)_i$  wurde berechnet nach:

$$P_{eff}(t)_i = E(P(t))_i \cdot \frac{Q_{Vol}}{VP} \quad 4.6$$

Der abflusswirksame Anteil des Niederschlags steht gleichverteilt über den Hangflächen der Distanzklasse  $i$  zum Abfluss an. Er bewegt sich mit gegebenem  $v_E$  entlang seines Fließweges in Richtung des Gewässers. Der Fließweg wird durch seinen Erwartungswert und die Standardabweichung im Modell beschrieben. Ist die Form des Fließweges, d.h. die Verteilung der Entfernungen zum Gewässer, bekannt, kann in Abhängigkeit von der Zeit der Anteil der entwässerten Hangflächen bestimmt werden. Bei einem Momentaneinheitsimpuls (d.h.  $P_{eff}(t)_i = 1$ ,  $VP = 1$ ;  $Q_{Vol} = 1$ ) entspricht dies dem GIUH.

Die empirische Verteilung der Fließwege wurde durch eine kontinuierliche Verteilungsfunktion, die Johnson<sub>SB</sub> Verteilung, abgebildet. Grundlage für diese Auswahl war die Analyse der Hangflächen im oberen Main, bei der diese Verteilung die beste Anpassung zeigte. Mit  $\Delta s = 1$  km wurde das gesamte Einzugsgebiet in 116 Distanzklassen unterteilt. In Abbildung 4-9 sind exemplarisch die empirischen Häufigkeiten der Hangfließlängen innerhalb der Distanzklassen 5 und 27 dargestellt. Zudem wird eine Auswahl von 4 Verteilungsfunktionen, die getestet wurden, dargestellt.



**Abbildung 4-9: Empirische Häufigkeitsverteilung von Hangfließlängen in zwei exemplarischen Distanzklassen des Einzugsgebietes oberer Main (links: 5km vom Auslass; rechts 27 km); Darstellung einer Auswahl von angepassten Verteilungsfunktionen**

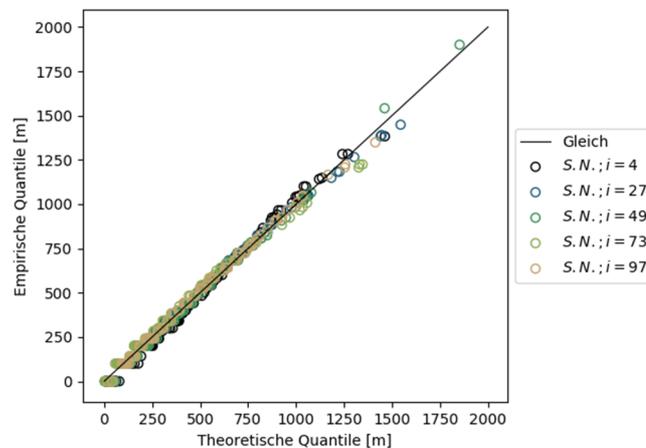
Normal- und Student-t-Verteilung können die Schiefe der empirischen Verteilung nicht abbilden. Daher musste die Wahl auf Verteilungen eingeschränkt werden, mit denen Rechtschiefe abgebildet werden konnte. Modifizierte Normalverteilungen stimmten dabei besser mit den empirischen Verteilungen der Fließwege überein als andere Verteilungsfunktionen. Hierbei vor allem die Johnsons<sub>SB</sub> und dreiparametrische Normalverteilung. Die Johnsons<sub>SB</sub> Verteilung war in einigen Distanzklassen in der Lage die geringere Häufigkeit der Hangdistanzklasse 0-100 m abzubilden, zu erkennen in Distanzklasse 27, rechte Darstellung in Abbildung 4-9. Zwar zeigte die schiefe Normalverteilung eine im Mittel gute Anpassung, die Anpassung erwies sich aber als numerisch ineffizient. Aufgrund der besseren Anpassungsergebnisse und der effizienteren Anpassung wurde die Johnsons<sub>SB</sub> Verteilung standardmäßig in allen Untersuchungsgebieten angewendet. Die Werte der kumulierten Johnsons<sub>SB</sub>-Verteilung berechnen sich nach folgender Gleichung:

$$F(x_h; a, b, l, s) = \Phi \left( a + b \cdot \ln \frac{x_h - l}{s + l - x_h} \right) \quad 4.7$$

Dabei ist  $\Phi$  der Wert der Standardnormalverteilung und  $x_h$  steht für eine Hangfließlänge. Die Parameter der Verteilungsfunktion,  $a$ ,  $b$ ,  $l$  und  $s$  werden iterativ über eine Maximum-Likelihood Zielfunktion für jede Distanzklasse bestimmt. Der Wert der kumulierten Verteilungsfunktion für ein gegebenes  $x_H$  kann als Anteil der Hänge der Distanzklasse interpretiert werden, deren Fließstrecken kleiner oder gleich  $x_H$  sind.

Abbildung 4-10 zeigt einen Quantil-Quantil Plot der empirischen und theoretischen Quantile verschiedener Distanzklassen, die einen Querschnitt durch das Gebiet darstellen sollen. Die theoretischen Quantile wurden mit der an die jeweilige Distanzklasse angepasste Johnsons<sub>SB</sub> Verteilung berechnet. Es ist zu erkennen, dass die empirischen und theoretischen Quantile nahezu identisch sind. Abweichungen sind vor allem an den Rändern der Verteilung zu beobachten. Zwar kann diese Analyse einen statistischen Test nicht ersetzen, es zeigt aber eine

grundsätzliche Übereinstimmung der empirischen und theoretischen Werte, die bei der Berechnung der Impulsantwort des GIUH verwendet werden.



**Abbildung 4-10: Quantil-Quantil-Plot für die Verteilung Hangflächen dargestellt mit Johnson<sub>SB</sub>-Verteilung für 5 Distanzklassen (4km; 27km; 49km; 79km; 97km) im Gebiet des oberen Mains.**

Die Berechnung der Ordinaten der geomorphologischen Impulsantwort  $u(\tau)_i$  zum diskreten Zeitschritt  $\tau$  in der Distanzklasse  $i$  wurden nach folgender Gleichung durchgeführt:

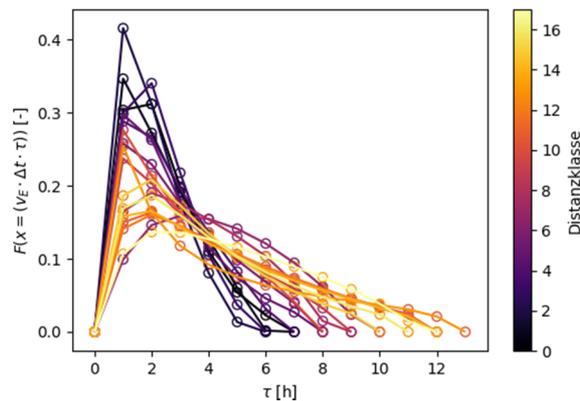
$$u(\tau)_i = F\left(x = v_{E,H} \cdot \Delta t \cdot \tau, a_i, b_i, s_i, l_i\right) \quad 4.8$$

Der Wert der geomorphologischen Impulsantwort ist damit abhängig von der Verteilung der Hangfließlängen, repräsentiert durch die Parameter der Johnson<sub>SB</sub>-Verteilung, dem Zeitschritt  $\tau$  sowie der Entwässerungsgeschwindigkeit des Hangs  $v_{E,H}$ . Die Entwässerungsgeschwindigkeit des Hangs  $v_{E,H}$  wird in der Basisvariante des Modells gleich der mittleren Entwässerungsgeschwindigkeit  $v_E$  gesetzt (eine Variation mit dem Hanggefälle wird in Kapitel 5.3 diskutiert).

Der Zeitschritt  $\tau$  wurde zur Berechnung des GIUH schrittweise um 1 erhöht, bis folgende Bedingung erfüllt war:

$$\sum_{\tau=1}^m u(\tau) \geq 1 - \varepsilon \quad 4.9$$

Gleichung 4.9 ist die Umsetzung einer der Grundannahmen der Einheitsganglinie, dass die Summe der Ordinaten 1 ergeben muss. Es werden also so lange GIUH Ordinaten erzeugt bis ihre Summe 1 ergibt. Ein Fehlerterm  $\varepsilon$  wurde eingefügt zur numerischen Lösbarkeit des Problems. Da eine unbeschränkte Verteilungsfunktion zur Beschreibung der Hangfließlängen verwendet wurde, ist theoretisch  $m = \infty$ . Um eine endliche Zahl von Ordinaten zu gewährleisten, wurde  $\varepsilon = 0.01$  gesetzt und damit ein Summenfehler von 0.1% zugelassen. In Abbildung 4-11 sind beispielhaft die Einheitsganglinien des Einzugsgebietes Bad Berneck am oberen Main dargestellt, berechnet mit  $\Delta s = 1\text{km}$  und  $v_E = 0.05\text{ m/s}$ .



**Abbildung 4-11: Geomorphologische Einheitsganglinien des Einzugsgebietes Bad Berneck mit  $\Delta s = 1\text{ km}$ ;  $v_{E,H} = 0.05\text{ m/s}$**

Mit bekannter Einheitsganglinie konnte der Hangabfluss der Distanzklasse  $i$  zum Zeitpunkt  $t$  wie folgt bestimmt werden:

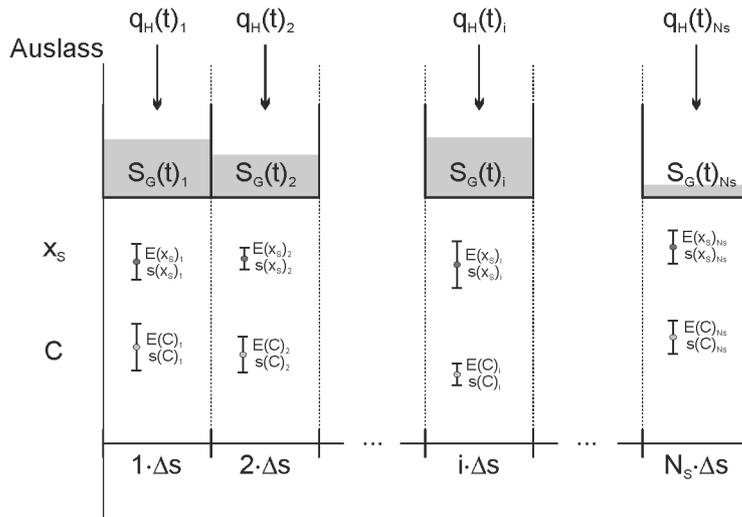
$$q_H(t)_i = W(i \cdot \Delta s) \cdot \sum_{\tau=1}^m P_{\text{eff}}(t-\tau)_i \cdot u(\tau) \quad 4.10$$

Die Menge des Effektivniederschlags, der sich noch auf den Hangflächen befindet und noch nicht an das Gewässer weitergegeben wurde, konnte wie folgt berechnet werden:

$$S(t)_i = W(i \cdot \Delta s) \cdot \sum_{\tau=1}^m (P_{\text{eff}}(t-\tau)_i - q_H(t-\tau)_i) \quad 4.11$$

#### 4.4.3 Gewässerabfluss

Wie die Hangflächen wurden alle Gewässer innerhalb einer Distanzklasse gemeinsam dargestellt und modelliert. Ein Gewässerobjekt wurde beschrieben durch die DFF-Werte der Gewässerfließstrecke  $x_S$  der jeweiligen Distanzklasse  $i$ . Auch im Gewässermodell können weitere Charakteristika mittels DFF eingebunden werden, um bspw. eine ungleichförmige Geschwindigkeitsverteilung zu erzeugen. Abbildung 4-12 zeigt den konzeptionellen Aufbau des Gewässermodells. In dieser Abbildung wurde das Gewässer als Abfolge von Speichern dargestellt. Dies soll aber nicht darüber hinweg täuschen, dass es sich um ein reines Translationsmodell handelt.



**Abbildung 4-12: Konzeptioneller Aufbau des Gewässermodells für ein Gebiet mit  $N_S$  Distanzklassen**

Zufluss erhalten die Gewässerobjekte, in Abbildung 4-12 und Abbildung 4-13 als Speicher dargestellt, als vertikalen Hangabfluss von den Hangflächen innerhalb der jeweiligen Distanzklassen. Zu Beginn eines Berechnungsschrittes  $t$  konnte damit zunächst das Wasservolumen im Gewässerabschnitt  $S_{G;temp}(t)_i$  bestimmt werden:

$$S_{G;temp}(t)_i = S_G(t-1)_i + q_H(t)_i \quad 4.12$$

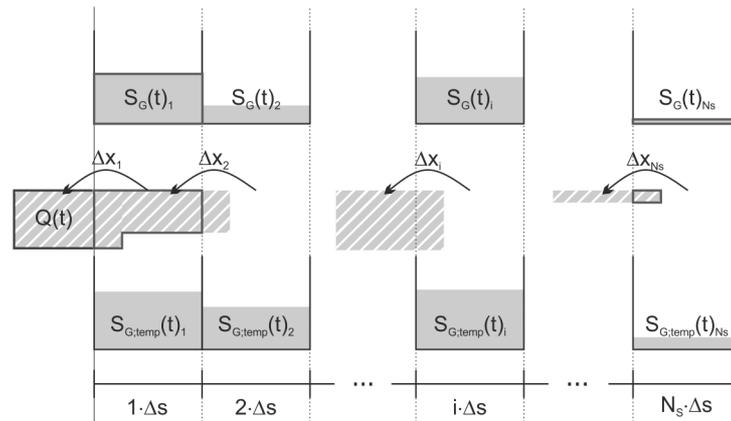
Bei bekannter Entwässerungsgeschwindigkeit der Gewässer  $v_{E;S}$  konnte in Abhängigkeit der Zeitschrittweite  $\Delta t$  die zurückgelegte Strecke des Wassers für jede Distanzklasse ermittelt werden. Normiert auf die gewählte räumliche Auflösung der Fließstrecken  $\Delta s$  ergibt dies den Anteil des Wassers im Gewässerabschnitt, der diesen verlassen hat:

$$\Delta x_i = \frac{v_{E;S;i} \cdot \Delta t}{\Delta s} \quad 4.13$$

$\Delta x_i$  ist für ein N-A Ereignis zeitlich invariant und variiert nur räumlich mit  $v_{E;S;i}$ . Die Weitergabe des Wassers erfolgt horizontal von Objekten höherer Ordnung (d.h. weiter entfernt vom Gebietsauslass) zu Objekten mit geringerer Ordnung (näher gelegene). In der Basisvariante des Modells wurde die Entwässerungsgeschwindigkeit der Gewässer  $v_{E;S}$  gleich der mittleren Entwässerungsgeschwindigkeit des Einzugsgebietes  $v_E$  gesetzt (eine räumliche Variation mittels der Strahler-Ordnung wird in Kapitel 5.3 diskutiert).

Diese Form der Weitergabe ist in Abbildung 4-13 dargestellt. Wasserpakete werden in der Konzeption des Modells von rechts nach links weitergegeben, das einer natürlichen Bewegung flussabwärts entspricht. Nach der Translation der Wasserpakete werden diese innerhalb der Distanzklassen summiert. In Abbildung 4-13 ist dieser Vorgang durch rote Umrandungen gekennzeichnet: die gekennzeichneten Volumina der Translation (blau schraffiert) entsprechen den Volumina, die sich am Ende des Zeitschrittes  $t$  im jeweiligen Gewässerabschnitt  $i$  befinden. Durch diesen Ausgleich, von ungleichmäßig zu gleichmäßig verteiltem

Wasser im Gewässerabschnitt wurde eine numerische Diffusion erzeugt, ähnlich der geomorphologischen Dispersion von Rinaldo, Marani & Rigon (1991).



**Abbildung 4-13: Konzeptuelle Darstellung der Wellenfortpflanzung im Gewässer mit gleichförmiger Entwässerungsgeschwindigkeit**

Die Normierung über  $\Delta s$  in Gleichung 4.13 und die konzeptionelle Annahme, dass das Wasser im Abschnitt gleichmäßig verteilt ist, ermöglichte die Verwendung von  $\Delta x_i$  zur Bestimmung der Zieldistanzklasse. Ausgehend von Distanzklasse  $i$  konnte die Ordnungszahl der Zieldistanzklasse mittels des Ganzzahlwertes von  $\Delta x$  bestimmt werden. Der horizontale Abfluss ausgehend von  $i$  in die Zieldistanzklasse  $i - \lfloor \Delta s \rfloor$  und dessen Oberlieger konnte mittels der Nachkommaanteile  $frac$  von  $\Delta s$  bestimmt werden:

$$\begin{aligned} qt_{i;i-\lfloor \Delta s \rfloor}(t) &= (1 - frac(\Delta s)) \cdot S_{G,temp}(t)_i \\ qt_{i;i-\lfloor \Delta s \rfloor+1}(t) &= frac(\Delta s) \cdot S_{G,temp}(t)_i \end{aligned} \quad 4.14$$

Die Indizes geben den Ursprung des Flusses und die Zielklasse an. Abschließend konnte die Wassermenge in jedem Gewässerabschnitt wie folgt bestimmt werden:

$$S_G(t)_i = \sum_{j=i+1}^{N_s} qt_{j,i}(t) \quad 4.15$$

Nach Gleichung 4.15 berechnete sich die Speicherfüllung zum Ende des Zeitschrittes  $t$  als Summe aller horizontalen Zuflüsse der Oberlieger  $j$  der Distanzklasse  $i$ .

Der Abfluss  $Q(t)$  aus dem betrachteten Einzugsgebiet ergibt sich analog zu Gleichung 4.15, wobei die Zieldistanzklasse den Index 0 trägt:

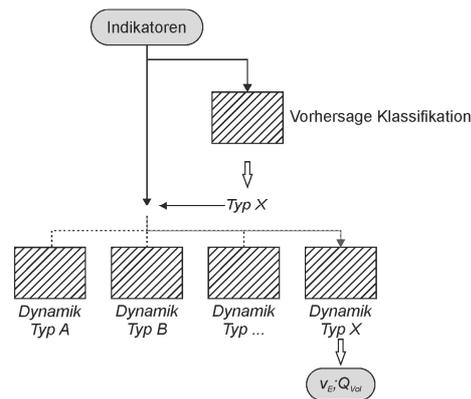
$$Q(t) = \sum_{j=1}^{N_s} qt_{j,0}(t) \quad 4.16$$

#### 4.4.4 Dynamikmodul

Wie aus Abbildung 4-5 und den vorhergehenden Kapiteln ersichtlich kommt dem Dynamikmodul eine wichtige Rolle zu. Vorhersagen dieses Moduls werden als ereignisweise Para-

metrisierung der übrigen Modellkomponenten verwendet. Eine möglichst genaue Vorhersage der Parameter  $v_E$  und  $Q_{Vol}$  war daher von eminenter Bedeutung.

Abbildung 4-14 zeigt die entwickelte parallele Struktur des Dynamikmoduls, das eigentlich aus einer Vielzahl von Modulen besteht, die je nach Ereignistyp aktiviert wurden. Im Anwendungsfall werden die Ereignisindikatoren (vgl. Kapitel 3.2) zunächst verwendet, um den Typ des anstehenden Abflussereignisses vorherzusagen. Basierend auf dieser Einschätzung werden dann das jeweilige Dynamikmodul aktiviert und mittels der Indikatoren eine Vorhersage von  $v_E$  und  $Q_{Vol}$  für das betreffende Ereignis erstellt.



**Abbildung 4-14: Konzept parallele Struktur Dynamikmodul für prozessorientierte Unterteilung der Daten/Ereignisse**

Jedes Modul ist ein Ensemble verschiedener datengetriebener Schätzer. Ausgegeben werden je nach Bedarf die Ergebnisse des besten Modells, ein gewichteter Mittelwert aller Schätzungen oder eine Ensemblevorhersage der Zielvariablen. Auswahl und Gewichtung wurden mit Hilfe des Bestimmtheitsmaßes zwischen angepasstem Modell und Trainingsdaten vorgenommen.

Bevor ein Dynamikmodul angewendet werden kann, muss es angelernt werden. Für ein Zielgebiet wurde ein regionaler Datensatz  $X$ , eine  $K \times N$  Matrix aus  $K$  Prädiktoren für  $N$  Ereignisse, mit bekannten Zielvariablen  $Y = [v_E; N; Q_{Vol}; N]$  aus beobachteten Gebieten gebildet. Diese Daten dienen als Lernmenge für die Module mit dem Ziel eine Beziehung zwischen Prädiktoren und Zielvariablen zu erstellen. Bei der Vorhersage des Ereignistyps ist die Zielvariable  $Y$  ein Vektor mit den Typenbezeichnungen (Kapitel 3.3). Die bekannten Ereignistypen wurden außerdem zur Unterteilung der Lernmenge verwendet, um die einzelnen Dynamikmodule individuell anzupassen.

Die verwendeten Schätzer lassen sich in drei Gruppen einteilen: multiple lineare Regression (MLR) als statistischer Einzelschätzer, einzelne ML-Schätzer und Ensemble Schätzer. Jeder Typ und die einzelnen verwendeten Algorithmen werden in den folgenden Unterkapiteln kurz erläutert. Die Algorithmen werden als Regressoren vorgestellt, die eingesetzt wurden um  $v_E$  und  $Q_{Vol}$  vorherzusagen. Die Algorithmen wurden im Modell aber als Klassifikatoren zur Vorhersage des Ereignistyps verwendet. Sofern nicht anders angegeben sind die folgen-

den Beschreibungen an Pedregosa et al. (2011), Kelleher, MacNamee & D'Arcy (2015), Alpaydin (2016), Alpaydin (2008) und Han & Kamber (2010) angelehnt.

#### 4.4.4.1 Multiple lineare Regression

Die Grundannahme dieses Schätzers ist, dass der mehrdimensionale Zusammenhang zwischen den Prädiktoren des Modells über  $K$  lineare Operatoren zum Ausdruck gebracht werden kann:

$$\hat{y} = \omega_0 + \omega_1 \cdot x_1 + \dots + \omega_k \cdot x_k + \dots + \omega_K \cdot x_K \quad 4.17$$

Eine Schätzung für die Variable  $\hat{y}$  wird mittels Summenprodukt aller Prädiktoren und ihrer Gewichte  $\omega_k$  gegeben.

Die Gewichte wurden mit Hilfe des Lerndatensatzes und Fehlerquadratminimierung bestimmt. Bei bekanntem  $y$  konnten die Residuen  $R$  der Regression wie folgt als Minimierungsproblem definiert werden:

$$R = \|X \omega - y\|_2^2 \rightarrow \min_{\omega} \quad 4.18$$

Für die Zielvariablen  $v_E$  und  $Q_{Vol}$  wurden zwei voneinander unabhängige MLR-Strukturen aufgebaut. Die Gewichte  $\omega$  wurden jeweils für  $y = v_E$  und  $y = Q_{Vol}$  bestimmt.

#### 4.4.4.2 Machine Learning Algorithmen

Vier grundlegende Algorithmen wurden in dieser Arbeit angewendet. Innerhalb eines Dynamikmoduls tauchen einige von ihnen in mehreren Formen auf. Im Folgenden wird kurz die grundlegende Arbeitsweise der Algorithmen beschrieben, eine detailliertere Darstellung ist in Anhang B gegeben.

##### **Support Vector Machines (SVM)**

Eine SVM, genauer ein Support Vector Regressor, versucht eine Regressionsgerade zwischen den Eingangsdaten  $X$  und einer, in der Lernphase bekannten, Ausgabevariable  $y$  anzupassen. Im Grunde funktioniert eine SVM wie eine MLR. Besteht ein linearer Zusammenhang zwischen den Prädiktoren  $X$  und Variable  $y$ , sind die Ergebnisse tatsächlich gleich. Es gibt aber fundamentale Unterschiede, die vor allem erkennbar werden, wenn der Zusammenhang nicht linear ist.

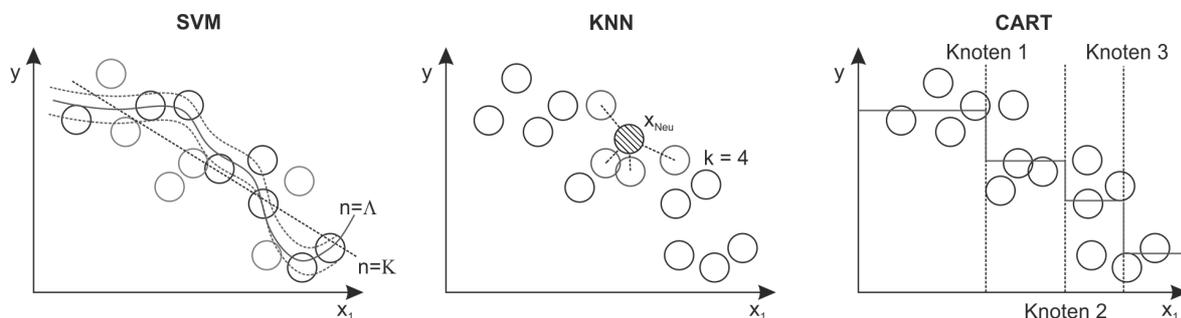
In diesem Fall wird das Problem in einen höher dimensionalen Raum projiziert, bspw. durch Linearkombination einzelner Prädiktoren, bis sich das Problem annähernd linear lösen lässt. Durch diese Lösung erscheint die Regression in der ursprünglichen Dimension der Daten nicht linear. Um Rechenzeit zu sparen, werden dabei Kernfunktionen eingesetzt, die es erlauben die Projektion in den höher dimensionalen Raum zu umgehen (siehe Anhang B). Die Arbeitsweise eines SVM-Regressors ist in Abbildung 4-15 dargestellt.  $K$  bezeichnet die ursprüngliche Dimension der Prädiktoren und  $\mathcal{A}$  die von der SVM verwendete Dimension zu Linearisierung des Problems. Ein weiterer Unterschied zur MLR ist Verwendung der na-

mensgebenden *support vectors*. Diese sind eine Auswahl der zur Verfügung stehenden Daten anhand derer die Regressionsgerade angepasst wird. In Abbildung 4-15 ist die Regressionsgerade der SVM in Rot mit einem umgebenden Fehlerbereich dargestellt. Die Datenpunkte repräsentieren die mehrdimensionalen Vektoren der Indikatoren. Punkte die außerhalb dieses Bandes liegen (Blaue Kreise), erzeugen den höchsten und zu minimierenden Fehler. Die Anpassung der Regressionsgeraden stützt sich auf diese ausgewählten Datenpunkte.

***k*-nächste Nachbarn (KNN)**

Der KNN-Regressor ist eine unparametrisierte Regressionsmethode. Im Gegensatz zu den anderen Modellen wird hier keinerlei Gleichung zur Schätzung von  $y$  mittels  $X$  angepasst. Stattdessen wird für einen neuen Satz von Prädiktoren  $\hat{X}$  eine Distanzanalyse, hier euklidische Distanz, zu den bekannten Prädiktoren des Lerndatensatzes durchgeführt. Basierend auf den berechneten Distanzen werden die  $k$  nächsten Nachbarn ausgewählt und der Mittelwert der zugehörigen  $y$ -Werte als Schätzung zurückgegeben. Dieses Vorgehen ist beispielhaft in der mittleren Abbildung 4-15 dargestellt.

Die Anzahl  $k$  der Nachbarn wurde innerhalb der Dynamikmodule zwischen 1 und 15 variiert und basierend auf dem maximalen Korrelationskoeffizienten ein  $k$  ausgewählt. Die Begrenzung auf 1 bis 15 mögliche Nachbarn zur Erstellung der Regression ist das Ergebnis einer Vorstudie mit dem KNN. Dabei fiel in rund 90% Fälle das optimale  $k$  in diesen Bereich (Darstellung in Anhang B).



**Abbildung 4-15: Beispielhafte Regression mittels SVM (links), KNN (Mitte) und CART (rechts) für Prädiktor  $x_1$  und Zielvariable  $y$**

***Entscheidungsbäume (CART)***

Entscheidungsbäume können für Klassifikationen und Regressionen eingesetzt werden (*Classification And Regression Trees*). In beiden Fällen wird sich einer klassischen Teile-und-Herrsche Strategie bedient: Ein nicht lineares Problem wird so lange unterteilt bis innerhalb der Teildatensätze eine akzeptable Lösung gefunden werden kann.

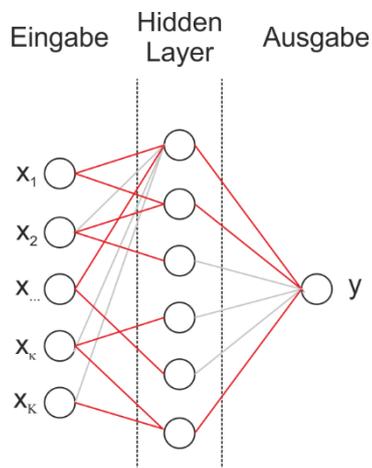
An jeder Verästelung eines Entscheidungsbaumes wird eine solche Unterteilung basierend auf einem, oder mehreren Grenzwerten eines, oder mehrerer Prädiktoren vorgenommen. Nach einer Verästelung  $m$  wird die die Lernmenge, die diese Verästelung  $X_m$  erreicht, unterteilt in zwei Datensätze  $X_{m+1}$  und  $X_{m+2}$ . Gleiches gilt für die zugehörigen Werte von  $Y_m$ . Die

Anpassung der Grenzwerte an einer Verästelung basiert auf der Schätzung von  $y$ , die gleich dem Mittelwert der unterteilten Datensätze  $Y_{m+1}$  und  $Y_{m+2}$  ist. Ziel ist es, den quadratischen Fehler an jeder Verästelung zu minimieren.

Ist der Fehler dieser Schätzung kleiner als eine definierte Fehlertoleranz, wird ein Blattknoten gebildet. An diesem Punkt wird keine weitere Unterteilung vorgenommen sondern der Schätzwert eingetragen. Einem Satz neuer Prädiktoren  $\hat{X}$ , der diesen Blattknoten erreicht, wird dann dieser Wert als Schätzung zurückgegeben. Ein Beispiel für eine Regression mittels CART ist in der rechten Abbildung 4-15 dargestellt. Um den Einfluss der Trainingsdaten auf das Ergebnis der Anpassung zu minimieren wird die Fehlertoleranz meist mit 0 angesetzt. Der Entscheidungsbaum damit so lange weiter verästelt bis alle Stränge rein sind, d.h. auf jeder Verästelung bleibt nur noch eine Stichprobe übrig und bildet einen eigenen Blattknoten. Um das Risiko einer Überanpassung zu mindern wird in der Regel eine maximale Anzahl von Verästelungen definiert. Im Dynamikmodul wird zwischen mehreren Bäumen mit und ohne Begrenzung der Verästelungen auf Grundlage ihrer Anpassungsergebnisse ausgewählt (Darstellung im Anhang B).

### **Künstliche neuronale Netze (ANN)**

Das in den Dynamikmodulen verwendete mehrlagige Perzeptron (*Multi-Layer Perceptron* MLP) ist eine der klassischen Varianten eines künstlichen neuronalen Netzes (*Artificial Neural Network* ANN). Es besteht aus mehreren Schichten, in denen sich eine vom Nutzer definierte Anzahl von Neuronen befindet. Ein einfaches MLP ist in Abbildung 4-16 dargestellt.



**Abbildung 4-16: Darstellung eines mehrlagigen Perzeptrons mit Eingabeebene, einem Hidden Layer und einer Ausgabeebene. Genutzte Verknüpfungen dargestellt in Rot, ungenutzte in Grau**

Die erste Schicht ist die Eingabeschicht. In dieser Schicht befinden sich genau  $K$  Neuronen, wobei jedes Neuron für einen Prädiktor steht. In der zweiten Schicht, dem *Hidden Layer*, befinden sich eine beliebige Anzahl an Neuronen. Jedes dieser Neuronen hat eine Verknüpfung zu jedem Neuron der vorherigen Schicht. Aus Gründen der Übersichtlichkeit sind in Abbildung 4-16 diese Verknüpfungen nur für das oberste Neuron des Hidden Layers darge-

stellt. Die letzte Schicht ist die Ausgabeschicht und beinhaltet in diesem Falle nur ein Neuron, das den Wert der Zielvariable  $y$  beinhaltet. Wie zuvor hat dieses Neuron Verknüpfungen zu allen Neuronen der vorherigen Schicht. Ein MLP kann aus mehreren aufeinander folgenden Hidden-Layern bestehen, deren Verknüpfungen analog zum dargestellten Aufbau gebildet werden.

Jede Verknüpfung symbolisiert einen Wichtungsfaktor  $\omega$ . Die Impulse der Eingangsneuronen entsprechen den Werten der jeweiligen Indikatoren eines Ereignisses. Der Wert des Neurons in der nächsten Schicht berechnet sich als Produktsumme aller Impulse und deren Wichtung. So gelangt das ANN von den Prädiktoren über den Hidden Layer zu einer Schätzung der Zielvariablen.

In einem angepassten ANN müssen nicht alle Verknüpfungen genutzt werden. In Abbildung 4-16 wird dies durch rote und graue Linien symbolisiert. Rote Linien symbolisieren eine Wichtung größer Null und graue Linien sind inaktive Verknüpfungen, also  $\omega$  gleich Null. Durch diese Struktur wird dem ANN ermöglicht nicht-lineare Regressionen durchzuführen. Seine Arbeitsweise ist also mit dem einer SVM (Abbildung 4-15, links) vergleichbar.

In den Ensembles der Dynamikmodule wurden zwei ANN-Konzepte angewendet. Ein MLP mit einem Hidden-Layer und mit zwei Hidden-Layer wurden angewendet. Die Anzahl der Neuronen wurden mittels der *Likelihood* aus einem Aufbau mit 10, 50 oder 100 Neuronen ausgewählt.

### 4.4.4.3 Ensemble Machine Learning Schätzer

Alle bisher vorgestellten Schätzer basieren auf der Anpassung eines einzelnen Modells und können ggf. eine Überanpassung an den Lerndatensatz aufweisen. Eine Methode diesem Problem entgegenzuwirken ist die Verwendung von mehreren Schätzern in einem Ensemble. Über eine Abstimmung oder Mittelwertbildung gelangt ein Ensemble-Schätzer zu einem eindeutigen Ergebnis.

#### **Random Forest (RF)**

Wie der Name bereits andeutet, besteht ein RF aus einer Vielzahl von Entscheidungsbäumen (Kap. 4.4.4.2), die auf das gleiche Problem angelernt wurden. Die einzelnen Entscheidungsbäume heißen in diesem Zusammenhang Basisschätzer. In dieser Anwendung wurde ein Ensemble von 100 Basisschätzern verwendet.

Jeder Entscheidungsbaum wird dabei auf eine unterschiedliche Teilmenge der zur Verfügung gestellten Trainingsdaten angelernt. Die Auswahl der Teilmengen erfolgt dabei zufallsbasiert. Durch das mehrfache Anlernen, basierend auf unterschiedlichen Teilmengen des ursprünglichen Datensatzes, gelten RF Schätzer als sicher gegen Überanpassung (Breiman, 2001).

In der gewählten Implementierung von Pedregosa et al. (2011) wird eine Schätzung durch Mittelwertbildung aller Entscheidungsbäume erstellt.

***AdaBoost (ADA)***

*Adaptives Boosting* (kurz: *adaboost*) ist eine erweiterte Technik zur Anpassung eines RF. Es werden mehrere Basisschätzer basierend auf unterschiedlichen Teilmengen des Trainingsdatensatzes auf das Ursprungsproblem angelernt.

Der Unterschied zum normalen RF ist, dass die Basisschätzer nacheinander angepasst werden und die Auswahl der Datenpaare (Prädiktoren und Zielvariable) auf den Fehlern des vorhergehenden Basisschätzers basiert. Genauer wird jedes Datenpaar mit einer Wahrscheinlichkeit versehen, die je nach Fehler des aktuellen Basisschätzers steigt oder sinkt. Diese Wahrscheinlichkeit bildet die Basis für die Zusammenstellung des nachfolgenden Samples. Für die Konstruktion der Teilmenge des nachfolgenden Basisschätzers steigt also die Wahrscheinlichkeit, dass sie die fehlerhaft vorhergesagten Datenpaare beinhaltet.

Die Schätzung eines *AdaBoost*-Schätzers wurde in den Dynamikmodulen als gewichteter Mittelwert aller Basisschätzer berechnet. Die Gewichte der Schätzer wurden angepasst an den jeweiligen quadratischen Fehler.

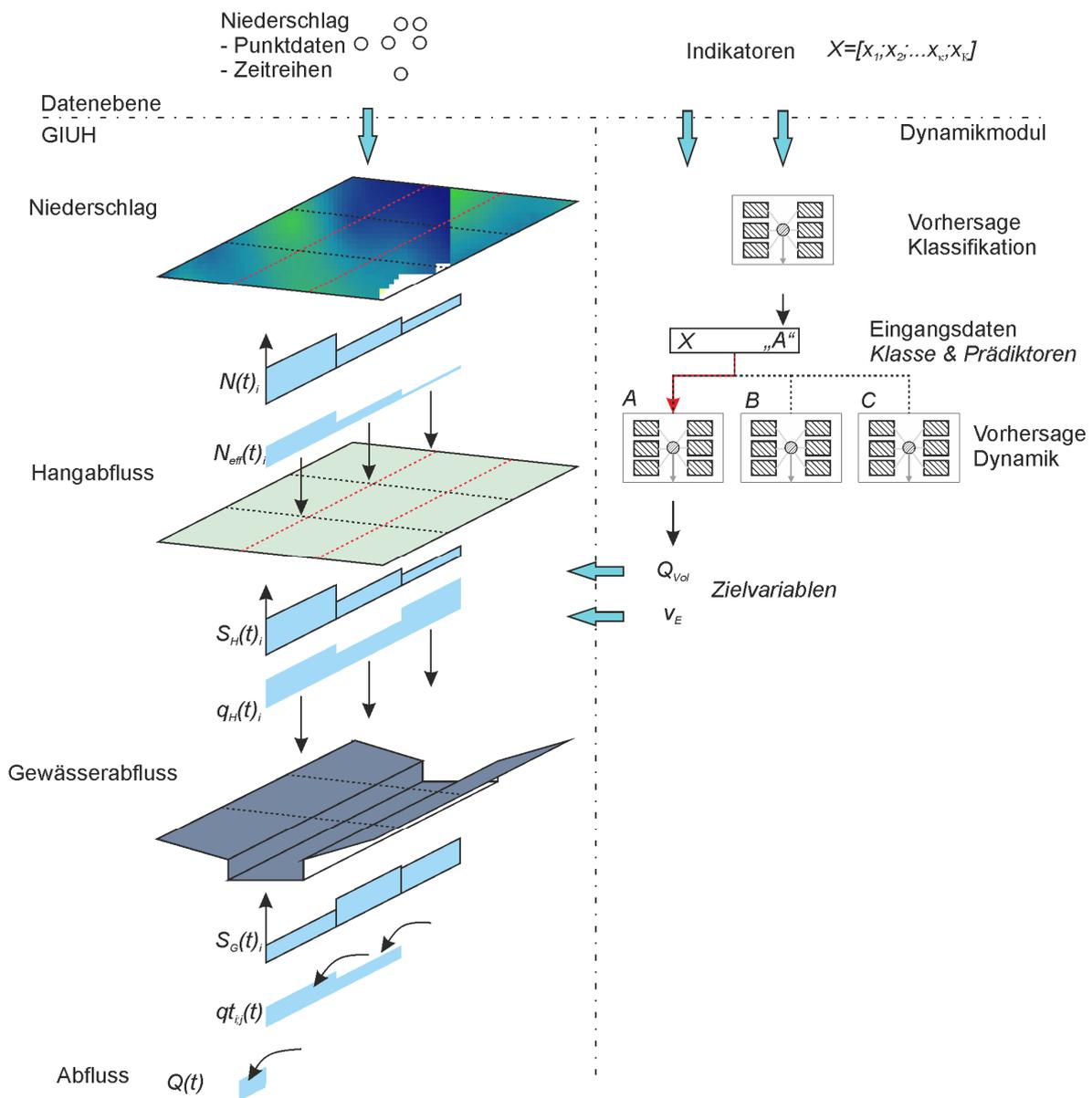
**4.5 Zusammenfassung Modellentwicklung**

An dieser Stelle soll noch einmal eine Übersicht über die entwickelte Modellstruktur gegeben werden. Außerdem soll auf die noch zu diskutierenden Punkte hingewiesen, bzw. auf deren Behandlung in dieser Arbeit verwiesen werden.

Abbildung 4-5 zeigt das grundlegende Konzept des gekoppelten Modells. Mit den Darstellungen der internen Prozessgleichungen und Module lässt sich diese Darstellung nun konkretisieren. In Abbildung 4-17 sind die Module und Ebenen, interne Wasser- und Informationsflüsse dargestellt.

Das neu entwickelte GIUH-ML Modell besteht aus zwei getrennten Modulen, der geomorphologischen Impulsantwort (GIUH-Komponente) auf der einen Seite und dem Dynamikmodul (ML-Komponente) auf der anderen Seite. Beide Module arbeiten grundlegend verschieden und es besteht lediglich ein Informationsaustausch zwischen ihnen. Der GIUH arbeitet zeitlich differenziert, sowohl Eingangs- wie auch Ausgangsdaten sind daher Zeitreihen. Das Dynamikmodul arbeitet mit Ereignisindikatoren und wird zur Berechnung eines Ereignisses nur einmal aktiviert. Die vom Dynamikmodul erstellten Schätzungen werden an den GIUH übergeben.

Interpolierte Niederschlagsdaten werden innerhalb des GIUH Moduls mittels DFF ausgewertet, per Abflussbeiwert zu effektiven Niederschlag transformiert und in vertikaler Fließrichtung an das Hangmodell weitergegeben. Auch innerhalb des Hangmodells gibt es keine Interaktion mit anderen Distanzklassen in horizontaler Richtung. Hangabfluss wird vertikal in das Gewässermodell überführt. Innerhalb der Gewässer ändert sich dann die Fließrichtung und Wasser wird horizontal in Richtung Gebietsauslass weitergegeben.



**Abbildung 4-17: Erweiterte Darstellung der Modellkonzeption mit internen Zuständen, Informationsaustausch und internen Flüssen**

Innerhalb des Dynamikmoduls werden die Ereignisinformationen (Indikatoren) zunächst verwendet, um eine Schätzung des Ereignistypen zu erstellen. Anschließend werden die Indikatoren dem passenden Vorhersageblock zur Schätzung von  $Q_{Vol}$  und  $v_E$  übergeben. Jeder Block umfasst mehrere ML-Algorithmen: eine multiple lineare Regression (MLR), ein ein- und zweilagiges künstliches neuronales Netz (ANN1 & ANN2), einen k-nächste Nachbarn Regressor (KNN), einen Entscheidungsbaum (CART), eine support vector machine (SVM) einen Random Forest (RF) und adaptives Boosting (ADA). Jeder Regressor nimmt auf Grundlage der gleichen Indikatoren eine individuelle Schätzung vor. Der Vorhersageblock kann entweder eine Ensemble-Schätzung aller Algorithmen ausgeben oder eine gemittelte Schätzung.

Die getrennte Struktur des Modells ermöglichte eine getrennte Analyse der Modellkomponenten. Zu klären war, ob die gewählte konzeptionelle Darstellung des Einzugsgebietes und des Abflussprozesses eine ausreichend genaue Repräsentation des natürlichen Abflussprozesses darstellte. In Kapitel 5 wird der Test des GIUH Moduls auf seine grundlegende Funktionsweise dargestellt. Die Fragestellung des Tests war: Bis zu welchem Grad kann das Modell bei bekanntem  $v_E$  und  $Q_{Vol}$  die zugehörige Ganglinie reproduzieren? Nachfolgend wurde die Annahme, dass  $v_E$  und  $Q_{Vol}$  stationär und gleichförmig sind, hinterfragt und verschiedene Alternativen zur Variation der Dynamikparameter getestet. Außerdem war der Einfluss der räumlichen Differenzierung des Modells, also die Wahl der DFF-Parameter, an diesem Punkt noch ungeklärt. In Kapitel 5.2 wird auf diesen Punkt eingegangen.

Kapitel 6 befasst sich mit Dynamikmodul und den ML-Schätzern. Diskutiert werden hier die Auswahl der besten Kombination von Indikatoren zur Vorhersage von  $v_E$  und  $Q_{Vol}$ , die beste Differenzierung der Eingangsdaten sowie die allgemeine Güte der ML-Vorhersagen. Ein bekanntes Problem im Zusammenhang mit ML ist der benötigte Umfang der Lernmenge (Solomatine & Ostfeld, 2008). In Kapitel 7 wird auf diese Thematik eingegangen. Ein relevanter Punkt war die Definition von *Region* zur Bildung der Regionalmodelle. Was definiert eine Region? Geographische Nähe, Naturräume oder hydrologische Ähnlichkeit? Daher werden in diesem Kapitel die Analysen zur selektiven Datenauswahl für die ML-Algorithmen mit verschiedenen Metriken zur Definition von Ähnlichkeit/Region vorgestellt.

## 5 Analyse Abflussmodell

---

In diesem Abschnitt wird die Untersuchung des geomorphologischen Einheitsganglinien (GIUH) Modells ohne die Verwendung des Dynamikmoduls dargestellt. Da es sich bei dem in Kapitel 4 vorgestellten Modell um eine Neuentwicklung handelt muss diskutiert werden wie das Modell funktioniert, ob es den Anforderungen genügt und wie es parametrisiert werden kann.

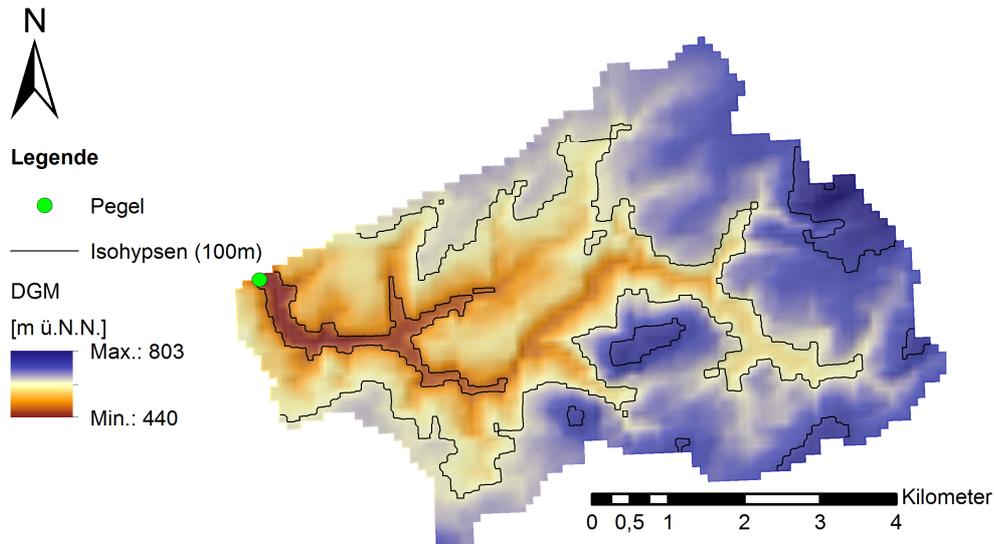
Zu einem Großteil kann das Modell direkt mit Beobachtungsdaten parametrisiert werden. Strukturparameter wie die Anzahl von Distanzklassen, mittlere Fließstrecken, etc. können direkt aus Geodaten mittels der Distanz-Faktor-Funktion (DFF) an das Modell übergeben werden. Diese Daten bilden die grundlegende Modellstruktur, die nur einen einzigen zu definierenden Parameter, die räumliche Auflösung der DFF  $\Delta s$ , besitzt. Zwei weitere Parameter, Entwässerungsgeschwindigkeit  $v_E$  und  $Q_{Vol}$ , werden benötigt, um Abflusswellen zu simulieren. Diese Parameter sind für jedes Ereignis erforderlich und müssen geschätzt werden. Analytisch können sie aus beobachteten Niederschlags- und Abflusszeitreihen ermittelt werden. In diesem Abschnitt soll unter anderem der Frage nachgegangen werden, ob das entwickelte Modell mit einer direkten Parametrisierung aus einer Ereignisanalyse Abfluss in ausreichender Modellgüte simulieren kann.

Bevor auf diese Frage eingegangen werden konnte, stand aber noch eine detaillierte Untersuchung des Modells aus. Zunächst sollte die Funktionsweise des Modells analysiert werden, um dann in der Analyse etwaige Fehler des Modells verstehen und deuten zu können. Dies umfasste außerdem eine Sensitivitätsanalyse des Modells bezüglich seiner Parameter sowie geomorphologischer Gegebenheiten. Zudem wurde eine räumlich-zeitliche Differenzierung der analytischen Parameter innerhalb der GIUH-Struktur entwickelt, erprobt und ausgewertet.

Zusammenfassend soll in diesem Kapitel ein grundlegendes Verständnis über die Simulation von Abfluss mittels des GIUH aufgebaut werden. Am Ende dieses Abschnittes soll außerdem eine Empfehlung für die vom Nutzer zu wählenden Parameter gegeben werden, ein erster Eindruck über den Gültigkeitsbereich des Modells bezüglich der Dynamikparameter und die Eignung des Modells mit analytischen Parametern Abflussganglinien zu simulieren gezeigt werden.

## 5.1 Funktionsweise Modellstruktur

Für diesen Teil der Untersuchung wurde ein reales Kopfeinzugsgebiet im Oberlauf des oberen Mains, Bernstein am Wald (a.W.), verwendet. Die Wahl fiel auf dieses Gebiet, da es eines der kleinsten Gebiete im Datensatz war und dadurch der numerische Aufwand klein gehalten werden konnte. Die zur Verfügung stehenden geomorphologischen Daten dieses Einzugsgebiets wurden als Grundlage zur Initialisierung des GIUH-Modells verwendet.



**Abbildung 5-1: Höhenstruktur des Einzugsgebietes Bernstein a.W.**

Das in Abbildung 5-1 dargestellte Einzugsgebiet umfasst eine Fläche von rund  $35 \text{ km}^2$  und wurde mit einer räumlichen Auflösung von  $\Delta s = 1 \text{ km}$  in 12 Distanzklassen unterteilt. Die mittlere Hangfließlänge beträgt  $384 \text{ m}$ , mit einem Variationskoeffizienten von  $0.7$ .

Ein synthetischer Blockregen mit stufenweiser ansteigender Intensität wurde zur Analyse verwendet. In den ersten  $50 \text{ h}$  fällt Niederschlag mit einer konstanten Intensität von  $1 \text{ mm/h}$ . Die Intensität erhöht sich danach auf  $2 \text{ mm/h}$ . Die Reaktion des GIUH auf diesen Modellregen, mit angenommener  $v_E = 0.1 \text{ m/s}$ , ist in Abbildung 5-2 als blaue Linie dargestellt und zeigt keinerlei ungewöhnliches Verhalten. Auf einen Anstieg der Niederschlagsintensität reagiert die Abflussganglinie stets mit einiger Verzögerung bis die Abflusshöhe exakt dem Wert der Niederschlagsintensität entspricht. Die Anstiegsintensität des Abflusses ist dabei abhängig von der Niederschlagsintensität.

Zum Vergleich wurden Abflussganglinien mit Hilfe eines Einzellinearspeichers (ELS) und einer linearen Speicherkaskade (LSK) berechnet. Der Abfluss  $Q_{ELS}$  und  $Q_{LSK}$  der Modelle wurde nach folgenden Gleichungen 5.1 und 5.2 bestimmt (Maidment, 1993):

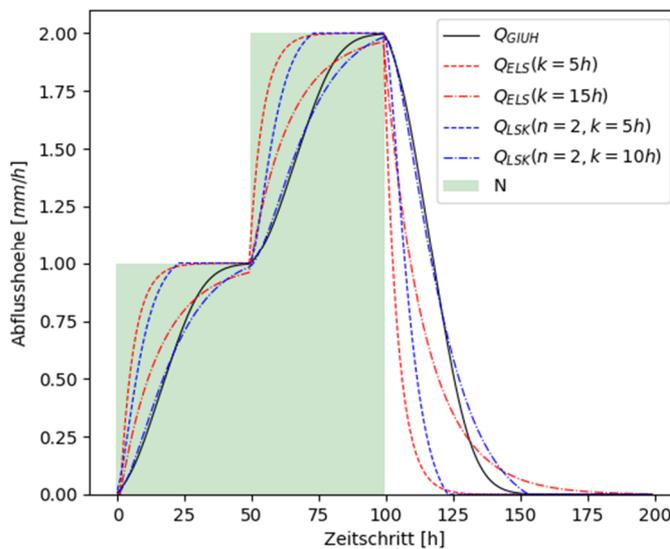
$$Q_{ELS}(t) = Q_{ELS}(t-1) \cdot e^{-\frac{\Delta t}{k}} + P(t) \cdot \left(1 - e^{-\frac{\Delta t}{k}}\right) \quad 5.1$$

$$Q_{LSK} = \sum_{i=1}^m P(t-i) \cdot h(i) \quad 5.2$$

mit:

$$h(\Delta t, t = i \cdot \Delta t) = \Delta t \cdot \frac{1}{k \cdot \Gamma(n)} \cdot \left( \frac{(i-0.5) \cdot \Delta t}{k} \right)^{n-1} \cdot e^{-\frac{(i-0.5) \cdot \Delta t}{k}} \quad 5.3$$

Dabei ist  $k$  die Speicherkonstante,  $n$  die Anzahl der Speicher und  $m$  die Anzahl der mit Gl. 5.3 berechneten Ordinaten  $h$ . In den durchgeführten Analysen wurden jeweils zwei Parametrisierungen getestet: jeweils mit  $k = 5$  h (gestrichelte Linien) und mit  $k = 10$  bzw.  $15$  h (Strichpunktlinien). Als initiale Parameterschätzung wurde  $n = 2$  angenommen. Die berechneten Abflussganglinien der beiden Modelle mit unterschiedlicher Parametrisierung sind in Abbildung 5-2 dargestellt. Die Erhöhung von  $n$  (für beide Modelle) stellt den ersten Schritt einer Anpassung der Modelle an die Abflussganglinie des GIUH dar.



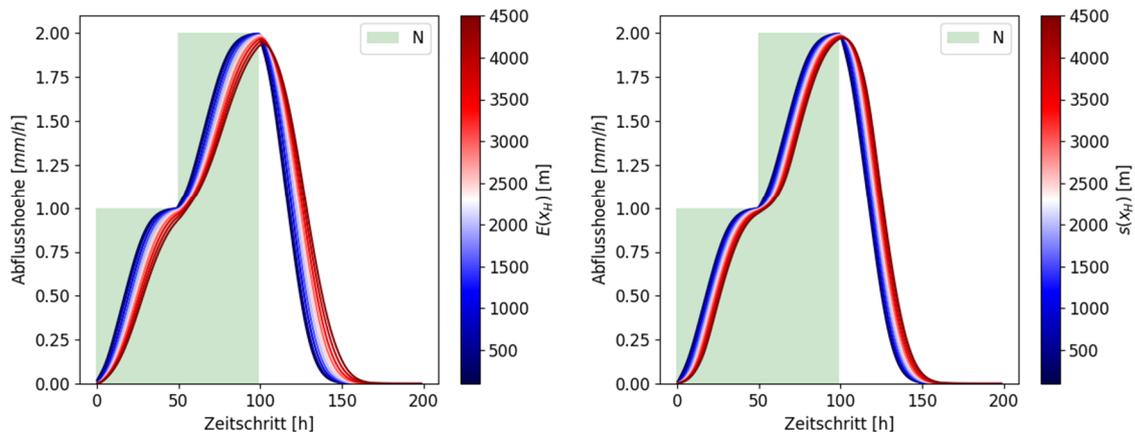
**Abbildung 5-2: Abflussreaktion auf einen synthetischen Blockregen des GIUH, ELS und der LSK mit unterschiedlichen Parametrisierungen**

Zu erkennen ist, dass die GIUH-Simulation im Vergleich zum ELS bei ähnlicher Verzögerung der Abflussreaktionen eine höhere Anstiegsintensität aufweist. Im Rückgang ist die GIUH-Reaktion verzögerter als die des ELS, flacht aber schneller ab. Größere Ähnlichkeit besteht zwischen der Abflussreaktion des GIUH und der LSK. Mit einer Speicherkonstante von  $k = 10$  h (grüne Strichpunktlinie) liegen die Ganglinien nahezu übereinander. Durch weitere Anpassung können die beiden Modelle nahezu komplett aufeinander abgestimmt werden (bspw.:  $k = 9.5$  h,  $n = 2$ ). Die grundlegende Reaktionsform des GIUH ähnelt demnach vor allem der Abflussreaktion einer LSK, also einer Aufreihung mehrerer ELS.

Die neu entwickelte Modellstruktur funktioniert also grundsätzlich wie bestehende Modellkonzepte des GIUH. Grimaldi, Petroselli & Nardi (2011) verwenden als Modellgleichung

die LSK und bestimmen die gesuchten Parameter  $n$  und  $k$  über eine empirische Beziehung zu Horton-Verhältnissen und einer angenommenen Fließgeschwindigkeit. Diese Übereinstimmung der simulierten Abflussreaktion des neu entwickeltem Konzept und bestehendem Konzept ist als positives Ergebnis zu werten. Es bestätigt zum einen die Funktionalität und Vergleichbarkeit des Modells. Zum anderen kann das Modell im Vergleich zum Modell von Grimaldi, Petroselli & Nardi (2011) wesentlich mehr Gebietscharakteristika berücksichtigen.

Als nächstes soll der Einfluss der Geomorphologie auf die simulierte Abflussreaktion analysiert werden. Die Parameter des in Abbildung 5-1 dargestellten Einzugsgebietes wurden dafür variiert und wurden somit ebenfalls zu einem synthetischen Datensatz. Variiert wurden zunächst der Erwartungswert  $E(x_H)$  sowie die Streuung der Hangfließstrecke  $s(x_H)$ , im Intervall von 100 m bis 4500 m. Allen Distanzklassen wurden dabei identische Werte für  $E(x_H)$  und  $s(x_H)$  zugewiesen. Die simulierten Abflussreaktionen sind in Abbildung 5-3 dargestellt. Ein höherer Erwartungswert der Hangfließstrecke sowie eine höhere Streuung führen zu einer langsameren Abflussreaktion, da die Entwässerungswege im Gebiet länger werden. Die Anzahl der Ordinaten des GIUH steigt, da die einzelnen Ordinaten kleiner werden. Des Weiteren ist zu erkennen, dass der Einfluss des Erwartungswertes  $E(x_H)$  höher ist, als der Einfluss der Streuung  $s(x_H)$ .

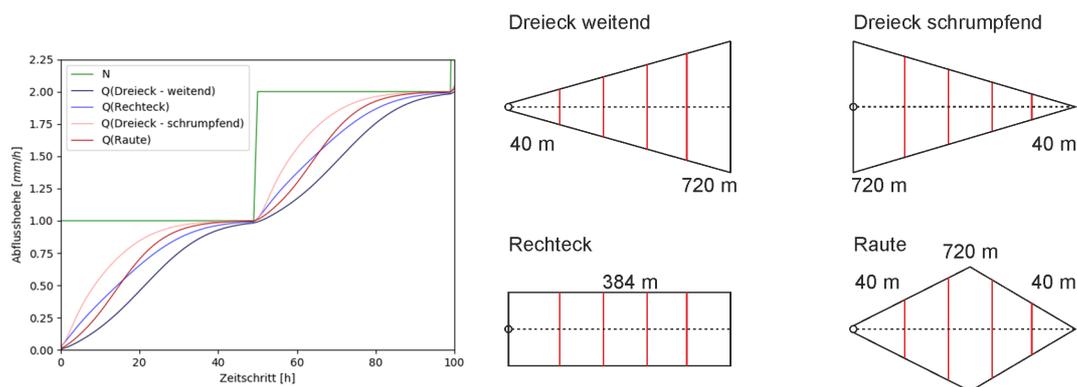


**Abbildung 5-3: Simulierte Abflussreaktion auf Blockregen in Abhängigkeit von  $E(x_H)$  (links) und  $s(x_H)$  (rechts)**

Neben dem tatsächlichen Wert Fließlänge spielt auch die Verteilung der Fließstrecken innerhalb eines Gebiets eine Rolle für die Prozesse der Abflusskonzentration. Um den Einfluss der Verteilung der Fließwege, also der Form des Einzugsgebietes, auf die Abflussreaktion zu überprüfen, wurden vier synthetische Einzugsgebiete erstellt. Der simpelste Aufbau ist eine Rechteckform. Jeder Distanzklasse wurde die gleiche Fläche und die gleichen Werte der Hangfließstrecken  $E(x_H)$  und  $s(x_H)$  zugewiesen. Im vorliegenden Fall wurde der Mittelwert von  $x_H$  des Gesamtgebietes von 384 m als  $E(x_H)$  verwendet und mittels des Variationskoeffizienten von 0.7 die Standardabweichung bestimmt. Als zweite Grundform wurde eine Dreiecksform gewählt. Diese wurde in zwei weiteren Varianten angewendet, eine vom Pegel

ausgehend weitend und eine zweite schrumpfend. Die Parametrisierung erfolgte erneut über die Daten des realen EZG Bernstein a.W.: Zunächst wurden  $s(x_H)$  und  $E(x_H)$  subtrahiert und der kleinste Werte dieser Subtraktion (40 m) als  $E(x_H)$  der Spitze des Dreiecks zugewiesen. Dem Fuß wurde das Maximum der Addition von  $E(x_H)$  und  $s(x_H)$ , 720 m, zugewiesen. Alle weiteren Werte wurden linear interpoliert und mittels Normalisierung der ursprüngliche Mittelwert des Gesamtgebietes von 384 m wieder hergestellt. Mit diesen ermittelten Werten wurde als vierte EZG-Form eine Raute gebildet, deren weiteste Klasse in der Mitte des Einzugsgebietes liegt und am Ober- und Unterlauf spitz zuläuft. Zu beachten ist, dass die Gewässerfließstrecke in allen synthetischen Gebieten gleich ist. Alle verwendeten Formen sind in Abbildung 5-4 dargestellt.

Eine schrumpfende Dreiecksform des EZGs weist die schnellste Reaktion auf, während die weitende Dreiecksform die langsamste Reaktion erzeugt. Die Rechteckform entwässert im Vergleich zu den Dreiecksformen relativ gleichmäßig. Zuletzt stellt die Rautenform die komplexeste Reaktion dar. Zunächst tritt der Abfluss verzögert auf, bedingt durch die weitende Dreiecksform nahe des Auslasses. Es folgt ein starker Anstieg des Abflusses der im Anschluss wieder schnell abflacht, bedingt durch das folgende Schrumpfen des Gebietes. Der Zeitpunkt des Hauptanstieges der Welle wird, bei gleichem Niederschlag, also durch die Form des Einzugsgebietes definiert.



**Abbildung 5-4: Abflussreaktion (links) synthetische EZG-Formen (rechts) mit Angabe der minimalen und maximalen  $E(x_H)$ , Gewässerachse (Punktlinie) und schematischen Distanzklassen (rot)**

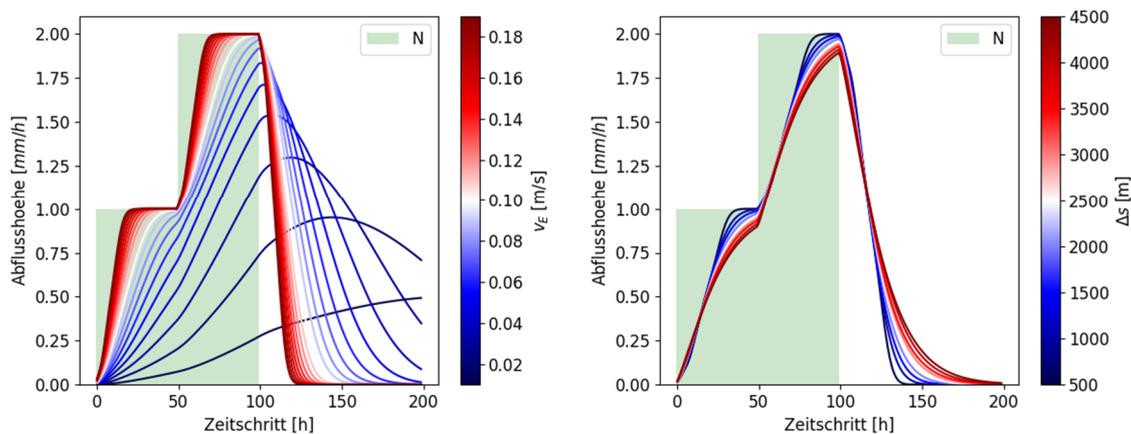
Zusammenfassend spiegeln die Ergebnisse dieser theoretischen Untersuchung die Erwartungen an die Modellstruktur wider. Durch den horizontalen Aufbau des Modells ist die Einheitsganglinie des GIUH Modells mit der einer LSK vergleichbar. Es konnte zudem gezeigt werden, dass das Modell in der Lage ist geomorphologische Gegebenheiten in den Abflussprozess mit einzubeziehen. Dadurch können Unterschiede in der Abflusskonzentration zwischen Einzugsgebieten alleine durch den Einbezug von Geokennwerten in der Modellstruktur abgebildet werden. Eine der Grundvoraussetzungen zur Übertragung in unbeobachtete Gebiete konnte also nachgewiesen werden: Anpassbarkeit des Modells an lokale Gegebenheiten ohne eine Notwendigkeit der Anpassung an beobachtete Ganglinien.

## 5.2 Sensitivität Parametrisierung

Ohne die Verwendung des Dynamikmoduls verfügt das GIUH Modell über drei Parameter: Entwässerungsgeschwindigkeit  $v_E$ , Abflussvolumen  $Q_{Vol}$  und die räumliche Auflösung der DFF-Funktion  $\Delta s$ . Dabei bleibt der letztere Parameter auch bei Verwendung des Dynamikmoduls als vom Nutzer zu definierender Parameter bestehen. In diesem Abschnitt soll auf den Einfluss dieser Parameter eingegangen werden.

Wie zuvor wurde das Einzugsgebiet Bernstein a.W. als Grundlage verwendet und der gleiche synthetische Blockregen als Eingangszeitreihe verwendet. In den Anwendungen in denen  $\Delta s$  nicht variiert wurde, wurde  $\Delta s = 1$  km gewählt.

Die Entwässerungsgeschwindigkeit  $v_E$  [m/s] geht an mehreren Stellen in die Berechnung der Abflussreaktion des GIUH ein. Zum einen wird sie zur Berechnung der Ordinaten des Hangabflusses (Gl. 4.8) und zum anderen zur Weitergabe von Wassereinheiten innerhalb des Gewässerabflusses (Gl. 4.13) verwendet. Ihr Einfluss auf die Form der Abflussreaktion ist daher nicht unmittelbar zu quantifizieren. Allerdings ist die Verwendung des Konzeptes der Entwässerungsgeschwindigkeit natürlich an Erwartungen gebunden. Wie schon in der ursprünglichen Entwicklung des GIUH (Rodríguez-Iturbe & Valdés, 1979) gezeigt, sollte eine veränderte Entwässerungsgeschwindigkeit, bzw. Wellengeschwindigkeit, unterschiedliche Wellenformen erzeugen.



**Abbildung 5-5: Einfluss von  $v_E$  (links) und  $\Delta s$  (rechts) auf simulierte Abflussreaktion**

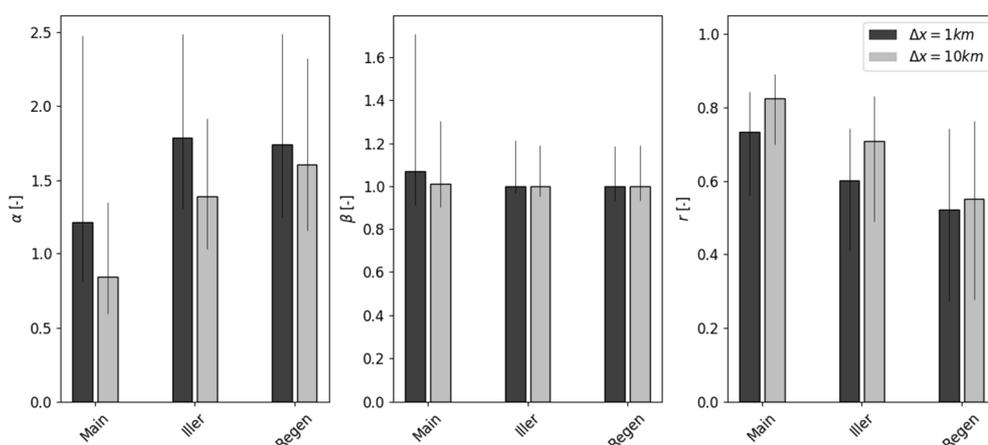
Zudem wird erwartet, dass sich Anstiegszeit und  $v_E$  antiproportional verhalten. Die Ergebnisse der Variation von  $v_E$  zwischen 0.01 und 0.2 m/s in Abbildung 5-5 (links) zeigen, dass die simulierten Abflussreaktionen des Modells den Erwartungen entsprechen. Je höher die Entwässerungsgeschwindigkeit, desto kürzer ist die Verzögerungszeit und desto steiler fällt der Anstieg und der Rückgang der Ganglinie aus. Je geringer die Geschwindigkeit desto länger ist der Anstieg, teilweise über die Niederschlagsdauer hinaus.

Auf der rechten Seite von Abbildung 5-5 ist der Einfluss des Strukturparameters  $\Delta s$  auf die simulierte Abflussganglinie dargestellt. Der Parameter  $v_E$  wurde in dieser Analyse konstant

mit 0.1 m/s angenommen. Die räumliche Auflösung  $\Delta s$  der DFF-Funktion (vgl. Abschnitt 4.2) geht nur indirekt in die Berechnung der Ganglinie mit ein. Sie definiert die Grenzwerte für die Gewässerfließstrecke  $x_s$  zur Abgrenzung der Distanzklassen. Der Parameter beeinflusst somit die strukturelle Grundlage des GIUH-Modells. Mit sinkendem  $\Delta s$  werden mehr Distanzklassen gebildet, da die Unterteilung der Gewässerfließstrecken  $x_s$  feiner wird.

Die Ergebnisse der Variation von  $\Delta s$  zeigen, dass sich die Anzahl von Distanzklassen und die Reaktionszeit des Modells proportional verhalten. Dies kann auf zwei Umstände innerhalb der Modellstruktur zurückgeführt werden. Zum einen steigt durch eine sinkende Anzahl von Klassen die Varianz der Kennwerte innerhalb der Klassen. Wie in Abschnitt 5.1 gezeigt, führt dies zu einem Anstieg der Entwässerungszeit, also einer Verlangsamung der Abflussreaktion. Zum anderen ist die numerische Diffusion im Gewässermodell stärker für Distanzklassen, deren Mittelwerte der Fließstrecken sich stärker unterscheiden. Wie in Abschnitt 4.4.3 beschrieben, wird nach einem Zeitschritt ein Wasserpaket, das sich auf der Grenze zwischen zwei Klassen befindet, dem Gewässerspeicher dieser beiden Klassen anteilig zugerechnet. Das Wasser, das zuvor einen Scheitel in der Nähe der Klassengrenze aufwies, wird gleichmäßig über die gesamte Länge, d.h. Fließdistanz der Klasse, verteilt. Je größer die Länge der Klasse, desto größer die Diffusion der Welle.

Zwar ist der Einfluss von  $\Delta s$  im Vergleich zum Einfluss von  $v_E$  wesentlich geringer, kann aber nicht vernachlässigt werden. Da aus diesen synthetischen Analysen nicht ersichtlich ist, welche räumliche Auflösung besser zu den Untersuchungsgebieten passt, wurde das GIUH-Modell mit analytisch ermittelten Parametern (vgl. folgendes Kapitel 5.3) für alle Ereignisse und alle Gebiete mit  $\Delta s = 1 \text{ km}$  und  $\Delta s = 10 \text{ km}$  angewendet. In Abbildung 5-6 sind die Veränderungen der Mittelwerte der Güteparameter und die Komponenten der Kling-Gupta-Effizient  $\alpha$ ,  $\beta$  und  $r$  der Untersuchungsgebiete dargestellt. Die Streuung der Werte wird durch Fehlerbalken angezeigt.



**Abbildung 5-6: Einfluss von  $\Delta s$  auf Simulationsgüte beobachteter Ganglinien, Verhältnis Standardabweichung (links), Mittelwert (Mitte) und Korrelationskoeffizient simulierter und beobachteter Ganglinie (rechts)**

Zum einen bewirkt die Erhöhung von  $\Delta s$  eine Senkung der Standardabweichung der simulierten Ganglinien aller Untersuchungsgebiete. Für die Ereignisse des Regens und der Iller ist dies ein positiver Aspekt da hier tendenziell eine Überschätzung vorlag, im Main eher ein negativer Aspekt. Dies ist erkennbar im linken Panel von Abbildung 5-6. Positiv ist, dass in allen Gebieten der Korrelationskoeffizient  $r$  steigt, erkennbar im rechten Panel von Abbildung 5-6. Nahezu keinen Einfluss hat  $\Delta s$  auf den Abflussmittelwert. Auf Grundlage dieser Werte wurde  $\Delta s = 10$  km, als Standardauflösung für weitere Untersuchungen verwendet.

Neben diesen Analysen kann die Courant-Bedingungen zur Überprüfung der räumlichen und zeitlichen Auflösung überprüft werden. Die Courant-Bedingung überprüft das Verhältnis zwischen der maximalen Strecke die eine Information im Modell zurücklegen kann, die abhängig von der zeitlichen Auflösung ist, zur räumlichen Auflösung (Maidment, 1993):

$$CR = \frac{v \cdot \Delta t}{\Delta s} \leq 1 \quad 5.4$$

Bei expliziten Lösungsverfahren, wie in dieser Arbeit angewendet, muss die Courant-Zahl  $CR$  kleiner gleich eins sein um numerisch stabile Ergebnisse liefern zu können. In dieser Arbeit sollen Ereignisse in stündlicher Auflösung berechnet werden, die mittleren Entwässerungsgeschwindigkeiten im Main liegen bei etwa 0,15 m/s, im Regen bei 1,04 m/s und in der Iller bei 0,35 m/s. Gemäß Gleichung 5.4 ergibt dies für  $\Delta s = 1$  km Courant-Zahlen von 0,54 (Main), 3,74 (Regen) und 1,26 (Iller). Die Courant-Bedingung ist mit  $\Delta s = 1$  km also nicht in allen Anwendungen erfüllt. Da es sich um Mittelwerte von  $v_E$  handelt muss davon ausgegangen werden, dass auch im Main Ereignisse vorliegen für die Gl. 5.4 nicht erfüllt ist. Für  $\Delta s = 10$  km ist sichergestellt, dass die Courant-Bedingung in nahezu allen Anwendungen erfüllt ist (Lediglich 12 Ereignisse weisen ein analytisches  $v_E$  auf, bei dem die Bedingung nicht erfüllt ist).

Auf eine detaillierte Darstellung des Einflusses von  $Q_{Vol}$ , bzw. dem daraus berechneten Koeffizienten aus  $Q_{Vol}$  und  $VP$  wurde an dieser Stelle verzichtet. Da eine sehr einfache Methode zur Simulation der Abflussbildung gewählt wurde, die mittels einer einzigen Gleichung durchgeführt wird (Gl. 4.6), kann der Einfluss des einzigen Parameters ohne Schwierigkeiten antizipiert werden: Zu variieren wäre der Quotient  $Q_{Vol}/VP$  zwischen 0 und 1. Daraus ergeben sich keinerlei Änderung im Abflusskonzentrationsprozess und daher auch keine Veränderung in der Form der Abflussreaktion. Lediglich das simulierte Abflussvolumen würde sich direkt proportional verändern. Wird der Quotient bspw. gleich 0.8 gesetzt, sinkt das simulierte Abflussvolumen auf 80% seines ursprünglichen Wertes. Dies ist bedingt durch Gl. 4.10 und Gl. 4.16, die keine weiteren Verlustterme aufweisen. Somit entspricht das simulierte Abflussvolumen der Summe von  $P_{eff}$  (Gl. 4.6).

### 5.3 Intra-Ereignis Variation analytischer Dynamikparameter

Für jedes Ereignis wurden Zielgrößen ermittelt, die zur Simulation der Ganglinie mittels GIUH als Dynamikparameter verwendet werden können. Die analytisch ermittelten Parameter stellen Ereignismittelwerte dar. Die Entwässerungsgeschwindigkeit sowie der Abflussbildungskoeffizient können entweder als konstant über die Dauer des entsprechenden Ereignisses und räumlich invariant über alle Distanzklassen angenommen werden oder zeitlich und räumlich variiert werden. Die Möglichkeit der räumlich-zeitlichen Variation der Entwässerungsgeschwindigkeit wurde bereits in den grundlegenden sowie den weiterführenden Arbeiten zum GIUH (Rodríguez-Iturbe & Valdés, 1979; Rigon et al., 2016) angemerkt bzw. umgesetzt.

In diesem Kapitel sollen verschiedene Möglichkeiten zur räumlichen Differenzierung der Dynamikparameter entwickelt werden und ihr Einfluss auf simulierte Abflussganglinien analysiert werden. Im Gegensatz zum vorgehenden Kapitel basiert die Evaluierung der Simulationsergebnisse auf dem Vergleich mit beobachteten Ganglinien. Daher lässt sich nicht nur die Wirkung der Differenzierungen auf die Abflussreaktion beurteilen, sondern auch, ob sie sinnvolle Ergänzungen zum vorgestellten Modellkonzept darstellen.

#### 5.3.1 Räumliche Variation

Durch die Verwendung der DFF können jeder Distanzklasse eines Einzugsgebietes eine beliebige Anzahl von zusätzlichen Kennwerten zugewiesen werden. Mit Hilfe dieser räumlich differenzierten Kennwerte wurde die räumliche Variation der Parameter realisiert. Mittels Ganglinienanalyse sind die Dynamikparameter  $Q_{Vol}$  und  $v_E$  nur als Ereignis- und Gebietsmittelwerte zu bestimmen. Die mittleren Parameter  $y_{Mid}$  wurden mit Hilfe der mittleren Gebietscharakteristika je Distanzklasse räumlich variiert:

$$y_i = y_{Mid} \cdot \frac{E(C)_i}{\frac{1}{N_S} \sum_{j=0}^{N_S} E(C)_j} \quad 5.5$$

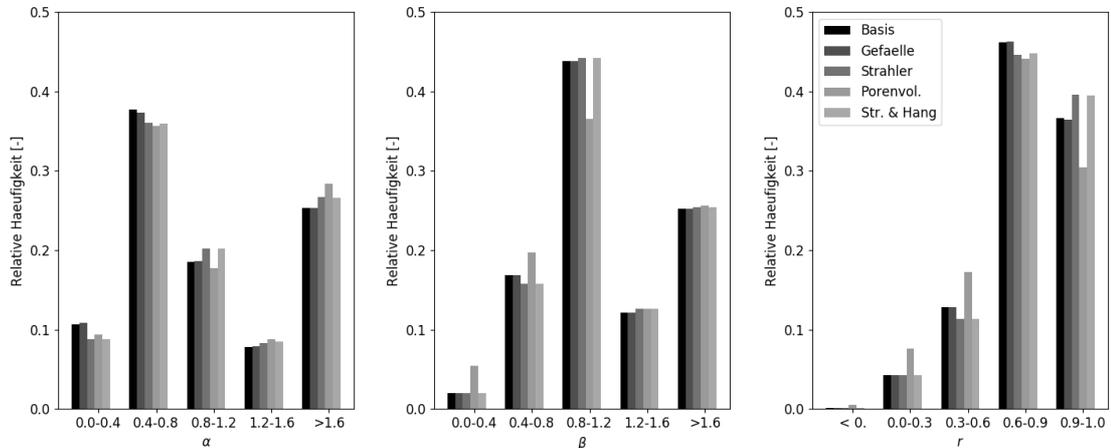
Die Dynamikparameter je Distanzklasse  $y_i$  ergeben sich dabei durch Multiplikation mit dem Quotienten aus Erwartungswertes des Kennwertes  $C$  und der Summe aller Erwartungswerte der verschiedenen Distanzklassen.

Eine Variation der Entwässerungsgeschwindigkeit wurde in der Literatur häufig in Verbindung mit der Strahler-Ordnung vorgenommen. Die Strahler Ordnung dient in diesen Fällen als Maßzahl für die Größe (im Sinne von Breite, Durchflussmenge, etc.) der Gewässer und liefert damit eine physikalische Begründung für die Variation von  $v_E$ : Größere Gewässer weisen in der Regel höhere Fließgeschwindigkeiten auf als kleinere Flüsse/Bäche. Die Entwässerungsgeschwindigkeit der Gewässer  $v_{E,S}$  wurde in dieser Arbeit mittels Gl. 5.5 räumlich differenziert. Der Wert von  $v_{E,S}$  für eine Distanzklasse  $i$  wurde als Produkt der mittleren Entwässerungsgeschwindigkeit  $v_E$  und dem Verhältnis des Erwartungswertes der Strahler Ordnung in Klasse  $i$  zum Mittelwert der Strahler-Ordnung im Einzugsgebiet bestimmt.

Als zweite räumliche Variation wurde die Entwässerungsgeschwindigkeit der Hangflächen  $v_{E;H}$  mit dem Gefälle variiert. Diese Variation folgte der Annahme, dass steile Hänge schneller entwässern als flache Hänge. Als Datengrundlage wurde mittels des DEM des Einzugsgebietes das Oberflächengefälle berechnet und mittels DFF ausgewertet. Die Entwässerungsgeschwindigkeit der Hänge  $v_{E;H}$  in Distanzklasse  $i$  berechnete sich gemäß Gleichung 5.5 als Produkt der mittleren Entwässerungsgeschwindigkeit  $v_E$  und dem Verhältnis des Erwartungswert des Hanggefälle in Klasse  $i$  zum Mittelwert des Hanggefälle im Einzugsgebiet.

Ist der Abflussbildungskoeffizient bekannt (in diesem Falle durch Ereignisanalyse), ist dem Einfluss der Vorfeuchte des Gebietes auf die Abflussbildung bereits Rechnung getragen. Bekannt ist der Anteil des effektiven Niederschlages am Gesamtniederschlag, bzw. wie hoch die Verlustmenge ist. Da das Modell kurzfristige Abflussprozesse simuliert, kann davon ausgegangen werden, dass ein Großteil dieser Verluste durch längerfristigen Rückhalt im Boden des Einzugsgebietes erzeugt wird. Die Aufnahmefähigkeit des Bodens ist begrenzt durch den zur Verfügung stehenden Porenraum (und dessen Füllung am Beginn des Ereignisses). Daher wurde eine räumliche Variation des Abflussbildungskoeffizient  $Q_{Vol}/VP$  mit den Erwartungswerten des Porenraums (GPV oder nFK) mittels Gleichung 5.5 vorgenommen.

Abbildung 5-7 zeigt die Güte der Simulationen des GIUH-Modells mit analytischen Parametern mit und ohne den vorgestellten räumlichen Variationen für das Untersuchungsgebiet Main. *Basis* steht dabei für die in Kapitel 4 vorgestellte Basisvariante des Modells ohne intra-Ereignis Variation der Parameter.



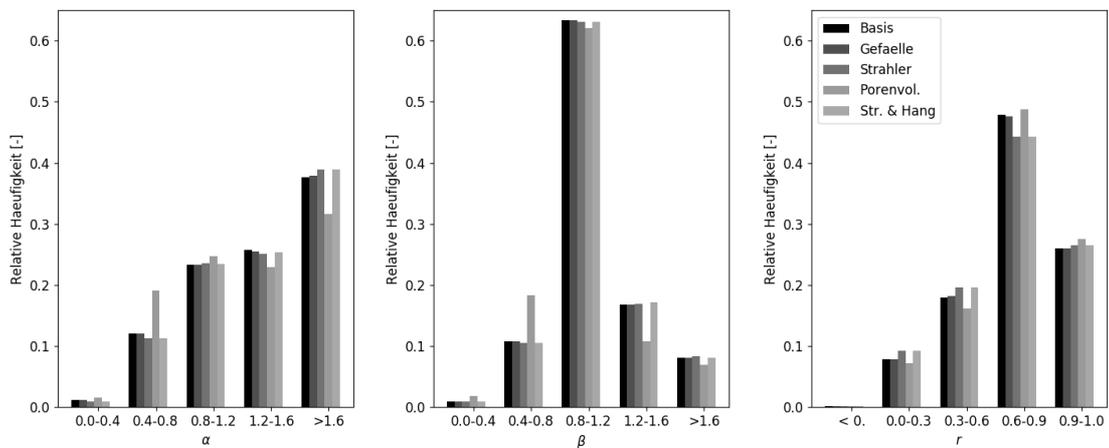
**Abbildung 5-7: Güteparameter der Simulation mit GIUH-Modell bei Verwendung analytischer Parameter, räumlich differenziert im Untersuchungsgebiet Main**

Die Standardabweichung der Ganglinien wird im Allgemeinen für rund 35% der Ereignisse unterschätzt. Demgegenüber steht eine Überschätzung bei rund 25% Ereignissen. Die Analyse von  $\beta$ , also des Mittelwertes zeigt, dass ein Großteil der Ereignisse gut abgebildet werden kann. Rund 35% der Ereignisse im Main weisen eine Korrelation zur Beobachtung über 0.9 auf, weitere 45% immerhin einen Wert zwischen 0.6 und 0.9.

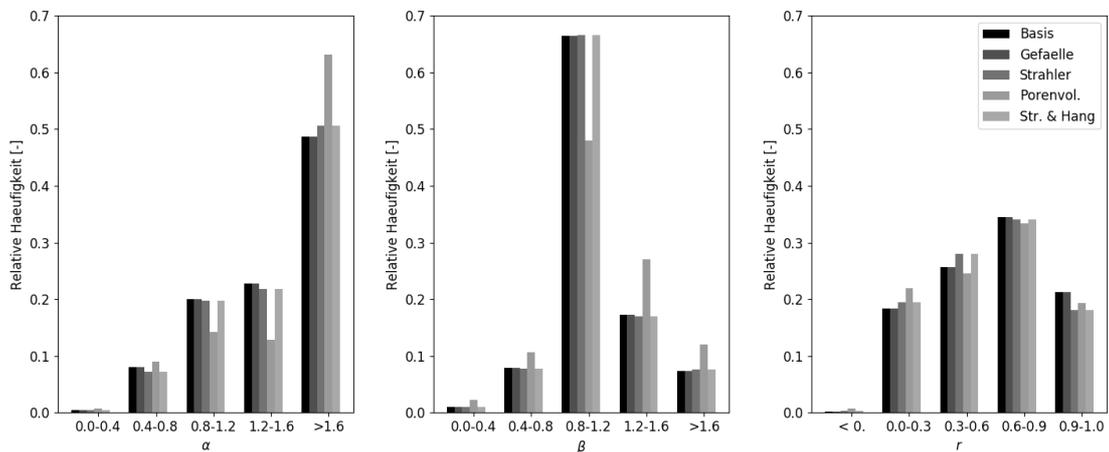
Ebenfalls zu erkennen ist, dass die räumliche Differenzierung die Verteilung der Simulationsgüte nur unwesentlich beeinflusst. Eine Verbesserung ist durch eine Differenzierung der

Entwässerungsgeschwindigkeiten der Hänge und der Gewässer mit der Strahler-Ordnung bzw. dem Hanggefälle, vor allem in Kombination, zu erkennen. Die Variation der Abflussbildung mit dem Porenraum verschlechtert die Simulationsergebnisse, sichtbar in allen Gütekriterien.

Im Vergleich zu den Ergebnissen des Mains zeigt sich bei der Betrachtung der Simulationsergebnisse im Untersuchungsgebiet Iller (Abbildung 5-8) eine andere Problematik. Hier besteht eine starke Überschätzung der Streuung. Mittelwert und Korrelation liegen auf einem vergleichbaren guten Niveau. Die Differenzierung nach Strahler-Ordnung und Gefälle hat hier keinen Einfluss, dafür kann der Anteil von Ereignissen mit großer Überschätzung der Streuung durch eine Differenzierung der Abflussbildung reduziert werden.



**Abbildung 5-8: Güteparameter der Simulation mit GIUH-Modell bei Verwendung analytischer Parameter, räumlich differenziert im Untersuchungsgebiet Iller**



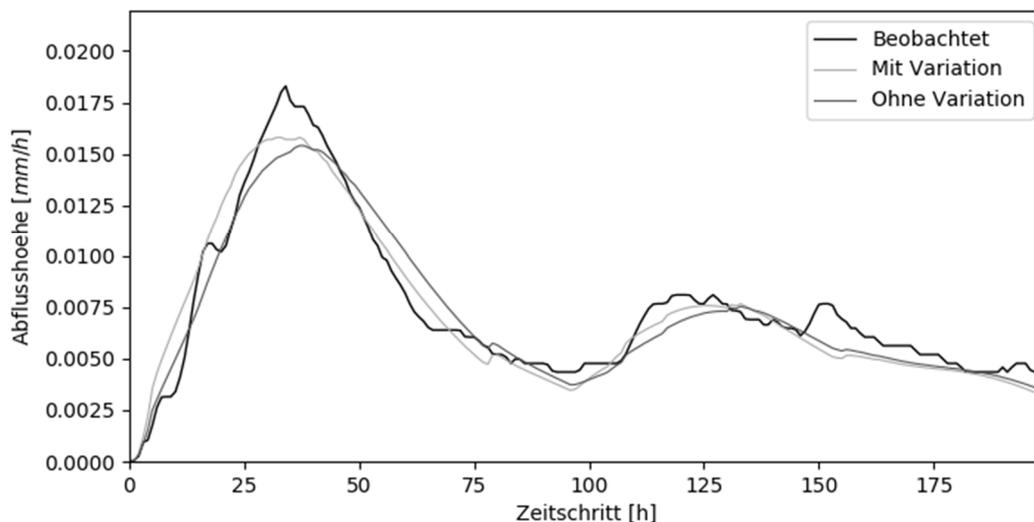
**Abbildung 5-9: Güteparameter der Simulation mit GIUH-Modell bei Verwendung analytischer Parameter, räumlich differenziert im Untersuchungsgebiet Regen**

Im Untersuchungsgebiet Regen (Abbildung 5-9) zeigt sich ein ähnliches Bild wie im Gebiet der Iller. Eine starke Überschätzung der Varianz, die in diesem Falle aber durch eine Variation der Abflussbildung noch vergrößert wird. Im Regen ist zudem auffällig, dass es starke

Probleme mit der Korrelation gibt. Keine der vorgestellten Differenzierungen zeigt in diesem Gebiet eine Verbesserung der Ergebnisse.

Zusammenfassend zeigte sich bei dieser Untersuchung, dass die räumliche Variation der Entwässerungsgeschwindigkeit nur geringfügige Auswirkungen auf die simulierten Ganglinien hatte. Bei gleicher Parametrisierung konnte durch die räumliche Variation des mittleren Kennwertes eine Steigerung der Ganglinienvarianz von 1-5% erzielt werden. Dass die Veränderung so gering ausfällt ist auch die nur moderate Variation des Parameters zurückzuführen. Es sind natürlich deutlich stärkere Maßnahmen möglich, die in weiterführenden Studien entwickelt und ausgewählt werden müssen. Festzustellen ist aber: die entwickelte Modellstruktur ist in der Lage eine Differenzierung der Abflussbildung und –konzentration abhängig vom Fließweg durchzuführen. Die in den gängigen GIUH-Modellen bestehende räumliche Einheit der Parameter wurde unter anderem von Blöschl & Sivapalan (1995) kritisiert. Im Rahmen dieser Arbeit wurde die vorgestellte moderate räumliche Variation der Entwässerungsgeschwindigkeit beibehalten und in den Fallstudien angewendet.

Als Beispiel für simulierte Ganglinien mit und ohne der hier vorgestellten Variation von  $v_E$  sind in Abbildung 5-10 die jeweils simulierten Ganglinien und das zugehörige beobachtete Ereignis dargestellt. Zu erkennen ist, dass durch die Variation der Geschwindigkeit die Reaktion des Gebiets schneller ausfällt und stärkere Peaks aufweist. Das GIUH-Modell kann also durch die Variation Ganglinien mit einer höheren Ganglinienvarianz simulieren. Aus den aufgeführten Gründen wurde in den Fallstudien die räumliche Variation mit Gefälle und Strahler-Ordnung angewendet.



**Abbildung 5-10: Vergleich beobachteter und simulierter Ganglinien, berechnet mit und ohne Variation von  $v_{E,H}$  und  $v_{E,S}$ , Ereignis vom 10.05.2003, Pegel Kemmern**

### 5.3.2 Zeitliche Variation

Das Konzept des Hangabflusses verbindet eine Vielzahl unterschiedlicher Prozesse inner-

halb der Übergangsphase von Niederschlag zu Gewässerabfluss. Die unterschiedlichen Zeitskalen dieser Prozesse, also schnelle Reaktion des Oberflächenabflusses, langsamere Reaktion von bodeninneren Abflüssen wie Interflow o.ä., werden im Parameter  $v_E$  vereint. Damit ist  $v_E$  auch ein Mittelwert aller Prozesse auf den Hangflächen. Aus diesem konzeptionellen Verständnis von  $v_E$  konnte eine Möglichkeit zur zeitlichen Differenzierung abgeleitet werden. Je höher die Niederschlagsintensität, desto wahrscheinlicher ist es, dass Oberflächenabfluss bedingt durch Infiltrationsüberschüsse auftritt. Steigt der Anteil von Oberflächenabfluss, steigt die mittlere Geschwindigkeit der Abflussprozesse eines Hanges. Demnach kann  $v_{E,H,t}$  mit der relativen Niederschlagsintensität  $I(P)$  zeitlich differenziert werden.  $I(P)$  wurde über den Mittelwert aller Intensitäten eines Ereignisses  $N_T$  normiert:

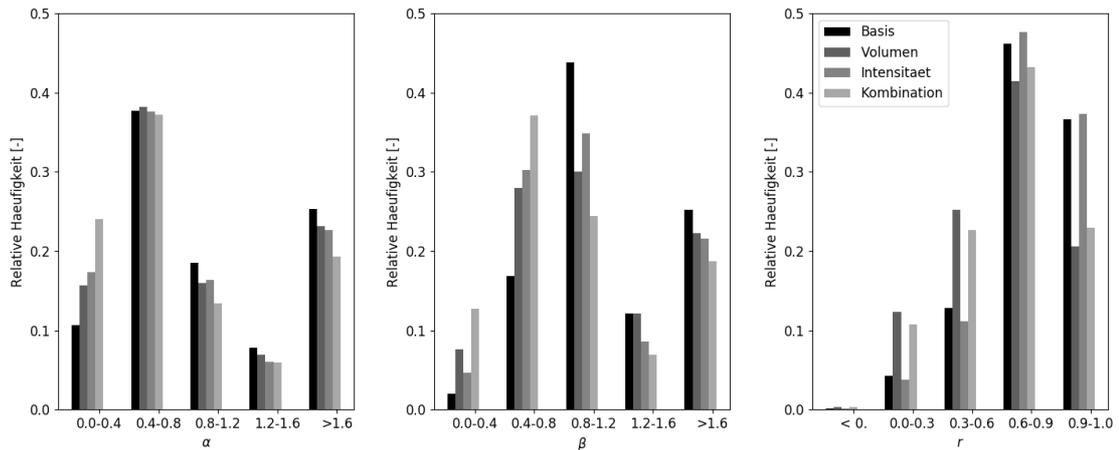
$$v_{E,H,t} = v_E \cdot \frac{I(P)_t}{\frac{1}{N_T} \sum_{i=1}^{N_T} I(P)_i} \quad 5.6$$

Wie bei der räumlichen Differenzierung von  $Q_{Vol}/VP$  beschrieben, hängt der Abflussbildungskoeffizient vom zur Verfügung stehenden Porenraum abzüglich des bereits gefüllten Anteils ab. Der Anteil des gefüllten Porenraumes ändert sich im Laufe eines Ereignisses, da mehr Wasser in den Boden infiltriert als in die Gewässer exfiltriert (da die vertikale Fließgeschwindigkeit des Bodens meist höher ist als die horizontale). Demnach verändert sich der Abflussbildungskoeffizient  $\psi$  (kurz für  $Q_{Vol}/VP$ ) über die Dauer des Ereignisses, bzw. mit der Summe des bereits gefallenen Niederschlages im Verhältnis zur Gesamtsumme  $VP$ :

$$\psi_t = \psi \cdot \frac{\sum_{j=1}^t P_j}{VP} \quad 5.7$$

In Abbildung 5-11 sind die Auswirkungen der zeitlichen Differenzierung auf die Modellergebnisse im Untersuchungsgebiet Main dargestellt. Beide Ansätze haben deutliche Auswirkungen auf die Modellgüte. Die Anzahl von Ereignissen mit einer zu hohen Streuung sank stark. Im Gegenzug erhöhte sich aber die Anzahl von Ereignissen mit einer zu geringen Streuung. Ähnlich waren die Auswirkungen auf  $\beta$  und  $r$ . Auf letzteres wirkte sich lediglich die Variation von  $v_E$  unter Berücksichtigung der Intensität positiv aus. Allerdings fiel dieser positive Aspekt im Vergleich zu den Verschlechterungen in den anderen Kriterien kaum ins Gewicht.

Die Auswirkungen der zeitlichen Differenzierungen in den anderen Gebieten verhielten sich ähnlich zu den in Abbildung 5-11 dargestellten Ergebnissen (Darstellung im Anhang). Vor allem in der Iller fiel die Reduktion der simulierten Varianz und des simulierten Volumens in ähnlich starkem Maße aus. Dagegen waren die Auswirkungen im Untersuchungsgebiet Regen zu vernachlässigen.



**Abbildung 5-11: Güteparameter der Simulation mit GIUH-Modell bei Verwendung analytischer Parameter, zeitlich differenziert im Untersuchungsgebiet Main**

Abschließend wurden die zeitlichen Variationen  $v_{E,H}$  mit der Niederschlagsintensität und von  $Q_{Vol}$  mit dem kumulierten Niederschlagsvolumen verworfen. Obwohl eine Reduktion der Varianz teilweise erforderlich wäre, sind die dadurch wachsenden Fehler im Bereich Korrelation und Mittelwert nicht akzeptabel.

Die hier dargestellten Analysen sind unabhängig von der räumlichen Auflösung. Zwar ändert sich mit  $\Delta s$  die grundlegende Verteilung der Gütekriterien, die Auswirkungen der zeitlich-räumlichen Differenzierung bleiben aber identisch. Darstellungen der Untersuchung können dem Anhang zu diesem Kapitel entnommen werden.

## 5.4 Fehleranalyse und Anpassung

In der vorhergehenden Analyse zeigte sich, dass mit Hilfe von analytischen Parametern nicht alle Ereignisse nachgebildet werden konnten. Dabei muss beachtet werden, dass ein Modell immer ein konzeptionelles Abbild der Realität ist. Die Verwendung von realen Kenngrößen als Modellparameter ist daher in den meisten Fällen nur bedingt möglich. Dabei nimmt die Schwierigkeit der Übertragung mit der Komplexität, d.h. in den meisten Fällen mit der Anzahl der Parameter, zu (u. a. Kirchner, 2006; Samaniego, Kumar & Attinger, 2010; Westberg et al., 2011). Daher kann mit den Ergebnissen der vorhergehenden Analysen nur bedingt Aussage darüber getroffen werden, inwieweit das Modell in der Lage ist die verwendeten Ereignisse nachzubilden.

Festzustellen bleibt aber, dass ein Teil der Ereignisse (ca. 41%) in ausreichender Güte simuliert werden konnte und das ohne weitere Kalibrierung. Alle Parameter wurden aus Geodaten und Niederschlags - Abfluss Messungen ermittelt. Als ausreichende Güte wurde in diesem Zusammenhang ein Ereignis bezeichnet, das bei allen Kriterien maximal  $\pm 0.4$  vom Zielwert abweicht.

Fraglich ist, ob die Fehler der zuvor dargestellten Simulation abhängig von Örtlichkeiten sind, also bedingt durch die Einzugsgebiete, oder systematisch. Lokale Abhängigkeiten deuten eher auf mangelnde Datenqualität hin, während systematische Fehler auf eine inadäquate Modellstruktur hinweisen können. Dieser Fragestellung wurde in diesem Kapitel mit zwei unterschiedlichen Analysen nachgegangen: Zum einen wurde die Verteilung der Simulationsgüte innerhalb der Untersuchungsgebiete verglichen. Da innerhalb eines Untersuchungsgebiets ein Großteil der verwendeten Ereignisse Aufzeichnungen von gleichen Zeitabschnitten sind, konnte über die Verteilung des Fehlers über die Teilgebiete zwischen lokal abhängigen Fehlern und systematischen Fehlern unterschieden werden.

Zum anderen wurde eine Kalibrierung des GIUH vorgenommen. Für jedes Ereignis wurde ein individuelles Parameterpaar ermittelt. Mit diesem Test sollte überprüft werden, ob das GIUH-Modell überhaupt in der Lage ist das entsprechende Ereignis zu modellieren. Wäre dies nicht der Fall, ließe dies auf einen systematischen Fehler schließen. Durch einen Vergleich der kalibrierten und analytischen Parameter konnte außerdem der Wertebereich der Dynamikparameter  $v_E$  und  $Q_{Vol}$  ermittelt werden. Als weitere Datengrundlage wurde parallel ein HBV-Modell ebenfalls für jedes Ereignis individuell angepasst. Durch diese Vergleichskalibrierung sollte ermittelt werden, ob die betreffenden Ereignisse überhaupt durch ein ereignisbasiertes N-A-Modell modelliert werden können. Die Wahl fiel auf das HBV-Modell, das sich in vielen Studien als äußerst flexibel und anpassungsfähig bewährt hat (u. a. Beck et al., 2016; Gharari et al., 2014).

### 5.4.1 Fehleranalyse Teilgebiete

In den folgenden Abbildungen wird die zuvor als Median der Untersuchungsgebiete dargestellte Simulationsgüte differenziert nach Teilgebieten dargestellt. Zusätzlich zum Median werden stets das 25- und 75% Quantil des jeweiligen Gütemaßes als Fehlerbalken angezeigt.

Im Main (Abbildung 5-12) zeigt sich eine ungleiche Verteilung der Fehler. Einige Gebiete, wie Adlerhütte und Bayreuth weisen überdurchschnittlich hohe Abweichungen im Bereich der Standardabweichung und dem Mittelwert der Ganglinien auf. Der Großteil der  $\alpha$ -Werte größer 1 (Abbildung 5-7) wird durch diese 3 Gebiete erzeugt. In den restlichen Gebieten liegt der  $\alpha$ -Median unter 1.

Ein gegensätzliches Bild zeigt sich im Gebiet der Iller (Abbildung 5-13). Auch wenn lokale Unterschiede bestehen, ist eine systematische Überschätzung der Standardabweichung ersichtlich. In allen Gebieten (mit Ausnahme von Thalkirchdorf) liegt  $\alpha$  deutlich über 1. Selbst die Streuung um den Median ist in allen Gebieten vergleichbar (vor allem in Relation zum Median, also bei Betrachtung des Variationskoeffizienten).

Hierzu vergleichbar verteilt sich der Fehler im Einzugsgebiet des Regens (Abbildung 5-14). Auch hier zeigt sich eine durchgehende (Ausnahme Kötzing) und damit als strukturell anzunehmende Überschätzung der Varianz des Abflusses. Deutlich stärker als im Main fällt aber die inter-Teilgebietsvarianz des Korrelationskoeffizienten aus.

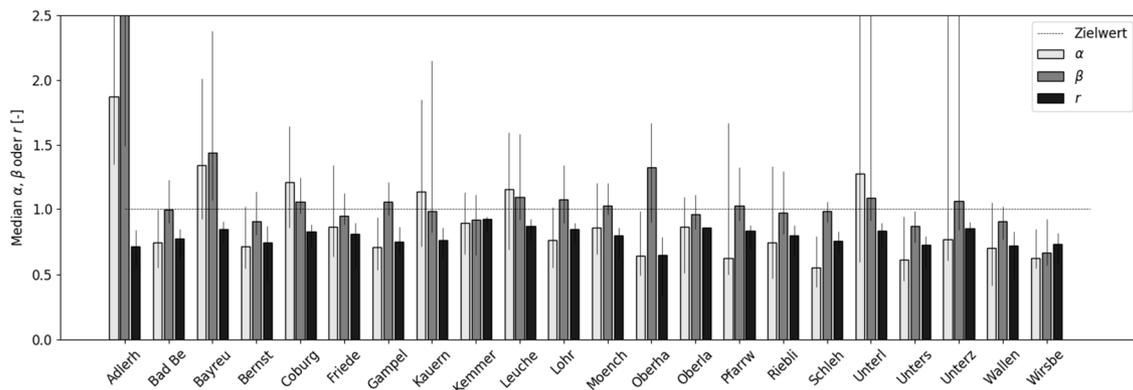


Abbildung 5-12: Mittlere Güteparameter pro Teilgebiet im Untersuchungsgebiet Main

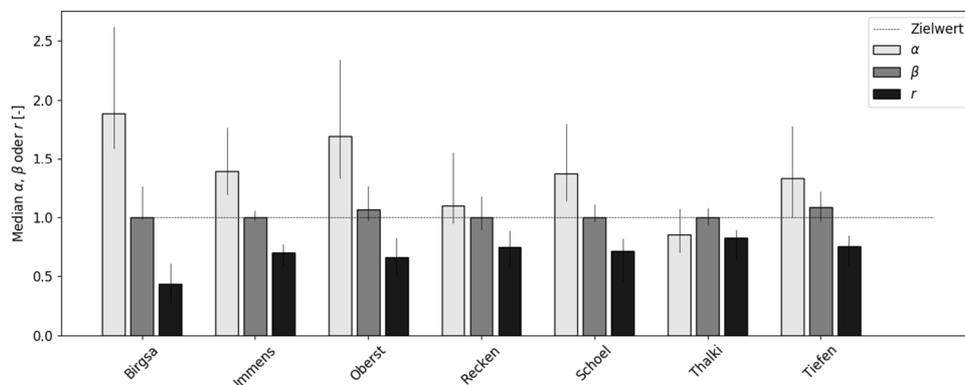


Abbildung 5-13: Mittlere Güteparameter pro Teilgebiet im Untersuchungsgebiet Iller

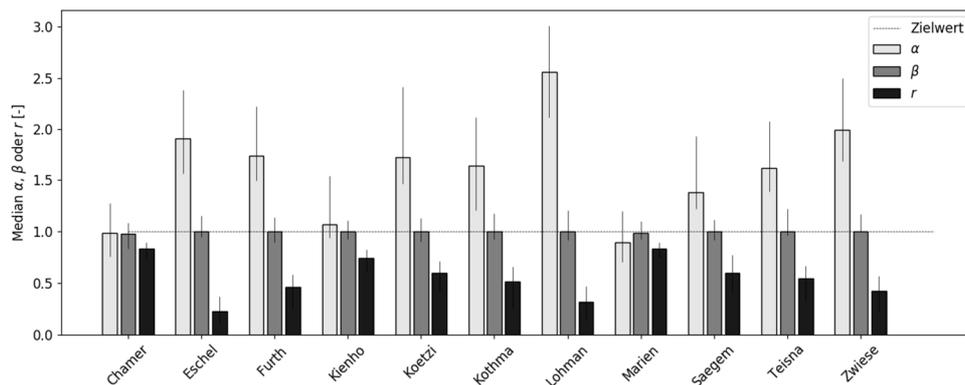


Abbildung 5-14: Mittlere Güteparameter pro Teilgebiet im Untersuchungsgebiet Regen

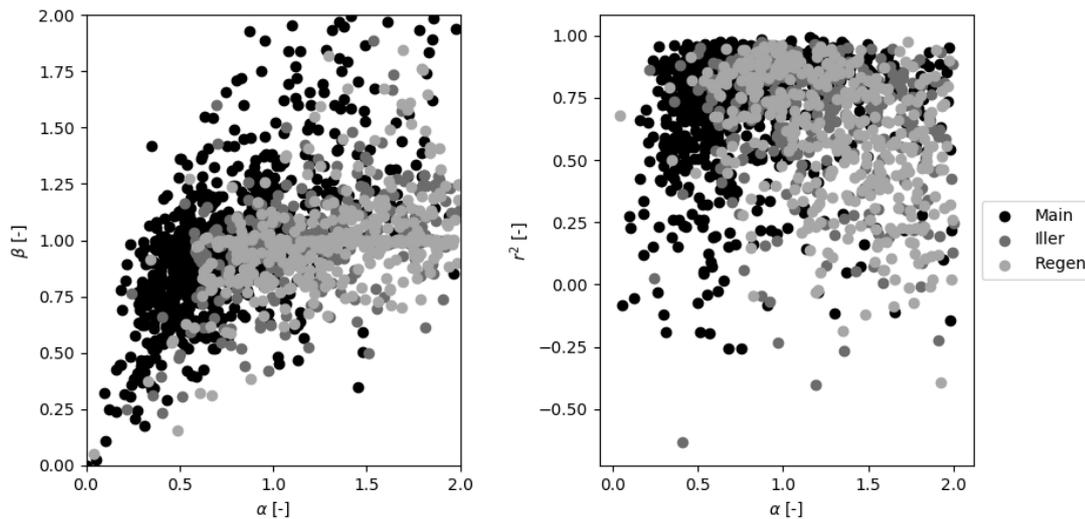
In dieser ersten Analyse zeigte sich, dass die Fehler innerhalb der Untersuchungsgebiete unterschiedlich verteilt sind. Im Main weist eine ungleiche Verteilung darauf hin, dass Fehler durch regionale Abhängigkeiten erzeugt werden, also entweder durch Datenfehler an den einzelnen Pegeln oder in den Niederschlagsaufzeichnungen. Dies bedeutet nicht, dass es

keine Ereignisse gibt, deren analytische Parameter generell nicht übertragbar waren. Ihr Anteil erscheint nur deutlich geringer als in den anderen Gebieten. In Iller und Regen deutet die ermittelte gleichmäßige Verteilung der Fehler darauf hin, dass es in diesen Gebieten strukturelle Probleme gibt. Entweder ist die Modellstruktur generell nicht in der Lage in diesen Gebieten angewendet zu werden, oder nur der Parameterübertrag zwischen Analyse und Modell ist nicht möglich. Auf diese Frage wird im folgenden Abschnitt eingegangen.

### 5.4.2 Anpassung der Modellparameter

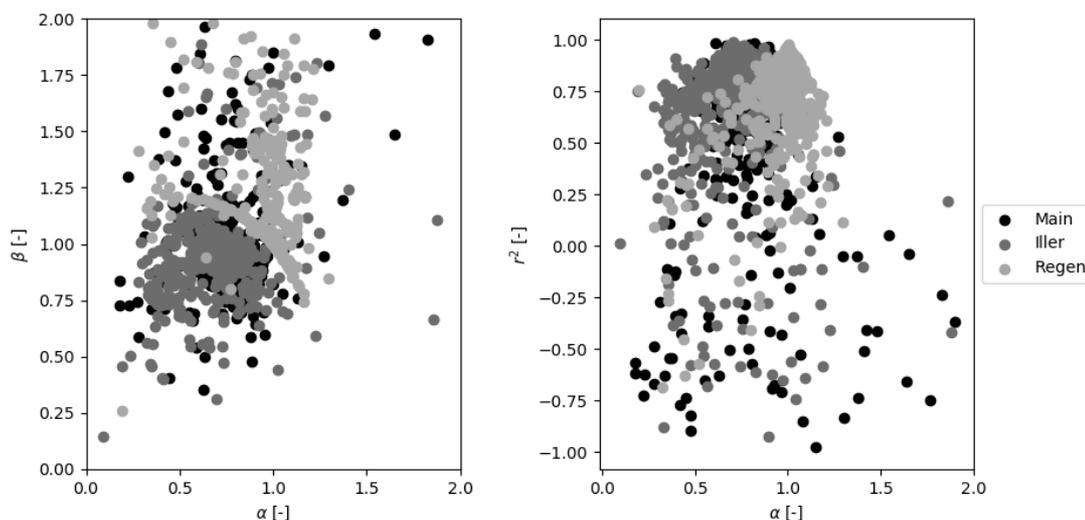
Für alle 2.249 Ereignisse des Datensatzes wurden individuelle Parameter angepasst. Simuliert wurde jedes Ereignis ab Niederschlagsbeginn für 200 h bis 300 h, je nach Ereignisdauer. Aus simulierter und beobachteter Ganglinie wurden alle KGE-Komponenten (Gl. 3.10) sowie der KGE (Gl. 3.11) selbst berechnet. Der KGE diente als Zielfunktion des gewählten Kalibrieralgorithmus *Bound Optimization by Quadratic Approximation* (BOBYQA) (Powell, 2009; Johnson, 2018), der  $v_E$  in den Grenzen [0.0001; 1.0] [m/s] und der Hilfsparameter  $q_{vr}$  in [0.5; 1.5] variiert, um ein Optimum der Zielfunktion zu erreichen. Der Hilfsparameter  $q_{vr}$  wurde als multiplikativer Parameter verwendet und mit dem jeweiligen Ereignisindikator  $Q_{Vol}$  verrechnet. Wie bei der Anwendung eines Kalibrieralgorithmus kann generell nicht ausgeschlossen werden, dass das erreichte Optimum nicht das globale Optimum der Zielfunktion darstellt. Daher wurde der Kalibriervorgang 5-mal wiederholt und nur das beste Ergebnis als finale Parametrisierung ausgewählt. Trotz der durchgeführten Mehrfach-Kalibrierung kann nicht ausgeschlossen werden, dass das globale Optimum nicht erreicht wurde. Lediglich die Chance wurde erhöht es erreicht zu haben.

Folgende Abbildung 5-15 zeigt eine Gegenüberstellung der Güteparameter  $\alpha$  und  $\beta$  sowie  $\alpha$  und  $r$  der Simulationen mit analytischen Parametern. Im Optimalfall sollten sich die Punkte (d.h. die Ereignisse) in der linken Abbildung in der Mitte zentrieren (also  $\alpha = 1$  und  $\beta = 1$ ), in der rechten Abbildung am mittig oberen Rand ( $\alpha = 1$  und  $r = 1$ ). Es ist deutlich zu erkennen, dass dies nicht der Fall ist. Die Ereignisse aus allen Gebieten weisen eine deutliche Streuung auf, vor allem für  $\alpha$  und  $r$ .



**Abbildung 5-15: Gegenüberstellung Güteparameter analytischer Parametrisierung,  $\alpha$  gegen  $\beta$  (links) und  $\alpha$  gegen  $r$  (rechts)**

Die Gegenüberstellung der Güteparameter der Simulationen mit kalibrierten Parametern ist in Abbildung 5-16 gegeben. Im Vergleich zu Abbildung 5-15 fällt eine deutlich höhere Zentrierung der Güteparameter ins Auge.



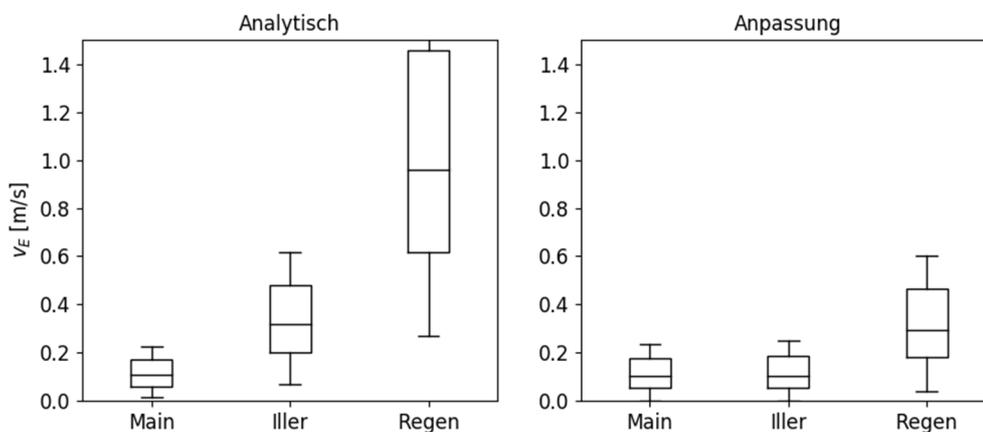
**Abbildung 5-16: Gegenüberstellung Güteparameter kalibrierter Parametrisierung,  $\alpha$  gegen  $\beta$  (links) und  $\alpha$  gegen  $r$  (rechts)**

Zwar sinkt die Streuung um die Zielwerte stark, aber eine Abweichung bleibt erkennbar. Ein Großteil der Ereignisse unabhängig vom Untersuchungsgebiet weist eine zu geringe Varianz auf ( $\alpha < 1$ ). Diese ist aber deutlich geringer als zuvor. Für den Korrelationskoeffizienten kann festgestellt werden, dass der  $r$ -Mittelwert aller Gebiete mit kalibrierten Parametern höher liegt, einzelne Ereignisse aber deutlich geringere Werte aufweisen als mit analytischen Parametern. Dieser Umstand kann in den meisten Fällen darauf zurückgeführt werden, dass

eine Reduktion von  $\alpha$  mit einer Reduktion von  $r$  einhergeht. Zusätzlich muss der nicht dargestellte Wertebereich erwähnt werden. Mit analytischen Parametern liegen 195 Ereignisse mit einem  $\alpha > 2$  und weitere 17 mit einem  $\beta > 2$  außerhalb des dargestellten Bereichs. Durch die Verwendung der kalibrierten Parameter liegen alle  $\alpha$ -Werte unter 2 und lediglich 11 Ereignisse verbleiben außerhalb des dargestellten  $\beta$  Bereichs.

Im Vergleich der Güteparameter vor und nach der Kalibrierung der Parameter ist erkennbar, dass die systematischen Unterschiede zwischen Main, Iller und Regen verschwinden. Die Gütekriterien streuen nach der Kalibrierung in vergleichbarem Maße um dieselben Erwartungswerte. Fraglich ist, ob die erreichte Annäherung der Gütekriterien zwischen den Gebieten durch eine Angleichung der Parameter erfolge. Daher wurden in einer folgenden Analyse, die Parameter vor und nach der Kalibrierung sowie der Einfluss der Veränderung auf die Modellgüte untersucht.

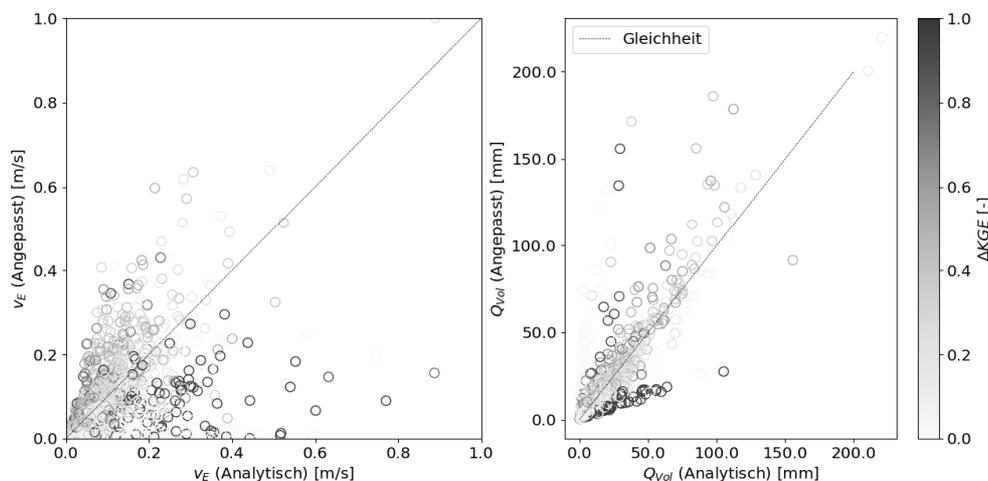
Zunächst wurden Median, 25- und 75% Quantil der analytischen und kalibrierten  $v_E$  Parameter in einem Boxplot nebeneinander gestellt (Abbildung 5-17). Zwei Tatsachen werden aus dieser Abbildung ersichtlich: Zum einen weichen die analytischen Parameter je nach Untersuchungsgebiet stark voneinander ab. Während im Main  $v_E$  Werte im Bereich zwischen  $\sim 0$  und 0.2 berechnet wurden, streuen die Werte in der Iller bis 0.6, im Regen zwischen 0.4 und über 1.4. Nach der Kalibrierung lagen diese Werte wesentlich näher beieinander, Main und Iller wiesen sogar einen sehr ähnlichen Median auf. Nur im Regen lagen die Werte von  $v_E$  noch etwas höher als in Main und Iller.



**Abbildung 5-17: Boxplot analytische und kalibrierte  $v_E$  Parameter**

Um die Anpassung der Parameter zu verstehen, wurde zunächst die Veränderung der Parameter im Gebiet Main untersucht. In diesem Gebiet konnte zuvor kein systematischer Fehler erkannt werden, die Anpassungen (zumindest von  $v_E$ ) in diesem Gebiet waren dadurch am geringsten ausgefallen. Tatsächliche lokale Anpassungen sollten in diesem Untersuchungsgebiet deutlicher zu erkennen sein als in den anderen Gebieten. Hier überlagerte eine systematische Anpassung der Parameter die individuelle Ereignisanpassung.

Für jedes Ereignis im Untersuchungsgebiet Main wurden in Abbildung 5-18 analytische und kalibrierte Parameter  $v_E$  und  $Q_{Vol}$  gegenübergestellt. Die Farbe des Punktes deutet an, wie stark die Anpassung den KGE verändert hat. Je dunkler bläulich der Punkt ist, desto stärker fiel die positive Veränderung des KGE aus. Auf der rechten Seite von Abbildung 5-18 wird die Veränderung von  $Q_{Vol}$  dargestellt. Für  $Q_{Vol}$  ist deutlich zu erkennen, dass nur geringe Veränderungen vorgenommen wurden. Wenn aber Änderungen vorgenommen wurden, hatten diese einen starken Einfluss auf das Kalibrierungsergebnis. Die Anpassung von  $v_E$  weist deutlich mehr Streuung, also stärkere Veränderungen auf. Hier ist auffällig, dass vor allem eine starke Senkung von  $v_E$  in den meisten Fällen mit einem starken Anstieg der Güte einherging.



**Abbildung 5-18: Vergleich analytische und kalibrierte Parameter mit Veränderung des KGE**

Würde in Abbildung 5-18 die farbliche Markierung anhand des finalen KGE vorgenommen (vgl. digitaler Anhang), wäre deutlich zu erkennen, dass die Punkte an den unteren Rändern, also die Ereignisse mit den stärksten Parameterveränderungen die Ereignisse mit dem geringsten KGE sind. Die Anpassung verbesserte zwar den KGE stark, das Ereignis kann aber dennoch nicht in ausreichender Güte modelliert werden. Die geringen Veränderungen von  $Q_{Vol}$ , die in ähnlichem, teils geringerem Umfang in den übrigen Gebieten ausfällt, lassen sich durch den simplen Aufbau der Abflussbildung erklären. Durch seine einfache 1-Parameter & 1-Gleichung Konzeption ist die unmittelbare Verwendung analytischer Parameter einfacher möglich als für  $v_E$ .

Für  $v_E$  zeigt sich in den übrigen Gebieten ein vergleichbares Bild (s. digitaler Anhang): Zwar findet eine stärkere Anpassung statt, die eine stärkere Veränderung des KGE zur Folge hat, aber auch hier finden die stärksten Anpassungen für die Ereignisse mit dem schlechtesten finalen KGE statt ( $KGE \leq 0$ ).

Die Kalibrierung des GIUH konnte zeigen, dass im Main eine Verwendung von analytischen Parametern eher möglich war als in den verbleibenden Gebieten. Dabei mussten die Ereignisse der Iller weniger angepasst werden als im Regen. Zu erklären ist dieser Unterschied

entweder durch eine bessere Datenlage, also bspw. eine höhere Dichte von Niederschlagsstationen, oder durch eine bessere Übereinstimmung zwischen Modellkonzeption und lokalen Gegebenheiten im Main als in Iller und Regen. Unabhängig von der Ursache konnte eine Anpassung der Parameter aber zeigen, dass es generell möglich ist die Ereignisse mit dem GIUH zu modellieren. Dafür wurden, vor allem im Gebiet der Iller, die Parameter auf einen ähnlichen Wertebereich von  $0.1 \pm 0.1$  [m/s] gebracht. Diese Tatsache spricht für einen konzeptionellen Unterschied zwischen den Gebieten.

Bereits bei der Fehleranalyse der Teilgebiete wurde festgestellt, dass es einzelne Teilgebiete gibt, die einen höheren Simulationsfehler aufweisen als andere. Selbst nach der Kalibrierung werden einige Ereignisse identifiziert, deren Simulationsgüte zwar gesteigert, aber nicht auf einen KGE-Wert  $> 0$  gebracht werden konnte. Diese nur unzureichend simulierten Ereignisse erzeugen die unterschiedliche Verteilung von Fehlern zwischen den Gebieten.

### 5.4.3 Ereignisanalyse

In den Analysen dieses Kapitels werden die N-A-Ereignisse mit bekannten Parametern simuliert. Entweder sind die Parameter mittels Ereignisanalyse analytisch bestimmt worden, oder sie wurden kalibriert. Dennoch zeigten einige Ereignisse mit einem  $KGE < 0$ , sehr schlechte Anpassungen. In diesem Abschnitt soll auf diese Ereignisse eingegangen werden und die Ursache für die schlechte Abbildung dieser Ereignisse ermittelt werden.

Zunächst wurden alle Ereignisse selektiert deren Simulationen sowohl mit analytischen als auch kalibrierten Parametern einen  $KGE < 0$  aufwiesen. Die Verwendung beider Gütekriterien soll die Auswahl auf Ereignisse beschränken, deren Prozesse nicht von der entwickelten Modellstruktur abgebildet werden können. Liegt der KGE-Wert mit analytischen Parametern unter 0 ist die Verwendung von analytischen Parametern nicht möglich. Liegt außerdem der KGE-Wert nach der Kalibrierung unter 0, konnte auch kein Parametersatz iterativ ermittelt werden. Sind diese beiden Voraussetzungen erfüllt, ist davon auszugehen, dass das Ereignis nicht durch das Modell simuliert werden kann.

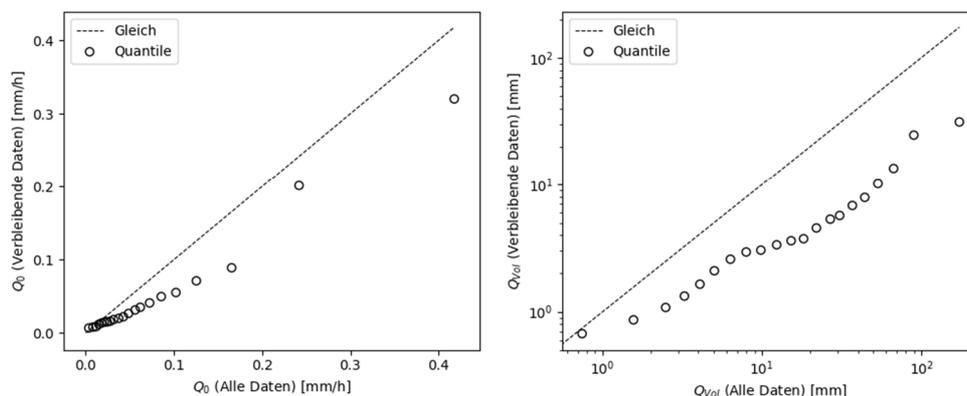
Für etwa 13% der Ereignisse des Datensatzes traf diese Bedingung zu. Untersucht wurden die Randbedingungen der Ereignisse und eine Vergleichskalibrierung wurde durchgeführt, um die Ereignisse genauer zu charakterisieren und damit die Ursache für die schlechte Modellierung zu ermitteln.

Die erste Ursache wurde mit Hilfe einer HBV-Vergleichskalibrierung identifiziert. Das HBV-Modell wurde analog zur GIUH-Kalibrierung individuell für jedes Ereignis mit BOBYQA (Powell, 2009; Johnson, 2018) angepasst. Ziel dieser Anpassung war zu überprüfen, ob die Ereignisse mit den vorhandenen Daten überhaupt modelliert werden konnten. 18,9% der selektierten Ereignisse erreichten auch mit dem HBV-Modell nur einen  $KGE < 0.1$ . Für diese Gruppe von Ereignissen muss angenommen werden, dass ihre Daten fehlerbehaftet sind. In dieser Gruppe tritt keine Häufung einzelner Gebiete auf, nahezu alle Teilgebiete weisen 1-2 Ereignisse in dieser Gruppe auf.

Weitere 19,3 % der ausgewählten Ereignisse sind durch starke Schneeakkumulation oder – ablation geprägt. Die Ereignisindikatoren SC,  $\Delta SC$  sowie sehr große Abflussbeiwerte (teilweise  $> 1.0$ ) ermöglichten die Identifizierung dieser Ereignisgruppe. In dieser Gruppe häufen sich Ereignisse aus den Teilgebieten der Iller. Vor allem für hoch gelegene Gebiete wie Oberstdorf und Birgsau konnte ein Großteil der ausgewählten Ereignisse in dieser Gruppe verortet werden. Das HBV-Modell konnte einen Großteil dieser Ereignisse gut abbilden, möglich durch eine Schneeroutine und mit kalibrierten Anfangszuständen. Ziel des GIUH-Modells ist es aber vor allem Hochwasserereignisse als Folge von Starkregen zu modellieren. Eine Schneeroutine wurde daher nicht vorgesehen. Ereignisse mit einer Beeinflussung aus Schneeschmelze müssen daher von den Analysen dieser Arbeit ausgeschlossen werden.

Eine Analyse der Korrelation zwischen interpoliertem Niederschlag und den Abflussganglinien zeigte, dass für weitere 18,9% der ausgewählten Ereignisse kein oder ein nur unzureichender Zusammenhang zwischen Niederschlag und Abfluss besteht. Diese unzureichende Korrelation kann mehrere Gründe haben, führt aber unabhängig von der Ursache dazu, dass der rein auf Direktabfluss ausgelegte GIUH die Ereignisse nicht abbilden kann. Anzumerken ist, dass auch bei den meisten Schneereignissen nur ein geringer Zusammenhang zwischen Niederschlag und Abfluss besteht. Auch hier sorgt u.a. die Möglichkeit von Anfangszuständen und Basisabflusssimulation des HBV<sub>96</sub>-Modells dafür, dass diese Ereignisse als grundsätzlich simulierbar identifiziert wurden.

Die Ereignisindikatoren der verbleibenden 115 ausgewählten Ereignisse (42%) deuten auf Ereignisse in trockenen Abflussmilieus hin. Die Größenordnungen der Vorfeuchte Indikatoren  $WI_{Pre}$ ,  $\Delta WI$  sowie  $Q_0$  liegen deutlich unter denen des Gesamtdatensatzes. Auf der rechten Seite von Abbildung 5-19 ist beispielhaft der Quantil-Quantil-Plot (QQ-Plot) des Basisabflusses  $Q_0$  dargestellt. Ersichtlich ist, dass der Basisabflussindikator  $Q_0$  der verbleibenden selektierten Daten deutlich geringer ist als im Gesamtdatensatz, also von der Gleichheit der Daten (gestrichelte Linie) abweicht.



**Abbildung 5-19: Quantil-Quantil-Plot Basisabfluss  $Q_0$  (links) und Abflussvolumen  $Q_{vol}$  (rechts) aller Daten und Ereignisse mit sehr geringer Abflussbildung**

Neben den Vorfeuchteindikatoren liegt auch die Größenordnung des Abflussvolumens deutlich unter der Verteilung im Gesamtdatensatz, dargestellt als QQ-Plot auf der linken Seite

von Abbildung 5-19. Daher wurde angenommen, dass bei diesen Niederschlagsereignissen hohe Infiltrationsverluste auftreten und daher kaum eine Abflussganglinie zustande kommt. Solche Ereignisse sind im Datensatz vorhanden (beispielhafte Ereignisse sind im digitalen Anhang zu diesem Kapitel dargestellt) und wurden über die Ereignis Selektion per Niederschlagsintensität (vgl. Kapitel 3) in den Datensatz aufgenommen. Diese Ereignisse wurden verwendet, um dem Dynamikmodul (vgl. Kapitel 6) zu ermöglichen Ereignisse zu identifizieren, die keine Hochwasserereignisse sind bzw. werden. Eine Betrachtung der Scheitelabflüsse (Darstellung ebenfalls im digitalen Anhang) bestätigt diese Annahme. Die Scheitelhöhen dieser Ereignisse liegen signifikant unter denen des gesamten Datensatzes.

Mittels Ereignisanalyse konnte ermittelt werden, welche Ursachen dafür verantwortlich sind, dass einige Ereignisse des Datensatzes mit nur unzureichender Güte modelliert werden konnten. Es wurden Ereignisse identifiziert, deren Daten fehlerbehaftet waren und somit für eine Modellstudie ausscheiden. Des Weiteren wurden Ereignisse identifiziert, die Niederschlagsereignisse sind und keine oder nur insignifikante Abflussereignisse nach sich ziehen. Für diese Ereignisse steht zur Diskussion, ob sie überhaupt von der GIUH-Modellstruktur simuliert werden müssen. Sollte das Dynamikmodul in der Lage sein diese Art von Ereignissen aus üblichen Ereignisindikatoren vorhersagen zu können, könnte auf eine Simulation verzichtet werden. Als strukturelles Defizit wurde das Fehlen einer Schneeroutine identifiziert, bzw. in einigen Aspekten die Möglichkeit der Basisabflusssimulation mit anpassbarer Vorfüllung. Diese Ergänzungen zur Modellstruktur stellen einen ersten Ansatz für kommende Modellerweiterungen dar. Da der Anteil dieser Ereignisse relativ gering ist, 2,5% aller Ereignisse, wurde auf eine Erweiterung im Rahmen dieser Arbeit verzichtet.

### 5.5 Zusammenfassung Abflussmodell

In diesem Kapitel wurde die GIUH-Komponente des in Kapitel 4 vorgestellten Modells untersucht. Die Aufgabe dieser Modellkomponente ist die Transformation von Niederschlagszeitreihen zu Abflusszeitreihen.

Die ersten Tests bezogen sich auf die allgemeine Funktionalität des Modells und die Sensitivität seiner Parameter. Es konnte gezeigt werden, dass der Aufbau des Modells die Form des Einzugsgebietes, in dem es angewendet wird, bei der Simulation von Abflussganglinien berücksichtigen kann. Die Form der Ganglinien wird durch die Geomorphologie des Gebietes definiert. Des Weiteren konnte gezeigt werden, dass die Entwässerungsgeschwindigkeit die Varianz der Ganglinien definiert. Als beste räumliche Auflösung des Modells, d.h. die Länge der Distanzklassen, wurde für eine Zeitschrittweite von  $\Delta t = 1\text{h}$  mit  $\Delta s = 10\text{ km}$  identifiziert. Mit einer Auflösung von 10 km, war in dieser Arbeit die Courant-Bedingung erfüllt. Für zukünftige Anwendungen sollte die hier ermittelte Empfehlung unter Berücksichtigung der vorliegenden Entwässerungsgeschwindigkeiten und der gewählten zeitlichen Auflösung überprüft werden.

Die wesentliche Neuerung der GIUH-Struktur ist die Möglichkeit die Parameter abhängig vom Fließweg zu variieren. Als beste Option zeigte sich in den Untersuchungsgebieten dieser Arbeit eine räumliche Variation der Entwässerungsgeschwindigkeit. Die nach Fließmedium (Hang  $x_H$ ; Gewässer  $x_S$ ) differenzierten Geschwindigkeiten wurden berechnet über die Multiplikation der mittleren Entwässerungsgeschwindigkeit  $v_E$  mit einem Verhältniswert des mittleren Hanggefälles, bzw. der mittleren Strahler-Ordnung je Distanzklasse. Die verwendete moderate Variation der Parameter resultierte in nur einer geringfügigen Erhöhung der Ganglinienvarianz.

Die Parameter des Modells,  $v_E$  und  $Q_{Vol}$ , wurden auf zwei Wegen bestimmt. Zum einen wurden sie mittels Ganglinienanalyse und zum anderen per Kalibrierung ermittelt. Anschließend wurden beide Parametersätze im GIUH-Modell angewendet. Der Vergleich der erzielten Anpassungsgüte zeigte, dass die analytischen Parameter aus Ganglinienanalyse durchaus verwendet werden können, eine Kalibrierung demnach nicht unbedingt erforderlich ist. Für eine direkte Verwendung sollten die Entwässerungsgeschwindigkeiten aber in einem Bereich von  $0.15 \pm 0.1$  m/s liegen. Mit kalibrierten Parametern ließ sich aber bessere Anpassung erreichen.

Abschließend wurde untersucht welche N-A Ereignisse nicht mit dem Modell simuliert werden können. Neben Ereignissen mit Datenfehlern, unzureichendem Zusammenhang zwischen Niederschlag und Abfluss und Ereignissen mit Schneeschmelz Einfluss wurden Ereignisse mit sehr geringer Abflussbildung als problematisch identifiziert. Diese Ereignisse stellen eine Besonderheit des verwendeten Datensatzes dar. Sie dienen lediglich der zweiten Modellkomponente um zu prognostizieren zu können ob ein Niederschlagsereignis ein Hochwasser produzieren wird oder nicht.

## 6 Analyse Dynamikmodul

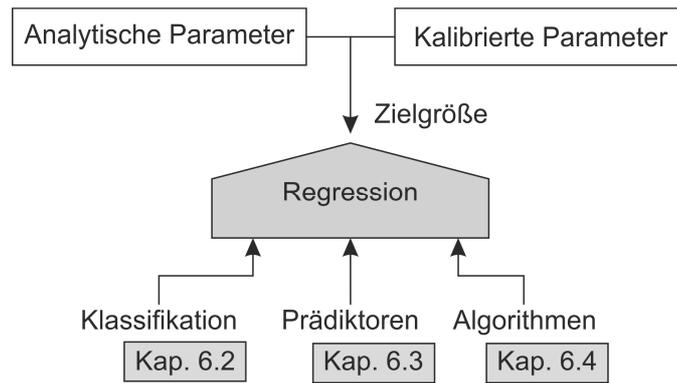
---

In unbeobachteten Gebieten können weder analytische Parameter berechnet werden, noch Parameter kalibriert werden. Um diesem Problem entgegen zu wirken, wurde in dieser Arbeit die Möglichkeit getestet  $v_E$  und  $Q_{Vol}$  mittels maschinellem Lernen (ML) vorherzusagen. Diese, auch als Daten-getriebene Modelle bezeichneten, Algorithmen erarbeiten selbstständig einen Zusammenhang zwischen Prädiktoren und Zielvariable(n).

Die Anwendung von Daten-getriebenen Modellen in unbeobachteten Gebieten erscheint zunächst paradox, da das Fehlen von Daten das eigentliche Problem darstellt. Eine zentrale, klassische Strategie zur Vorhersage von Ganglinien in unbeobachteten Gebieten stellt die Regionalisierung von beobachteten Kennwerten, Parametern oder Modellen dar. Dieser Ansatz wurde auf die Verwendung von ML-Algorithmen übertragen. Zur Regression von Parametern wurden ML-Algorithmen in Einzugsgebieten mit ausreichend zur Verfügung stehenden Daten trainiert und anschließend auf unbeobachtete Gebiete übertragen. Hinter diesem Vorgehen steht die Annahme, dass die Algorithmen selbstständig Prozessgleichungen aus den Daten ableiten, die räumlich übertragbar sind.

In diesem Kapitel soll die Regression von Parametern in unbeobachteten Gebieten und ihre Einflussfaktoren diskutiert und detailliert dargestellt werden. Die Regression ist die Methode des Dynamikmoduls, die im entwickelten N-A-Modell für die Vorhersage von  $v_E$  und  $Q_{Vol}$  verwendet wird. Zudem soll die Leistungsfähigkeit des Dynamikmoduls im Hinblick auf die Genauigkeit der Vorhersage von analytischen und kalibrierten Parametern untersucht und dargestellt werden.

Zu Beginn der Entwicklung des Dynamikmoduls stand aber eine Vielzahl von Möglichkeiten für den Aufbau der Regression zur Auswahl. Beginnend mit der in Kapitel 4.1 formulierten These, dass das Modell eine prozessorientierte Unterteilung benötigt, musste diesbezüglich eine Auswahl der Klassifikation getroffen werden. Es wurden drei unterschiedliche Klassifikationen durchgeführt, um die aufgestellte These zu testen. Neben dieser Auswahl standen 15 Prädiktoren in unterschiedlichen Kombinationen sowie 8 verschiedene ML-Algorithmen zur Auswahl. Abbildung 6-1 fasst all diese Kombinationsmöglichkeiten zusammen und zeigt in welcher Reihenfolge die Auswahlmöglichkeiten in dieser Arbeit diskutiert werden.



**Abbildung 6-1: Übersicht Auswahlmöglichkeiten bei der Regression von Parametern in unbeobachteten Gebieten**

Neben diesen Darstellungen wird in Kapitel 6.1 noch auf die Vorhersage der Klassifikationen und der Ereignisselektion, die im Laufe der Kalibrierung durchgeführt wurde eingegangen. Denn zunächst sollte geklärt werden, ob die Klasse eines anstehenden Ereignisses im Anwendungsfall ausreichend genau vorhergesagt werden kann. Sollte dies nicht der Fall sein, wäre eine Verwendung einer Klassifikation zwecklos, da das falsche Modell verwendet werden könnte. Die Vorhersage der Selektion hingegen folgt einer anderen Überlegung: Wie zuvor dargestellt, konnten Ereignisse identifiziert werden, die mit der verwendeten Modellstruktur nicht simuliert werden können. Steht ein solches Ereignis an, wäre es erstrebenswert, dass das Modell eine Warnung ausgibt, dass die Vorhersage nicht zuverlässig ist, bzw. das Modell dieses Ereignis nicht nachbilden kann.

Allen folgenden Analysen, mit Ausnahme von Kapitel 6.1, liegt eine identische Arbeitsweise zugrunde: eine modellbasierte Auswertung. Mit einer *Leave-one-out* Strategie wurde über alle Teilgebiete eines Untersuchungsgebietes iteriert. Ein ausgewähltes Gebiet wurde als „unbeobachtet“ behandelt und seine Ereignisse zur Validierung verwendet. Die Daten aller verbleibenden Teilgebiete bildeten den Trainingsdatensatz. Die Algorithmen wurden mit Hilfe des Trainingsdatensatzes angepasst und auf das ausgewählte Gebiet übertragen. Für jedes Ereignis wurden Vorhersagen berechnet und mit den zurückgehaltenen beobachteten Daten der MAE beider Zielvariablen berechnet.

## 6.1 Vorhersage Klassifikation & Selektion

In der in Kapitel 4 beschriebenen Kombination von GIUH-Modell und ML übernimmt das Dynamikmodul die Aufgabe der ereignisbezogenen Parametrisierung des GUIH-Modells. Das Dynamikmodul verfügt dabei über eine parallele Struktur zur Schätzung der Parameter. Für jeden Ereignistyp stehen individuell angepasste ML-Algorithmen zur Verfügung. In der praktischen Anwendung des vorgeschlagenen Modells, ist der Typ eines bevorstehenden N-A-Ereignisses unbekannt. Alle verwendeten Klassifikationsmethoden (für *tq\_regional*, *tq\_lokal* und *FuzzyEB*) benötigen Abflussinformationen, die vor Aktivierung des Dynamikmoduls noch unbekannt sind. Um den Typ eines bevorstehenden Ereignisses zu schätzen wurde auf ML-Klassifikatoren zurückgegriffen. Die verwendeten Klassifikatoren arbeiten

analog zu den in Kapitel 4.4.4 vorgestellten Algorithmen, mit dem Unterschied, dass sie diskrete Zielvariablen zurückgeben.

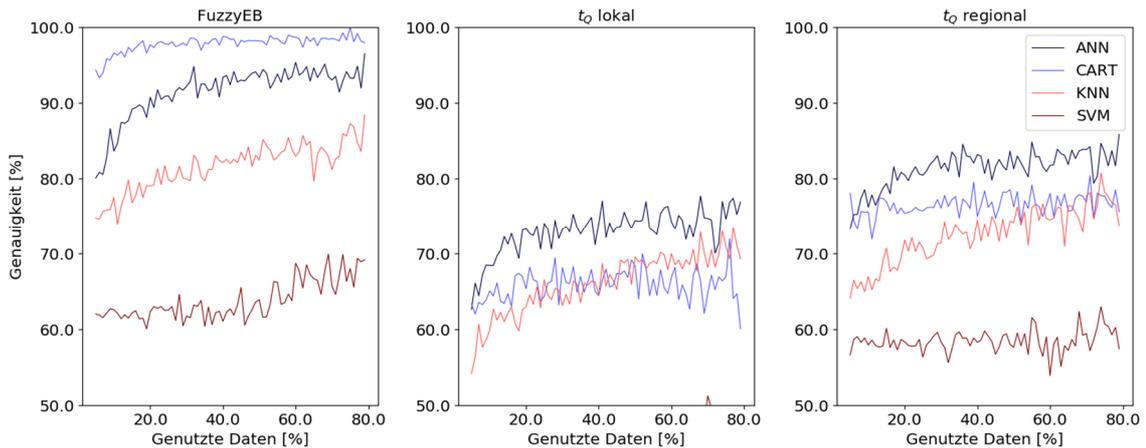
Festzustellen war ob eine Vorhersage des Ereignistyps in ausreichender Güte möglich ist und wie groß die dafür erforderliche Datenmenge ist. Um den Umfang der benötigten Trainingsdaten zu ermitteln, wurde diese Analyse als Lernstudie durchgeführt. Alle Daten innerhalb eines Untersuchungsgebietes wurden zu einem regionalen Datensatz zusammengefasst. Ein Anteil von 20% der Daten wurde als Validierungsdaten zufallsbasiert ausgewählt und die verbleibenden Daten sukzessive dem Trainingsdatensatz zugefügt. Um den Einfluss von Zufallseffekten zu minimieren, wurde die Analyse 10-Mal wiederholt.

In Abbildung 6-2 sind beispielhaft die Ergebnisse der Tests im Untersuchungsgebiet Regen dargestellt. Getrennt nach Klassifikation und verwendetem Algorithmus wurde die mittlere Genauigkeit der Vorhersage über die durchgeführten 10 Iterationen dargestellt. Genauigkeit bedeutet in diesem Zusammenhang zu wieviel Prozent die Vorhersage exakt mit den tatsächlichen Clusterzuordnungen übereinstimmte.

Zu erkennen ist, dass fast alle Algorithmen eine Lernkurve aufweisen. Der einzige Algorithmus, auf den dies nicht zutrifft, ist die Support Vector Machine (SVM), die durchgehend schlechte Ergebnisse liefert und auch keine signifikante Verbesserung aufweist. Dies kann mehrere Ursachen haben, bspw. verwendete Kernel, oder dieses Klassifikationsproblem lässt sich nicht mit einer SVM nachbilden. Die besten Ergebnisse ließen sich in dieser Anwendung mittels eines künstlichen neuronalen Netzes (ANN) (2-lagig) und eines Entscheidungsbaums (CART) erzielen. Letzterer lieferte vor allem für die *FuzzyEB*-Klassifikation besonders gute Ergebnisse. Dieser Umstand ist auf die Baumstruktur des Fuzzy-Entscheidungsbaums und des CART-Algorithmus zurückzuführen. ANN-Klassifikatoren erweisen sich bei diesem Problem als schnellste Lerner, sie erreichten im Vergleich zu den anderen Algorithmen mit weniger Daten ein Level von 80% Genauigkeit für alle Klassifikationen.

Die hier dargestellten Ergebnisse wiederholten sich in den anderen Untersuchungsgebieten. Mit ANN-Klassifikatoren konnten 77% und 76% Genauigkeit in Iller und Main für *tq\_lokal*, 82% und 85% für *tq\_regional* erreicht werden. Mit CART konnte ein Maximum von 90% in der Iller, 97% Genauigkeit für *FuzzyEB* im Main erreicht werden.

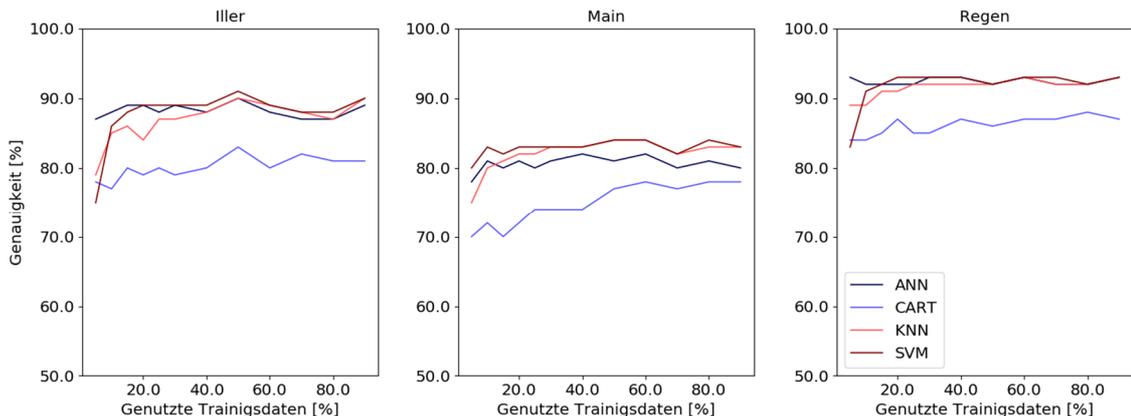
Diese Ergebnisse zeigen, dass die Klassifikationen mit hoher Genauigkeit vorhergesagt werden konnten. Vor allem die *FuzzyEB*-Klassifikation wäre bei gleichem Einfluss auf das Regressionsergebnis aufgrund ihrer hohen Genauigkeit zu bevorzugen. Doch auch die *tq*-Klassifikationen konnten in ausreichendem Maße vorhergesagt werden. Zwar muss mit einem Anteil von 15-18% fehlerhaften Klassifikationen gerechnet werden, der Einfluss einer Fehlklassifikation wird aber im Folgenden (vgl. Kapitel 8.5) noch diskutiert.



**Abbildung 6-2: Entwicklung Vorhersagegüte Klassifikationen nach Algorithmen**

Das zweite Label-Prognose Problem, das in diesem Abschnitt behandelt wird, ist die Vorhersage der Ereigniseselektion aus Kapitel 5.4.3. Wie zuvor wurde diese Analyse als Lernstudie durchgeführt und ML-Klassifikatoren verwendet. Zielvariablen waren in diesem Falle die Selektionslabel.

Die Ergebnisse der Analysen aller Gebiete sind in Abbildung 6-3 dargestellt. In diesem Fall erreichen ANN, KNN und SVM in allen Gebieten in etwa gleiche Genauigkeiten von 85% im Main und 90% in Regen und Iller. Etwas schlechtere Ergebnisse werden in dieser Anwendung mit dem CART erreicht. Auch diese Anwendung zeigt, dass nicht ein einzelner Algorithmus per se als universal bester Algorithmus angesehen werden kann. Wie zuvor können mit einem ANN mit den wenigsten Daten die besten Ergebnisse erzielt werden. Aber generell scheint für dieses Klassifikationsproblem eine geringe Datenbasis ausreichend zu sein, um eine gute Vorhersagequalität zu erreichen.



**Abbildung 6-3: Entwicklung Vorhersagegüte Selektion nach Algorithmen**

Die Ergebnisse zeigen, dass auch die Vorhersage der Ereigniseselektion möglich ist. Die in Kapitel 5.4.3 definierte Selektion unterschied die Ereignisse der Datenbank aufgrund ihrer Eignung zur Evaluierung des Modells. Als geeignet wurden Abflussereignisse aus Starkregen kategorisiert. Als ungeeignet wurden Ereignisse erfasst, deren Aufzeichnungen anscheinend fehlerhaft waren, die unter Schneeeinfluss standen, eine unzureichende Korrelation

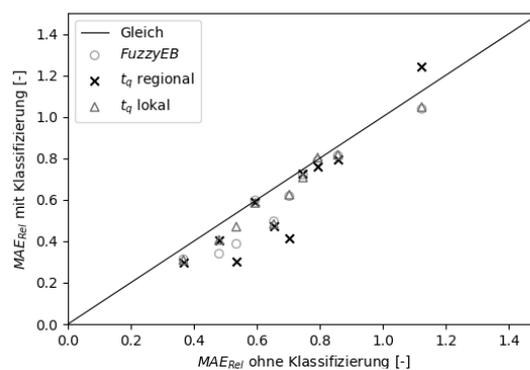
zwischen Niederschlag und Abfluss aufwiesen oder bei denen nur sehr geringe Abflussreaktionen auftraten. Die Ergebnisse der ML-Klassifikation zeigen, dass das Modell aus den Ereignisindikatoren erkennen kann, ob das zu erwartende Ereignis zu diesen Arten von Ereignissen gehören wird. Damit kann das Modell seine Eignung zur Simulation des zu erwartenden Ereignisses prognostizieren.

## 6.2 Prozessorientierte Unterscheidung

Um zu überprüfen, ob eine prozessorientierte Unterscheidung der Ereignisse verwendet werden soll und welche Klassifikation dafür in Frage kommt, wurde eine modellgestützte Analyse durchgeführt.

Wie zuvor beschrieben wurde eine PUB-Regression mit allen drei Klassifikationen sowie ohne Klassifikation durchgeführt. Die MAE-Werte der Simulationen mit unterschiedlichen Algorithmen wurden in allen Teilgebieten anschließend als Mittelwerte pro Untersuchungsgebiet berechnet. Wie in Kapitel 3.4 dargelegt, wurde der MAE verwendet, da er mehrere Aspekte der Anpassung zwischen Beobachtung und Simulation vereint und dabei eindeutig interpretierbar bleibt.

In Abbildung 6-4 stellt die relativen MAE-Werte, also MAE bezogen auf den jeweiligen Gebietsmittelwert, mit und ohne Klassifikation gegenüber. Der relative MAE wurde verwendet, um die Vorhersage von  $v_E$  und  $Q_{Vol}$  gemeinsam darzustellen. Sind der relative MAE, bspw. für  $v_E$ , der Simulation mit und ohne Klassifikation in einem Gebiet, bspw. Regen, gleich so befindet sich die Markierung auf der schwarzen Diagonalen. Ist der Fehler ohne Klassifikation höher, befindet sie sich unterhalb der Diagonalen.



**Abbildung 6-4: Gegenüberstellung relativer MAE ( $v_E$  und  $Q_{Vol}$ ) mit und ohne Prozessunterscheidung für PUB-Regression in allen Untersuchungsgebieten**

Bis auf eine Anwendung mit  $tq\_regional$  liegen alle Punkte in Abbildung 6-4 unterhalb der Diagonalen. Den im Mittel größten Abstand zur Diagonalen, gleichbedeutend mit dem geringsten Fehler, weisen die Simulationen mit Klassifikation  $tq\_regional$  auf, gefolgt von  $FuzzyEB$ .

Der hier dargestellten Untersuchung liegen die Simulationen von jeweils 2 Zielvariablen für 2.249 Ereignisse aus 42 Teilgebieten zu Grunde. Im Mittel dieser 4498 Stichproben konnte mit *tq\_regional* die beste Vorhersagegüte erreicht werden. Zu beachten ist aber, dass alle Klassifikationen zu einer Verbesserung führten. Diese Untersuchung reicht zwar nicht aus, um die These der Prozessunterscheidung zu bestätigen, aber sie konnte durch diese umfangreiche Untersuchung auch nicht verworfen werden. Aufgrund dieser und der vorhergehenden Untersuchung wurde *tq\_regional* als prozessorientierte Unterscheidung für das Dynamikmodul und damit für die folgenden Analysen ausgewählt.

### 6.3 Prädiktoren

Insgesamt standen für diese Arbeit 15 Ereignisindikatoren als mögliche Prädiktoren für die ML-Algorithmen zur Verfügung. Es ist unklar, ob alle Indikatoren benötigt werden, um ausreichend Informationen über die Zielvariablen an den Regressor zu vermitteln, oder ob zu viele Prädiktoren redundante und daher störende Daten in den Regressionsprozess einfließen lassen. Daher wird in diesem Abschnitt eine Auswahl von adäquaten Prädiktoren ermittelt. Zunächst wird auf die Indikatoren und ihre Kombinationen eingegangen. Anschließend wird die Güte der PUB-Regressionen mit den unterschiedlichen Kombinationen dargestellt.

#### 6.3.1 Kombinationen

Die Ereignisindikatoren sind teilweise miteinander korreliert. Es ist a priori anzunehmen, dass beispielsweise die mittlere mit der maximalen Niederschlagsintensität korreliert ist. Um diese Vermutung und die anderen Zusammenhänge sichtbar zu machen, sind in Tabelle 6-1 die Spearman's  $\rho$  Werte, also die rangbasierten Korrelationskoeffizienten zwischen allen berechneten Ereignisindikatoren, dargestellt. Höhere Werte sind hervorgehoben.

**Tabelle 6-1: Spearman's  $\rho$  zwischen den Ereignisindikatoren**

	DP	I <sub>0</sub>	VP	VRI	$\psi$	T <sub>E</sub>	T <sub>PRE</sub>	Q <sub>0</sub>	W <sub>IE</sub>	W <sub>IPRE</sub>	$\Delta$ WI	I <sub>MAX</sub>	D <sub>50</sub>	SC	$\Delta$ SC
DP	1.00	0.39	0.26	0.04	0.02	0.17	0.14	0.05	0.01	0.01	0.07	0.15	0.05	0.11	0.03
I <sub>0</sub>	0.39	1.00	0.53	0.02	0.01	0.16	0.14	0.00	0.08	0.10	0.10	0.87	0.07	0.17	0.13
VP	0.26	0.53	1.00	0.02	0.01	0.02	0.02	0.01	0.02	0.04	0.03	0.71	0.04	0.03	0.05
VRI	0.04	0.02	0.02	1.00	0.00	0.01	0.01	0.01	0.30	0.30	0.15	0.03	0.02	0.04	0.03
$\psi$	0.02	0.01	0.01	0.00	1.00	0.25	0.26	0.03	0.03	0.06	0.16	0.04	0.00	0.35	0.04
T <sub>E</sub>	0.17	0.16	0.02	0.01	0.25	1.00	0.93	0.05	0.34	0.32	0.22	0.16	0.05	0.75	0.02
T <sub>PRE</sub>	0.14	0.14	0.02	0.01	0.26	0.93	1.00	0.06	0.39	0.38	0.17	0.14	0.08	0.79	0.13
Q <sub>0</sub>	0.05	0.00	0.01	0.01	0.03	0.05	0.06	1.00	0.04	0.04	0.05	0.01	0.05	0.07	0.05
W <sub>IE</sub>	0.01	0.08	0.02	0.30	0.03	0.34	0.39	0.04	1.00	0.96	0.20	0.05	0.06	0.36	0.14
W <sub>IPRE</sub>	0.01	0.10	0.04	0.30	0.06	0.32	0.38	0.04	0.96	1.00	0.01	0.07	0.07	0.37	0.15
$\Delta$ WI	0.07	0.10	0.03	0.15	0.16	0.22	0.17	0.05	0.20	0.01	1.00	0.08	0.01	0.13	0.04
I <sub>MAX</sub>	0.15	0.87	0.71	0.03	0.04	0.16	0.14	0.01	0.05	0.07	0.08	1.00	0.20	0.18	0.10
D <sub>50</sub>	0.05	0.07	0.04	0.02	0.00	0.05	0.08	0.05	0.06	0.07	0.01	0.20	1.00	0.09	0.14
SC	0.11	0.17	0.03	0.04	0.35	0.75	0.79	0.07	0.36	0.37	0.13	0.18	0.09	1.00	0.31
$\Delta$ SC	0.03	0.13	0.05	0.03	0.04	0.02	0.13	0.05	0.14	0.15	0.04	0.10	0.14	0.31	1.00

Ersichtlich ist das Vorhandensein von Multikolarität. Mehrere Indikatoren sind stark miteinander korreliert, verfügen also über größtenteils gleiche Informationen. Die Verwendung von korrelierten Prädiktoren kann unter Umständen zu einer Überanpassung der Modelle führen. Diese Gefahr besteht vor allem für MLR-Modelle. Daher wurden auf Grundlage von Tabelle 6-1 Gruppen von korrelierten Indikatoren gebildet. Anschließend wurde eine erste, variable Prädiktorkombination „*Select*“ entwickelt, bei der die Algorithmen sich jeweils einen der Indikatoren aus den folgenden Gruppen wählen:

- Niederschlag (NC): DP, I<sub>0</sub>, VP, I<sub>MAX</sub>
- Zustand (CS): VRI, WI<sub>E</sub>, WI<sub>Pre</sub>
- Schnee (SC): T<sub>E</sub>, T<sub>Pre</sub>, SC

Zudem wurden der *Select* Kombination alle unkorrelierten Indikatoren,  $\psi_{Pre}$ ,  $\Delta WI$ ,  $D_{50}$ ,  $\Delta SC$  und  $Q_0$  zugewiesen.

Zwar sind einige der Indikatoren korreliert, dennoch fügen sie einen, wenn auch kleineren, weiteren Teil Informationen in den Regressionsprozess hinzu. ML-Algorithmen wie ANN, RF u.a. sollen allgemein durch ihre Struktur gegen eine Überanpassung durch Multikolarität gefeit sein (Han & Kamber, 2010). Daher wurde für die Erstellung der übrigen Kombinationen eine andere Strategie gewählt.

Viele Indikatoren werden aus den gleichen Daten abgeleitet. Beispielsweise ist für die Berechnung der Indikatoren VP, DP, I<sub>0</sub>, I<sub>MAX</sub>,  $D_{50}$  und VRI lediglich eine Niederschlagszeitreihe nötig. Stehen diese Daten zur Verfügung, besteht kein Unterschied im Arbeitsaufwand für die Bestimmung eines einzelnen Indikators oder aller Indikatoren. Folgend dieser Logik wurden die Prädiktorkombinationen basierend auf der Datengrundlage entwickelt. Wie zuvor beschrieben, umfasst Kombination *N* alle Indikatoren aus Niederschlagsdaten. Kombination *T* umfasst die Temperaturindikatoren T<sub>E</sub> und T<sub>Pre</sub> und *S* die aus MODIS Daten gewonnenen Schneedaten SC und  $\Delta SC$ . Kombination *W* fasst die Vorfeuchteindikatoren WI<sub>E</sub>, WI<sub>PRE</sub> und  $\Delta WI$  sowie  $\psi_{Pre}$  zusammen, da für all diese Indikatoren Zeitreihen des Abflusses zur Verfügung stehen müssen. Letztlich wurde  $Q_0$  als eigene Kombination aufgenommen, da dieser Indikator eine Momentaufnahme des Abflusses darstellt und keine Aufzeichnung der Vergangenheit.

Kombination *N* wurde als einzige Kombination als alleinige Prädiktorkombination verwendet, da das Vorhandensein von Niederschlagszeitreihen als Voraussetzung zur Anwendung des Modells angesehen wurde. Alle weiteren Kombinationen wurden in unterschiedlichen Zusammensetzungen in der PUB-Regression getestet. Bspw. beschreibt NQT die Verwendung der Kombinationen *N*,  $Q_0$  und *T* als Prädiktoren der ML-Algorithmen zur Vorhersage der Zielvariablen. Um die unübersichtliche Anzahl von möglichen Zusammensetzungen von Kombinationen etwas zu reduzieren, werden im folgenden Abschnitt nur die besten 9 Kombinationen gezeigt (die Anzahl 9 ergibt sich aus Tabelle 6-2).

### 6.3.2 Anwendung der Kombinationen

Mit jeder Zusammensetzung von Kombinationen wurde in allen 3 Untersuchungsgebieten für beide Zielvariablen, analytisch wie kalibriert, mit allen 4 Klassifikationen eine Regression durchgeführt und die mittlere Güte berechnet. Jeder der erhaltenen 48 Werte stand für die mittlere Güte einer Kombination über alle Teilgebiete und ML-Algorithmen innerhalb eines Untersuchungsgebietes und einer Klassifikation. Basierend auf dieser Simulationsstudie wurde eine Ranganalyse durchgeführt. Es wurden sowohl analytische als aus kalibrierte Zielvariablen verwendet, da die gewählte Prädiktorkombination für beide Fälle die beste Wahl darstellen sollte.

Innerhalb einer Simulationsstudie (bspw. Regen; *tq\_regional*; analytisch  $v_E$ ) wurde jeder Kombination abhängig vom MAE eine Rangzahl zugeordnet. Der geringste MAE erhielt die Rangzahl 1, der höchste die Rangzahl 9. Dadurch konnten die MAE-Werte zwischen Variablen und Teilgebieten vergleichbar gemacht werden. In Tabelle 6-2 sind die Häufigkeiten der vergebenen Rangzahlen je Kombination dargestellt. In der letzten Zeile ist die nach Häufigkeit gewichtete mittlere Rangzahl der Kombinationen gegeben.

**Tabelle 6-2: Ranganalyse der besten 9 Prädiktorkombinationen**

Rang(MAE)	R = 1	R = 2	R = 3	R = 4	R = 5	R = 6	R = 7	R = 8	R = 9	SP
NQTS	4	5	11	5	5	4	6	4	4	3.83
NQT	6	6	5	7	6	5	4	7	2	4.43
NQ	3	6	6	7	10	4	4	5	3	4.52
NT	7	3	4	7	2	12	7	3	3	4.83
NQWS	7	7	7	2	4	2	3	7	9	5.00
Select	4	7	3	6	10	5	2	2	9	5.07
NWS	6	5	5	2	4	6	13	5	2	5.13
NW	8	4	3	3	2	8	6	9	5	5.47
N	3	5	4	9	5	2	3	6	11	5.87

Kombination NQTS weist demnach den geringsten mittleren Rang auf, erreicht in Mittel also den geringsten Fehler. Dennoch ist Vorteil von NQTS nicht eindeutig. Zwar erreicht diese Kombination im Mittel den geringsten Fehler, bei 4 Simulationen erzeugt sie aber den höchsten Fehler (Rangzahl = 9). Allerdings kann auch keine andere Kombination dieses Defizit vermeiden. Zudem wird in Tabelle 6-2 die Wichtigkeit des Prädiktors  $Q_0$  ersichtlich, denn die drei besten Kombinationen vereinen alle die Niederschlagsindikatoren und den Abfluss zu Beginn des Ereignisses. Zu beachten ist außerdem, dass eine schlechte Gesamtbeurteilung nicht zwangsläufig auf schlechte Ergebnisse in allen Anwendungen hinweist. Bspw. die Kombinationen N und NW, die in der gesamt Betrachtung am schlechtesten abschnitten wurden in 8, bzw. 2 Anwendungen als beste Alternative gewählt. Zu beachten ist auch, dass die hier schlechtesten Kombinationen N oder NW, teilweise die beste Kombination waren.

Tabelle 6-2 stellt den Mittelwert über alle Algorithmen dar. Es ist dabei fraglich, ob die Kombinationen für alle Algorithmen gleich gut funktionieren oder ob einige Algorithmen andere Kombinationen bevorzugen. Daher wurde die durchgeführte Ranganalyse, getrennt

nach den verwendeten Algorithmen, erneut durchgeführt. Als Ergebnis wurde in Tabelle 6-3 die Kombination mit dem geringsten gewichteten Rang angegeben (das vollständige Ranking kann dem digitalen Anhang entnommen werden).

Die Ranganalyse zeigt, dass Regressionsalgorithmen mit nur einem Schätzer Kombinationen mit wenigen Prädiktoren präferieren, wie bspw. MLR, CART, ANN und besonders SVM. In diesem Falle führt eine Reduktion der Prädiktoren zur Vermeidung von Überanpassung. Anders KNN, der ein Ähnlichkeitsschätzer ist. Im Falle dieses Algorithmus wird die Regression auf der Ähnlichkeit der Datensätze aufgebaut. Je mehr Prädiktoren eine Kombination hat desto detaillierter kann die Ähnlichkeitsanalyse ausfallen. Somit führten umfangreichere Kombinationen in diesem Fall zu besseren Ergebnissen. Die Ensemble Regressoren RF und ADA präferieren ebenfalls Kombinationen mit vielen Prädiktoren, da sie durch ihre Struktur mit mehreren Schätzern bereits gegen Überanpassung gefeit sind. Auffällig ist zudem, dass ANN1 und ANN2 unterschiedliche Präferenzen aufweisen, obwohl sie sich lediglich durch die Anzahl der *Hidden-Layer* unterscheiden. Ein Blick auf das vollständige Ranking zeigte, dass sich die präferierten Kombinationen auf den vorderen Rängen sehr ähnelten, durch die geringen Unterschiede in der Güte aber ihre Reihenfolge unterschiedlich war.

**Tabelle 6-3: Ergebnis Ranganalyse getrennt nach Algorithmus, Veränderung des mittleren Fehlers bei Verwendung der Kombination NQTS**

Algorithmus	Beste Kombination	$\Delta MAE_{Rel}$ (NQTS)	
		$Q_{Vol}$	$v_E$
MLR	NQ	-4%	-4%
CART	Select	-7%	-4%
KNN	NQWS	-1%	-1%
SVM	N	-3%	-3%
ANN1	NWS	0%	-4%
ANN2	NQ	0%	-5%
RF	NQTS	0%	0%
ADA	NQTS	0%	0%

Angestrebt wurde, das Dynamikmodul mit einer einheitlichen Prädiktorkombination zu betreiben. Daher wurde in Tabelle 6-3 ermittelt welche Veränderung des MAE durch die Auswahl von NQTS für die einzelnen Algorithmen zu erwarten ist. Die Verschlechterung lag im Bereich zwischen 1% und maximal 7%. Diese relativ geringen Veränderungen spiegeln den geringen Einfluss der Kombinationen wieder. Der Variationskoeffizient des MAE für  $Q_{Vol}$  betrug 0,08 und für  $v_E$  0,02. Zum Vergleich: Der Variationskoeffizient des MAE zwischen den Untersuchungsgebieten lag bei 0,7 ( $Q_{Vol}$ ) und 1,0 ( $v_E$ ).

Eine der Schwächen der hier aufgeführten Analysen ist die Betrachtung der Ränge des MAE. Eine Fehlerkompensation der möglichen Prädiktorkombinationen bleibt unberücksichtigt. In der durchgeführten Analyse wurden Fälle in denen der MAE sehr nahe beieinander lag und Fälle in denen sehr große Unterschiede zwischen den MAEs bestanden gleichwertig behandelt. Wie zuvor erwähnt viel der Unterschied zwischen den MAEs der Kombinationen sehr

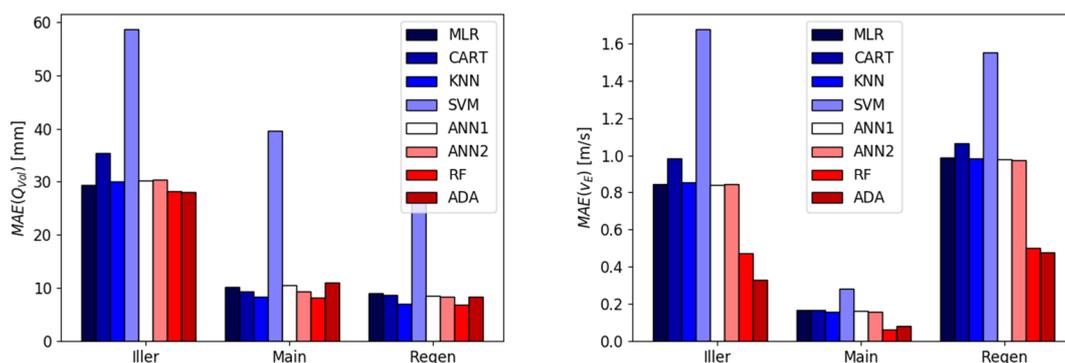
gering aus, daher wäre eine zusätzliche Gewichtung nach MAE-Abstand im Endeffekt auf eine Gleichgewichtung hinausgelaufen. Dennoch, bei zukünftigen Anwendungen ist dieser Aspekt bei der Auswahl der Prädiktorkombination zu berücksichtigen.

Aufgrund der hier dargestellten Untersuchungen wurde NQTS als Prädiktorkombination für die folgenden Analysen ausgewählt. Zum einen wies diese Kombination den geringsten mittleren Fehler auf und zum anderen ist die Steigerung des MAE gegenüber dem Aufwand der Verwendung individueller Kombinationen je Algorithmus vertretbar. Des Weiteren hat diese Kombination den Vorteil keine separate Wasserhaushaltssimulation zu benötigen.

## 6.4 Algorithmen

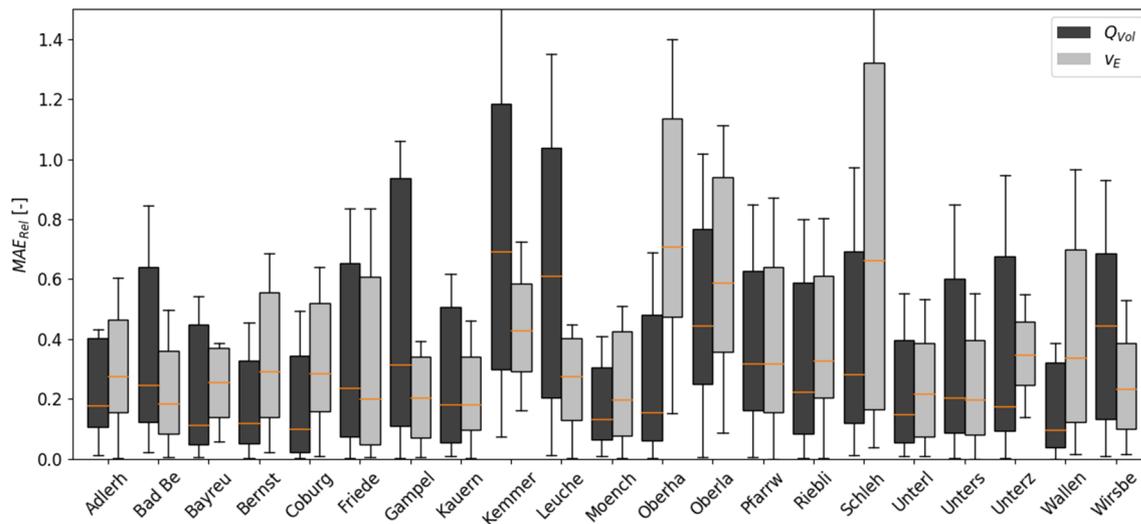
Welcher Fehler ist bei Anwendung der verschiedenen Algorithmen mit der ausgewählten Klassifikation und Prädiktorkombination zu erwarten? Gibt es einen einzelnen Algorithmus, der stets zuverlässige Prognosen erzeugt? In diesem Abschnitt soll auf diese Fragen eingegangen werden.

Abbildung 6-5 zeigt den MAE für die Simulation von analytischem  $Q_{Vol}$  (links) und  $v_E$  (rechts) getrennt nach verwendetem Algorithmus. Es zeigt sich in allen Untersuchungsgebieten eine ähnliche Rangfolge: Die Regression mit SVM erzeugt den höchsten Fehler. Gefolgt von den übrigen Einzelregressoren, die in etwa gleiche Fehler aufweisen. Dabei sticht CART in der Iller etwas heraus mit deutlich höherem MAE. Die Ensemble-Methoden ADA und RF weisen den geringsten Fehler auf. RF ist meist besser als ADA. Der Vorteil der Ensemble Methoden gegenüber den anderen Algorithmen war bei  $v_E$  deutlich größer als bei  $Q_{Vol}$ , teilweise mit einem 50% geringeren MAE als die Einzelregressoren.



**Abbildung 6-5: MAE je Untersuchungsgebiet und Algorithmus, links:  $Q_{Vol}$ , rechts:  $v_E$**

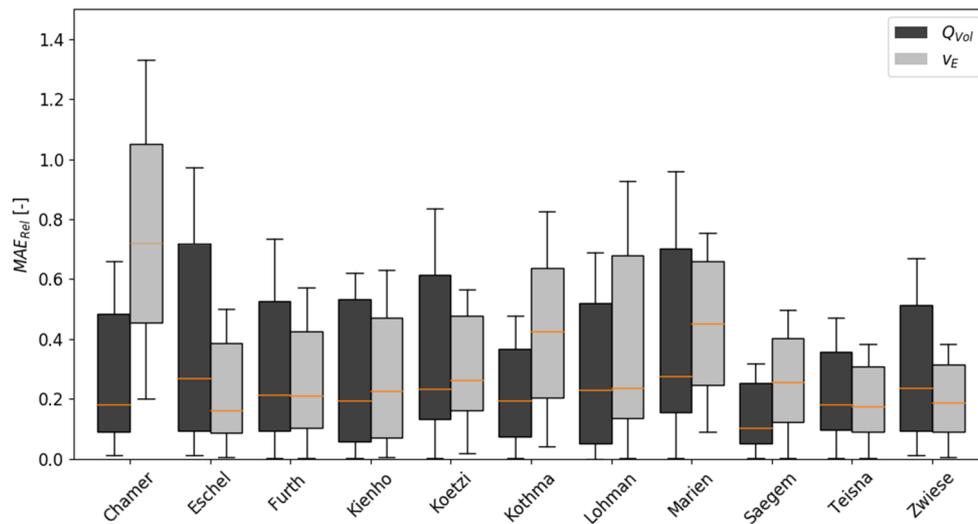
Zur genaueren Untersuchung der Regressionsfehler wurden die PUB-Regressionen der einzelnen Teileinzugsgebiete analysiert. Zunächst wurde die Vorhersage mittels RF betrachtet, da mit diesem Algorithmus, folgend Abbildung 6-5, die geringsten Fehler zu erwarten sind. Abbildung 6-6 zeigt den relativen MAE der Teilgebiete für die Vorhersage der analytischen  $Q_{Vol}$  und  $v_E$  Werte mittels RF.



**Abbildung 6-6: Relativer MAE  $Q_{Vol}$  und  $v_E$  je Teilgebiet am oberen Main berechnet mit RF**

Im Falle von  $Q_{Vol}$  liegt für 13 Gebiete der Median des relativen MAE unter 0.2, in 4 Gebieten in etwa bei 0.4 und 2 Gebiete weisen einen relativen Fehler größer gleich 0.6 auf. Einige Gebiete wie Gampelmühle weisen zudem eine große Streuung auf. Bei  $v_E$  erreichen 8 Gebiete einen Median unter 0.2, 10 liegen noch unter 0.4 und 3 darüber. Allerdings ist die Streuung hier geringer. Wie beim GIUH-Test ist der Fehler des Modells im oberen Main ungleich verteilt. Zwar gibt es einen systematischen Fehler (er wird nie null) aber die Höhe des Fehlers variiert zwischen den Teilgebieten. Gebiete wie Kemmern scheinen besonders schwierig für den RF zu sein, während andere Gebiete wie Bernstein a.W. oder Unterlangenstadt für beide Parameter relative MAE Werte unter 0.2 aufweisen. Zu erkennen ist zudem, dass der Fehler innerhalb einiger Teilgebiete zwischen den Parametern sehr stark variiert. Ein Beispiel hierfür ist Oberhammer: der Median für  $Q_{Vol}$  liegt unter 0.2 während für  $v_E$  der Median bei etwa 0.7 liegt.

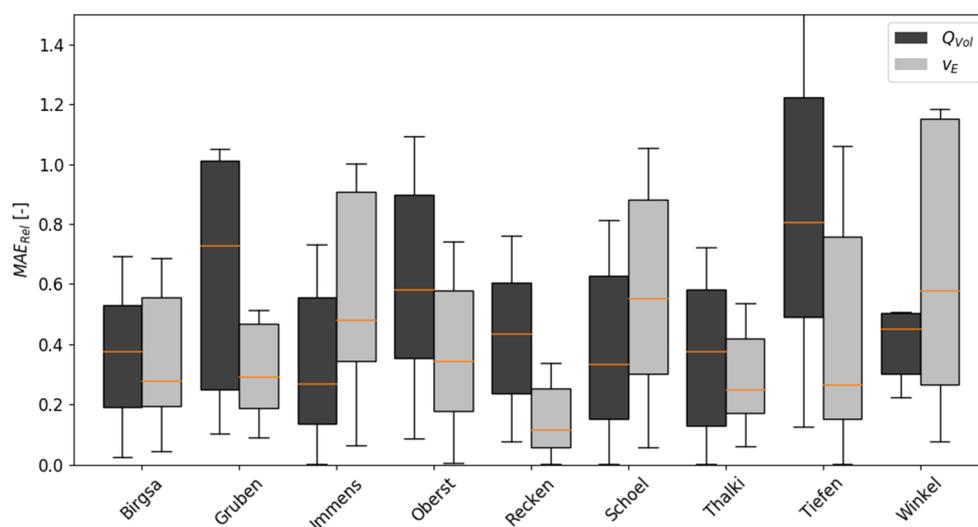
Im Untersuchungsgebiet Regen zeigt sich ein anderes Bild (Abbildung 6-7). Zwar sind auch hier die Ergebnisse mit einer deutlichen Streuung behaftet, doch liegen zumindest für  $Q_{Vol}$  fast alle Median Werte im Bereich von 0.2. Mit Ausnahme von Kothmaißling, Marienthal und Chamerau gilt dies auch für  $v_E$ . Der Fehler der Vorhersage ist in diesem Gebiet deutlich homogener verteilt. Dennoch können auch hier regionale Unterschiede festgestellt werden.



**Abbildung 6-7: Relativer MAE  $Q_{Vol}$  und  $v_E$  je Teilgebiet am Regen berechnet mit RF**

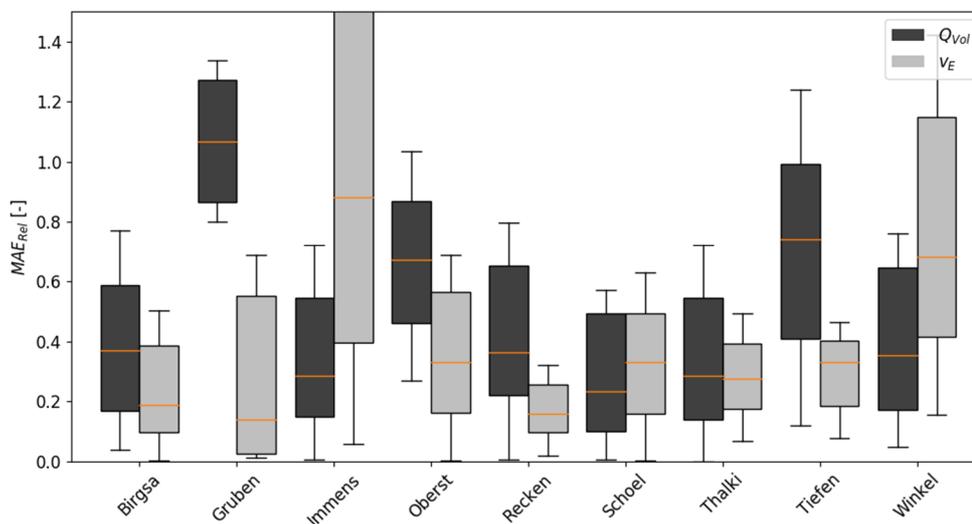
In der Iller liegen die Median-Werte des relativen MAE auf einem etwas höheren Niveau als im Regen, in etwa bei 0.4. Abbildung 6-8 zeigt die Teilgebietsauswertung der RF-Prognose in der Iller. Die größten Fehler bei der Regression von  $Q_{Vol}$  sind in Gruben und Tiefenbach zu erkennen, im Falle von  $v_E$  sind es Schöllang und Immenstadt. Zu beachten ist, dass in der Iller im Vergleich zum Regen eine stärkere Varianz des relativen MAE zu beobachten ist. Diese ist eher vergleichbar mit den schlechteren Ergebnissen einiger Teilgebiete des Mains.

In der Mittelung der MAE-Werte aller Teilgebiete (Abbildung 6-5) ist zu erkennen, dass in diesem Untersuchungsgebiet RF nicht der präferierte Algorithmus ist. ADA weist hier einen im Mittel geringeren Fehler auf. Daher wurde in Abbildung 6-9 die Teilgebietsauswertung für die Iller mit ADA erneut dargestellt.



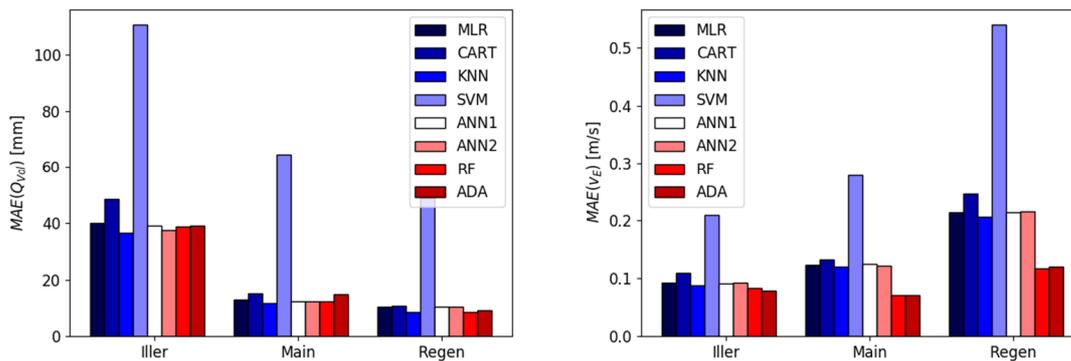
**Abbildung 6-8: Relativer MAE  $Q_{Vol}$  und  $v_E$  je Teilgebiet am Iller berechnet mit RF**

Das Boosting, also die fehlerorientierte Optimierung der Entscheidungsbaumstrukturen des ADA, sorgt dafür, dass die Fehlervarianz sinkt. Diese Beobachtung gilt für alle Untersuchungsgebiete, denn auch in Regen und Main sinkt bei Verwendung des ADA die Streuung der Fehler (Darstellung im digitalen Anhang). Im Falle der Iller führt dies gleichzeitig zu einer Reduktion des Median MAE, zumindest in einigen Teilgebieten. Dazu im Widerspruch stehen die Ergebnisse im Teilgebiet Immenstadt. Während mit dem RF der MAE zwar hoch, aber noch in einem akzeptablen Rahmen lag, steigt er auf über 1.0 im Median mit signifikant erhöhter Streuung. In diesem Falle bewirkte das Boosting eine Überanpassung an die Trainingsdaten, die nicht zu den Entwässerungsgeschwindigkeiten des Gebiets Immenstadt passen. Diese Tatsache ist besonders interessant, da der Pegel Immenstadt im Ablauf eines Sees positioniert ist. Im Untersuchungsgebiet gibt es keinen weiteren Pegel mit dieser Besonderheit.



**Abbildung 6-9: Relativer MAE  $Q_{Vol}$  und  $v_E$  je Teilgebiet am Iller berechnet mit ADA**

Die bislang gezeigten Ergebnisse bezogen sich auf die analytischen Parameter. Der Vollständigkeit halber werden in Abbildung 6-10 die mittleren MAE-Werte je Untersuchungsgebiet für die kalibrierten Zielvariablen dargestellt.



**Abbildung 6-10: MAE je Untersuchungsgebiet und Algorithmus, kalibrierte Parameter**  
links:  $Q_{vol}$ , rechts:  $v_E$

Für  $Q_{vol}$  ist im Vergleich zu den analytischen Werten eine leichte Erhöhung des Fehlers in der Iller erkennbar, Main und Regen liegen auf demselben Niveau. Im Falle von  $v_E$  sind die mittleren MAE-Werte deutlich gesunken, vor allem in Iller und Regen. Dies ist zum einen auf den veränderten Wertebereich der Parameter zurückzuführen. Im Falle des Regens bleibt der relative MAE in etwa konstant. Zum anderen macht sich aber auch die Selektion bemerkbar. Der relative MAE sinkt hier auf 0.25. Auf eine weitere detaillierte Darstellung der einzelnen Algorithmen und relativen Fehler wird an dieser Stelle verzichtet, da sich die Ergebnisse nicht signifikant von denen der analytischen Parameter unterscheiden. Festzuhalten bleibt aber, dass auch die kalibrierten Parameter mit den verwendeten Algorithmen vorhergesagt werden können. Die Güte der Vorhersage ist vergleichbar zur Vorhersage der analytischen Parameter.

Abschließend kann festgestellt werden, dass die Ensemble-Methoden RF und ADA im Mittel bessere Ergebnisse lieferten als die Einzelschätzer-Methoden. Boosting, wie es beim ADA vorgenommen wird, führte in den meisten Teilgebieten zu einer Überanpassung und daher zu Einbußen in der Validierungsgüte. Dennoch weist ADA eine geringere Streuung in seiner Vorhersage auf.

Von einer Festlegung auf einen einzelnen Algorithmus als Grundlage für das Dynamikmodul wird abgesehen, da keiner der verwendeten Algorithmen gleichbleibende Ergebnisse erzielen konnte. Je nach Teilgebiet und Ereignis variiert die Leistung der Algorithmen. Zu beachten ist, dass auch die Einzelschätzer-Methoden, wie bspw. ANN2, teilweise sehr gute Ergebnisse erzielen konnten. Hinzu kommt, dass Regressionen mit RF und ADA stets in Richtung der Mittelwerte des Trainingsdatensatzes tendieren werden. Daher bieten Einzelschätzer die Möglichkeit davon abweichende Ereignisse erfassen zu können.

Um die Übereinstimmung, bzw. die Divergenz der Vorhersagen zu beurteilen, wurden lineare Korrelationskoeffizienten zwischen den MAE-Werten der Regressionen mit verschiedenen Algorithmen berechnet. Tabelle 6-4 listet die Koeffizienten für die Anwendung in der Iller für  $Q_{vol}$  auf. Die dargestellte Tabelle ist beispielhaft für alle Anwendungen, die kaum von den hier gezeigten Werten abweichen (vollständige Darstellung im digitalen Anhang).

Zunächst wird eine grundlegende Vermutung bestätigt, dass die beiden neuronalen Netze übereinstimmende Aussagen treffen. Die ANNs sind außerdem mit den Ergebnissen des KNN, MLR und CART korreliert. Die Koeffizienten liegen für  $v_E$  bei etwa 0.8, für  $Q_{Vol}$  sind die Werte geringer und etwas differenzierter, im Bereich zwischen 0.4 und 0.8. SVM-Regressionen divergieren am stärksten von den anderen Algorithmen. Für  $Q_{Vol}$  sind nur sehr schwache Korrelationen zu RF und ADA ersichtlich, bei  $v_E$  steigt die Korrelation auf 0.4 – 0.8 zu MLR, KNN, ANN und CART. Ebenfalls zu erwarten war die starke Korrelation zwischen ADA und RF, zu allen anderen Algorithmen weisen sie eine Korrelation zwischen 0.4 - 0.8 auf.

**Tabelle 6-4: Lineare Korrelationskoeffizienten der MAE( $Q_{Vol}$ )-Werte verschiedener Algorithmen im Untersuchungsgebiet Iller**

	MLR	CART	SVM	KNN	ANN1	ANN2	RF	ADA
MLR	1.00	0.48	0.22	0.77	0.90	0.84	0.72	0.76
CART	0.48	1.00	0.15	0.44	0.45	0.45	0.54	0.51
SVM	0.22	0.15	1.00	0.27	0.18	0.14	0.39	0.43
KNN	0.77	0.44	0.27	1.00	0.80	0.70	0.75	0.81
ANN1	0.90	0.45	0.18	0.80	1.00	0.82	0.69	0.78
ANN2	0.84	0.45	0.14	0.70	0.82	1.00	0.66	0.74
RF	0.72	0.54	0.39	0.75	0.69	0.66	1.00	0.83
ADA	0.76	0.51	0.43	0.81	0.78	0.74	0.83	1.00

Diese Analyse bestätigt, dass die Einzelschätzer eine sinnvolle Ergänzung zu den Ensemble-Methoden darstellen. Ihre Schätzungen sind zwar untereinander korreliert, weichen aber von denen von RF und ADA ab. Dies ist von Vorteil für Ereignisse, die nicht mit RF und ADA abgebildet werden können, da mit den Einzelschätzern weitere, strukturell verschiedene, Methoden zur Verfügung stehen und somit die Wahrscheinlichkeit erhöhen die richtige Prognose zu erstellen.

## 6.5 Zusammenfassung Dynamikmodul

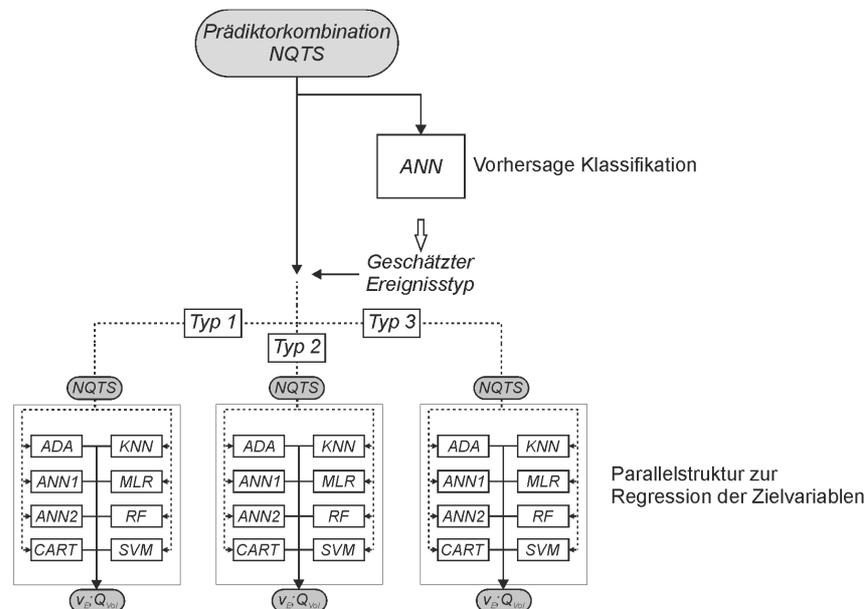
In diesem Kapitel wurde das Dynamikmodul des entwickelten gekoppelten GIUH-ML Modells untersucht. Die Aufgabe des Moduls ist die ereignisweise Parametrisierung der GIUH-Komponente.

Durch eine modellgestützte Auswahl wurde die Kombination NQTS als beste Prädiktorkombination ausgewählt. Alle Vorhersagen des Modells werden in den Fallstudien basierend auf Kennwerten des Niederschlags (Dauer DP, Volumen VP, zeitliche Verteilung  $D_{50}$ , mittlere und maximale Intensität  $I_0$  &  $I_{Max}$ , Vorregenindex VRI), dem initialen Abfluss  $Q_0$ , der Temperatur vor Ereigniseintritt  $T_{Pre}$  sowie der Schneebedeckung SC und der Änderung der

Schneedecke vor Ereignisbeginn  $\Delta SC$ , produziert. Die Analyse zeigte aber, dass die Auswahl der Prädiktoren, im Vergleich zu anderen Einflussfaktoren der Regression, einen nur geringen Einfluss auf den zu erwartenden Fehler hat. Daher kann die vorgeschlagene Methodik auch bei schlechterer Datenlage, wenn bspw. nur Niederschlagsdaten vorliegen angewendet werden.

Abbildung 6-11 zeigt das Dynamikmodul, wie es auf Grundlage der in diesem Kapitel dargestellten Ergebnisse, in den Fallstudien angewendet wurde. Zunächst wird mit einem ANN-Klassifikator der Ereignistyp des zu erwartenden Ereignisses prognostiziert.

Der ANN Klassifikator zeigte in den durchgeführten Tests gleichmäßig gute Ergebnisse in allen Untersuchungsgebieten und für alle Klassifikationen. Für 70% – 90% aller Ereignisse konnte eine richtige Prognose des Ereignistyps gegeben werden. Die Auswirkungen von Fehlklassifikationen auf das gesamte Simulationsergebnis können nur im Kontext des gesamten Modells evaluiert werden. Daher muss auf die Fehlerdiskussion durch Fehlklassifikation in Kapitel 8.5 verwiesen werden.



**Abbildung 6-11: Darstellung des Dynamikmoduls mit ausgewählter Prädiktorkombination und prozessorientierter Differenzierung der Ereignistypen**

Die vom ANN prognostizierten Ereignistypen entsprechen der *tq\_regional* Klassifikation. Mit dieser Klassifikationsmethode wurden die Ereignisse basierend auf ihrem Füllen-Scheitel Verhältnis auf regionaler Ebene differenziert. Diese Klassifikation setzte sich in einer weiteren modellgestützten Auswahl gegenüber den anderen Klassifikationen durch. Zudem konnte gezeigt werden, dass die parallele Struktur des Dynamikmoduls sich positiv auf den Vorhersagefehler auswirkte.

Die parallele Struktur sieht drei separate Blöcke zur Regression der Zielvariablen  $Q_{Vol}$  und  $V_E$  vor. Abhängig vom geschätzten Ereignistyp wird der entsprechende Vorhersageblock verwendet. Innerhalb der Blöcke befinden sich acht, individuell angepasste Algorithmen. Die Ensembleschätzer RF und ADA zeigten in den durchgeführten Analysen den geringsten

Regressionsfehler. Etwas höher lag der zu erwartende Fehler für die Einzelschätzer ANN1, ANN2, KNN und MLR. Etwas schlechtere, aber dennoch vergleichbare, Ergebnisse wurden mit dem CART-Algorithmus erreicht. Die Fehler des SVM-Regressors fielen am schlechtesten aus. Der MAE war in allen Anwendungen etwa doppelt so als die der anderen Einzelregressoren.

Dennoch konnte kein Algorithmus als eindeutig besser gegenüber den anderen identifiziert werden. Auch wenn RF und ADA im Mittel die besten Ergebnisse erzielten, lieferten für einige Ereignisse andere Algorithmen bessere Ergebnisse. Diese Ergebnisse decken sich mit den Erfahrungen aus anderen Anwendungen in der Literatur (bspw. Elshorbagy et al., 2010a) und sind dem *No-free-lunch* Theorem (Wolpert & Macready, 1997) geschuldet. Das Theorem besagt, dass es keine einheitliche Lösung für alle Probleme geben kann. Ein Algorithmus, bzw. eine Methode, der ein bestimmtes Problem zufriedenstellend lösen kann, wird ähnliche Probleme ebenfalls lösen können, diese aber mit geringerer Güte. Daher wurde das Dynamikmodul als Ensemble-Schätzer konzipiert, alle hier dargestellten Algorithmen werden verwendet, um  $Q_{Vol}$  und  $v_E$  zu schätzen. Die berechneten Ganglinien können dann gewichtet nach der Güte der Algorithmen (berechnet am Trainingsdatensatz) einen Unsicherheitsbereich der Ganglinie Vorhersage abbilden (Beven & Binley, 1992).

## 7 Datenkomposition

---

Die Eignung eines datengetriebenen Modells hydrologische Prozesse innerhalb eines Einzugsgebietes wiederzugeben, hängt direkt von den zur Verfügung gestellten Trainingsdaten ab. Sind die Daten unzureichend, oder schlicht unzutreffend, kann der dem Modell zugrunde liegende Algorithmus keine adäquate Beziehung zwischen den Prädiktoren und den Zielvariablen herstellen. Eine Regionalisierung kann daher nur erfolgreich sein, wenn im unbeobachteten Zielgebiet dieselben hydrologischen Prozesse dominant sind wie in den beobachteten Gebieten, aus denen die Trainingsdaten stammen. Eine der grundlegenden Annahmen zur Übertragung von Parametern aus beobachteten zu unbeobachteten Gebieten ist, dass in ähnlichen Gebieten ähnliche Prozesse relevant sind (Wagener, Wheater & Gupta, 2004; Wagener et al., 2007; Sawicz et al., 2011; He, Bárdossy & Zehe, 2011a). Zu definieren ist in diesem Zusammenhang, wie die Ähnlichkeit natürlicher Einzugsgebiete definiert werden kann. In der Fachliteratur wird diese Thematik häufig im Zusammenhang mit Gebietsklassifikation diskutiert, stets mit dem Ziel, dass für ähnliche Gebiete ähnliche hydrologische Gesetzmäßigkeiten angenommen werden können. Dabei dominieren zwei grundsätzliche Herangehensweisen (He, Bárdossy & Zehe, 2011b): Zum einen eine distanzbasierte Definition von Ähnlichkeit und zum anderen die sog. hydrologische Ähnlichkeit. Einer der bekanntesten Ansätze zur Regionalisierung, basierend auf distanzbasierter Ähnlichkeit, ist das Top-Kriging Verfahren (Skøien, Merz & Blöschl, 2006). Wesentlich mehr Aufmerksamkeit wurde der hydrologischen Ähnlichkeit basierend auf der Idee von funktionalen Einheiten (Sawicz et al., 2011) gewidmet. Grundsätzlich wird bei diesem Ansatz versucht durch die Ähnlichkeit einzelner Indikatoren der Klimatologie, des Abflussregimes, Kennwerten der EZG-Oberfläche oder Landnutzung etc. auf EZG-Gruppen zu schließen, die ein einheitliches Abflussverhalten aufweisen. Eingeführt von Winter (2001) wurde dieser Ansatz mit einer Vielzahl von Kennwerten, unterschiedlichen Distanz-Metriken und in den verschiedensten Regionen getestet und weiterentwickelt (u.a. Dunn & Lilly, 2001; Soulsby et al., 2006; Müller et al., 2009; Nobre et al., 2011; Gharari et al., 2011; Casper, Gronz & Gemmar, 2015; Rodriguez et al., 2016). Die Vielzahl der Anwendungen deutet bereits auf das Fehlen einer einheitlichen Metrik, bzw. Definition von hydrologischer Ähnlichkeit hin (Wagener et al., 2007).

Die fehlende Allgemeingültigkeit der in der Literatur vorgestellten Methoden macht eine von der Modellstruktur und vom Anwendungsfall abhängige Auswahl der Methoden notwendig. Für die von Casper, Gronz & Gemmar (2015) sowie Dunn & Lilly (2001) verwendeten Methoden müssten bestehende Klassifikationen zur Verfügung stehen, die es für die

betrachteten Untersuchungsgebiete nicht gab. Zudem müsste ein flächendifferenziertes Modell verwendet werden. Ähnliches gilt für die HAND-Methodik von Gharari et al. (2011), die einen Zusammenhang zwischen der Höhe einer Fläche über dem entwässernden Gewässer in Beziehung zu Prozesses setzt. Diese Klassifikation ist maßgeschneidert für das Flex-Topo Modell (Savenije, 2010) und daher kaum auf andere Modellstrukturen übertragbar. Distanz-basierte Ähnlichkeit ist das einzige universell anwendbare Konzept, dass unabhängig von Modellstrukturen und Datenlage angewendet werden kann. Die Nachteile sind aber offensichtlich: unmittelbar benachbarte Gebiete sind nicht per se ähnlich. Im Umkehrschluss sind aber auch weit voneinander entfernte Gebiete nicht per se unähnlich.

Zusammenfassend muss das Konzept hydrologisch ähnlicher Gebiete also an die Prozessannahmen des verwendeten Modells angepasst werden. Daher wurde in diesem Abschnitt eine Ähnlichkeitsmetrik, basierend auf der DFF-Funktion zweier Gebietskennwerte entwickelt und getestet. Zum Vergleich wurden zwei Distanz-basierte Metriken verwendet. Diese Metriken wurden verwendet um eine optimale Lernmenge zu definieren.

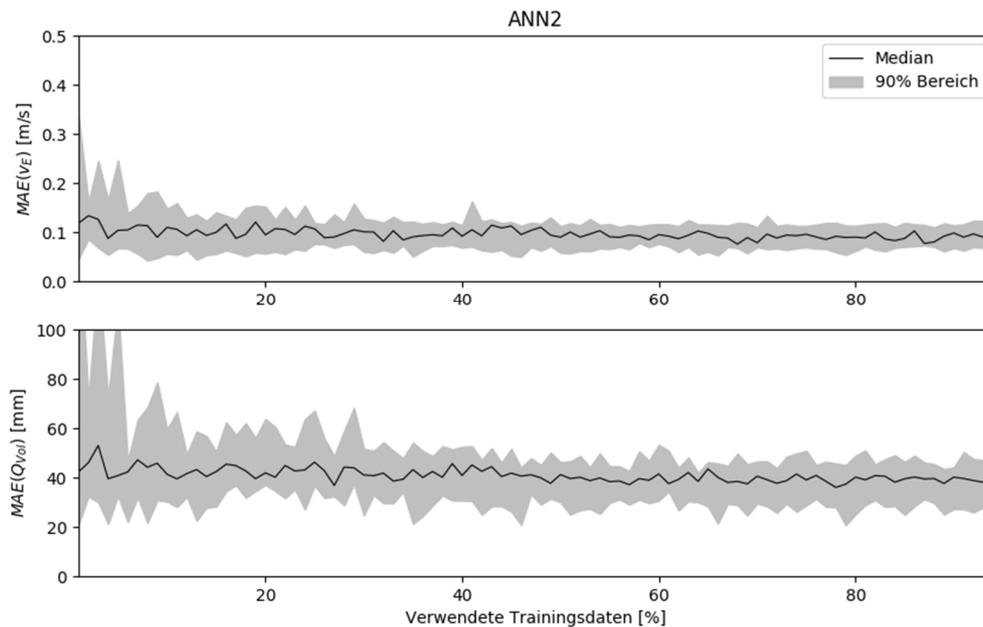
Als Ausgangspunkt der Analysen in diesem Kapitel wurde eine Lernstudie mit dem Dynamikmodul durchgeführt. Ermittelt wurde ob durch die Verwendung aller zur Verfügung stehenden Daten eine Über- bzw. Fehlanpassung erzeugt wurde. Eine optimale Lernmenge würde nur Daten in die Anpassung der Modelle einbeziehen, die eine Verbesserung der Anpassung erzielt. Daten von unähnlichen Gebieten, die eine Fehlanpassung des Modells erzeugen müssen von einer optimalen Lernmenge ausgeschlossen werden. Darüber hinaus minimiert sie den Aufwand der Datenaufbereitung.

Hilfe der Ähnlichkeits-Metriken wurden Reihenfolgen für jedes Teilgebiet ermittelt, aus denen unterschiedliche Kompositionen der Trainingsdaten abgeleitet wurden. Abschließend wurden die verschiedenen Rankings im Hinblick auf die daraus resultierenden Lernkurven evaluiert und eine Auswahl getroffen.

### **7.1 Einfluss der Lernmenge auf Regionalisierung**

Um den Einfluss des Umfangs des Trainingsdatensatzes auf die Validierungsgüte der Algorithmen zu überprüfen, wurde eine Lernstudie durchgeführt. Basierend auf dem Vorgehen der in Kapitel 6 beschriebenen PUB-Regression wurde wie folgt vorgegangen: Ein Teileinzugsgebiet wurde als „unbeobachtet“ ausgewählt und die Daten dieses Gebietes zur Validierung dem Datenpool entzogen. Aus den verbleibenden Daten wurden anschließend per Zufall die Trainingsdaten zur Anpassung der Modelle ausgewählt. Beginnend bei 5% der zur Verfügung stehenden Daten wurde anschließend der Anteil der genutzten Daten schrittweise bis zur vollständigen Nutzung aller Ereignisse erhöht. Die ausgewählten Daten wurden dem Trainingsdatensatz hinzugefügt und die Modelle angepasst. Nach jeder Anpassung wurde das Modell validiert.

In Abbildung 7-1 sind als beispielhaftes Ergebnis die Validierungsgüte der Teileinzugsgebiete im Untersuchungsgebiet der Iller mit einem 2-lagigen künstlichem neuronalem Netz (ANN2) dargestellt. Dargestellt ist der Bereich in den 90% aller MAE-Werte in der Lernstudie im Untersuchungsgebiet Iller gefallen sind. Zusätzlich ist der Verlauf der Median-MAE dargestellt. Im digitalen Anhang können die individuellen Verläufe der Teilgebiete entnommen werden.



**Abbildung 7-1: Einfluss der Lernmenge auf MAE( $v_E$ ) (oben) und MAE( $Q_{Vol}$ ) (unten) für ANN2 im Untersuchungsgebiet Iller**

Die Varianz des MAE beider Zielvariablen nahm in dieser Anwendung mit zunehmender Lernmenge ab. Der Lernfortschritt für  $v_E$  war dabei schneller als für  $Q_{Vol}$ . Unsicherheit ist in diesem Fall dargestellt durch die Varianz zwischen den Teilgebieten eines Untersuchungsgebiets. Die Inter-Ereignisvarianz verhielt sich aber analog zu den hier dargestellten Beobachtungen.

Ergebnisse aller Algorithmen und Untersuchungsgebiete können dem digitalen Anhang entnommen werden. Die Ergebnisse des ANN2 stehen in Abbildung 7-1 repräsentativ für alle anderen Anwendungen. Lediglich Unterschiede im Rahmen der MAE-Varianz bei sehr geringem Umfang des Trainingsdatensatzes sind in den meisten Fällen sichtbar. Die generelle Tendenz ist aber in allen Gebieten und für alle Algorithmen identisch. Eine länger anhaltende Lernkurve, also ein stetiges Sinken des Median MAE und eine weitere Reduktion der Streuung, ist in den Ergebnissen des adaptiven Boostings (ADA) sichtbar. Dieses Verhalten ist auf die steigende Datenmenge zur gezielten Optimierung des Vorhersageensembles zurückzuführen. Die multiple lineare Regression (MLR) wies zwar dieselbe Tendenz wie das ANN2 auf, auf eine sichtliche Instabilität muss aber an dieser Stelle hingewiesen werden.

Punktuell steigt der MAE sprunghaft an und sinkt nach zufügen weiterer Daten auf ein ursprüngliches Maß zurück. Dies zeigt die Anfälligkeit der MLR gegenüber Ausreißern in den Lerndaten. Einzig die *Support Vector Machine* (SVM) weist einen komplett gegensätzlichen Trend auf. Mit steigendem Umfang steigt für beide Zielvariablen der MAE.

Im Falle der SVM deutet dies auf eine klare Überanpassung mit steigendem Umfang hin. Für diesen Algorithmus ist eine Reduktion der Daten also zwingend erforderlich. Für die anderen Algorithmen ist die wichtigste Beobachtung, dass ab Erreichen eines je nach Algorithmus und Zielgröße variierenden Schwellenwertes kein Lernfortschritt mehr erzielt wurde. Im dargestellten Beispiel (Abbildung 7-1) lagen die Schwellenwerte bei 20% ( $v_E$ ) und 40% ( $Q_{Vol}$ ). Diese Tatsache deutet darauf hin, dass sich positive, wie negative Aspekte des wachsenden Datensatzes in diesem Falle ausgleichen. Eine optimale Lernkurve würde zunächst eine Reduktion des Fehlers zeigen und ab Erreichen des Schwellenwertes wieder eine Steigerung, also eine U-Form aufweisen. Dem Modell würden in diesem Falle die Daten geordnet nach ihrem Informationsgehalt zugeführt werden. Die Ergebnisse zeigen, dass tendenziell zu viele Daten zur Anpassung der Algorithmen zur Verfügung stehen, bzw. eine bessere Komposition der Lernmenge erforderlich ist.

## 7.2 Metriken zur Schätzung von Ähnlichkeit

Für jedes Zielgebiet wurden Rankings aller verbleibenden Teilgebiete erstellt, basierend auf vier Metriken. Je zwei Metriken basierten dabei auf einer distanz-basierten Metrik sowie einer Metrik der hydrologischen Ähnlichkeit.

### 7.2.1 Centroid

Der geometrische Schwerpunkt eines Einzugsgebietes, oder auch Centroid, wurde als Mittelwert aller Koordinaten eines EZG berechnet. Die Koordinaten des Schwerpunktes  $c$  wurden für ein EZG  $J$  wie folgt berechnet:

$$c_J = \frac{1}{A_J} \int x dA_J \quad 7.1$$

Die Fläche des EZG wird in Gl. 7.1 durch  $A_J$  repräsentiert,  $x$  repräsentiert die Koordinatenpaare. Die Distanz zwischen den Schwerpunkten zweier Teilgebiete  $J$  und  $K$  wurde als euklidische Distanz  $d_{JK}$  berechnet:

$$d_{JK} = |c_J - c_K| \quad 7.2$$

Mittels der berechneten Distanzen wurde für jedes EZG ein Ranking erstellt. Dem EZG mit dem geringsten Abstand zum Zielgebiet wurde die Ordnungszahl 1 zugewiesen, dem höchsten Abstand die höchste Ordnungszahl.

### 7.2.2 Top-Kriging

Speziell für die Regionalisierung von abflussbezogenen Größen wurde das Topologische-Kriging von Skøien, Merz & Blöschl (2006) entwickelt. Zur geostatistischen Interpolation einer beliebigen Größe werden dabei Interpolationsgewichte mit Hilfe eines Semi-Variogramms abgeleitet.

Das Semi-Variogramm beschreibt die Abhängigkeit zwischen euklidischer Distanz der bekannten Punkte und der Semivarianz der beobachteten Werte. Beim Top-Kriging sind dies der Centroid-Abstand und die Semivarianz der geographischen Positionen der betrachteten Gebiete. Somit soll nach Skøien, Merz & Blöschl (2006) neben der Distanz zwischen den Gebieten die Verschachtelung von Teilgebieten zu berücksichtigen werden. Dahinter steckt die Annahme, dass eine Unter- / Oberlieger-Beziehung zwischen zwei Teilgebieten einen starken Zusammenhang bezogen auf die Zielgröße impliziert. Die Semivarianz zwischen zwei Gebieten  $\gamma$  wird nach folgender Gleichung berechnet:

$$\gamma_{1;2} = \frac{1}{A_1 A_2} \int_{A_1} \int_{A_2} \gamma_p(|x_1 - x_2|) dx_1 dx_2 - 0.5 \cdot \left[ \frac{1}{A_1^2} \int_{A_1} \int_{A_1} \gamma_p(|x_1 - x_2|) dx_1 dx_2 + \frac{1}{A_2^2} \int_{A_2} \int_{A_2} \gamma_p(|x_1 - x_2|) dx_1 dx_2 + \right] \quad 7.3$$

In Gleichung 7.3 stehen die Indizes 1 und 2 für zwei betrachtete Gebiete. Neben den Koordinatenvektoren  $x$  kann die Semivarianz einer weiteren Zielgröße  $\gamma_p$  bei der Berechnung der Top-Kriging Semivarianz berücksichtigt werden. Da neben den Topologischen Gegebenheiten kein weiterer Aspekt berücksichtigt werden sollte wurde  $\gamma_p = 1$  gewählt. Der erste Term in Gl. 7.3 beschreibt die Varianz zwischen den Gebieten, berechnet über die Differenz aller Punkte der jeweiligen EZGs. Von diesem Wert abgezogen wird die mittlere Varianz innerhalb der jeweiligen EZGs. Um diese innere Varianz zu berechnen, wird die Differenz zwischen allen Koordinaten innerhalb des EZG integriert.

Um den Zusammenhang zwischen Distanz  $d$  (Gl. 7.2) und der empirischen Semivarianz  $\gamma$  (Gl. 7.3) zu beschreiben, wird mittels Fehlerquadratminimierung ein theoretisches Variogramm angepasst. Das verwendete exponentielle Variogramm verfügt über drei Parameter (a, b und c):

$$\gamma_i = a + b \cdot \left( 1 - e^{-\frac{d}{c}} \right) \quad 7.4$$

Parameter a beschreibt den *Nugget*-Effekt, also Abweichungen zwischen Beobachtungen mit 0-Distanz, c ist der *Range* Parameter und beschreibt die Einflussbereich einer Messung bei der Interpolation und b ist der Schwellenwert (auch *Sill*). Ab einer Entfernung zwischen zwei Punkten  $d$ , die größer als c ist, nimmt das Variogramm den Schwellenwert c an.

Ausgehend von der theoretischen Semivarianz (Gl. 7.4) kann folgende Gleichung aufgestellt und gelöst werden, um die Kriging-Gewichte  $\lambda$  zu schätzen:

$$\sum_{i=1}^n \lambda_i \cdot \gamma_t(d_{i,j}) + \mu = \gamma_t(d_{i,z}) \quad i=1,2,\dots,n \quad 7.5$$

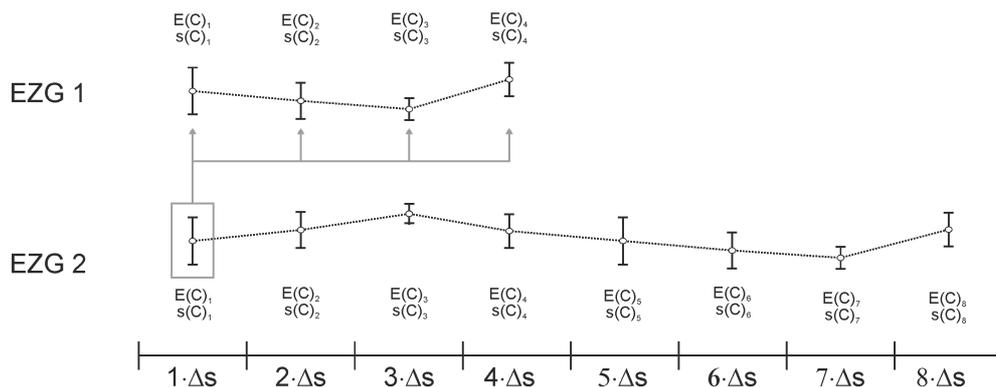
Dabei ist  $\mu$  der Lagrange Parameter,  $\gamma_t(d_{i,j})$  steht für die Anwendung des theoretischen Variogramms für die Distanz zwischen allen bekannten Datenpunkte, also Gebieten die zum Datenpool der Lernmenge gehören.  $\gamma_t(d_{i,z})$  steht für die theoretische Semivarianz der Distanz zwischen Zielgebiet und den bekannten Datenpunkten. Nach Lösung von Gleichung 7.5 steht ein Vektor  $\lambda$  zur Verfügung, der die Gewichte jedes Gebietes im Datenpool der Lernmenge beinhaltet.

Die erhaltenen Kriging Gewichte  $\lambda_i$  wurden im Anschluss zum Ranking verwendet. Dem höchsten  $\lambda_i$  wurde die Ordnungszahl 1 und dem geringsten  $\lambda_i$  die höchste Ordnungszahl zugewiesen.

### 7.2.3 DFF-Ranking

Eine Metrik zur Schätzung hydrologischer Ähnlichkeit erfordert eine a-priori Definition von Prozessen und mit welchen Kennwerten diese verknüpft werden können. Das in dieser Arbeit verwendete Modell basiert auf der geomorphologischen Organisation des Einzugsgebietes. Es lag daher nahe eine Metrik basierend auf dieser grundlegenden Annahme zu entwickeln. Den Darstellungen aus Kapitel 4.2 folgend wurde der Ansatz der Distanz-Faktor-Funktion (DFF) als Grundlage zur Erfassung der geomorphologischen Struktur verwendet.

Abbildung 7-2 zeigt schematisch die DFF zweier hypothetischer Einzugsgebiete. *EZG 1* ist in diesem Falle der Oberlieger von *EZG 2*. Dargestellt sind die Erwartungswerte und Standardabweichungen eines Kennwertes  $C$  in den jeweiligen Distanzklassen, also geordnet nach ihrer Distanz zum Gebietsauslass. Angedeutet ist, dass die Kennwerte der 4 Distanzklassen von *EZG 1* den Werten der Distanzklassen 5 – 8 aus *EZG 2* entsprechen.



**Abbildung 7-2: Schematische Berechnung Semivarianz DFF-Ranking**

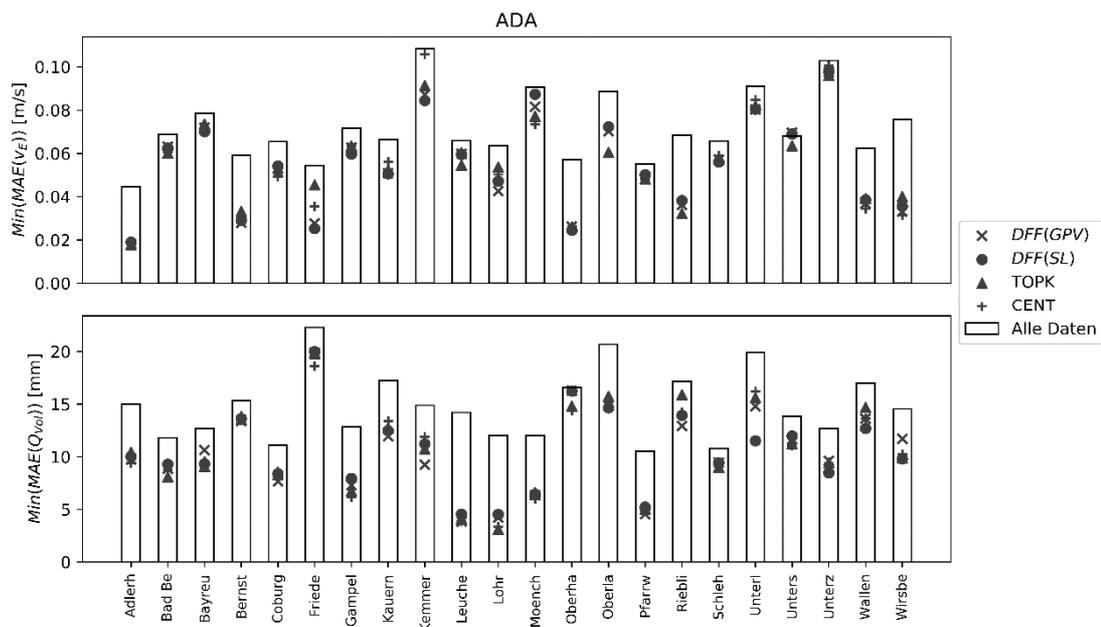
Um die Ähnlichkeit der DFFs zu bestimmen, wurde die Semivarianz zwischen den DFFs gemäß Gl. 7.3 berechnet. In diesem Fall stand  $x$  für einen Vektor mit drei Ordinaten: Distanzklasse,  $E(C)$  und  $s(C)$ . Demnach berechnet sich die Semivarianz über die Varianz der geomorphologischen Struktur zwischen und innerhalb der Gebiete.

Der Vorteil dieses Vorgehens besteht in der Elimination der Distanz aus dem Ähnlichkeitsmaß. Im Gegensatz zum Top-Kriging wurden beim DFF-Ranking keine Kriging Gewichte berechnet, da kein Bezug zwischen Semi-Varianz und Distanz in das Ranking einfließen sollte. Daher wurde die empirische Semivarianz  $\gamma$  zwischen dem Zielgebiet und allen verbleibenden EZGs berechnet und basierend auf diesen Werten das Ranking durchgeführt.

Dem Einzugsgebiet mit der geringsten absoluten  $\gamma$  wurde die Ordnungszahl 1 zugewiesen, dem höchsten  $\gamma$  die höchste Ordnungszahl. DFF-Ranking wurde mit zwei Unterschiedlichen Kennwerten  $C$  durchgeführt: mit dem Gesamtporenvolumen (GPV) und dem Gefälle (SL).

### 7.3 Selektive Datenkomposition

Die erstellten Rankings wurden im Rahmen einer PUB-Regression angewendet. Dabei wurden dem Trainingsdatensatz sukzessive, dem jeweiligen Ranking folgend, die einzelnen Datensätze hinzugefügt. Berechnet wurde zunächst der MAE für beide Zielvariablen über alle Ereignisse innerhalb eines Teilgebietes je ML-Algorithmus und verwendeter Metrik.



**Abbildung 7-3: Vergleich MAE( $v_E$ ) (oben) und ( $Q_{Vol}$ ) (unten) bei Verwendung aller Daten (Balken) und optimaler Datenkomposition (Markierungen) mit ADA im Main**

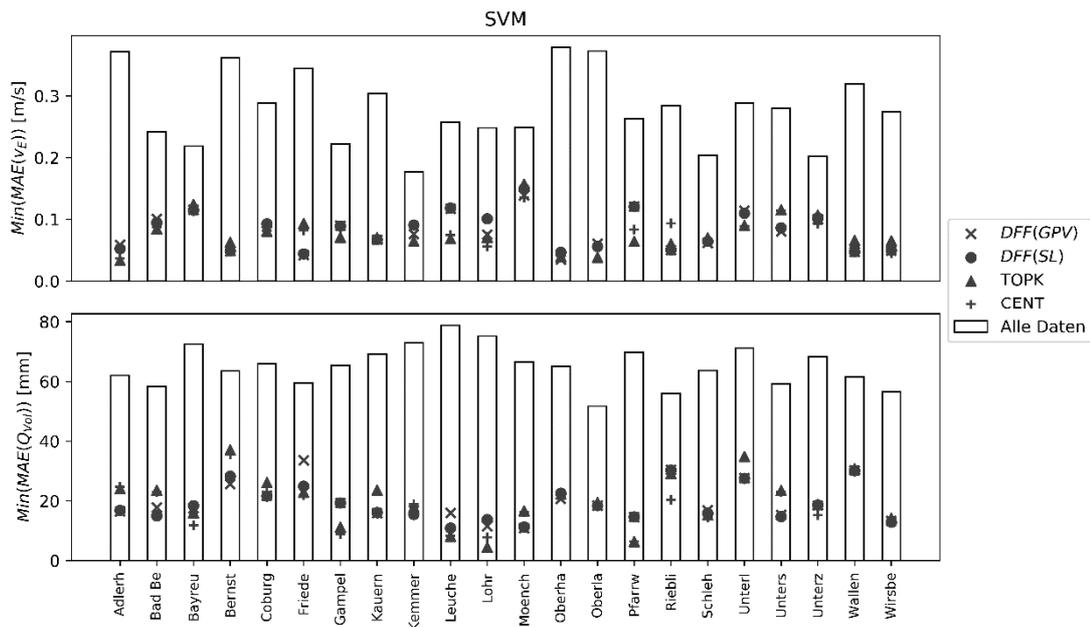
Abbildung 7-3 zeigt den mit ADA erreichten MAE für  $v_E$  und  $Q_{Vol}$  bei Verwendung aller Daten im Vergleich zum minimal erreichbaren MAE bei Verwendung der optimalen Datenkompositionen im Untersuchungsgebiet Main. (Alle weiteren Anwendungen sind dem digitalen Anhang zu entnehmen).

Einige EZGs wie Adlerhütte (kurz: Adlerh) und Oberlauter (kurz: Oberla) profitierten für beide Zielvariablen. Der  $MAE_{(v_E)}$  in Adlerhütte wurde halbiert,  $Q_{Vol}$  um  $1/3$  des ursprünglichen MAE reduziert, in Oberlauter wurde der MAE beider Variablen jeweils um  $1/4$  reduziert. Zwar profitieren beide EZGs in unterschiedlichem Maße, dennoch ist die Reduktion beider MAE relativ hoch. Dies ist aber eher die Ausnahme, bei den meisten EZGs profitiert eine Variable deutlich stärker als die andere: bspw. Leucherhof (kurz: Leuche) profitiert stark bei  $Q_{Vol}$  aber kaum bei  $v_E$ . Am deutlichsten zu erkennen ist, dass keiner Metrik eine eindeutige Präferenz zukommt. Des Weiteren ist auffällig, dass in den meisten Fällen ein ähnlicher minimaler MAE erzielt wurde.

Das hier dargestellte und beschriebene Bild wiederholt sich in den meisten anderen Untersuchungsgebieten und für die anderen Algorithmen. Einzig die Höhe der Reduktion sowie deren Verteilung variiert. Allerdings ist zu beobachten, dass der MAE lokal für alle Algorithmen auf ein vergleichbares Maß reduziert werden konnte. Während bei Verwendung aller Daten der MAE in bspw. Adlerhütte noch stark mit dem verwendeten Algorithmus variiert, liegt er mit selektiver Datenkomposition bei  $7,5 - 10$  mm. Abhängig vom Ausgangsfehler der Algorithmen variiert dann die Reduktion.

Einzig die SVM zeigt in allen Gebieten eine starke Reduktion der MAE beider Variablen. In Abbildung 7-4 ist die Auswertung der SVM im Main dargestellt. Der Fehler der Vorhersage wird in allen Teilgebieten auf etwa  $1/3$  des ursprünglichen Fehlers reduziert. Dennoch verbleibt der Fehler auf einem höheren Niveau als der Durchschnitt der anderen Algorithmen. Ein weiterer, hier nicht dargestellter, Unterschied zwischen SVM und den anderen Algorithmen besteht in der Anzahl verwendeter Gebiete. Die SVM verwendet durchweg 1-2 Datensätze, um einen minimalen MAE zu erreichen, während die Anzahl aller weiteren Algorithmen zwischen 2 und allen Datensätzen variiert.

Die stark variierende Anzahl von Datensätzen warf die Frage auf, ob eine einheitliche Anzahl von Datensätzen definiert werden kann und ob eine Auswahl der besten Metrik möglich ist.

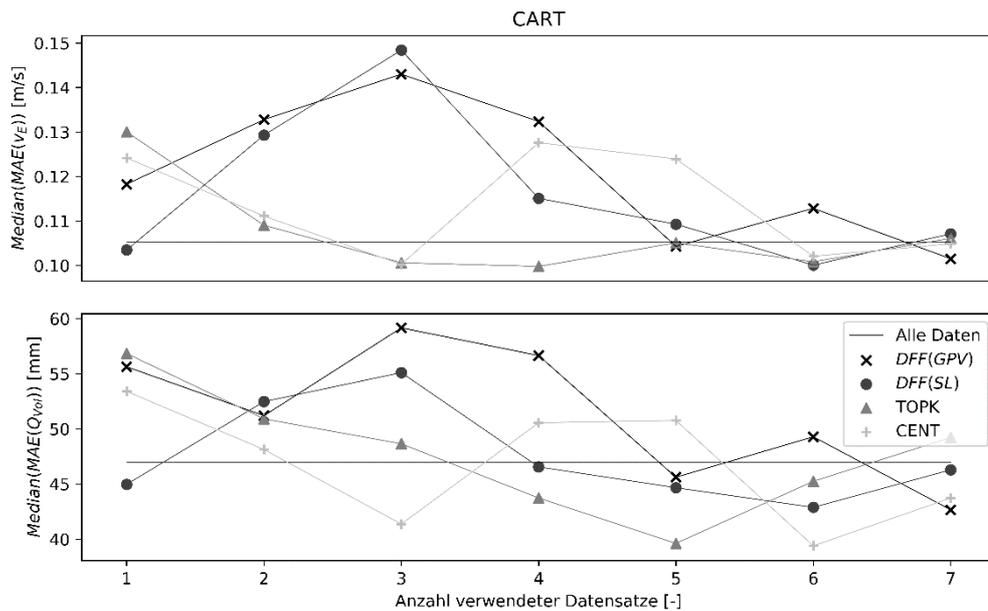


**Abbildung 7-4: Vergleich MAE( $v_E$ ) (oben) und ( $Q_{Vol}$ ) (unten) bei Verwendung aller Daten (Balken) und optimaler Datenkomposition (Markierungen) mit SVM im Main**

Um eine einheitliche Anzahl von Datensätzen zu bestimmen, wurde die Auswertung der zuvor durchgeführten Lernstudie mit selektiver Datenkomposition erneut ausgewertet. Für jede Anzahl zugefügter Datensätze wurde der MAE-Mittelwert über alle Teilgebiete je Algorithmus und Untersuchungsgebiet berechnet.

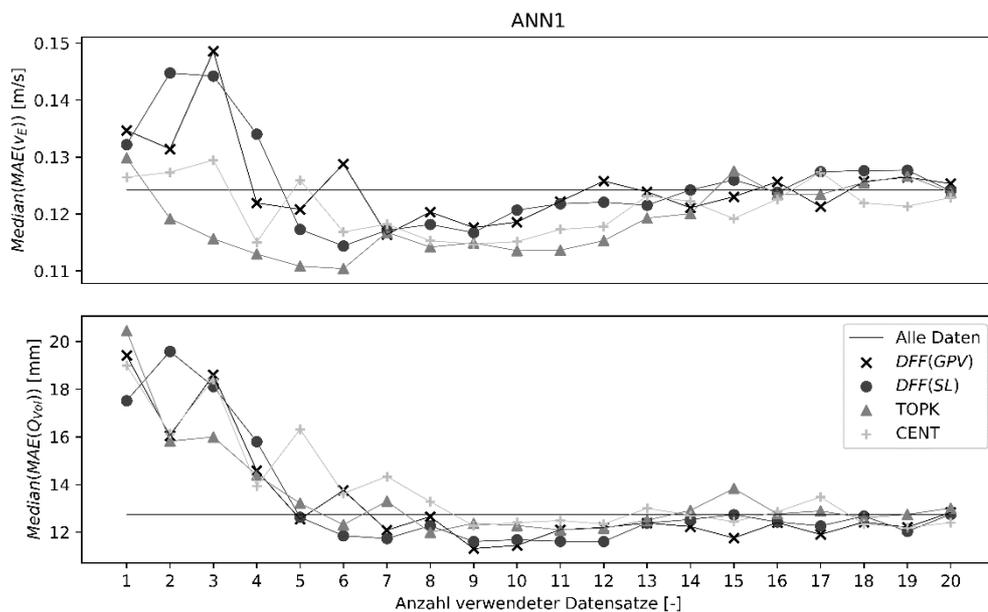
Die MAE-minimierende Anzahl kann als die optimale Anzahl von Datensätzen zur Komposition des Datensatzes definiert werden. Zudem musste außerdem das optimale Ranking ermittelt werden. Dies wurde über eine Analyse der Lernkurven realisiert: Wie zuvor definiert weist eine optimale Lernkurve eine U-Form auf. Tritt eine U-Form auf, wurden dem Algorithmus zuerst alle relevanten Informationen zugeführt und dann die verbleibenden redundanten, irrelevanten oder fehlerhaften Daten. Eine Metrik, die eine U-Kurve erzeugt und mit weniger Datensätzen zu besseren Ergebnissen führt, ist gegenüber den anderen Metriken zu bevorzugen.

Als erstes Beispiel der durchgeführten Auswertung sind in Abbildung 7-5 die Lernkurven des CART Algorithmus im Untersuchungsgebiet der Iller dargestellt. Eindeutig sind die Lernkurven mit Top-Kriging und Centroid gegenüber den DFF-Kurven zu präferieren. Beide weisen eine ungefähre U-Form auf, während die DFF-Kurven das genaue Gegenteil zeigen. Top-Kriging erweist sich hier als beste Variante, da die U-Kurven des Centroid bei beiden Variablen kurzzeitige Anstiege des MAE im Bereich von 4-5 verwendeten Datensätzen aufweisen.



**Abbildung 7-5: Lernkurven CART im Untersuchungsgebiet Iller mit selektiver Datenkomposition**

So deutlich wie in diesem Fall war die Auswertung nicht in allen Fällen. Vor allem der deutliche Zusammenhang von  $v_E$  und  $Q_{Vol}$  war nur in einigen Fällen gegeben. In der Mehrzahl der Fälle wiesen die Variablen unterschiedliche Präferenzen auf. Als Beispiel für diese Beobachtung sind in Abbildung 7-6 die Lernkurven des ANN1 im Main dargestellt.



**Abbildung 7-6: Lernkurven ANN1 im Untersuchungsgebiet Main mit selektiver Datenkomposition**

Wieder ist zu erkennen, dass für  $v_E$  Top-Kriging präferiert wurde. Eine deutliche U-Form ist zu erkennen und das Minimum des MAE wurde mit dieser Metrik erreicht. Für  $Q_{Vol}$  wurde hingegen DFF(SL) präferiert. DFF(SL) weist trotz anfänglicher Schwankungen die stabilste U-Form auf. Top-Kriging und DFF(GPV) weisen stärkere Schwankungen im Tal der U-Kurve auf, dem Bereich, der für die Auswahl der Anzahl am wichtigsten war. Daher wurde in diesem Falle DFF(SL) der Vorzug gegeben. Dennoch ist die Auswahl nicht so eindeutig wie zuvor.

Da die hier dargestellten Verläufe deutlich häufiger zu beobachten waren als die in Abbildung 7-5 dargestellten, wurde eine nach Variable getrennte Auswertung durchgeführt. Ermittelt wurde die präferierte Metrik je Untersuchungsgebieten und Algorithmen. Die Ergebnisse der Auswertung sind in Tabelle 7-1 zusammengefasst.

**Tabelle 7-1: Auswertung präferierter Metriken nach Algorithmen und Untersuchungsgebieten (I = Iller; M = Main; R = Regen)**

Alg.	$v_E$				$Q_{Vol}$			
	TOPK	CENT	DFF(SL)	DFF(GPV)	TOPK	CENT	DFF(SL)	DFF(GPV)
ADA	R	I	M			I		MR
ANN1	M		R	I			MR	I
ANN2	IM			R	I		M	R
CART	IM		R			I	MR	
KNN	IM		R			I	MR	
RF	I		MR		I		MR	
SVM	MR	I			M	I	R	

Zur Komposition der optimalen Lernmenge, werden nach Tabelle 7-1 im Untersuchungsgebiet Iller distanzbasierte Metriken bevorzugt, unabhängig von den Zielvariablen. Top-Kriging wird dabei deutlich häufiger präferiert. Ähnlich, wenn auch mit umgekehrtem Ergebnis, verhält es sich im Regen. Optimale Datenkompositionen werden nahezu ausschließlich über DFF-Metriken für beide Variablen erstellt. Der Unterschied zwischen den Kurven von GPV und SL waren im Regen sehr gering, mit einem wie dargestellten leichten Vorteil für SL. Im Main hingegen scheint eine Differenzierung zwischen den Variablen notwendig. Für  $v_E$  dominierten die distanzbasierten- für  $Q_{Vol}$  die DFF-basierten Metriken zur Erstellung der Datenkomposition.

Dass in unterschiedlichen Gebieten unterschiedliche Metriken präferiert wurden, erscheint bei der Betrachtung der Gebiete sinnvoll. In der Iller sind (fast) alle Teilgebiete unabhängig voneinander (vgl. Abbildung 3-6), es liegt nur eine geringe Verschachtelung vor. Dadurch definiert sich die Ähnlichkeit der Gebiete hier in höherem Maß über Form und Nähe als im Regen. Zudem ist die Dominanz der Topologie in diesem Gebiet hervorzuheben. Die deutlichen Talstrukturen und das Fehlen von Bodenbedeckung und Porenraum in dem alpinen Gebiet scheint die Rolle der Topologie zu stärken. Im Regen liegt hingegen ein hoher Grad von Verschachtelung vor, nur 3 Gebiete sind Kopfgebiete (vgl. Abbildung 3-4). Aufgrund

der häufigen Unter- und Oberlieger Beziehungen ist anzunehmen, dass die distanzbasierten Verfahren hier von Nachteil waren. Zur Verdeutlichung: Im Regen gibt es nur wenige Kopfgebiete (vgl. Abbildung 3-4), die alle in höher gelegenen Regionen liegen. Durch die längliche Struktur des Einzugsgebietes sind ihre Centroide aber weiter voneinander entfernt als ihre Unterlieger. Obwohl anzunehmen ist, dass die Prozesse in den Kopfgebieten ähnlicher sind, als in den jeweiligen, größeren Unterliegergebieten. Im Main liegt durch die hohe Anzahl von Teilgebieten beides vor: ein hoher Anteil von Kopfgebieten und parallele Unter- und Oberlieger Beziehungen wie sie in den anderen Gebieten nicht vorliegen. Auch der Wechsel zwischen höheren Lagen und Flachland deutet darauf hin, dass in diesem Gebiet eine stärkere Differenzierung notwendig ist.

Dass für  $Q_{Vol}$  und  $v_E$  unterschiedliche Datensätze zur Anpassung der Modelle verwendet werden sollen (durch unterschiedliche Metriken), widerspricht auf dem ersten Blick den gängigen Methoden der Kalibrierung. Dennoch, die Trennung von Abflussbildung und Abflusskonzentration in den meisten hydrologischen Modellen impliziert bereits die notwendige Trennung der Prozesse. Wenn für diese unterschiedlichen Prozesse getrennte Modelle existieren, liegt der Schluss nahe, dass eine getrennte Kalibrierung, auch auf Grundlage anderer Daten, sinnvoll sein kann. Die Ergebnisse der hier durchgeführten Untersuchungen weisen darauf hin.

Abschließend wurden die Lernkurven über alle Algorithmen innerhalb der Untersuchungsgebiete gemittelt. Dies wurde auf Grundlage der Ergebnisse in Tabelle 7-1 durchgeführt, die nur geringe Unterschiede in den Präferenzen der Algorithmen zeigten. Ausgehend von der durchgeführten Mittelung wurde die MAE-minimierende Anzahl von Datensätzen ermittelt. Die Ergebnisse sind in Tabelle 7-2 zusammengefasst, die Lernkurven sind dem digitalen Anhang zu entnehmen.

**Tabelle 7-2: Auswahl optimaler Anzahl Datensätze und Ranking**

Gebiet	Metrik	Anzahl Datensätze
Iller	Top-K.	4
Regen	DFF(SL)	4
Main ( $v_E$ )	Top-K.	6
Main ( $Q_{Vol}$ )	DFF(SL)	11

## 7.4 Alternative Methoden

Mit den in diesem Kapitel verwendeten Metriken ist natürlich noch nicht das volle Maß an Möglichkeiten ausgeschöpft, wie hydrologische Ähnlichkeit definiert werden kann. Wagner et al. (2007) sowie He, Bárdossy & Zehe (2011b) geben einen Überblick über weitere Möglichkeiten. Alternativ könnte auch per Monte-Carlo Simulation die jeweils optimale Reihenfolge und Anzahl der Datensätze ermittelt werden. Ob sich die Ergebnisse aber durch weitere Datensatz-Rankings weiter verbessern lassen, ist zu bezweifeln. Bei Auswahl der minimalen Fehler, individuell nach Algorithmus und Teilgebiet, zeigte sich, dass mit allen

Metriken ein ähnliches Minimum erreicht werden konnte. Lediglich die Form der Lernkurve und damit die Anzahl der benötigten Datensätze variierten mit den Metriken.

Als weitaus vielversprechender wird ein Ranking der Ereignisse eingeschätzt. Der Informationsgehalt eines Trainingsdatensatzes hängt von der Häufigkeit bestimmter Ereignistypen ab, bzw. dem Spektrum der Prädiktoren und Zielvariablen, die von den Ereignissen abgedeckt werden. Befinden sich im Trainingsdatensatz nur sehr ähnliche Ereignisse, ist der Informationsgehalt gering. Sind die Ereignisse hingegen sehr unterschiedlich, ist der Informationsgehalt wesentlich höher. Anstatt also alle Ereignisse eines Datensatzes zu verwenden, könnte mit Hilfe eines *Maximum-Entropy* Ansatzes (Li, Singh & Mishra, 2012; Gong et al., 2014; Cui & Singh, 2015; Chacon-Hurtado, Alfonso & Solomatine, 2017) eine informationsmaximierende Auswahl der Ereignisse durchgeführt werden. Der Informationsgehalt einer Variable kann nach Shannon (1948) über die Entropie, also das Integral ihrer Häufigkeitsverteilung, berechnet werden. Da Entropie-Werte additiv sind, können die Informationsgehalte aller Prädiktoren und Zielvariablen summiert werden. Der Trainingsdatensatz wird dann so komponiert, dass die Entropie des Datensatzes maximal ist. Neben einer strukturierten Suche nach der besten Komposition der Ereignisse ist auch in diesem Falle die Möglichkeit einer Monte-Carlo Simulation gegeben.

## 8 Ergebnisse der Fallstudien

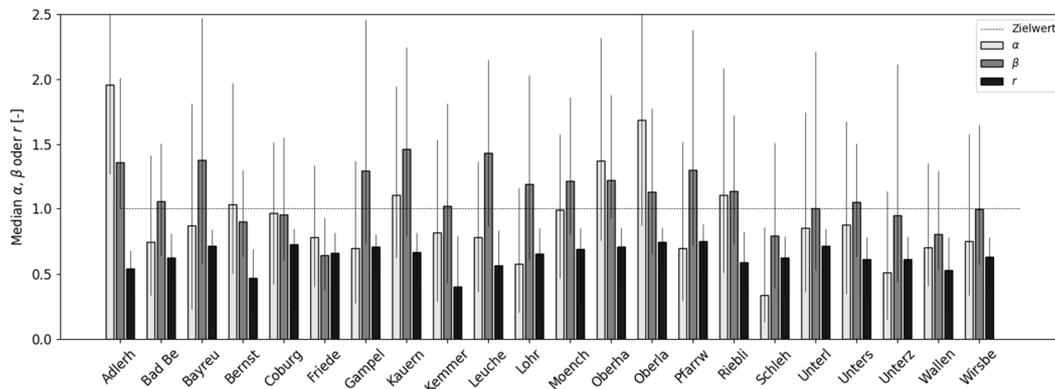
---

Die vorgestellte aus Dynamikmodul und GIUH-Modell kombinierte Modellstruktur wurde in drei Fallstudien angewendet. Wie in Kapitel 6 wurde in Main, Regen und Iller mit einer *Leave-one-out* Strategie die Anwendung des Modells in unbeobachteten Gebieten getestet.

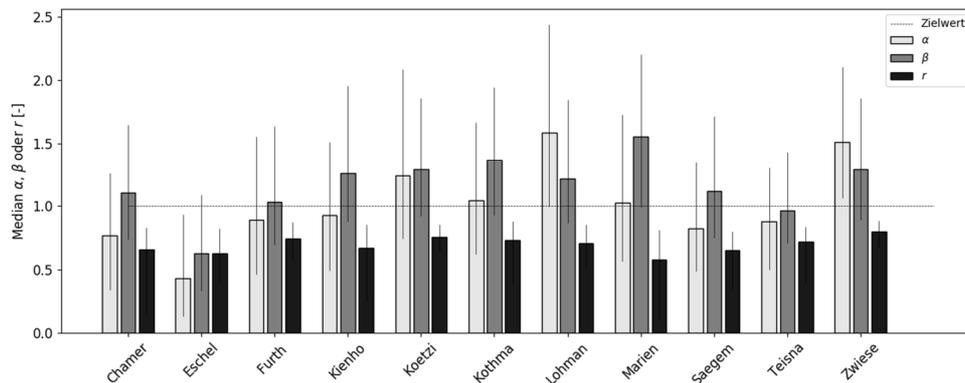
Dabei wurde jeweils ein Gebiet als unbeobachtet ausgewählt. Mit Hilfe der verbleibenden Daten des Untersuchungsgebietes wurde das Dynamikmodul trainiert. Das trainierte Modell wurde im Anschluss mit den Ereignissen des „unbeobachteten“ Gebiets validiert. Das Dynamikmodul erstellt dabei mit Hilfe der Indikatoren für jedes Ereignis eine Prognose über Ereignistyp, Entwässerungsgeschwindigkeit  $v_E$  und zu erwartendes Volumen  $Q_{Vol}$ . Die Parameter werden vom GIUH-Modell verwendet, um mittels geomorphologischer Impulsantwortfunktion den Ereignisniederschlag in eine Abflussganglinie zu transformieren. Zur Bewertung des Modells wurden die simulierten Ganglinien mit den beobachteten verglichen und mit den eingeführten Gütekriterien bewertet.

Zu beachten ist, dass die Ergebnisse der vorhergehenden Analysen in den Fallstudien berücksichtigt wurden. Als Prädiktoren des Dynamikmoduls wurden eine Kombination aus Niederschlags-, Temperatur und Schneedaten verwendet sowie der Basisabfluss am Zielort zu Ereignisbeginn (vgl. Kapitel 6.3). Als Zielvariablen wurden die kalibrierten Werte von  $Q_{Vol}$  und  $v_E$  verwendet (vgl. Kapitel 5.4). Die Trainingsdaten um das Dynamikmodul anzupassen wurden auf Grundlage einer ausgewählten Ähnlichkeitsmetrik gewählt. Je nach Gebiet und Zielvariable wurden die 4-6 ähnlichsten Gebiete, ermittelt mit Top-Kriging oder der DFF-Funktion des Gefälles, verwendet um die ML-Algorithmen anzupassen (vgl. Kapitel 7). Darüber hinaus verfügte das Dynamikmodul über eine parallele Struktur, um unterschiedliche Ereignistypen differenziert prognostizieren zu können. Dafür wurden die Indikatoren basierend auf einer Klassifikation der zugehörigen N-A-Ereignisse nach dem Füllen-Scheitel-Verhältnis (*tq\_regional*) in drei Gruppen aufgeteilt. Für jede Gruppe stand ein eigenes Ensemble von 8 ML-Algorithmen zur Verfügung, die parallel angepasst wurden (vgl. Kapitel 6.2). Zur Prognose eines Ereignisses wurde mit einem trainierten zwei-lagigem ANN, der zu erwartende Ereignistyp geschätzt und die jeweilige Gruppe von Algorithmen zur Vorhersage von  $Q_{Vol}$  und  $v_E$  verwendet (vgl. Kapitel 6.1). Diese mittleren Ereignisparameter wurden dann an das GIUH-Modell übergeben. Die Entwässerungsgeschwindigkeit des Hangs wurde als  $v_{E,H}$  räumlich differenziert nach dem Hanggefälle angesetzt und die Geschwindigkeit der Gewässer als  $v_{E,S}$  räumlich differenziert nach der Strahler-Ordnung (vgl. Kapitel 5.3). Die ermittelten Geschwindigkeiten gingen dann in die Ermittlung der geomorphologischen Impulsantwort ein. Aus punktuell verfügbaren Niederschlagszeitreihen wurden im Modell die finalen Prognosen der Abflussganglinien erstellt (vgl. Kapitel 4.4.).

Als erstes Ergebnis der Fallstudie im Untersuchungsgebiet Main zeigt Abbildung 8-1 den Median der KGE-Komponenten sowie deren Unsicherheit über die 25- und 75% Quantile. Dabei wurden die Ergebnisse aller Ereignisse, aller Cluster und aller Algorithmen gemeinsam ausgewertet.



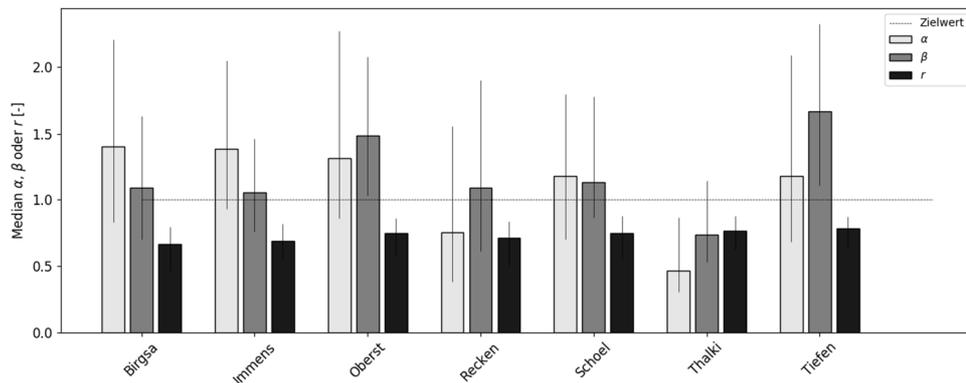
**Abbildung 8-1: Median und Unsicherheitsbereiche KGE-Komponenten der Fallstudie im Untersuchungsgebiet Main, differenziert nach Teilgebiet**



**Abbildung 8-2: Median und Unsicherheitsbereiche KGE-Komponenten der Fallstudie im Untersuchungsgebiet Regen, differenziert nach Teilgebiet**

Der Mittelwert im gesamten Untersuchungsgebiet liegt für  $\alpha$  bei 0.92, für  $\beta$  bei 1.10 und für  $r$  bei 0.63. Davon abweichend traten in einigen Teilgebieten nur deutliche höhere Werte, bspw. in Adlerhütte, oder deutlich geringere, bspw. Schlehenmühle, auf. Ein besonders niedriger  $r$  wurde im Teilgebiet Kemmern erreicht. Wie in den Kapiteln zuvor, spiegeln auch diese Ergebnisse eine regionale Abhängigkeit der ML-Algorithmen wieder.

Ähnliche Ergebnisse wurden in den anderen Fallstudien Regen (Abbildung 8-2) und Iller (Abbildung 8-3) erreicht. Die Medianwerte lagen im Regen für  $\alpha$  bei 1.01, für  $\beta$  bei 1.17 und  $r$  bei 0.70. Im Untersuchungsgebiet Iller für  $\alpha$  bei 1.10, für  $\beta$  bei 1.17 und  $r$  bei 0.73.



**Abbildung 8-3: Median und Unsicherheitsbereiche KGE-Komponenten der Fallstudie im Untersuchungsgebiet Iller, differenziert nach Teilgebiet**

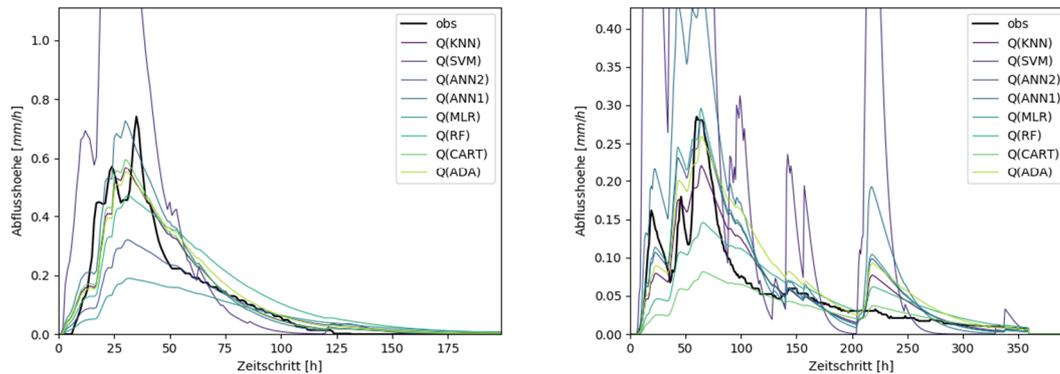
Auch in diesen Untersuchungsgebieten weichen die Teilgebiete zum Teil erheblich von diesen Medianwerten ab. Im Gebiet der Iller ist unter anderem Thalkirchdorf (kurz: Thalki) zu nennen. Varianz und Mittelwert der Ganglinien werden in diesen Gebieten eher unterschätzt, erkennbar durch  $\alpha$  und  $\beta < 1$ . Oberstdorf (kurz: Oberst) hingegen weist deutlich höhere  $\alpha$  und  $\beta$  Werte auf. Bemerkenswert ist die gleichmäßige Verteilung von  $r$  über alle Teilgebiete. Auch im Regen können vom Mittelwert abweichende Teilgebiete leicht identifiziert werden. In Eschelkam (kurz: Eschel) liegen die Gütekriterien deutlich unter den Medianwerten, in Zwiesel (kurz: Zwiese) deutlich darüber.

In allen Fallstudien liegt  $\beta$  bei etwa 1.1, was als eine 10% Überschätzung des Abflussvolumens interpretiert werden kann. In Iller und Regen liegt  $\alpha$  etwa bei 1, im Main leicht darunter. Auch hier liegen rund 10% Abweichung vor, in diesem Falle die Varianz der Ganglinie betreffend. Die simulierte Wellenform ist demnach etwas glatter als die beobachtete.

Die teils signifikante Abweichung der Gütekriterien zwischen den Gebieten, aber auch innerhalb der Gebiete, die durch die Unsicherheitsbalken in Abbildung 8-1, Abbildung 8-2 und Abbildung 8-3 dargestellt sind, ist indes nicht zu vernachlässigen. Zur Varianz der Kriterien zwischen den Teilgebieten kann auf die Diskussionen der einzelnen Modellkomponenten in den Kapiteln 5 und 6 verwiesen werden. Die Analysen zeigten, dass die Methoden unterschiedliche Eignung für die verwendeten Teilgebiete aufweisen. Als Beispiel sei an dieser Stelle Schlehenmühle angeführt: In der durchgeführten Fallstudie konnten in diesem Gebiet mit  $\alpha = 0.34$ ,  $\beta = 0.80$  und  $r = 0.62$ , vor allem in Hinblick auf  $\alpha$  sehr schlechte Güte erreicht werden. Verglichen mit den Ergebnissen der mittleren Prognosefehler der ML-Algorithmen im oberen Main in Abbildung 6-6, zeigt sich, dass Schlehenmühle die höchsten Prognosefehler für  $v_E$  aufwies. Der Fehler in der Parameterprognose bedingt also zu einem Großteil den Fehler der prognostizierten Ganglinien.

Innerhalb der Teilgebiete wurde die Streuung der Kriterien zum einen durch die Ereignisse erzeugt, zum anderen durch die 8 verwendeten Algorithmen. Abbildung 8-4 zeigt beispiel-

haft zwei Ereignisse am Pegel Bad Berneck (Main). Zu erkennen ist die Bandbreite der Prognosen der unterschiedlichen Algorithmen.



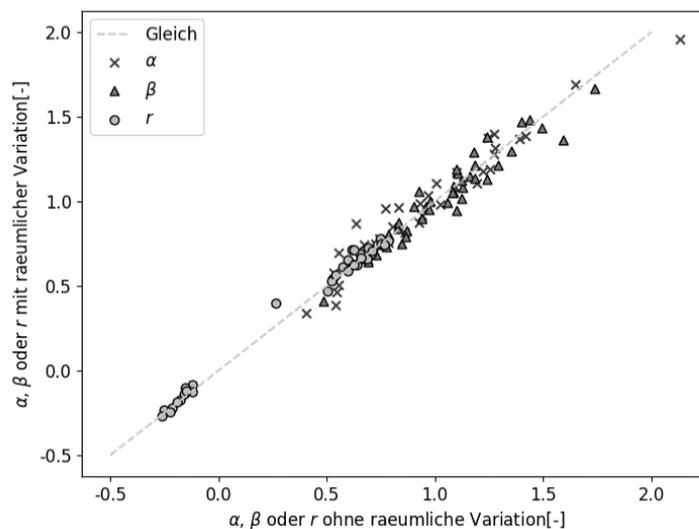
**Abbildung 8-4: Ganglinien Prognose Bad Berneck (links) 30.11.2002;(rechts) 14.12.2003**

Dieses Beispiel legt in Verbindung mit den Ergebnissen aus Kapitel 6.3.2 nahe, dass nicht die einzelnen Prognosen der Algorithmen zu verwenden sind sondern eine Ensemblevorhersage. Im folgenden Kapitel 8.2 wird auf die gesamte Prognosegüte des Ensembles eingegangen sowie eine Methodik vorgestellt, wie die einzelnen Prognosen zu einer Ensembleprognose mit Unsicherheitsband verarbeitet werden können.

Zunächst wird kurz auf den Einfluss der Modellstruktur, d.h. die räumliche Variation der Parameter eingegangen. In Kapitel 8.3 wird dann auf die Güte der Prognosen der einzelnen Algorithmen analysiert. Eine erweiterter Vergleich der Ergebnisse der Fallstudie mit den Ergebnissen der Kalibrierung wird in 8.4 vorgestellt und eine abschließende Fehlerdiskussion in den Kapiteln 8.5 und 8.6 geführt.

## 8.1 Einfluss Modellstruktur

Bereits bei der Entwicklung der räumlichen Variation der Entwässerungsgeschwindigkeit wurde auf die daraus resultierende Erhöhung der Ganglinienvarianz eingegangen. In den in diesem Kapitel dargestellten Ergebnissen wurde daher stets die vorgeschlagene Variation mit der Hanggefälle und Strahler-Ordnung verwendet. In diesem Abschnitt soll aber noch kurz ein Blick auf den tatsächlichen Einfluss auf die Simulation der Ganglinien geworfen werden. In Abbildung 8-5 wurden die Median KGE-Komponenten aller Teilgebiete der Simulation mit und ohne Variation der Parameter gegenübergestellt.



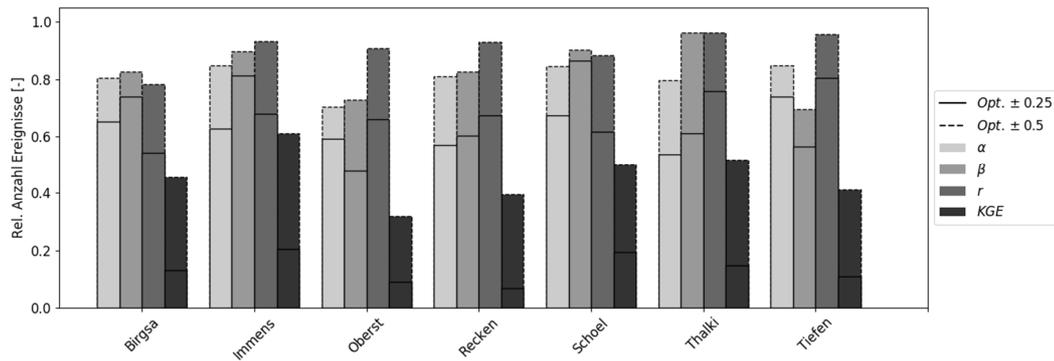
**Abbildung 8-5: Vergleich Median Gütekriterien mit und ohne räumlicher Variation der Dynamikparameter**

Abbildung 8-5 zeigt, dass keine eindeutige Tendenz vorliegt. Auch die Betrachtung der mittleren Komponenten zeigt lediglich eine geringe Verbesserung der Ergebnisse. Unter Einbezug der Variation verbessert sich die Güte um lediglich 0.01 bzw. 0.02 (bspw. erhöht sich das mittlere  $\alpha$  von 0.93 auf 0.94). Dennoch, für einzelne Ereignisse und einzelne Teilgebiete ist die Variation vorteilhaft. Da sie zudem einen der Kritikpunkte am GIUH behandelt, dass die Parameter ohne Variation auf das gesamte Gebiet angewendet werden (Blöschl & Sivapalan, 1995), wurde die Variation der Parameter beibehalten.

## 8.2 Ensemble Performance

Neben der mittleren Anpassungsgüte stellt sich bei den vorliegenden Ergebnissen die Frage, wie häufig eine Ganglinie mit guter Anpassung produziert werden konnte. Dazu muss „gut“ definiert werden: als erste Definition wurden alle Ganglinien, deren Güte um 0.25 vom Optimum abweichen, als „gut“ angenommen. Für die zweite Definition wurde der Grenzwert auf 0.5 gesenkt. Die verwendeten Grenzen stellen natürlich eine subjektive Auswahl dar, dienen an dieser Stelle aber auch lediglich zur Übersicht über die Ergebnisse.

In folgender Abbildung 8-6 ist für das Untersuchungsgebiet Iller dargestellt, für wie viele Ereignisse, normiert und differenziert nach Teilgebiet, mindestens eine simulierte Ganglinie als „gut“ beurteilt wurde. Für einen Grenzwert von maximal 0.25 Abweichung vom Optimum sind dies zwischen 60-70% aller Ereignisse, für  $\alpha$ ,  $\beta$  und  $r$ . Für einen Grenzwert von 0.5 steigt der Anteil auf 80-90%.

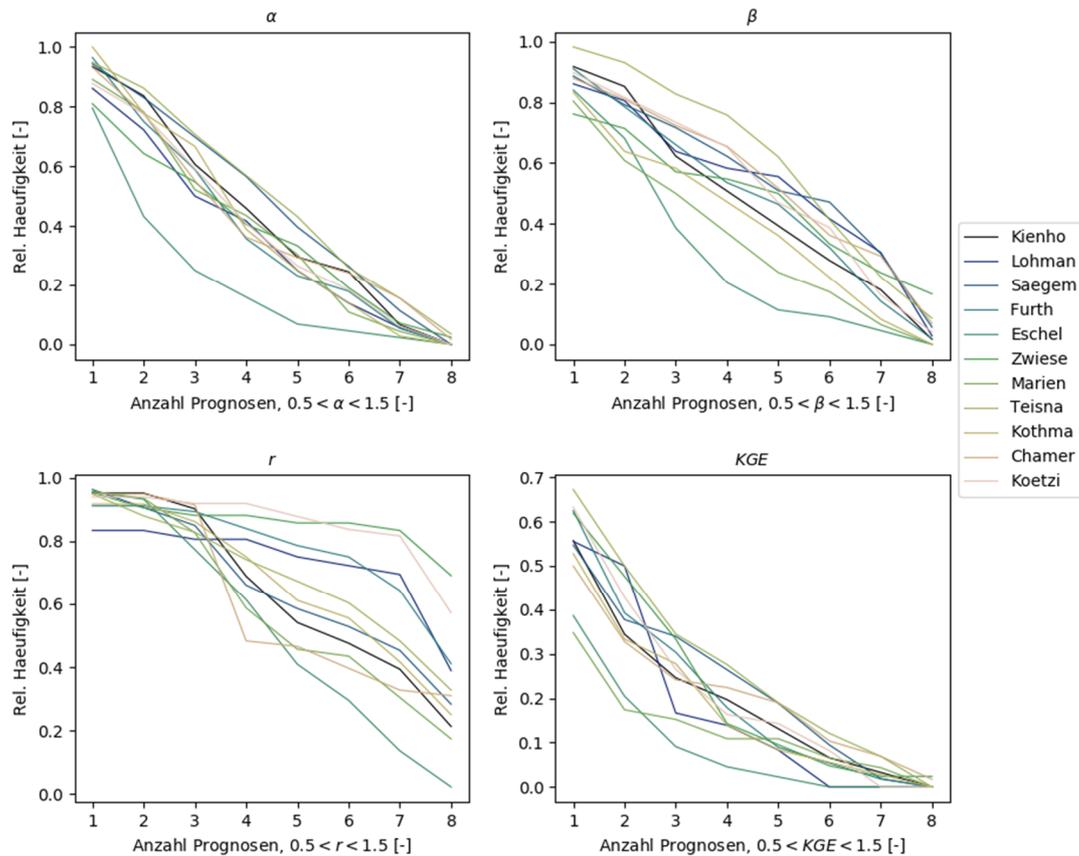


**Abbildung 8-6: Normierte Anzahl von Ereignissen für die mindestens eine simulierte Ganglinie max. 0.25/0.5 vom Optimum der Güte abweicht, Untersuchungsgebiet Iller**

Im Main und Regen wurden ähnliche Werte bestimmt, die zugehörigen Darstellungen können dem digitalen Anhang entnommen werden. Im Main sind die mittleren Anzahlen in etwa identisch zu den zuvor dargestellten Ergebnissen der Iller. Etwas höher liegen die Anteile im Regen, bei rund 70-80% für einen Grenzwert von 0.25, 90% bei 0.5. Die Ergebnisse deuten darauf hin, dass bei Anwendung in unbeobachteten Gebieten die Wahrscheinlichkeit relativ hoch ist (60-80%, bzw. 80-90%, je nach geforderter Güte), dass zumindest eine der simulierten Ganglinien die Standardabweichung, Mittelwert und/oder Korrelation der tatsächlichen Ganglinie abbilden kann.

Ziel muss sein, dass so viele Ganglinien des Ensembles wie möglich eine gute Prognose abgeben. Daher wurde die Mindestanzahl von prognostizierten Ganglinien, deren Gütekriterien maximal 0.5 vom Optimum abweichen, schrittweise erhöht und in jedem Schritt die (normierte) Anzahl von Ereignissen bestimmt, auf die diese Bedingung zutrifft. Die Ergebnisse für das Untersuchungsgebiet Regen sind in Abbildung 8-7 dargestellt.

Auf der Abszisse abgetragen ist Anzahl von vorhergesagten Ganglinien je Ereignis, die der zuvor formulierten Definition von „gut“ entsprachen. Die maximale Anzahl ist acht, da acht Algorithmen verwendet wurden. Auf der Ordinate ist der Anteil der Ereignisse, je Teileinzugsgebiet, aufgetragen für die eine entsprechende (auf der Abszisse aufgetragenen) Anzahl von guten Prognosen vorlagen. In den vier Paneelen von Abbildung 8-7 sind die Verläufe für die vier unterschiedlichen Gütekomponenten ( $\alpha$ ,  $\beta$ ,  $r$  und KGE) dargestellt. Ein kurzes Beispiel zur Handhabung der Abbildung: Abzulesen ist, dass für rund 80% der Ereignisse Prognosen vorlagen (in mittel für alle Teileinzugsgebiete), die die Ganglinienvarianz gut abbilden konnten, d.h.  $0.5 < \alpha < 1.5$ . Fünf gute Prognosen im Hinblick auf  $\alpha$  lagen nur noch für rund 40% der Ereignisse vor, der Anteil sinkt auf 0% für acht Prognosen. Dies bedeutet, dass für kein Ereignis alle verwendeten Algorithmen gleichzeitig „richtig“ lagen. Analog zu diesem Beispiel können alle anderen Paneele ausgewertet werden.



**Abbildung 8-7: Relative Häufigkeit von Ereignissen für die eine (oder mehr Prognosen) vorlagen, deren Güteparameter maximal 0.5 vom Optimalwert abwichen (oben links:  $\alpha$ ; rechts:  $\beta$ ; unten links:  $r$ ; rechts: KGE)**

Die vorherrschende lokale Variation der Ergebnisse ist auch in dieser Auswertung deutlich. Für eine feste Anzahl von geforderten „guten“ Ganglinien liegen die dargestellten Linien auf unterschiedlichen Höhen, der Anteil von Ereignissen mit „guten“ Simulationen variiert also mit den Teilgebieten. Mit steigender Anzahl von geforderten „guten“ Simulationen fällt zudem der Anteil der Ereignisse in einigen Teilgebieten stärker als in anderen. Die Ganglinien sind hier also deutlich verschiedener.

Eine zweite Beobachtung bezieht sich auf Differenzen des Rückgangs zwischen den Gütekriterien. Je nach Kriterium liegt eine andere Form der Kurven vor. Am stärksten fallen die KGE-Kurven. Dies war teilweise zu erwarten, da der KGE die Kombination der drei übrigen Komponenten darstellt und sehr sensitiv auf den Abfall einer einzelnen Komponente reagiert. Während die  $\beta$ -Kurven einer Winkelhalbierenden ähneln, scheint das Gefälle der  $\alpha$ -Kurven etwas stärker als bei  $\beta$  zu sein. Bei  $r$  bleibt der Anteil der Ereignisse hingegen mit steigender geforderter Anzahl von Ganglinien dagegen länger auf einem konstanten Niveau.

Die Ergebnisse zeigen, dass bei einem Großteil der Ereignisse einzelne Ganglinien, oder teilweise mehrere, gute Abbildungen der realen Ganglinien darstellen. Dennoch, nicht alle

Mitglieder des Ensembles liefern bei allen Ereignissen gute Ergebnisse. Ergebnisse vorheriger Kapitel legen zudem nahe, dass keine Korrelation zwischen den Gütekomponenten der Vorhersagen einzelner Ereignisse, berechnet mit unterschiedlichen Algorithmen, besteht. Diese Ergebnisse decken sich bspw. mit den Ergebnissen von Elshorbagy et al. (2010b).

Die hier dargestellten Ergebnisse können als repräsentativ für die anderen Fallstudien angesehen werden. Die dargestellten Kurven vereinen die in den Gebieten Iller und Main vorliegenden Kurvenformen. In der Iller verbleiben alle Teilgebiete für  $r$  bis zu 7-8 geforderten Ganglinien auf einem hohen Niveau, das Gefälle für KGE fällt geringer aus.

Basierend auf diesen Ergebnissen und den Empfehlungen in der Literatur erscheint es sinnvoll das Modellensemble zu verwenden, um einen Vorhersagebereich zu bestimmen, in dem ein anstehendes Ereignis zu erwarten ist. Die in dieser Arbeit verwendeten 8 Algorithmen bilden zwar noch keine umfangreiche Datengrundlage für eine fundierte Definition von Unsicherheitsbändern, die generelle Möglichkeit soll aber an dieser Stelle beleuchtet werden.

Zwei Ansätze wurden im Rahmen dieser Arbeit getestet. Zum einen wurde der Unsicherheitsbereich des gesamten Ensembles über das 10% und 90% - Quantil aller simulierten Ganglinien bestimmt. Zum anderen wurde ein GLUE Ansatz (Beven & Binley, 1992) verwendet, bei dem die *Likelihood* der einzelnen Algorithmen berücksichtigt wurde.

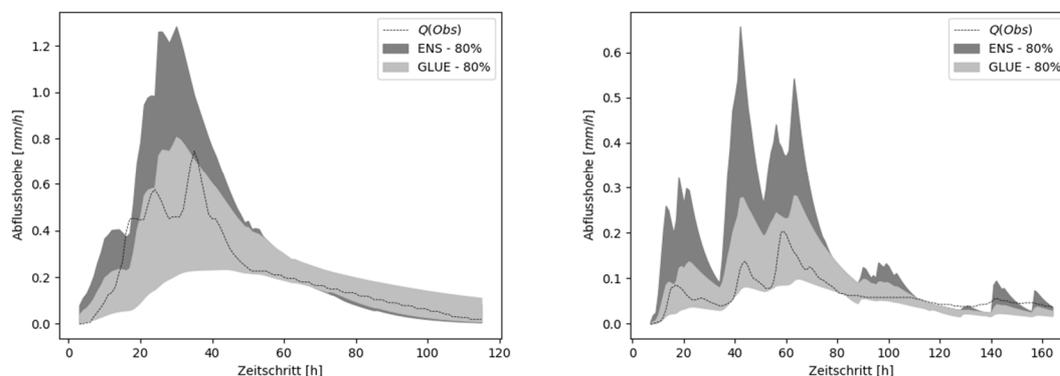
Folgend den Definitionen von Beven & Binley (1992) sollte die verwendete *Likelihood*-Funktion die Güte der Modelle, mit denen die betreffenden Ganglinien berechnet wurden, repräsentieren. Außerdem sollte sie Werte zwischen 0 und 1 annehmen. Im vorliegenden Fall wurde die *Likelihood* über den Korrelationskoeffizienten zwischen Prognose und Testdatensatz der ML-Algorithmen in der Anlernphase geschätzt. Dafür wurde dem Trainingsdatensatz der Algorithmen ein Anteil von 10% der zur Verfügung gestellten Daten vorenthalten und für eine nachfolgende Vor-Validierung verwendet. Die Korrelationskoeffizienten zwischen vorhergesagtem und beobachtetem  $v_E$  sowie  $Q_{Vol}$  wurden somit bestimmt. Nach der Anlernphase wurden 2 mal 8 Korrelationskoeffizienten zurückgegeben, die wie folgt in *Likelihood*-Werte umgerechnet wurden:

$$L(m) = L_{v_E}(m) \cdot L_{Q_{Vol}}(m) = \frac{\max_M(r_{v_E}) - r_{v_E}(m)}{\max_M(r_{v_E}) - (\max_M(r_{v_E}) - a)} \cdot \frac{\max_M(r_{Q_{Vol}}) - r_{Q_{Vol}}(m)}{\max_M(r_{Q_{Vol}}) - (\max_M(r_{Q_{Vol}}) - a)} \quad 8.1$$

Dabei berechnete sich die *Likelihood* des Algorithmus  $m$  als Produkt der *Likelihood*-Werte von  $v_E$  und  $Q_{Vol}$ . Diese wurden wiederum berechnet als lineare Interpolation zwischen dem maximalen  $r$  aller Algorithmen  $M$  und dem maximalem  $r$  abzüglich eines Grenzwertes  $a$  der in diesem Test zu 0.4 angenommen wurde. Zu beachten ist, dass die untere Grenze der Interpolation nie kleiner als 0 gesetzt wurde. Die berechneten *Likelihood*-Werte wurden als normierte Werte  $L_{norm}(m)$  gespeichert. Der hier eingeführte Grenzwert  $a$  dient zur Beschränkung der *Likelihood* und schließt Prognosen von der Definition des Unsicherheitsbandes aus, die durch Algorithmen produziert wurden, deren Anpassungsgüte signifikant schlechter

war als der Rest des Ensembles. Der gewählte Wert ist wie in den Analysen zuvor das Ergebnis einer subjektiven Entscheidung des Anwenders. In diesem Test ist er in Anlehnung an die zuvor verwendete maximale Abweichung vom Optimum von 0.5 gewählt worden. Allerdings etwas geringer, da die erreichten maximalen Gütekriterien nur selten an den Optimum von 1 heranreichen. Eine Reduzierung des Wertes kann eine Reduzierung der Bandbreite ermöglichen und sollte für zukünftige Anwendungen in Betracht gezogen werden.

Um die Unsicherheitsbänder zu berechnen, wurden nach der Simulation aller Ganglinien für jeden simulierten Zeitschritt die berechneten Abflüsse zunächst ihrer Größe nach geordnet. Die zugehörigen  $L_{Norm}$ -Werte wurden dann der Sortierung der Abflüsse folgend kumuliert. Aus den kumulierten *Likelihood* Werten konnten dann das 10% und 90% Quantil ermittelt werden. Abbildung 8-8 zeigt zwei Beispiele für Unsicherheitsbänder, die aus diesen Quantil-Werten mittels GLUE und des gesamten Ensembles (ENS) berechnet wurden.



**Abbildung 8-8: Ensemble Unsicherheitsbänder berechnet mit gesamtem Ensemble (ENS) und GLUE für Ereignisse vom 30.11.2002 Bad Berneck (links); 14.12.2003 Wirsberg (rechts)**

Um die Modellgüte der Ensemble-Vorhersage zu bewerten, wurde ermittelt, in wie vielen Fällen die beobachteten Ereignisse innerhalb des berechneten Unsicherheitsbandes lagen. Eine Übereinstimmung von 2/3 aller Zeitschritte eines Ereignisses wurde vorausgesetzt, um die Übereinstimmung als positiv zu bewerten. Die Ergebnisse werden in Tabelle 8-1 als Mittelwerte der Untersuchungsgebiete aufgelistet. Zwar lagen auch für diese Ergebnisse lokale Abweichungen vor, entsprachen aber den bereits diskutierten regionalen Gegebenheiten.

Neben der relativen Anzahl von Ereignissen, die innerhalb der Bänder lagen, sind die Bandbreite des Unsicherheitsbandes berechnet mit GLUE gegeben sowie das Verhältnis zwischen den Bandbreiten von GLUE und ENS. Die Bandbreiten wurden als mittlere Differenz zwischen dem 10% und 90% Quantil berechnet und normiert über die Standardabweichung des jeweiligen Ereignisses.

**Tabelle 8-1: Anteil Ereignisse die innerhalb des Unsicherheitsband lagen, berechnet mit ENS und GLUE; Verhältnis der Bandbreiten GLUE/ENS und normierte Bandbreite GLUE**

Gebiet	GLUE	ENS	Verh. Bandbreite	Normierte Bandbreite GLUE
Main	40.9%	63.0%	42.3%	44.1%
Regen	44.7%	64.9%	43.0%	44.9%
Iller	13.5%	37.8%	51.1%	15.7%

In den Untersuchungsgebieten Main und Iller ergaben sich sehr ähnliche Ergebnisse. Im Mittel lagen hier 40.9% (Main), bzw. 44.7% (Regen) aller Ereignisse innerhalb des GLUE-Unsicherheitsbandes. Ohne Verwendung einer *Likelihood*-Funktion (ENS) steigt der Anteil auf 63% bis 65%. Allerdings sind die ENS-Bänder signifikant breiter als die GLUE-Bänder. Die Bandbreite von GLUE beträgt im Mittel 44% der Standardabweichung des Ereignisses und umfasst lediglich 42%-43% der Breite eines ENS Bandes.

Starke Abweichungen dieser Ergebnisse lagen im Flussgebiet Iller vor. Hier lagen nur rund 13.5% der Ereignisse innerhalb der GLUE-Bänder und nur 37.8% innerhalb des ENS. Dies kann durch eine geringe Breite der GLUE-Bänder von lediglich 15.7% der Ereignis-Standardabweichung erklärt werden. Auch das ENS-Band ist wesentlich schmaler als in Main und Regen. Zwar ist es fast doppelt so breit wie das GLUE-Band, bleibt aber unter den Breiten in den anderen Untersuchungsgebieten zurück.

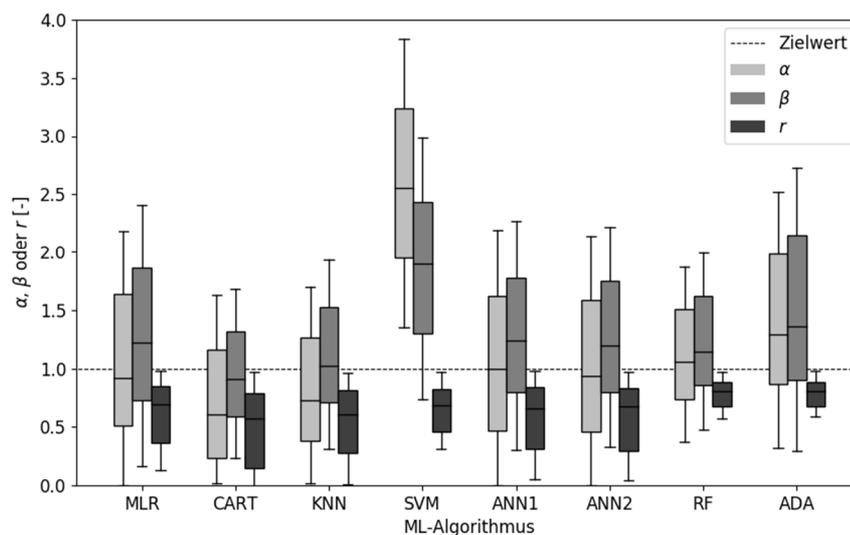
Dieses Ergebnis deutet bereits darauf hin, dass die individuellen Prognosen innerhalb der Iller sehr ähnlich sind. Durch die geringere Varianz der Prognosen wurde dann ein sehr enges Unsicherheitsband erzeugt. Dies ist eines der Symptome, resultierend aus dem Aufbau des Dynamikmoduls: der mangelnden Anzahl von Schätzern. Ein weiteres Symptom ist die Abhängigkeit der GLUE-Unsicherheitsbänder zu den besten Ensemblemitgliedern. In den meisten Fällen sind dies RF und ADA. Die Bänder richten sich größtenteils nach ihrer Prognose. Fälle, in denen ein anderes Ensemble-Mitglied richtig liegt, bleiben häufig außerhalb des GLUE-Unsicherheitsbandes.

In beiden Fällen kann eine Erhöhung der Anzahl von Ensemble-Mitgliedern die Prognose verbessern. Eine Möglichkeit wäre weitere Algorithmen aufzunehmen, eine andere mehrere Strukturen gleicher Algorithmen zu verwenden. Im Falle des ANN ist dies bereits der Fall, es wird ein- sowie zweilagiges MLP (*multi-layer-perceptron*) angewendet. Die Anzahl von Neuronen innerhalb der Layer wurde aber in der Anlernphase per *Trial-&-Error* bestimmt. Ähnlich wurden KNN und CART parametrisiert. Anstatt eine Auswahl zu treffen, könnten die unterschiedlichen Parametrisierungen in das Ensemble aufgenommen werden.

### 8.3 Güte individueller Prognosen

Die Auswertung der Ensembles wies bereits darauf hin, dass die Güte der Prognosen je Algorithmen in Main und Regen ähnlich seien und Abweichungen zur Iller aufweisen. Zur Überprüfung dieser Vermutung wurde die Simulationsgüte aller Ereignisse innerhalb der Untersuchungsgebiete, differenziert nach Algorithmen, ermittelt. Abbildung 8-9 zeigt Median, 25% und 75% Quantile sowie die Standardabweichung im Untersuchungsgebiet als Box-Whisker-Plot.

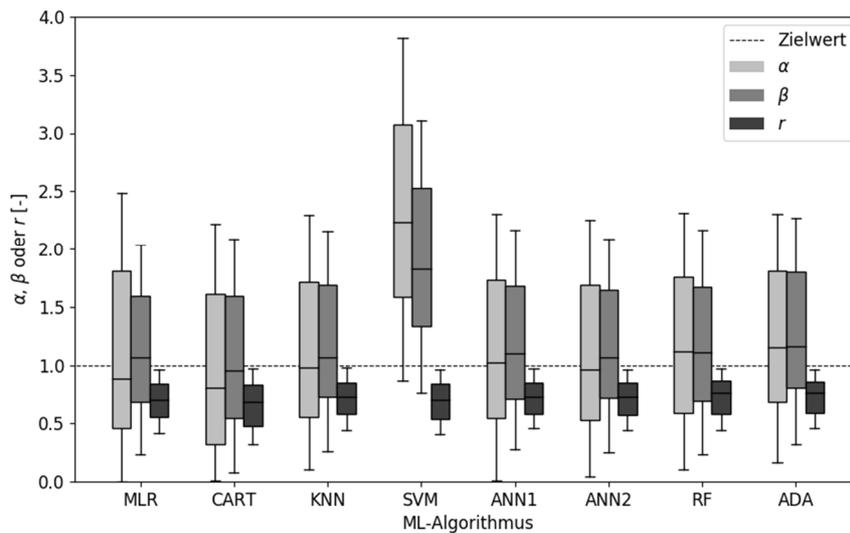
In diesem Untersuchungsgebiet lag der Median aller KGE-Komponenten des RF am nächsten an den optimalen Wert von 1. Auch die Streuung der Werte war mit diesem Algorithmus am geringsten. ADA zeigte hingegen  $\alpha$ - und  $\beta$ -Werte über 1, also eine Überschätzung von Ganglinienvarianz und –mittelwert. Besonders auffällig ist SVM: die Prognosen dieses Algorithmus lagen weit ab vom  $\alpha$ - und  $\beta$ -Optimum, die Korrelation der Ganglinien lag aber auf einem, zu den anderen Algorithmen, vergleichbaren Niveau. Die anderen Algorithmen, wie CART und die ANNs, wiesen aber durchaus gute Ergebnisse auf. Zwar ist die Streuung der Güte dieser Methoden höher als beim RF, im Mittel sind sie aber besser als die von ADA.



**Abbildung 8-9: Güte Vorhersagen je ML-Algorithmus, Untersuchungsgebiet Regen**

Ein vergleichbares Bild zeigt sich im Main (Darstellung im digitalen Anhang). Auch in diesem Flussgebiet liegt die Güte der SVM weit hinter den der anderen Algorithmen. Die Ensemble-Methoden RF und ADA weisen wie im Regen die geringste Streuung auf, etc. Die Ergebnisse weichen nur in der Weite der Boxen, also der Streuung der Güte von denen im Regen ab, im Main lag sie etwas höher.

Wie erwartet ist die Verteilung der Modellgüte über die Algorithmen in der Iller wesentlich gleichmäßiger. Zwar fällt auch hier die SVM deutlich aus dem Rahmen, alle anderen Algorithmen zeigen aber sehr ähnliche Medianwerte und Varianz der Güte. Dies bestätigt die Annahme, dass die Prognoseergebnisse nur sehr gering voneinander abweichen.

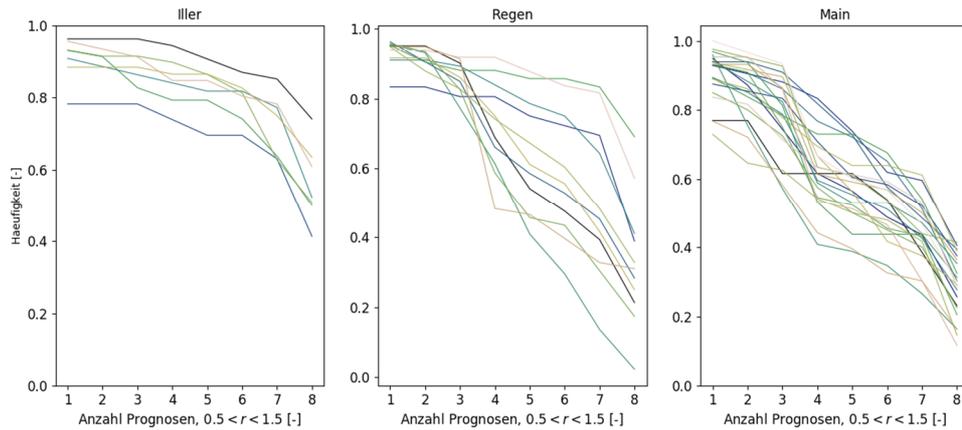


**Abbildung 8-10: Güte Vorhersagen je ML-Algorithmus, Untersuchungsgebiet Iller**

Die hier dargestellten Ergebnisse decken sich im Falle von Main und Regen mit den in Kapitel 6.4 durchgeführten Analysen des Dynamikmoduls. Eine Varianz in den prognostizierten Dynamikparametern eines Ereignisses resultierte demnach in einer Varianz der berechneten Ganglinien. Im Falle der Iller wurde ebenfalls eine Variation der prognostizierten Dynamikparameter festgestellt, die Varianz der Ganglinien fiel aber wesentlich geringer aus. Eine Ursache dafür mag die Verwendung der kalibrierten Parameter als Lernmenge der Zielvariablen sein. In Abbildung 5-18 wurden die analytischen und kalibrierten Parameter der Untersuchungsgebiete verglichen. Dabei wurde festgestellt, dass durch die Kalibrierung vor allem die Entwässerungsgeschwindigkeit in der Iller stark vermindert wurde. Daraus resultierte ebenfalls, dass der absolute Fehler der Algorithmen sank, auch wenn er relativ gesehen gleich blieb. Eine geringere absolute Varianz von  $v_E$  resultierte dann in einer geringeren Varianz der Ganglinien, die mittels des GIUH berechnet wurden.

Einen Hinweis auf die Gültigkeit dieser Erklärung kann über die Anzahl von Kurven mit „gutem“ Ergebnis gegeben werden. Bei einer Gegenüberstellung der Anzahl von Prognosen, deren  $r$  über 0.5 liegt (wie in Abbildung 8-7), mit dem Anteil von Ereignissen bei denen diese Bedingung erfüllt ist werden die gesuchten Unterschiede sichtbar (Abbildung 8-11).

In der Iller (linke Abbildung), verbleibt der Anteil von Ereignissen mit steigender Anzahl von geforderten „guten“ Prognosen auf einem gleichbleibenden Niveau. Im Main (rechte Abbildung) fällt der Anteil hingegen deutlich schneller. Die Teilgebiete des Regens (mittlere Abbildung) zeigen hingegen beide Arten von Verläufen. Dieses Muster zeigt die Ähnlichkeit der Kurven. Da die Kurven in der Iller sehr ähnlich sind, fällt der Anteil „guter“ Kurven nur sehr langsam. Im Regen wurden die  $v_E$ -Parameter ebenfalls durch die Kalibrierung gesenkt, verbleiben aber auf einem höheren Niveau als in der Iller. Daher sind in diesem Gebiet beide Arten von Verläufen erkennbar.

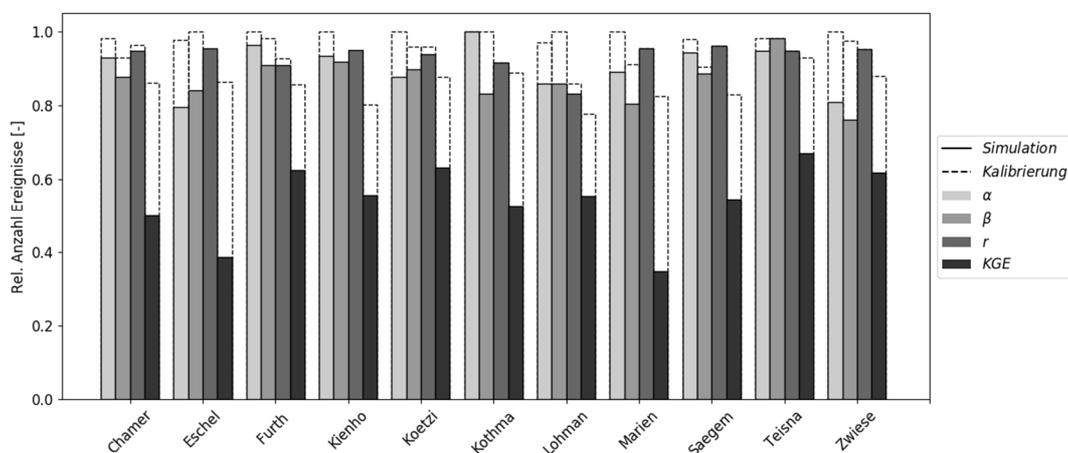


**Abbildung 8-11: Relative Häufigkeit von Ereignissen für die eine (oder mehr Prognosen) vorlagen, mit  $r \geq 0.5$ ; Iller (links), Regen (Mitte), Main (rechts)**

### 8.4 Vergleich Kalibrierungsergebnisse

Bei der Fehlerdiskussion der Fallstudien muss berücksichtigt werden, dass nicht alle Ereignisse mit der GIUH-Modellstruktur bei bekannter Parametrisierung simuliert werden konnten. Zwar wurden im Rahmen der Kalibrierung Ereignisse aus dem Datensatz entfernt die nicht dem Fokus dieser Arbeit entsprachen (vgl. Kapitel 5.4), doch sind weiterhin Ereignisse im Datensatz enthalten, die auch mit bekannten Parametern nur schlecht simuliert werden konnten. Daher muss die Referenzmenge von Ereignissen auf die Anzahl von „guten“ Simulationen während der Kalibrierung reduziert werden.

In folgender Abbildung 8-12 sind zum einen die relativen Anzahlen von Ereignissen des Regen dargestellt, die in der Kalibrierung maximal 0.5 vom Optimum abwichen (gestrichelte, weiße Balken). Zum anderen die Anzahlen von Ereignissen mit mindestens einer guten Prognose in der Fallstudie.



**Abbildung 8-12: Vergl. Kalibrierung und Fallstudie: Normierte Anzahl von Ereignissen mit mind. eine Prognose deren Güte max. Opt.  $\pm 0.5$ ., Untersuchungsgebiet Regen**

In Abbildung 8-12 schwanken die Unterschiede zwischen Kalibrierung und Fallstudie im Hinblick auf  $\alpha$  und  $\beta$  in einem stärkeren Umfang als  $r$ . Die Minderung der Anzahl „guter“ Simulationen durch die Verwendung des Dynamikmoduls beträgt (für den gewählten Grenzwert) zwischen 5-10% im Mittel über alle Teilgebiete, maximal 20% in Zwiesel. Für den KGE fällt dieser Rückgang deutlich höher aus, zwischen 30% und 40% im Mittel, am stärksten in Marienthal, hier fast 50%. Für  $r$  hingegen ist, wenn überhaupt, nur ein marginaler Rückgang zu erkennen.

Der lineare Korrelationskoeffizient  $r$  kann als Kennwert für zeitlichen Zusammenhang von simulierter und beobachteter Ganglinie betrachtet werden. In der entwickelten Modellstruktur wird die Form der Ganglinie durch die Fließwege und die Entwässerungsgeschwindigkeit definiert. Die Prognosefehler von  $v_E$ , die vom Dynamikmodul gemacht wurden, konnten anscheinend durch diese doppelte Abhängigkeit und der validen Abbildung der Geomorphologie abgefangen werden.

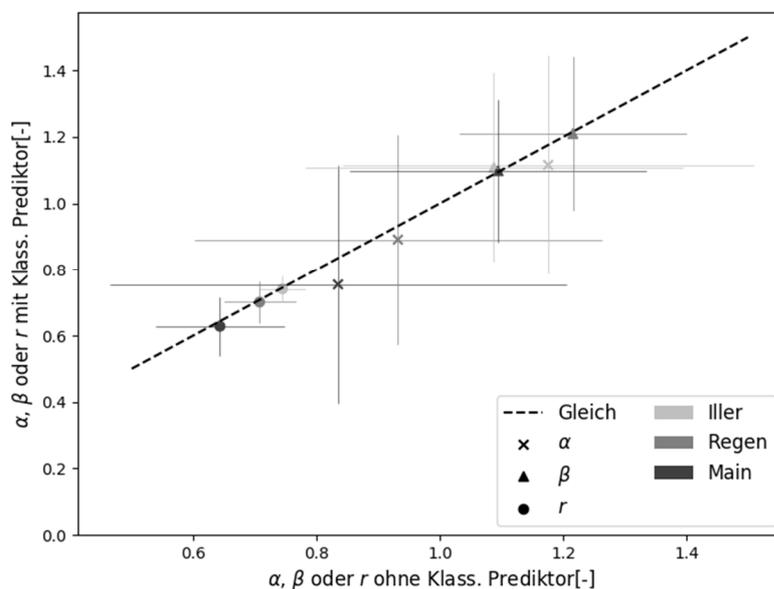
Ein weiteres Problem der Prognosen wird durch den starken Rückgang von Ganglinien mit einem KGE  $> 0.5$  deutlich. Der KGE verrechnet alle anderen Komponenten und reagiert empfindlich auf den Rückgang einer einzelnen Komponente. Da  $\alpha$  und  $\beta$  in geringerem Umfang als der KGE zurückgehen, muss davon ausgegangen werden, dass die prognostizierten Ganglinien häufig nur im Hinblick auf eine einzelne Gütekomponente als „gut“ eingeordnet werden können. Dass eine Prognose alle Gütekriterien erfüllt, ist seltener der Fall. Dies deutet auf die Spezialisierung der Algorithmen hin. Nur selten kann ein Algorithmus die komplette Bandbreite der aktiven hydrologischen Prozesse abbilden. Dieses Ergebnis unterstreicht die Notwendigkeit von Ensemblevorhersagen.

Die hier genannten absoluten Zahlen sind abhängig vom gewählten Grenzwert zur Definition eines „guten“ Ereignisses. Daher besitzen sie keine Allgemeingültigkeit, können aber zur Abbildung der Verhältnisse zwischen Kalibrierung und Fallstudie verwendet werden. Dennoch lässt sich aus diesen Ergebnissen das Verhältnis der systematischen Fehler zwischen GIUH-Modellstruktur und Dynamikmodul ableiten. Gemittelt über alle Komponenten liegen mit dem kalibrierten GIUH-Modell für 93.5% der Ereignisse „gute“ Simulationen vor. Die Fehlerquote lag also bei 6.5%. Bei Verwendung des Dynamikmoduls senkte sich die relative Anzahl auf 81.3%, eine Minderung um 12.2%. Der systematische Fehler des Dynamikmoduls ist demnach doppelt so hoch wie der des GIUH-Modells.

In dieser Analyse bestanden nur vernachlässigbare Unterschiede zwischen den Untersuchungsgebieten, daher wurden die Ergebnisse des Untersuchungsgebietes Regen hier exemplarisch dargestellt (alle weiteren Abbildungen sind dem digitalen Anhang zu entnehmen).

## 8.5 Einfluss Fehlklassifikationen

Für jedes Ereignis der unbeobachteten Gebiete wurde eine Prognose über den zu erwartenden Ereignistyp erstellt. Abhängig von dieser Prognose wurde das entsprechend trainierte datengetriebene Modell zur Prognose von  $v_E$  und  $Q_{Vol}$  verwendet. Die Untersuchung der Typvorhersage (Kapitel 6.1) zeigte, dass bei der Vorhersage der Typen in etwa 10 bis 20% fehlerhaft klassifizierte Ereignisse zu erwarten sind. Um den Einfluss dieser fehlerhaft klassifizierten Ereignisse auf die Gesamtgüte des Modells zu beurteilen, wurden die Fallstudien erneut mit bekannter Klassifikation durchgeführt. Abbildung 8-13 stellt die Median- und Standardabweichung der KGE-Komponenten als Mittelwerte der Untersuchungsgebiete gegenüber. Zusätzlich ist eine Linie dargestellt, die die Positionen der Median Werte bei Wertgleichheit darstellt.



**Abbildung 8-13: Median und Standardabweichung KGE-Komponenten in Iller, Main und Regen bei Simulation mit und ohne Klassifikationsprediktor**

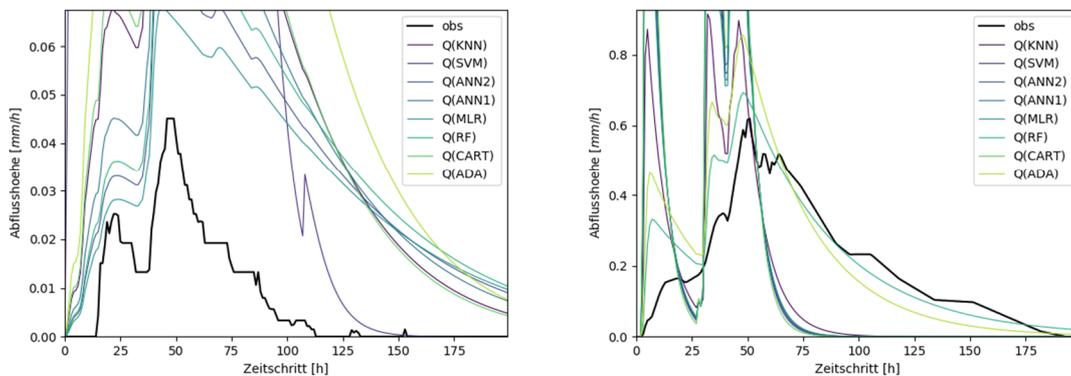
Die Abweichung zwischen den Simulationen mit und ohne Klassifikationsprediktor kann dieser Darstellung nach vernachlässigt werden. Die Median-Werte befinden sich alle nah an der Wertgleichheit und auch der Vergleich der Streuung, durch Fehlerbalken dargestellt, zeigt keine signifikanten Unterschiede. Vorhandene Abweichungen liegen im Rahmen von zufallsbedingten Abweichungen der Lernprozesse der Algorithmen.

Der hier festgestellte geringe Einfluss von Fehlklassifikationen kann zwei Ursachen haben. Entweder sind die Prognoseergebnisse der unterschiedlichen Clustermodelle sehr ähnlich, oder die Fehlklassifikation findet an den Randbereichen der Cluster statt. Die Eindeutigkeit der Klassifikation in diesen Bereichen ist nicht in allen Fällen gegeben. Da in Kapitel 6.2 gezeigt werden konnte, dass eine prozess-orientierte Unterscheidung innerhalb des Dynamikmoduls sinnvoll ist, muss davon ausgegangen werden, dass die Prognosen der Modelle

unterschiedlich ausfallen. Daher erscheint die Uneindeutigkeit der Clusterzuordnung im Randbereich als die wahrscheinlichere Erklärung.

## 8.6 Abhängigkeit Modellgüte zu Ereignisgröße

Abbildung 8-14 zeigt zwei Beispiele für typische Fehler, die in den Fallstudien aufgetreten sind. Bei der Vorhersage des Volumens auf der linken Seite scheint eine starke Überschätzung durch die ML-Algorithmen aufgetreten zu sein. Zwar passt die Wellenform, die Höhe der Welle wurde aber deutlich überschätzt. Auf der rechten Seite lag hingegen eine fehlerhafte Prognose der Entwässerungsgeschwindigkeit vor. Da  $v_E$  zu hoch angesetzt wurde, zumindest von den meisten Algorithmen, fiel der Wellenanstieg zu steil, schnell und zu hoch aus. Der Rückgang der Ganglinien war folglich zu kurz. In dieser Abbildung ist aber auch zu erkennen, dass ADA und RF bessere Prognosen abgaben als der Rest des Ensembles.



**Abbildung 8-14: Fehlerhafte Simulationen (links) Volumenfehler, Ereignis vom 13.02.2005 Bad Berneck; (rechts)  $v_E$  Fehlprognose, Adlerhütte am 28.01.2002**

Dies sind nur zwei mögliche Fehlerarten, aber sie zeigen, welche Probleme bei den Ereignissimulationen auftreten können. In diesem Abschnitt soll untersucht werden, ob sich abhängig von Ereignischarakteristiken Fehler in der Prognose bzw. Simulation häufen. Dazu wurde zunächst der Zusammenhang der Modellgüte und der Ereignisgröße betrachtet. Als Größe eines Ereignisses wurden die Höhe des Abflussscheitels und das Abflussvolumen definiert.

Jedem Ereignis wurde im Kontext aller verwendeten Ereignisse eines Pegels eine Rangzahl für das Abflussvolumen  $Q_{Vol}$  und eine weitere für den Abflussscheitel  $Q_P$  vergeben. Die über die Anzahl der Ereignisse am Pegel normierten Rangzahlen werden im Folgenden als empirische Quantile  $R(Q_{Vol})$  und  $R(Q_P)$  bezeichnet. Der Zusammenhang zwischen Modellgüte und  $R(Q_{Vol})$  sowie  $R(Q_P)$  wurden dann über die Rangkorrelation, Spearman's  $\rho$ , unter Berücksichtigung aller Daten des Untersuchungsgebietes berechnet.

Als erstes Resultat dieser Analyse zeigte sich, dass sich die berechneten Abhängigkeiten der Güte zu  $R(Q_{Vol})$  und  $R(Q_P)$  sehr ähnlich verhalten. Daher sind in Tabelle 8-2 die berechneten Spearman's  $\rho$  – Werte zwischen den Gütekriterien und  $R(Q_P)$  dargestellt. Direkt ersichtlich

sind die mit Algorithmus und Untersuchungsgebiet variierenden Korrelationswerte.

**Tabelle 8-2: Rangkorrelation zwischen KGE-Komponenten der Fallstudien und Abflussscheitel differenziert nach Untersuchungsgebiet und ML-Algorithmus**

Gütemaß	MLR	CART	KNN	SVM	ANN1	ANN2	RF	ADA
<b>Main</b>								
$\alpha$	-0.11	0.09	-0.01	-0.86	-0.08	-0.13	-0.52	-0.78
$\beta$	-0.36	-0.05	-0.22	-0.85	-0.31	-0.36	-0.50	-0.77
$r$	0.37	0.45	0.42	0.03	0.32	0.33	0.24	0.21
<b>Regen</b>								
$\alpha$	-0.20	0.10	-0.05	-0.88	-0.10	-0.10	-0.50	-0.73
$\beta$	-0.51	-0.26	-0.46	-0.87	-0.44	-0.47	-0.57	-0.77
$r$	0.38	0.45	0.45	0.19	0.37	0.35	0.24	0.21
<b>Iller</b>								
$\alpha$	-0.49	-0.19	-0.55	-0.75	-0.51	-0.52	-0.62	-0.68
$\beta$	-0.60	-0.31	-0.62	-0.68	-0.64	-0.62	-0.59	-0.66
$r$	0.18	0.20	0.18	0.18	0.18	0.20	0.10	0.13

Zwischen den Teilgebieten bestanden zwar auch Unterschiede, diese aber lediglich in den absoluten Werten, die generelle Tendenz war hier gleich. Daher wurde diese Analyse auf der Ebene der Untersuchungsgebiete durchgeführt.

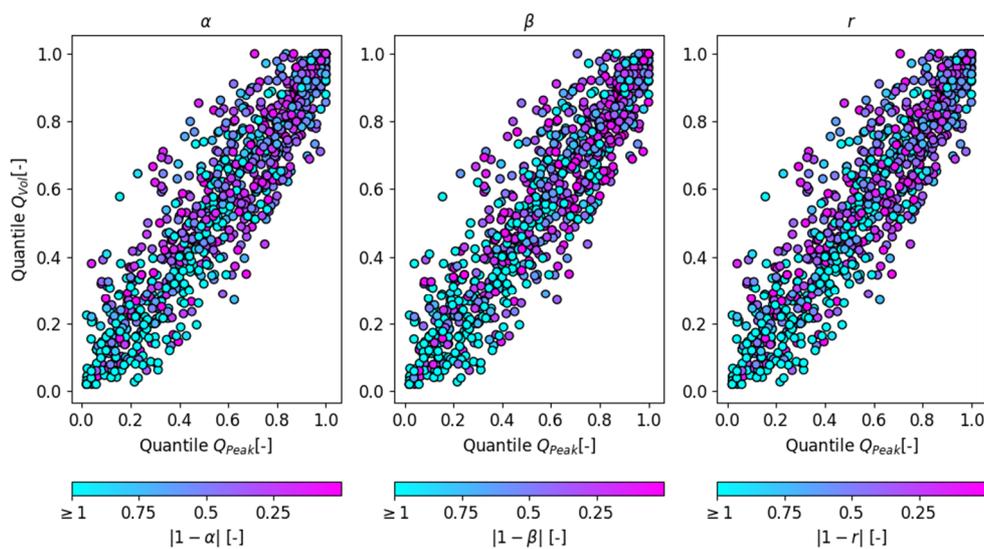
Die Gütekomponenten der Ensemble Methoden RF und ADA sowie SVM, als einziger Einzelschätzer, waren gebietsunabhängig negativ korreliert mit dem Ereignisscheitel. Die Ergebnisse des RF und ADA wiesen zudem eine schwach positive Korrelation zwischen  $Q_P$  und  $r$  auf. Wie in den Analysen des Dynamikmoduls (Kapitel 6.4) wiesen KNN, beide ANN-Strukturen und die MLR ähnliche Abhängigkeitsstrukturen auf. Im Main und Regen hatten die Ergebnisse aller 4 Algorithmen keine, oder nur schwache Korrelation zwischen  $\alpha$  und  $Q_P$ . Dafür waren eine leicht negative Korrelation zwischen  $\beta$  sowie eine leicht positive Korrelation zwischen  $r$  und  $Q_P$  zu sehen. In der Iller verschwindet diese leicht positive Korrelation, dafür tritt hier eine leicht negative Abhängigkeit von  $\alpha$  auf. Die Abhängigkeitsstrukturen von CART ähneln zum Großteil den der anderen Einzelschätzer, lediglich eine leicht positive Korrelation zu  $\alpha$  in Main und Regen weicht von anderen Schätzern ab.

Eine negative Korrelation zwischen  $Q_P$  und  $\alpha$ , bzw.  $\beta$  bedeutet eine reziproke Abhängigkeit der Variablen. Steigt der Scheitelabfluss, sinkt das simulierte Volumen und die simulierte Streuung im Vergleich zur vorhandenen Streuung/Mittelwert. Eine positive Korrelation, wie sie vor allem für  $r$  beobachtet werden kann, bedeutet hingegen, dass die zeitliche Übereinstimmung der Ganglinien mit steigendem Scheitelabfluss steigt. Die hier dargestellten Zusammenhänge gelten in gleicherweise für  $Q_{Vol}$ .

Dass  $r$  mit Ereignisgröße steigt, zeigt eine Stärke des Modells in Hinblick auf zeitliche Übereinstimmung für größere Ereignisse. Ob dies hinsichtlich Volumen und Form der Ganglinien gilt, ist ohne weiteres nicht zu beurteilen. Sollte eine generelle Überschätzung vorliegen,

wären mit Ereignisgröße steigende  $\alpha$ - und  $\beta$ -Werte von Nachteil. Um dies zu beurteilen, wurden in Abbildung 8-15 die empirischen Quantile  $R(Q_P)$  und  $R(Q_{Vol})$  aller Ereignisse im Untersuchungsgebiet Main gegenübergestellt. Die farbliche Markierung der Punkte zeigt den Abstand der Simulationsgüte des entsprechenden Ereignisses vom Optimum des Güte Maßes an.

Das Ergebnis ist zwar keineswegs eindeutig, erkennbar ist aber eine Verdichtung an den Rändern. Die Dichte der dunkel blauen Punkte, also Simulationen, deren Güte nahe am Optimum lag, ist bei höheren Quantilen am höchsten. Umgekehrt ist die Dichte der weißen Punkte, also Güte weit entfernt vom Optimum, bei den niedrigeren Quantilen am höchsten. Die Wahrscheinlichkeit eine gute Simulation zu erhalten steigt also mit der Ereignisgröße.

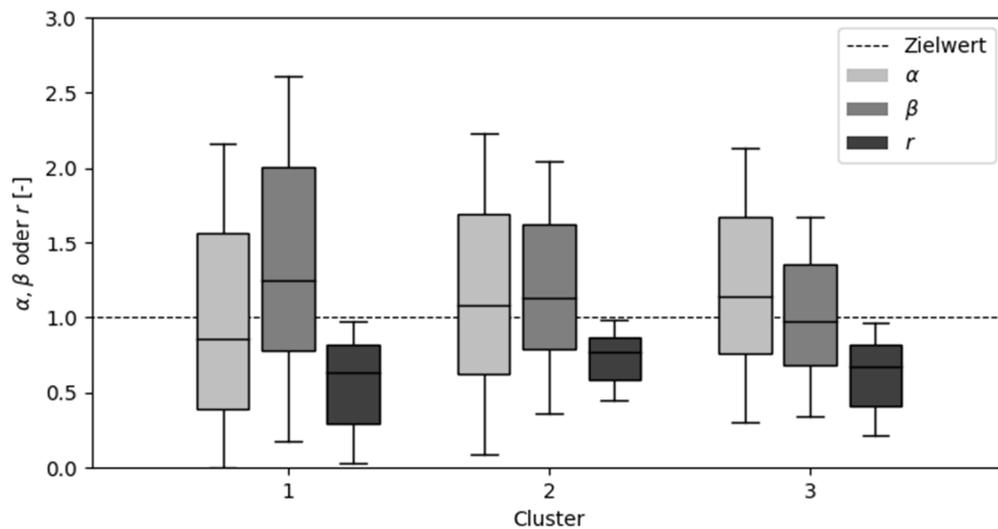


**Abbildung 8-15: Gegenüberstellung empirischer Quantile  $Q_{Vol}$  und  $Q_P$ , farbliche Markierung stellt Abstand zu optimaler Simulationsgüte dar, Ergebnisse des Untersuchungsgebiets Main**

Eine interessante Ergänzung zu dieser Analyse stellt die Ermittlung der mittleren Güte nach Ereignistyp dar. In Abbildung 8-16 sind die Ergebnisse für das Untersuchungsgebiet Regen, gemittelt über alle Algorithmen und Teilgebiete, dargestellt.

Die Clusternummern stehen für die regionale Klassifikation nach Volumen-Scheitelverhältnis. Ereignisse im Cluster 1 weisen ein hohes Volumen im Vergleich zu ihrem Scheitel auf, Ereignisse des Clusters 3 einen hohen Scheitel im Vergleich zu ihrem Volumen. Cluster 2 beinhaltet Ereignisse mit einem ausgeglichenen Verhältnis zwischen diesen Größen.

Mit der Clusternummer steigt der Median von  $\alpha$ , der Median von  $\beta$  fällt hingegen. Die höchste Korrelation wird für Ereignisse des zweiten Clusters erreicht, hier ist auch die  $r$  Varianz minimal. Die hier erkennbaren Strukturen treten so ebenfalls in Main und Iller auf (siehe digitaler Anhang).



**Abbildung 8-16: Mittlere Ereignisgüte je Cluster, Untersuchungsgebiet Regen**

Der Anstieg von  $\alpha$  mit der Clusternummer kann entweder durch einen Anstieg der simulierten Varianz erzeugt werden, oder durch ein Sinken der beobachteten Varianz. Tatsächlich nimmt die beobachtete Varianz der Ereignisse von Cluster 1 bis 3 zu. Daher steigt die simulierte Streuung tatsächlich überproportional an. Das Absinken von  $\beta$  kann, wie zuvor, entweder durch kleiner werdende Werte der Simulation, oder durch steigende beobachtete Werte mit der Clusternummer erzeugt werden. Die beobachteten Mittelwerte der Ereignisse steigen von Cluster 1 bis 3, dies aber in einem wesentlich stärkeren Umfang als der Rückgang von  $\beta$  ausfällt. Das simulierte Volumen fällt also tatsächlich. Dass im Cluster 2 die höchsten Korrelationswerte erreicht werden, kann an dieser Stelle auf die Ergebnisse der vorherigen Analyse zurückgeführt werden. Da die Mehrzahl der höchsten Ereignisse aus diesem Cluster stammt, ist die Wahrscheinlichkeit einer guten Simulation in diesem Cluster höher als in den anderen.

Zusammenfassend konnte festgestellt werden, dass das Modell größere Ereignisse besser nachbilden kann. Eine Fehlerhäufung bei Ereignissen mit geringen Abflussvolumina und Scheitelabfluss konnte identifiziert werden. Im Umkehrschluss stieg die Wahrscheinlichkeit einer guten Simulation mit der Ereignisgröße. Zudem konnte festgestellt werden, dass das Verhältnis zwischen Scheitel und Volumen ebenfalls das Simulationsergebnis beeinflusste. Ereignisse mit einem hohen Volumen und einem geringen Scheitel wurden im Mittel mit einer geringeren Ganglinienstreuung simuliert. Die simulierten Ganglinien sind demnach etwas glatter als die beobachteten Ganglinien. Die Wahrscheinlichkeit eines Volumenfehlers ist bei diesen Ereignissen ebenfalls höher als bei anderen.

## 8.7 Ereignischarakteristika

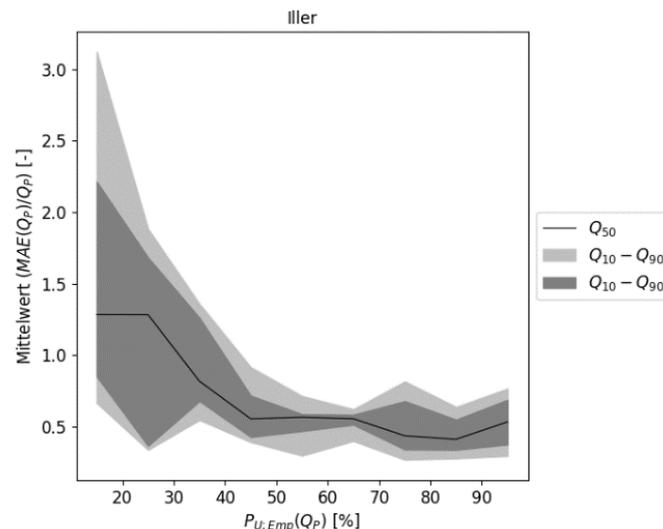
Für praktische Anwendungen von N-A-Modellen ist neben der Güte in Hinblick auf die Ganglinien auch die Abbildung verschiedener Charakteristika der N-A-Ereignisse von Bedeutung. Dem Abflussscheitel kommt dabei meist eine besondere Rolle zu, da er für viele Bemessungsaufgaben die maßgebende Größe ist. Für die operationelle Hochwasservorhersage ist zudem die Anstiegszeit eines Ereignisses von Bedeutung, da sie entscheidend für Maßnahmen- und Evakuierungspläne ist. Außerdem von Bedeutung ist das Abflussvolumen. Da das Abflussvolumen eine Zielvariable der Algorithmen war, wurde die Wiedergabe dieser Charakteristik einer Hochwasserwelle bereits analysiert. Gemäß den Ergebnissen aus Kapitel 6.4 ist für das Volumen ein mittlerer Fehler von 40% zu erwarten.

Daher soll in diesem Kapitel auf die Wiedergabe von Abflussscheitel und Anstiegszeit eingegangen werden. Definiert wurden sie wie folgt: Der Abflussscheitel  $Q_P$  ist der höchste Abflusswert innerhalb des Simulationszeitraumes. Als Anstiegszeit  $t_A$  wurde die Zeitspanne zwischen Beginn des Niederschlags und Erreichen des Scheitels verwendet. Bei mehrgipfligen Ereignissen besteht die Gefahr, dass bei der Auswertung der beobachteten Ganglinie ein anderer Abflussscheitel untersucht wird als bei der simulierten Ganglinie. Ist bspw. bei einer beobachteten Ganglinie der erste Scheitel höher als alle folgenden, ist dieser Scheitel maßgebend für  $Q_P$  und  $t_A$ . Ist in der Simulation der zweite Scheitel höher als der erste (und alle nachfolgenden), ist dieser Scheitel maßgebend. Im Falle von  $Q_P$  ist dies durchaus vertretbar, da in der Bemessungspraxis i.d.R. der höchste Wert verwendet wird. Im Falle von  $t_A$  wird der Fehler aber überschätzt, weil die zeitliche Differenz zwischen erstem und zweitem Scheitel mit in die Berechnung der zeitlichen Abweichung von  $t_A$  eingeht. Um die Analyse genauer zu machen, müsste ein Abfluss-Grenzwert eingeführt werden, ab dem bspw. im operationellen Betrieb eine Warnung ausgegeben würde. Überschreitet der Abfluss diesen Grenzwert würde dieser Zeitpunkt als  $t_A$  gewertet werden. Um die Einführung eines weiteren Grenzwertes zu vermeiden, wurde für diese einfache Auswertung der Ergebnisse das zuvor beschriebene Vorgehen verwendet. Daher muss bei der Auswertung von  $t_A$  berücksichtigt werden, dass der ermittelte Fehler durchaus geringer sein kann als berechnet.

Der mittlere absolute Fehler (MAE) zwischen simuliertem und beobachtetem  $Q_P$ , normiert auf den Wert des beobachteten  $Q_P$ , beträgt 53%, gemittelt über alle Teileinzugsgebiete und Ereignisse. Ergänzend wurde der  $BIAS(Q_P)$  analysiert (nicht normierte mittlere Abweichung), mit dem Ergebnis, dass die Werte der Teileinzugsgebiete um einen BIAS von null schwanken. Über- und Unterschätzung traten demnach in gleichem Maße auf. Dieser sehr hohe zu erwartende Fehler ist eine klare Schwäche des Modells. Es ist aber zu betonen, dass dieses Ergebnis für die Vorhersage in unbeobachteten Gebieten ermittelt wurde, ein größerer Fehler daher zu erwarten war. Erzeugt wird dieser Fehler vor allem durch das Abflussbildungsmodell und die zu erwartenden Fehler bei der Schätzung des Abflussbildungsparameters. Durch die einfache multiplikative Abflussbildungskomponente des Modells beeinflusst das Abflussvolumen allein die Ganglinienhöhe ohne ihre Form zu beeinflussen. Ein zu hoch

geschätztes Volumen wird bei richtiger Form der Welle einen Fehler der gleichen Größenordnung in der Schätzung des Scheitelabflusses verursachen. Wie zuvor festgestellt, lag der zu erwartende Fehler bei Schätzung von  $Q_{Vol}$  bei rund 40%, also in einer vergleichbaren Größenordnung.

Wie im vorhergehenden Kapitel 8.6 beschrieben, hängt die Modellgüte von der Ereignisgröße ab. Mit folgendem Vorgehen wurde untersucht, ob der  $MAE(Q_P)$  ebenfalls von der Ereignisgröße abhängt: Für jedes Ereignis wurde die empirische Unterschreitungswahrscheinlichkeit  $P_U$  im betreffenden Teileinzugsgebiet berechnet. Anschließend wurden die Ereignisse auf der Ebene der Untersuchungsgebiete basierend auf diesen Unterschreitungswahrscheinlichkeiten gruppiert. Zusammengefasst wurden Ereignisse, deren Unterschreitungswahrscheinlichkeiten bspw. zwischen 10% und 20% lagen. Jede Gruppe umfasste also 10% der verfügbaren Ereignisse, je nach Untersuchungsgebiet etwa 70 Ereignisse. Abbildung 8-17 stellt die Quantile des relativen  $MAE(Q_P)$  der Gruppen den zugehörigen Unterschreitungswahrscheinlichkeiten von  $Q_P$  gegenüber.



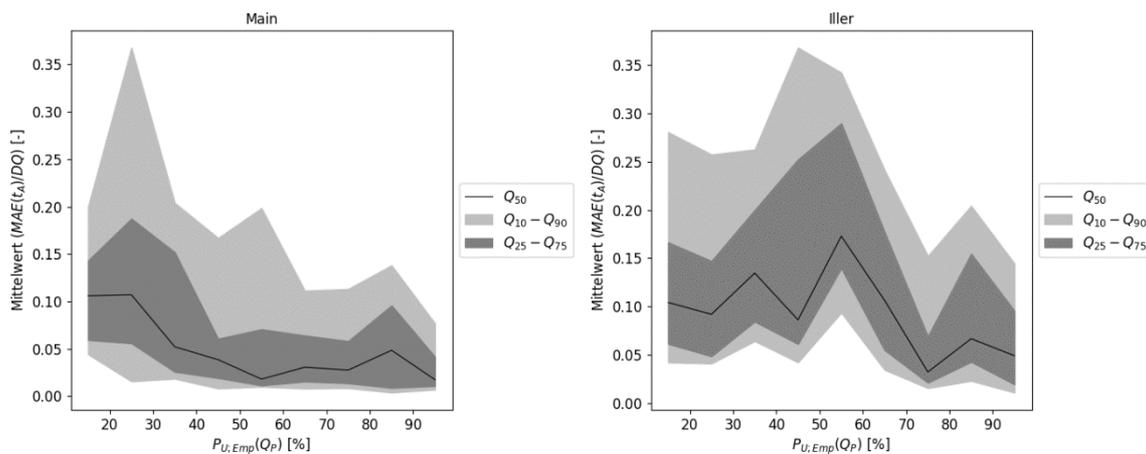
**Abbildung 8-17: Relativer  $MAE(Q_P)$  gegenüber der empirischen Unterschreitungswahrscheinlichkeit des Scheitels  $Q_P$ , Untersuchungsgebiet Iller**

Zu erkennen ist, dass im dargestellten Gebiet sowohl der Unsicherheitsbereich wie auch der Median-Fehler mit zunehmender Ereignisgröße abnehmen. Dies bestätigt die Vermutung, dass das Modell für Hochwasserereignisse, also Ereignisse mit einer höheren Unterschreitungswahrscheinlichkeit des Scheitels, bessere Ergebnisse liefert als für kleine und Kleinstereignisse. Der hier dargestellte Verlauf kann als repräsentativ für alle drei Untersuchungsgebiete angesehen werden die Darstellungen der anderen Gebiete können dem digitalen Anhang entnommen werden.

Deutlich geringer fiel der Fehler der Anstiegszeit aus. Normiert über die Ereignisdauer ergab sich mittlerer Fehler von rund 10%. Bei einer Ereignisdauer von bspw. 50 Stunden ist demnach mit einem Fehler in der Anstiegszeit von  $\pm 5$  Stunden zu rechnen. Die Anstiegszeit wird

vor allem definiert über die Schätzung der Entwässerungsgeschwindigkeit und die Geomorphologie. Bereits der hohe Korrelationskoeffizient beim Vergleich der Ganglinien deutete eine gute zeitliche Übereinstimmung zwischen Beobachtung und Simulation an. Die Schätzung der Anstiegszeit ist demnach als Stärke des Modells zu betonen.

Abschließend wurde untersucht, ob wie für  $Q_P$ , auch für  $t_A$  eine Abhängigkeit zur Ereignisgröße besteht. Daher wurde die zuvor beschriebene Analyse der Unterschreitungswahrscheinlichkeiten von  $Q_P$  erneut durchgeführt. In Abbildung 8-18 ist der mittlere Fehler von  $t_A$  normiert auf die Ereignislänge  $DQ$  den Unterschreitungswahrscheinlichkeiten gegenüber gestellt.



**Abbildung 8-18: MAE( $t_A$ ) normiert auf Ereignisdauer, gegenüber der empirischen Unterschreitungswahrscheinlichkeit des Scheitels  $Q_P$ , Untersuchungsgebiet Main (links) & Iller (rechts)**

Im Untersuchungsgebiet Main zeigte sich ein ähnliches Bild, wie zuvor für  $Q_P$  beobachtet wurde. Der Median und die Unsicherheitsbänder nehmen mit steigender Unterschreitungswahrscheinlichkeit ab. Im Main fällt der Median Fehler ab ca.  $P_U = 40\%$  auf unter 5%. Im Untersuchungsgebiet Iller und Regen (letzteres im digitalen Anhang dargestellt) steigt der Unsicherheitsbereich hingegen in der Mitte der Abbildung an. In einem Bereich von  $P_U = 40\% - P_U = 70\%$  ist dieser Anstieg zu verorten. Danach fällt auch in diesen Gebieten der zu erwartende Fehler auf unter 5%. Eine mögliche Erklärung für diesen Anstieg kann das vermehrte Vorhandensein von mehrgipfligen Ereignissen in diesem Bereich der Unterschreitungswahrscheinlichkeit sein.

Die hier dargestellten Analysen wurden mit dem RF durchgeführt, da sich dieser Algorithmus in den vorhergehenden Studien als bester, einzelner Algorithmus bewährt hat. Alternativ kann die Analyse mit dem GLUE-Median des Ensembles durchgeführt werden. Die Ergebnisse verändern sich unwesentlich. Der mittlere Fehler für  $Q_P$  lag bei 56%, für die Anstiegszeit bei 16%. Beide Kriterien lagen durch den Einbezug weiterer Algorithmen etwas höher. Für den GLUE-Median ergaben sich aber die gleichen Abhängigkeiten wie in diesem Kapitel dargestellt. Die Abbildungen zur Auswertung des GLUE-Median können dem digitalen Anhang entnommen werden.

Zusammenfassend kann festgehalten werden, dass die Anstiegszeiten von N-A-Ereignissen mit dem GIUH-ML Modell sehr gut vorhergesagt werden können. Vor allem für Hochwasserereignisse reduziert sich der mittlere Fehler auf unter 5% der Ereignisdauer. Die Vorhersage des Abflussscheitels ist mit einem mittleren Fehler von rund 53% als sehr unsicher einzustufen. Auch für diesen Kennwert nimmt die Güte mit steigender Ereignisgröße zu.

## 9 Weiterführende Analysen

---

In diesem Kapitel werden die Ergebnisse zweiter Analysen, die über die zuvor dargestellten Fallstudien hinausgehen vorgestellt. Untersucht wurde zum einen, ob die Ergebnisse des in dieser Arbeit vorgestellten Modells, gegenüber eines etablierten Modells und einer üblichen Regionalisierungsstrategie vergleichbare Ergebnisse liefern kann. Die Ergebnisse der komparativen Modellstudie sind im folgenden Kapitel 9.1 dargestellt. Zum anderen wurde untersucht, ob die trainierten ML-Modelle Rückschluss auf hydrologische Prozesse zulassen. In Kapitel 9.2 wird die Untersuchung und Deutung der datengetriebenen Modelle vorgestellt.

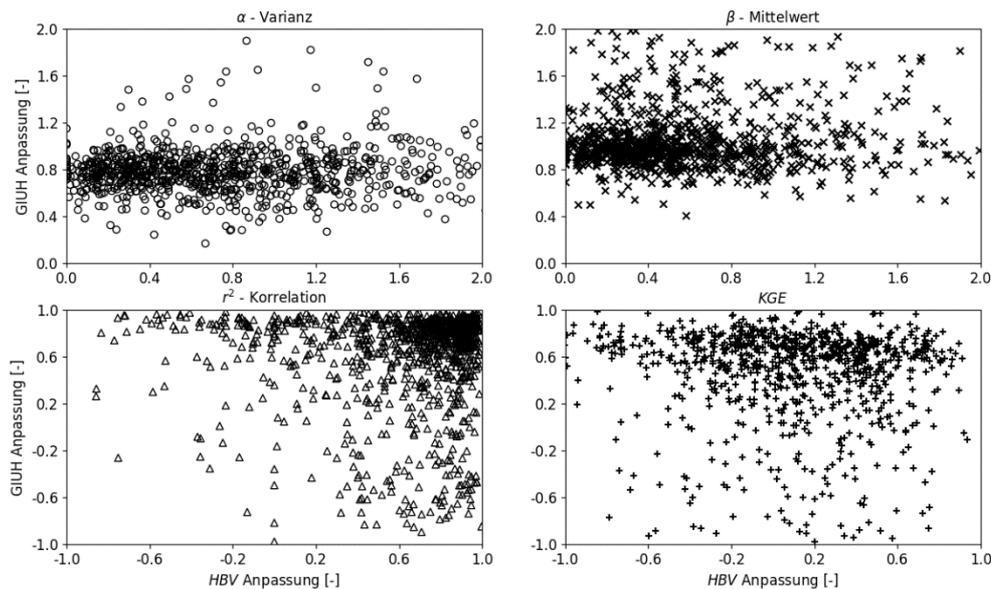
### 9.1 Komparative Modellstudie

Um die Simulationsergebnisse des entwickelten GIUH-ML Modells einordnen zu können, wurde eine weitere Modellstudie im Untersuchungsgebiet Main durchgeführt. Verwendet wurde das HBV Modell (Lindström et al., 1997) in der Umsetzung von Tyralla & Schumann (2013). Um die Ergebnisse vergleichbar zu halten, wurden dieselben Daten und eine ähnliche Kalibrierstrategie verwendet.

Zunächst wurden für jedes Einzugsgebiet drei individuelle Modelle angepasst, eines für jedes Cluster der regionalen  $t_Q$ -Klassifikation. Innerhalb der Teilgebiete wurde eine einfache, an der Landnutzung orientierte zonale Unterteilung vorgenommen: Je eine Zone für Landwirtschaft, Forst und unbedeckte Flächen. Siedlungsflächen wurden der Zone für unbedeckte Flächen hinzugerechnet. Kalibriert wurden 18 Parameter je Zone und 6 je Teilgebiet, insgesamt wurden also 60 Parameter angepasst. Verwendet wurde wie in Kapitel 5.4.2 der Optimierungsalgorithmus BOBYQA (Johnson, 2018; Powell, 2009). Als zu maximierende Zielfunktion wurde der KGE aller N-A Ereignisse des jeweiligen Teilgebietes und Clusters verwendet, um im Mittel alle Ereignisse mit einem einzigen Parametersatz abbilden zu können. Die KGE-Werte und -Komponenten jedes Ereignisses der HBV-Kalibrierung sind in Abbildung 9-1 der Anpassungsgüte des GIUH-Modells gegenübergestellt.

Deutlich wird, dass die individuellen Anpassungsergebnisse von HBV deutlich stärker streuen als die des GIUH-Modells. Dies hat natürlich zwei einfache Gründe: das HBV-Modell ist in seiner grundsätzlichen Konzeption nicht für die Ereignissimulationen vorgesehen. Es ist vielmehr für die Simulation von langen Zeitreihen unter Berücksichtigung aller Wasserhaushaltskomponenten ausgelegt. Die guten Ergebnisse der Vergleichskalibrierung aus Kapitel 5.4.3 sowie der lange Betrachtungszeitraum der Simulation trugen aber dennoch

zur Auswahl dieses Modells bei. Zwar wurden nur kurze Abschnitte simuliert, diese Abschnitte lagen aber in einem langjährigen Kontext. Daher musste die Modellstruktur, für die ein einheitlicher Parametersatz gefunden werden sollte, ein breites Spektrum von Prozessen abbilden können. Das gewählte Modell versprach diesen Anforderungen gerecht zu werden. Der zweite Grund ist, dass die Parameter des GIUH-Modells zeitlich verteilt sind. Für jedes Ereignis wurden individuelle Parametersätze definiert. Der Vergleich ist an dieser Stelle also nur bedingt zulässig.

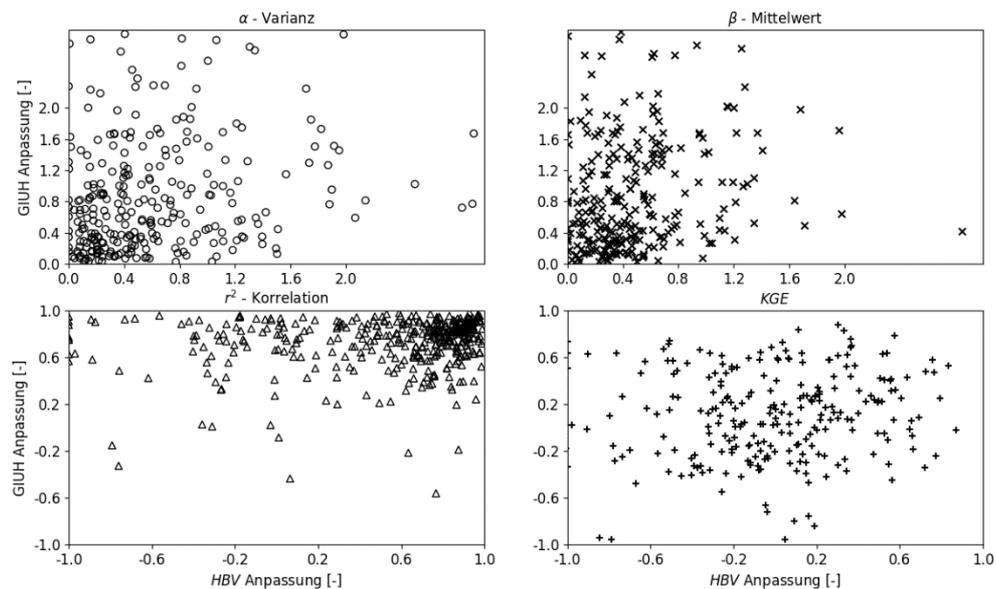


**Abbildung 9-1: Gegenüberstellung der KGE-Komponenten nach der Kalibrierung des GIUH Modells und HBV**

Zu beachten bleibt dennoch, dass durch die wenigen Parameter des GIUH-Modells eine Anzahl von kalibrierten Parametern in der gleichen Größenordnung liegt wie für das HBV-Modell. Sind es bei HBV 60 Parameter je Teileinzugsgebiet sind es beim GIUH im Mittel 100 Parameter je Teilgebiet (2 Parameter für ~50 Ereignisse je Teilgebiet). Diese Notwendigkeit, der ereignisbasierten Parameterschätzung gilt als eine der Schwächen des GIUH-Modells (bspw. Rigon et al. (2016)). Die vorgeschlagene Kopplung des GIUH-Modells mit einer ML-basierten Schätzung der ereignisbezogenen Parameter bietet einen Lösungsansatz für diese Problematik.

Um einen valideren Modellvergleich durchzuführen, wurde eine Anwendung des GIUH-ML Modells auf überwachte Einzugsgebiete durchgeführt. Dabei wurden 60% der verfügbaren Ereignisse zum Trainieren des RF-Algorithmus verwendet, die verbleibenden 40% wurden zu Validierung verwendet. Ausgewählt wurde der RF, weil mit dem RF in den Fallstudien, die besten Ergebnisse eines einzelnen Algorithmus erzielt werden konnten. Der Grenzwert 60% wurde gewählt, um die mittlere Anzahl der Parameter mit der Anzahl von HBV-Parametern je Einzugsgebiet anzugleichen. Die Gütekriterien der validierten Ereignisse sind in Abbildung 9-2 denen der Kalibrierung des HBV-Modells gegenübergestellt.

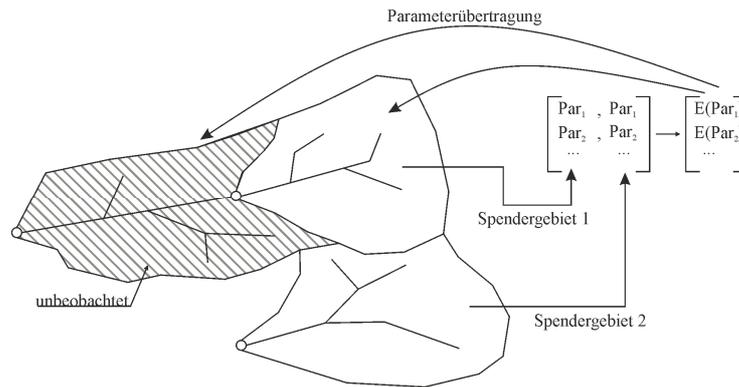
Erkennbar ist, dass die Streuung der Kriterien nun in ähnlichem Umfang ausfällt. Die Mediane der KGE-Komponenten zeigen, dass die Varianz der Ereignisse mit GIUH mit  $\alpha = 1.02$  näher am Optimum liegt als mit dem HBV Modell, das einen Wert von  $\alpha = 0.59$  erzielt. Für  $\beta$  und  $r$  verhält es sich ähnlich, mit dem GIUH-Modell liegen die Werte bei  $\beta = 1.35$  und  $r = 0.79$ . Mit dem HBV Modell liegen sie bei  $\beta = 0.48$  und  $r = 0.72$ . Der Median KGE ist für HBV mit  $-0.06$  besser als der des GIUH-Modells mit  $-0.18$ .



**Abbildung 9-2: Gegenüberstellung der KGE-Komponenten regionales GIUH-RF Modell und HBV Kalibrierung**

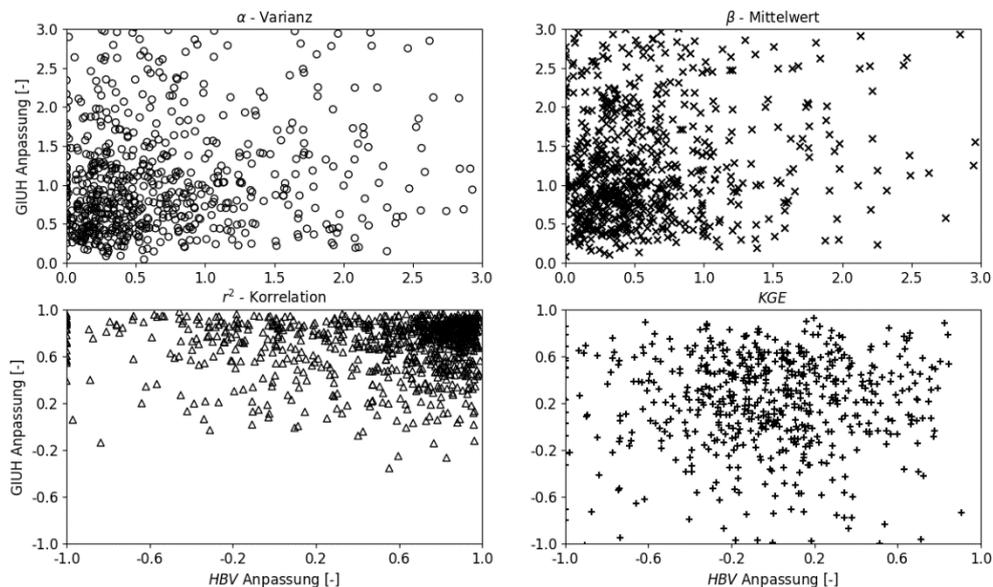
Zwar sind die Gütekriterien, vor allem der KGE, teilweise sehr weit entfernt vom Optimum, bieten aber einen ersten Eindruck über die Nützlichkeit des GIUH-ML Modells. Im Vergleich zum etablierten HBV-Modell wurde in dieser Anwendung auf 396 Ereignisse eine vergleichbare Anpassungsgüte erzielt, in einzelnen Komponenten teilweise eine bessere. Die Ergebnisse erhalten noch mehr Gewicht, wenn berücksichtigt wird, dass die HBV-Ergebnisse aus der Kalibrierungsphase stammen und die GIUH-ML-Ergebnisse aus der Validierung des Modells. Zudem wurde nur ein einziger Algorithmus verwendet, keine selektive Datenkomposition durchgeführt und die Vorteile des Modellensembles blieben in dieser kurzen Studie ungenutzt.

Der nächste Schritt dieser vergleichenden Modellstudie war die Regionalisierung der HBV-Parameter und der anschließende Vergleich der Modellgüte in unbeobachteten Gebieten. Ein häufig verwendetes Vorgehen zur Regionalisierung von HBV-Parametern ist der Übertrag eines mittleren Parametersatzes von Spendergebieten auf das unbeobachtete Gebiet.



**Abbildung 9-3: Schematische Darstellung HBV-Regionalisierung**

In Abbildung 9-3 ist das Vorgehen der Regionalisierung schematisch dargestellt. Einem Spendergebiet, das ein Oberlieger des Zielgebietes ist, werden dabei ebenfalls die regionalisierten Parameter zugewiesen. Somit soll eine verbesserte Vergleichbarkeit zu den Ergebnissen des GIUH-Modells hergestellt werden. Da keine Teilgebietsdifferenzierung im GIUH-Modell vorgesehen wurde und damit die Parameter von Oberliegern nur indirekt in das ML-Modell einfließen, wurden die HBV Parameter aller Oberlieger des Zielgebietes ebenfalls angepasst. Im Rahmen dieser Vergleichsstudie, wurde die in der Literatur häufig verwendete Ähnlichkeitsdefinition über Nähe gewählt. Um Unter- und Oberlieger Beziehungen mit zu berücksichtigen, wurde das Ranking mit Top-Kriging (vgl. Kapitel 7.2.2) verwendet und nur die Parameter des ähnlichsten Gebietes verwendet. Die Ergebnisse der Anwendung der regionalisierten Parameter sind denen des GIUH Modells mit RF als Parameterschätzer in Abbildung 9-4 gegenübergestellt.



**Abbildung 9-4: Gegenüberstellung KGE-Komponenten GIUH-RF und HBV bei Anwendung in un beobachteten Gebieten**

Anzumerken ist, dass das HBV-Modell einige Ereignisse in dieser Anwendung nicht modellieren konnte. Durch ungünstige Parametrisierung tritt in einigen Gebieten vor allem bei Ereignissen mit kleinen Volumina und Scheitelabflüssen überhaupt kein Abfluss auf. Die Ereignisse sind in Abbildung 9-4 jeweils auf der Ordinatenachse eingetragen, am besten erkennbar im Panel von  $r$  (unten links). Wieder liegen die Mediane des GIUH-RF mit  $\alpha = 1.08$ ,  $\beta = 1.24$  und  $r = 0.76$  wesentlich näher am Optimum als die HBV Median Werte von  $\alpha = 0.60$ ,  $\beta = 0.45$  und  $r = 0.70$ . Die KGE des GIUH-RF Modells liegt mit  $KGE = 0.06$  in dieser Anwendung knapp höher als des HBV-Modells mit  $KGE = -0.08$ .

Zu beobachten ist, dass beide Modelle im Vergleich zur vorherigen Anwendung auf beobachtete Gebiete eine vergleichbare Güte erreichten. Die Median KGE-Komponenten des HBV-Modells veränderten sich kaum, die des GIUH-Modells haben sich sogar verbessert. Letzteres kann unter anderem durch die Verwendung der selektiven Datenkomposition erklärt werden. Wieder sind die Prognosen des GIUH-ML Modells einzelne Gütekomponenten betreffend besser als die des HBV-Modells, in der gesamt Güte liegen sie aber gleich auf mit dem HBV-Modell.

In dieser Analyse konnte gezeigt werden, dass die entwickelte Kopplung von GIUH-Modell mit *Machine Learning* eine sinnvolle Ergänzung zur dynamischen Schätzung der Parameter darstellt. Im Vergleich zum HBV-Modell konnten, trotz ungenutzter Potentiale des GIUH-ML Modells, teils bessere, teil gleichwertige Ergebnisse in beobachteten und unbeobachteten Gebieten erreicht werden. Die Ergebnisse der Vergleichsstudie sind natürlich anfechtbar. Die Auswahl des Modells, Kalibrierstrategie, Übertragungsfunktion etc. sind subjektive Entscheidungen, die weitergehend diskutiert werden können. Die getroffenen Entscheidungen spiegeln aber gängige Methoden aus der wissenschaftlichen Praxis wieder und sollten die Vergleichbarkeit der Ergebnisse ermöglichen.

## 9.2 Prozessabbildung der ML-Algorithmen

Eine der wesentlichen Neuerungen dieser Arbeit ist die Verwendung von *Machine Learning* Regressoren, um hydrologische Prozesse zu imitieren. Dies steht im Gegensatz zur konzeptionellen Herangehensweise eine Prozessannahme über den funktionalen Zusammenhang von Niederschlag und Abfluss in einem Modell nachzubilden. ML-Ansätze stehen daher im Ruf lediglich eine funktionale Ergänzung darzustellen (Hrachowitz et al., 2013; Blöschl et al., 2013) und keine Einsicht in hydrologische Prozesse zu ermöglichen. Shen et al. (2018) hingegen sehen ML-Ansätze als einen alternativen Weg, um Erkenntnisse über Prozesse zu gewinnen. Das Fehlen einer Vorab-Einschätzung der Prozesse sehen sie gerade als Vorteil, da Abhängigkeiten aufgedeckt werden könnten, die sonst unbeachtet blieben.

In diesem Abschnitt soll ein erster Schritt in die Richtung Erkenntnisgewinn durch die gezielte Auswertung der ML-Modelle getan werden. Es soll untersucht werden, wie die Modelle einen Zusammenhang zwischen den Prädiktoren und den Zielvariablen herstellen und

welche der Faktoren den höchsten Einfluss hatten. Das Hauptproblem ist, dass die angelegten Strukturen, vor allem die der ANNs, nach menschlicher Logik nicht zu interpretieren sind (Goodfellow, Bengio & Courville, 2016). Daher musste ein alternativer Ansatz entwickelt werden.

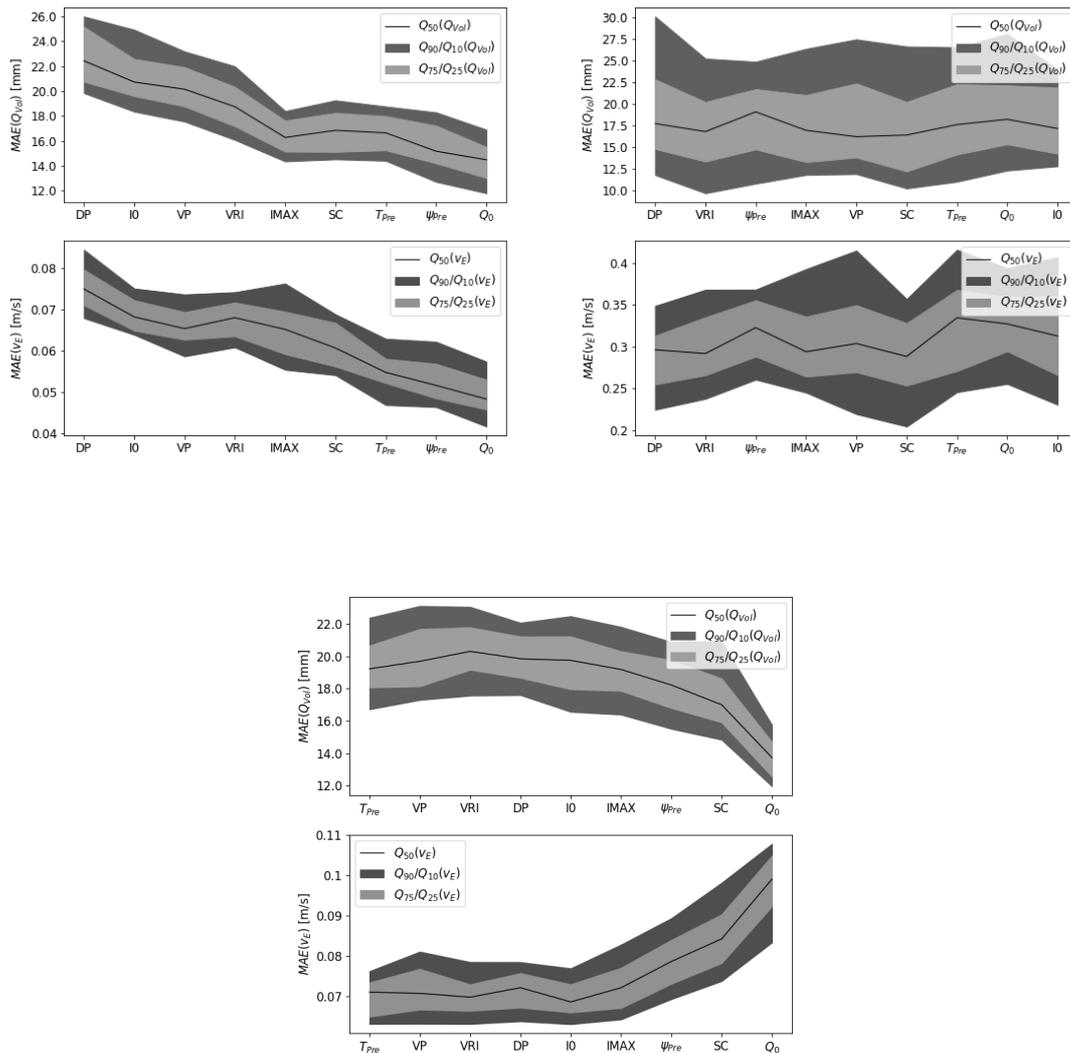
Die Analyse findet auf der Ebene der Modellanwendung statt, da diese Ergebnisse interpretierbar sind. In einer iterativen Studie sollte ermittelt werden, welcher Ereignisindikator den MAE der Prognose von  $Q_{Vol}$  und  $v_E$  am stärksten senkt. Denn ein Prädiktor mit hohem Informationsgehalt über eine Zielvariable müsste einen geringeren MAE in einer Regression erzeugen als ein Prädiktor ohne nennenswerte Information. Daher wurde für die Prädiktoren eine Reihenfolge ermittelt, geordnet nach der Reduktion des MAE der durch ihre Verwendung erzielt wurde.

Ermittelt wurde diese Reihenfolge mit folgendem Vorgehen: Um den wichtigsten Prädiktor zu finden, wurde eine Regression zwischen einem der Prädiktoren und den Zielvariablen durchgeführt. Die Residuen der Regression, also der MAE, wurden gespeichert. Dies wurde für alle Prädiktoren durchgeführt und anhand der gespeicherten MAE Werte, der Prädiktor der den MAE minimierte gewählt. Der zweitwichtigste Prädiktor nach demselben iterativem Schema ermittelt, nur das bei der Regression nun zwei Prädiktoren verwendet wurden: der wichtigste Prädiktor plus ein weiterer Prädiktor. Mit jedem Schritt vergrößert sich somit die Anzahl der Prädiktoren in der Regression. Am Ende dieser iterativen Suche ist die Reihenfolge aller Prädiktoren, je Untersuchungsgebiet und Algorithmus, zur effektivsten Senkung des MAE bekannt.

In jedem Iterationsschritt wurde die Anpassung und Validierung der Algorithmen mit den ausgewählten Prädiktorensatz 50-Mal wiederholt. Dabei wurden 40% der Daten zur Validierung verwendet, die individuellen Validierungsereignisse wurden zufallsbedingt in jedem Durchgang von neuem bestimmt. Als Entscheidungsvariable zurückgegeben wurden der Median der berechneten MAE-Werte sowie die 10-, 25-, 75- und 90%-Quantile. Abbildung 9-5 stellt exemplarisch drei Verläufe des MAE für  $Q_{Vol}$  (blau) und  $v_E$  (rot) in drei unterschiedlichen Anwendungen dar. Mittels empirischer Quantile konnten neben der Median-Kurve Unsicherheitsbänder dargestellt werden.

Über alle Untersuchungsgebiete und Algorithmen traten die in Abbildung 9-5 dargestellten Formen auf. Formtyp I (links oben) zeigt ein gleichmäßiges Lernen, der MAE sinkt kontinuierlich mit wachsender Anzahl von Prädiktoren. Die Form ist vor allem für RF und ADA zu beobachten. Typ II (rechts oben) beschreibt einen gleichbleibenden MAE. Trotz Zunahme weiterer Prädiktoren stellt sich keine signifikante Verbesserung der Prognose ein. Typ III (unten) zeigt einen gegenläufigen Lernprozess. Dabei nimmt der MAE von  $Q_{Vol}$  mit zunehmenden Prädiktoren ab, während der MAE von  $v_E$  steigt. Dieser Typ ist besonders deutlich bei ANN2, tritt aber vermehrt in den Untersuchungsgebieten Main und Regen auf.

Die in Abbildung 9-5 zuerst aufgeführten Prädiktoren sind die Parameter, die den MAE am stärksten minimierten. Daher kann ihnen der höchste Informationsgehalt zu den Zielvariablen zugerechnet werden. Nun sind einige der Prädiktoren korreliert, in Kapitel 6.3.1 wurden sie daher in Gruppen zusammengefasst: *N* für Indikatoren der Niederschlagscharakteristika, *C* für die Charakterisierung von Vorfeuchtebedingungen und *S* für Schneeindikatoren. Für eine vergleichende Auswertung erschien es sinnvoll die Prädiktoren nach diesen Gruppen zu erfassen, da sie teilweise zufallsbedingt gegeneinander ausgetauscht werden konnten.



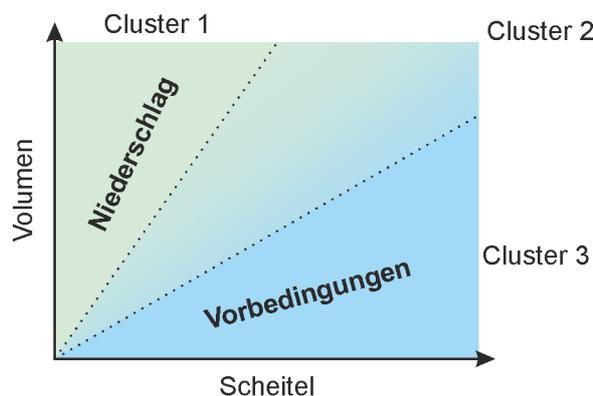
**Abbildung 9-5: Veränderung des MAE (Rot:  $v_E$ ; Blau:  $Q_{voi}$ ) durch geordnete Hinzunahme von Prädiktoren. (links oben) RF Main, Cl. 1; (rechts oben) KNN Regen, Cl. 2; (unten) ANN2 Main, Cl. 1.**

Tabelle 9-1 stellt für drei unterschiedliche Algorithmen die jeweilige Gruppe der Prädiktoren auf den ersten 4 Rängen der ermittelten Ordnungen dar. Ermittelt wurde diese Reihenfolge getrennt für alle Algorithmen, Untersuchungsgebiete und Cluster. Zudem ist die Form der Lernkurven aufgelistet.

**Tabelle 9-1: Ausgewählte Parametergruppe auf Position 1-4 und Form der MAE-Kurven, differenziert nach Untersuchungsgebiet, Algorithmus und Cluster. Parametergruppen: N – Niederschlag, C – Vorfeuchte, S-Schnee**

Gebiet	Algorithmus	Indikatorgruppen			Form		
		Cl. 1	Cl. 2	Cl. 3	Cl. 1	Cl. 2	Cl. 3
Iller	RF	NNSN	NCCN	NCNC	I	I	I
	ANN2	NCCN	CSNN	CCNC	II	II	II
	KNN	NNSN	NCCS	SCCC	I	I	I
Main	RF	NNCN	NNNC	CCSN	I	I	I
	ANN2	NNCN	SNCN	CNNS	III	III	III
	KNN	NNCC	NNCC	NCCN	III	III	II
Regen	RF	NCCN	NNNC	SCCC	I	I	I
	ANN2	NNNS	NCCN	SCNN	III	III	III
	KNN	NCNN	NCNN	NCCN	I	III	II

Zwar ist das dargestellte Ergebnis nicht eindeutig, lässt aber grundlegende Rückschlüsse zu. Wird die Häufigkeit der Gruppen je Cluster betrachtet, sind in Cluster 1 vermehrt N-Indikatoren zu finden und in Cluster 3 vermehrt C. Dies bedeutet, dass in Cluster 1 häufig Niederschlagscharakteristika den höchsten Zusammenhang zu den Zielvariablen aufwiesen. In Cluster 3 besteht ein stärkerer Zusammenhang zwischen den Vorbedingungen des Gebietes und den Zielvariablen. Wird S als Zustandsbeschreibung eines EZGs vor Ereignisbeginn angesehen, wird die Präferenz in Cluster 3 noch deutlicher. Cluster 1 steht für Ereignisse mit einem hohen Volumen im Vergleich zum Scheitel. Zur Beschreibung dieser volumendominierten Ereignisse sind Niederschlagscharakteristika die entscheidende Größe. Ereignisse, die einen hohen Scheitel im Vergleich zu ihrem Volumen aufweisen, hier als Cluster 3 bezeichnet, können eher über die Vorfeuchte Bedingungen beschrieben werden. Dieser Zusammenhang wird in Abbildung 9-6 schematisch zusammengefasst.



**Abbildung 9-6: Unterschiedliches Erklärungspotential der Indikatorgruppen abhängig vom Volumen/Scheitel Verhältnis**

Dieser Schlussfolgerung liegt zwar eine Vielzahl von Simulationen zugrunde, die Ergebnisse reichen aber nicht aus, um sie eindeutig zu belegen. Die Analyse der Algorithmen liefert an dieser Stelle eher eine Anfangsvermutung über funktionale Zusammenhänge. Ob nun Niederschlagscharakteristika mit hohen Volumina oder geringen Scheiteln zusammenhängen und wie dieser Zusammenhang aussieht, muss durch weitere Untersuchungen erörtert werden. Bspw. könnte über Entropie-Betrachtungen (wie in May et al. (2008) oder Nearing & Gupta (2015)) der Zusammenhang der Variablen genauer untersucht werden. Auch eine getrennte Betrachtung der Parameterreihenfolgen könnte weitere Klarheit schaffen.

Eine Trennung der Variablen erscheint vor allem durch das häufige Auftreten von gegenläufigen MAE-Kurven (Typ III) sinnvoll. Es ist anzunehmen, dass die Form auf die Komplexität der Variablen hinweist und ob ein Algorithmus in der Lage ist die Komplexität der dahinter stehenden Prozesse zu imitieren. Es ist auffällig, dass die Ensemble-Methoden, in Tabelle 9-1 durch RF repräsentiert, unabhängig von Gebiet und Cluster eine stetige Senkung des MAE erzielen, während bei den anderen Algorithmen diese Form nur selten auftritt. Dieser stetige Lernprozess erklärt die geringeren MAE-Werte in den Fallstudien (vgl. Kapitel 6.4). ANN und KNN zeigen zudem Unterschiede zwischen den Untersuchungsgebieten. In der Iller kann KNN noch einen leichten Lernprozess aufweisen, ANN2 stagniert hier. In Main und Regen treten dagegen vermehrt Typ III Kurven auf. KNN zeigt zudem noch Unterschiede je Cluster. Im Regen konnte im Cluster I noch ein stetiger Abfall von MAE erkannt werden, Cluster III zeigt einen gleichbleibenden MAE und Cluster II die gegenläufige Entwicklung.

Diese Beobachtung zeigt die unterschiedliche Komplexität der Zielvariablen  $Q_{Vol}$  und  $v_E$ . Das ANN2 kommt, bspw. im Main (unteres Panel in Abbildung 9-5), mit wenigen Prädiktoren auf seinen minimalen MAE für  $v_E$ . Zwei Prädiktoren reichen hier aus, um den mittleren Fehler und die Unsicherheit zu minimieren. Alle relevanten Informationen sind mit diesen Prädiktoren dem Trainingsdatensatz hinzugefügt worden, alle weiteren tragen redundante oder keinerlei Informationen bei und verschlechtern das Regressionsergebnis. Gegenteilig verhält es sich für  $Q_{Vol}$ . Es sind wesentlich mehr Prädiktoren erforderlich, um den MAE für  $Q_{Vol}$  auf sein Minimum zu bringen. Dasselbe gilt für den KNN, der im Gegensatz zum ANN keine Regression vornimmt sondern seine Regressionsentscheidung auf Ähnlichkeit aufbaut. Beim KNN treten aber die Unterschiede zwischen den Gebieten wieder deutlicher in den Vordergrund. In der Iller konnte der KNN die zusätzlichen Informationen der zusätzlichen Prädiktoren nutzen, um die Beziehung zwischen Ähnlichkeit der Prädiktoren und der Zielvariablen zu verbessern. Hier erscheinen beide Variablen einen ähnlichen Grad an Komplexität aufzuweisen, da auch das ANN hier keine gegenläufigen MAE-Kurven zeigt. Im Regen scheinen die Komplexitäten mit dem Cluster zu variieren. KNN kann zusätzliche Daten im Cluster 1 noch in eine leichte MAE-Reduktion umsetzen,  $Q_{Vol}$  und  $v_E$  konnten also durch mehrere Prädiktoren besser beschrieben werden. Im Cluster 2 ist dies nicht der Fall, hier ist  $v_E$  wieder durch weniger Prädiktoren zu beschreiben.

Die Algorithmen scheinen mit der unterschiedlichen Komplexität der Prozesse unterschiedliche umgehen zu können. Der RF konnte durch seine Ensemble-Struktur redundante Informationen verarbeiten ohne eine Überanpassung zu erzeugen. Während der Entwicklung der RF-Technik wurde diese Stärke der RFs, der strukturelle Schutz gegen Überanpassung, bereits hervorgehoben (Breiman, 2001) und ist in dieser Analyse erneut erkennbar. ANNs scheinen hingegen häufiger zu einer Überanpassung durch redundante oder irrelevante Daten zu tendieren und bedürfen daher einer stärkeren Prädiktorselektion als der RF oder KNN.

Zusammenfassend kann an dieser Stelle festgehalten werden, dass über die Analyse der angepassten ML-Algorithmen eine fundierte Annahme über prozessorientierte Zusammenhänge von Anfangs- und Randbedingungen eines Ereignissen und Abflusscharakteristika formuliert werden konnte. Außerdem konnte die unterschiedliche Komplexität dieses Zusammenhangs zu den unterschiedlichen Zielvariablen deutlich gemacht werden. Ein Vergleich der Untersuchungsgebiete offenbarte zudem regionale Unterschiede in der Komplexität der Prozesse.

## 10 Zusammenfassung

---

In dieser Arbeit wurde eine Kopplung eines geomorphologischen Einheitsganglinien (GIUH) Modells mit einem Machine Learning (ML) Modul entwickelt und im Kontext der Simulation von Niederschlags-Abfluss Ereignissen in unbeobachteten Gebieten getestet. Machine Learning übernimmt in dieser Kopplung die ereignisbasierte Parameterschätzung der Entwässerungsgeschwindigkeit und eines Abflussbildungsparameters für das GIUH-Modell. Das ML-Modul, auch Dynamikmodul genannt, verwendete ein Ensemble verschiedener Regressoren: multiple lineare Regression (MLR), Entscheidungsbaum (CART), ein- und mehrlagige künstliche neuronale Netze (ANN), Support Vector Machine (SVM), nächster Nachbar-Schätzer (KNN), Random Forest (RF) ohne und mit Boosting (ADA). Basierend auf einer Datengrundlage von 2250 Ereignissen wurden die Algorithmen trainiert, um mit Hilfe von 9 Indikatoren für Niederschlag, Vorfeuchte und Schnee die Modellparameter des GIUH zu prognostizieren. Das GIUH-Modell verwendet die geschätzten Parameter, um über die Verteilung von Fließwegen und der Entwässerungsgeschwindigkeit eine ereignisbezogene Einheitsganglinie zu bestimmen und die Abflussbildung des Ereignisses zu simulieren. Die Geomorphologie eines Einzugsgebietes wurde in der implementierten GIUH-Struktur durch eine fließstreckenbasierte Auswertung der Entwässerungswege, ausgewählter Boden- und Oberflächenkennwerte erfasst und die Modellstruktur so an die Struktur des natürlichen Einzugsgebietes angepasst.

Gegenüber bestehenden GIUH-Ansätzen hat die entwickelte Modellstruktur mehrere Vorteile. Die Schätzung der Ereignisparameter mittels des ML-Dynamikmoduls stellt eine mögliche Lösung eines der grundlegenden Probleme von GIUH-Modellen dar. Ein weiteres Vorteil ist die fließwegorientierte Modellstruktur. Sie ermöglicht räumliche Daten im Modell ohne Kompression auf 1-dimensionale Zeitreihen o.ä. zu verwenden. Dies betrifft zum einen die Niederschlagsdaten, die zur Ereignissimulation verwendet werden, und zum anderen weitere Gebietsgrößen wie Bodenspeicher oder Gefälle, die zur räumlichen Differenzierung von Entwässerungsgeschwindigkeit und Abflussvolumen verwendet werden können.

Es wurde die Annahme getroffen, dass eine prozessorientierte Differenzierung der Ereignisse erforderlich ist, um zu adäquaten Modellergebnissen zu gelangen. Es wurden drei Ansätze zur Klassifizierung der Ereignisse in unterschiedliche Ereignistypen getestet. In der Modellstruktur wurden die unterschiedlichen Typen durch parallele ML-Modelle berücksichtigt, je ein Dynamikmodul für jeden Cluster. In den Fallstudien wurde diese Annahme durch den Vergleich von Simulationsergebnissen, die mit Differenzierung nach Cluster und ohne Differenzierung getestet. Eine Differenzierung nach Füllen-Scheitel Klassifikation auf

regionaler Ebene zeigte eine bessere Modellanpassung sowohl gegenüber den anderen Klassifikationen als auch gegenüber einem Modell ohne Klassifikation.

Die GIUH-Struktur wurde auf ihre grundlegende Funktionalität in synthetischen Fallstudien überprüft. Dabei konnte gezeigt werden, dass die simulierten Ganglinien des Modells gemäß ihrer Struktur den Ganglinien einer linearen Speicherkaskade ähneln. Die Modellstruktur konnte darüber hinaus die Form und die Verteilung der Fließstrecken innerhalb des Einzugsgebietes in der Nachbildung der Abflussreaktion berücksichtigen. Oberflächengefälle wurde verwendet, um eine Variation der Entwässerungsgeschwindigkeit auf den Hangflächen zu erzeugen. Eine Variation der Entwässerungsgeschwindigkeit im Gewässer wurde mittels der Verteilung der Strahler-Ordnungszahlen realisiert. In einer Simulationsstudie mit analytischen Parametern erzeugte diese Variation der Parameter eine Erhöhung der Ganglinienvarianz, die sich in den später durchgeführten Fallstudien als positiver Effekt herausstellte.

Als analytische Parameter werden dabei Entwässerungsgeschwindigkeit und Abflussvolumen bezeichnet, die mittels Ganglinienanalyse ermittelt wurden. Es wurde getestet ob die analytisch ermittelten Parameter ohne weitere Anpassung zur Simulation verwendet werden konnten. Dies hätte eine vollständige Parametrisierung des Modells aus Daten ermöglicht. Die analytische Parametrisierung wurde in drei Untersuchungsgebieten getestet. Im Einzugsgebiet des oberen Mains standen im Mittel jeweils 48 Niederschlags-Abfluss Ereignisse an 22 verschiedenen Pegeln in stündlicher Auflösung zur Verfügung, im Einzugsgebiet des Regens an 11 Pegeln je 63 Ereignisse und im Einzugsgebiet der Iller an 10 Pegeln 57 Ereignisse. Die analytische Parametrisierung des Modells ergab in den drei Untersuchungsgebieten ein unterschiedliches Bild. Im Main variierten die Ergebnisse mit den Pegeln aufgrund von lokalen Gegebenheiten um die optimale Anpassung. Es lagen teils Unterschätzungen, teils Überschätzungen, vor allem der Abflussdynamik vor. In Iller und Regen war hingegen eine systematische Überschätzung der Abflussvarianz erkennbar. Diese hatte auch Auswirkung auf die zeitliche Übereinstimmung von simulierten und beobachteten Ganglinien.

Zum Vergleich wurden die Parameter der Ereignisse zusätzlich kalibriert. Wie zu erwarten, sank die Varianz der Modellgüte und die systematische Überschätzung der Varianz konnte behoben werden. Die Wertebereiche der angepassten Parameter wurden durch die Kalibrierung vergleichmäßig. Vor der Kalibrierung waren die Entwässerungsgeschwindigkeiten in der Iller doppelt so hoch wie im Main, im Regen teilweise fast 10-Mal so hoch. Nach der Kalibrierung sind die Wertebereiche von Main und Iller identisch und auf dem Niveau des analytischen Wertebereiches des Mains. Im Regen konnten die Werte auf ein 1.5-faches der Werte von Main und Iller reduziert werden. Zwar konnten mit den kalibrierten Parametern bessere Ergebnisse in der Gangliniensimulation erzeugt werden, dennoch konnten im Main auch mit den analytischen Parametern zufriedenstellende Ergebnisse erzielt werden. Generell scheint der optimale Bereich der Entwässerungsgeschwindigkeit für das GIUH-Modell bei etwa 0.13 m/s zu liegen. Liegt der analytische Wertebereich auf diesem Niveau  $\pm 0.05$  m/s, scheint die Verwendung analytischer Parameter möglich.

Dennoch stellt die Kalibrierung der Parameter keinen nennenswerten Mehraufwand gegenüber der Verwendung analytischer Parameter dar. Aufgrund des simplen Aufbaus des Modells sind für beide Wege der Parametrisierung dieselben Daten erforderlich. Da mit den kalibrierten Parametern bessere Gangliniensimulationen möglich waren und auch die Vorhersage der kalibrierten Parameter keinen signifikanten Verlust in der Vorhersagegüte bedeutete, wurden die kalibrierten Parameter im weiteren Verlauf der Arbeit verwendet.

Im Dynamikmodul wurde die angestrebte prozessorientierte Unterscheidung der Ereignisse umgesetzt. Für jeden Ereignistyp stand ein individuelles Dynamikmodul zur Verfügung, das spezialisiert für den jeweiligen Typ angepasst wurde. In der operationellen Anwendung steht natürlich zu Ereignisbeginn keine Klassifikation zur Verfügung, daher muss das Dynamikmodul vor der Bestimmung der Parameter den zu erwartenden Ereignistyp vorhersagen. Mittels ANN-Klassifikation konnte in einem Test 90% der Ereignisse richtig klassifiziert werden. Eine spätere Fehleranalyse im Rahmen der Fallstudien zeigte das der Einfluss der zu erwartenden 10% fehlerhaft klassifizierten Ereignisse nur marginal war. Die Fehlklassifikationen betrafen demnach vor allem Ereignisse an den Rändern der Cluster an denen die Zuordnung der Ereignisse nicht eindeutig war.

Die im Dynamikmodul verwendeten Algorithmen wurden, wie die GIUH-Struktur, in den drei Untersuchungsgebieten getestet, um ihre generelle Eignung zur Vorhersage der benötigten Parameter zu testen. Dabei konnten mit RF und ADA die besten Ergebnisse erzielt werden. Mit Hilfe dieser Algorithmen konnte der Fehler der Vorhersage mit einer mittleren Abweichung von 20% für die Entwässerungsgeschwindigkeit und rund 40% für das Abflussvolumen beziffert werden. Die Abweichungen für gemessene und kalibrierte Parameter waren dabei nahezu identisch. Leicht höhere mittlere Fehler und größere Fehlervarianz traten bei den Prognosen der Regressoren mit nur einem Schätzer, ANN, KNN, CART und MLR auf. Durchweg die schlechtesten Ergebnisse wurden mit einem SVM Regressor erreicht. Trotz dieser Bewertung des mittleren Fehlers konnte kein Algorithmus identifiziert werden, der in allen Fällen die besten Prognosen erstellen konnten. Vielmehr stellten die Algorithmen eine gegenseitige Ergänzung dar. Daher wurden in den Fallstudien alle Algorithmen als Ensemble verwendet.

Einer der kritischen Punkte bei der Verwendung von Machine Learning ist der notwendige Datenumfang, um stabile Ergebnisse zu erhalten. Eine Studie des benötigten Umfangs der Daten zeigte, dass ca. 25% bis 50% der verfügbaren Daten ausreichte, um ein stabiles Regionalmodell zu etablieren. Je nach Algorithmus variierte die erforderliche Datenmenge in den diesen Grenzen. Bei im Mittel 50 Ereignissen je Teilgebiet bedeutet dies einen minimalen Umfang von 25 Ereignissen je Teilgebiet, um stabile Modelle betreiben zu können. Neben der Menge der Daten ist aber auch die Datenkomposition von Bedeutung. Werden dem Algorithmus zu viele Daten und damit redundante Informationen zugeführt oder Daten aus Gebieten in denen völlig andere Prozesse dominant sind, kann die Modellgüte sinken. Eine Strategie zur selektiven Datenkomposition wurde daher für die Regionalisierung des Dynamikmoduls entwickelt.

Um ähnliche Gebiete zu identifizieren, deren Daten dann die Grundlage für eine Ganglinienvorhersage in einem unbeobachteten Zielgebiet bilden, wurden vier Metriken zur Definition von Ähnlichkeit getestet. Der Abstand der Gebietscentroide und Top-Kriging wurden als Vertreter von distanzbasierten Metriken verwendet. Des Weiteren wurde die Semivarianz zwischen den Distanz-Faktor-Funktionen (DFF) des Oberflächengefälles und des Porenvolumens als geomorphologisches Ähnlichkeitsmaß verwendet. In der Iller stellte sich Top-Kriging als beste Metrik zur selektiven Datenkomposition heraus, im Regen die Semivarianz der DFF-Gefälle. Im Main wurde eine getrennte Datenkomposition für Entwässerungsgeschwindigkeiten mit Top-Kriging und für das Abflussvolumen mit DFF-Gefälle vorgenommen. Basierend auf den gewählten Ähnlichkeitsmetriken wurden für jedes Teilgebiet ein Ranking der ähnlichsten Gebiete erstellt und für die Vorhersage in diesen Gebieten nur die 4-6 ähnlichsten Gebiete verwendet.

Die entwickelte Modellstruktur wurde dann in einer *Leave-one-out* Fallstudie zur Vorhersage von Abflussganglinien in unbeobachteten Gebieten angewendet. Jeweils ein Gebiet wurde als unbeobachtet behandelt und die Daten dieses Gebietes als Validierungsdatensatz verwendet. Alle übrigen Daten des jeweiligen Untersuchungsgebietes dienten, folgend der ermittelten Rankings zur Datenkomposition, als Trainingsdaten zur Anpassung des Dynamikmoduls. Die GIUH-Struktur wurde über die DFF des DGMs des unbeobachteten Gebietes an dessen Struktur angepasst.

Der Median der Gütekriterien über alle Untersuchungsgebiete und alle Algorithmen lag bei  $\alpha = 1.09$ ,  $\beta = 1.14$ ,  $r = 0.69$  und  $KGE = -0.01$ . Vor allem  $\alpha$ ,  $\beta$  und der KGE wiesen dabei erhebliche Varianz auf. Diese wird teils durch unterschiedliche Anpassungsgüte an die Vielzahl der Ereignisse erzeugt, teils durch die variierende Güte der ML-Algorithmen. Die Diskrepanz zwischen den Werten der einzelnen Komponenten und des gesamten KGE verdeutlicht, dass keiner der Algorithmen die gesamte Heterogenität hydrologischer Prozesse abbilden konnte. Daher müssen die Prognosen als Ensemble-Vorhersage ausgewertet werden. Es konnte festgestellt werden, dass für 70-80% aller Ereignisse mindestens eine Simulation des Ensembles bis auf 0.5 Abweichung vom Optimum der KGE-Komponenten herankommt. Darüber hinaus liegen 63% der Ereignisse in Main und Regen innerhalb der Unsicherheitsbänder, die durch das Ensemble aufgespannt wurden. Um die Breite des Unsicherheitsbandes zu minimieren, wurde mittels GLUE ein Unsicherheitsband basierend auf der Anpassungsgüte der Algorithmen definiert. Die GLUE-Methodik konnte die Breite der Bänder um 60% reduzieren, der Anteil von ihnen erfasster Ereignisse sank aber auf rund 42%. Schlechter fielen die Ergebnisse in der Iller aus. Hier konnten nur 38% der Ereignisse durch das gesamte Ensemble erfasst werden. Die Unsicherheitsbänder in diesem Gebiet fielen aber auch 2/3 kleiner aus als in Main und Regen. Es konnte gezeigt werden, dass die starke Senkung der Entwässerungsgeschwindigkeit durch die Kalibrierung zu einer starken Reduktion der absoluten Prognosevarianz führten, die wiederum die Angleichung der Ganglinienprognosen bedingte. Abschließend konnte dieses Ergebnis den Eindruck bestätigen, dass das Ensemble der ML-Algorithmen in dieser Arbeit noch zu klein gewählt wurde. Auch die GLUE-

Methodik erfordert noch weitaus mehr Prognosemodelle, als in der aktuellen Modellstruktur Anwendung gefunden haben.

Dennoch, die mittleren Gütekomponenten sowie der hohe Anteil von erfassten Ereignissen in der Ensemble Vorhersage sprechen für die Anwendbarkeit des GIUH-ML Modells in unbeobachteten Gebieten. Zum Vergleich der erreichten Modellgüte wurde ein HBV-Modell mit denselben Daten und derselben Zielstellung angepasst und die Modellparameter mittels Top-Kriging regionalisiert. Die Ergebnisse dieser Vergleichsstudie zeigten, dass das GIUH-ML teils bessere, teils dem HBV gleichwertige Ergebnisse erzielte. Dieser Schluss wurde für die Anwendung in unbeobachteten Gebieten gezogen und für die Anwendung zur Ereignissimulation in überwachten Gebieten.

Abschließend wurde die Möglichkeit überprüft aus der Funktionsweise der Algorithmen Rückschlüsse auf die Prozesse innerhalb der von ihnen abgebildeten hydrologischen Systeme zu ziehen. In der Analyse wurde die Rangfolge der verwendeten Indikatoren basierend auf ihrem Informationsgehalt über die Zielvariablen ermittelt. Die Analyse wurde getrennt nach Untersuchungsgebieten, Algorithmen und Clustern durchgeführt. Die unterschiedlichen Präferenzen der Algorithmen in den verschiedenen Clustern zu den Indikatorgruppen ließen einen Rückschluss über funktionale Zusammenhänge zu. Ereignisse mit einem niedrigen Scheitel im Vergleich zu ihrem Volumen konnten vorrangig über Niederschlagscharakteristika beschreiben werden. Im Gegenzug konnten Ereignisse mit einem hohen Scheitel im Vergleich zu ihrem Volumen besser über die Anfangszustände der Einzugsgebiete beschrieben werden. Die Analyse konnte zudem die unterschiedliche Komplexität der hydrologischen Prozesse hinter der Entstehung von Abflussvolumen und der Entwässerungsgeschwindigkeit aufdecken. Demnach sind die Prozesse für das Abflussvolumen im Main und teilweise im Regen komplexer als die für die Entwässerungsgeschwindigkeit relevanten Prozesse. Die ML-Algorithmen konnten mit dieser Diskrepanz unterschiedlich gut umgehen. Während RF die negativen Effekte abfangen konnte, reagierten ANN und KNN wesentlich stärker auf diese Diskrepanz.

## 11 Fazit und Ausblick

---

Die Kombination zweier grundlegend verschiedener Modellansätze erwies sich in den durchgeführten Fallstudien und Analysen als gute gegenseitige Ergänzung der Modelle. Das verwendete Machine Learning (ML) Modul konnte genau an der Stelle das konzeptionelle Modell ergänzen, an dem fehlendes Prozessverständnis eine Begrenzung des Modells darstellte. Dadurch konnte das Konzept der geomorphologischen Einheitsganglinie (GIUH) einen Schritt näher an die operationelle Anwendung gebracht werden. Die benötigte ereignisbezogene Parametrisierung von GIUH-Modellen stellte bislang die Begrenzung der Anwendbarkeit der Modelle in zeitlicher wie räumlicher Extrapolation dar.

Auch das in dieser Arbeit entwickelte GIUH-Modell hat wesentliche Vorteile gegenüber klassischen konzeptionellen Modellen und anderen GIUH-Ansätzen. Über den fließwegorientierten Aufbau des Modells und die individuelle Anpassung der Modellstruktur an das betrachtete Einzugsgebiet kann das Modell dessen räumliche Organisation strukturell berücksichtigen. So konnte das Modell ohne weitere Parametrisierung die Form des Einzugsgebietes bei der Bestimmung der Abflussreaktion des Gebietes berücksichtigen. Von Blöschl et al. (2013) wurde mit „*read the landscape*“<sup>1</sup> eine der Voraussetzungen bei der Auswahl von Modellstrukturen formuliert. Die verwendete Distanz-Faktor-Funktion zeigte, dass mit ihrer Hilfe räumliche Daten des Einzugsgebietes mit minimalem Informationsverlust in die Modellstruktur einfließen können und das Modell damit diese Anforderung erfüllen konnte.

Die Aufgabe, die dem ML-Ensemble in dieser Arbeit gestellt wurde, konnten die Algorithmen zufriedenstellend lösen. Als Regionalmodell zeigten sie in überwachten, wie in unbeobachteten Gebieten gute Simulationsergebnisse. Im Vergleich zu einem konzeptionellen Modell konnten sie teils bessere, teils gleichwertige Ergebnisse erzielen. Dennoch fällt ein Großteil der Fehler in der Ganglinienprognose auf die fehlerhafte Schätzung der Ereignisparameter zurück. Die Ergebnisse dieser Arbeit decken sich mit den bisherigen Erkenntnissen zur Anwendung von ML im Kontext hydrologischer Modellierung: Ein einzelner ML-Schätzer ist nicht in der Lage alle Facetten natürlicher Heterogenität abzubilden. So kann und muss an dieser Stelle die Empfehlung, ML-Schätzer stets als Ensemble zu verwenden, unterstützt werden.

---

<sup>1</sup> Seite 241; Kapitel: „*Prediction of runoff hydrographs in ungauged basins*“

Das gekoppelte GIUH-ML Modell kann zum derzeitigen Stand der Entwicklung als Brückentechnologie angesehen werden. So lange kein konzeptionelles Modell die Kopplung zum ML überflüssig macht, stellt sie aber eine wesentliche Weiterentwicklung dieses vielversprechenden Modellansatzes dar. Darüber hinaus ist das ML-Ensemble aber nicht nur Brückentechnologie sondern auch eine Methode zur Untersuchung der von ihr imitierten natürlichen Prozesse.

Natürlich besteht für alle Komponenten des GIUH-ML Modells enormes Entwicklungspotenzial. Bislang ist die GIUH-Komponente des Modells auf die Simulation von kurzzeitigen Niederschlags-Abfluss-Ereignissen beschränkt und verfügt über keinerlei Schnee-, Grundwasser- oder Verdunstungsroutinen. Um das Modell in einem breiteren Spektrum von Ereignissen und in weiteren Regionen anwendbar zu machen, ist eine Ergänzung dieser Komponenten erforderlich. Außerdem muss die Abflussbildungskomponente des Modells weiter diskutiert und -entwickelt werden. Sie wurde bislang sehr einfach gehalten. Dies hat Vorteile, aber die Nachteile überwiegen. Da keinerlei räumliche Informationen des Einzugsgebiets genutzt wurden, stellte die Schätzung des Abflussbildungsparameters eine wesentliche komplexere Aufgabe dar, als die Schätzung des Abflusskonzentrationsparameters. Bei der Weiterentwicklung der Berechnungsroutinen wird ein Kompromiss zwischen einer Erweiterung der zu schätzenden Parameter und Erweiterung der Anwendbarkeit gefunden werden müssen. Zudem wurde das GIUH-ML Modell im Rahmen dieser Arbeit nur auf Gesamtgebiete angewendet. Die Einführung von Teilgebieten kann zum einen die Fehler durch die ML-Selbstkalibrierung des Modells verbessern. Zum anderen reduziert dies den Fehler durch Mittelung der interpolierten Niederschlagswerte innerhalb der Distanzklassen.

Um die Ensemblevorhersagen zu verbessern, wurde im Laufe der Arbeit der Vorschlag entwickelt das Ensemble zu vergrößern. Dies könnte durch die Hinzunahme weiterer Algorithmen geschehen oder durch die Verwendung der bereits verwendeten Algorithmen mit unterschiedlichen Parametrisierungen. Bislang ungenutzt blieben auch die Möglichkeiten des *Deep Learning*. Zwar bieten die verwendeten Algorithmen ein hohes Maß an Flexibilität, um sich an die Daten anzupassen, dennoch sind ihre Lernstrukturen vorgegeben. Bspw. muss für ein ANN die Anzahl der Lagen und Perzeptoren innerhalb der Lagen vorab definiert werden. Beim *Deep Learning* ist die anzulernende Struktur des Algorithmus selbst Bestandteil des Lernprozesses. Diese neue Technologie verspricht die Möglichkeiten der Prognose von Parametern durch datengetriebene Modelle noch zu verbessern.

Als weitere Verbesserungsmöglichkeit wird die getrennte Betrachtung der Zielvariablen angesehen werden. In dieser Arbeit wurde sich um ein einheitliches Vorgehen bemüht. Es sollte bei einer Anwendung stets ein einheitlicher Prädiktorensatz verwendet werden. Die Analysen zeigten aber, dass die Prozesse hinter den Zielvariablen unterschiedlich komplex sind und eine unterschiedliche Anzahl und womöglich auch unterschiedliche Prädiktoren benötigt werden, um die jeweilige Zielgröße zu ermitteln. Für einige Algorithmen war diese Diskrepanz von Nachteilen. Daher sollte in einer weiterführenden Studie untersucht werden, wie sich die Prädiktoren, maßgeschneidert für die einzelnen Zielvariablen unterscheiden und

ob so die Ergebnisse noch verbessert werden könnten.

Ungenutzt blieb im Rahmen dieser Arbeit auch die Möglichkeit die Modellgüte zu prognostizieren. In Kapitel 6.1 konnte festgestellt werden, dass das Modell in der Lage ist Ereignisse zu erkennen, die mit der entwickelten Modellstruktur nicht modelliert werden können. Da solche Ereignisse in den nachfolgenden Analysen ausgeschlossen wurden, war eine Verwendung dieser Schätzer nicht mehr erforderlich. In der operationellen Anwendung wäre es aber von Vorteil, wenn das Modell zusätzlich vermutete fehlerhafte Vorhersagen anzeigt. Ungenutzt blieb zudem das vollständige Potential der Ereignisklassifikation. In Rahmen dieser Arbeit wurde mit einer exklusiven Klassifikation der Ereignisse gearbeitet und die Ergebnisse eines Algorithmen-Ensembles für die Simulation der Abflussganglinie verwendet. Stattdessen kann mit Typwahrscheinlichkeiten gearbeitet werden, wie bereits umgesetzt für die Klassifikation *FuzzyEB*, und basierend auf den Zugehörigkeitswerten eine De-Fuzzifizierung der individuellen Typprognosen vorgenommen werden.

Der angestoßene Prozess aus der Funktionsweise der angelegten Algorithmen Prozessverständnis über die Abflussgenese zu erlangen muss in weiteren Arbeiten vertieft und verfeinert werden. Die in dieser Arbeit angewendete Methodik konnte nur eine Arbeitshypothese aus den Ergebnissen der Analysen ableiten. Ziel muss sein mit gewonnen Erkenntnissen eine Theorie zu etablieren, die die Kopplung der Modelle letztlich überflüssig machen wird.

## 12 Literaturverzeichnis

---

- Abbott, M., Bathurst, J., Cunge, J., O'Connell, P. und Rasmussen, J. 1986. An introduction to the European Hydrological System — Systeme Hydrologique Europeen, “SHE”, 1. History and philosophy of a physically-based, distributed modelling system. *Journal of Hydrology*, **87** (1-2), S. 45–59.
- Addor, N., Nearing, G., Prieto, C., Newman, A., Le Vine, N. und Clark, M. 2018. A Ranking of Hydrological Signatures Based on Their Predictability in Space. *Water Resources Research*, **54** (11), S. 8792–8812.
- Allen, R. 1998. *Crop evapotranspiration. Guidelines for computing crop water requirements*. Rome. (FAO irrigation and drainage paper; 56).
- Alpaydin, E. 2016. *Machine Learning. The New AI*. Cambridge: The MIT Press. (The MIT Press Essential Knowledge series).
- Alpaydin, E. 2008. *Maschinelles Lernen*. München: Oldenbourg.
- Alvisi, S., Mascellani, G., Franchini, M. und Bárdossy, A. 2006. Water level forecasting through fuzzy logic and artificial neural network approaches. *Hydrology and Earth System Sciences*, **10** (1), S. 1–17.
- Bárdossy, A. 2007. Calibration of hydrological model parameters for ungauged catchments. *Hydrology and Earth System Sciences*, **11** (2), S. 703–710.
- Bárdossy, A., Huang, Y. und Wagener, T. 2016. Simultaneous calibration of hydrological models in geographical space. *Hydrology and Earth System Sciences*, **20** (7), S. 2913–2928.
- Bayerisches Landesamt für Umwelt. 2018. *Gewässerkundlicher Dienst Bayern*. Augsburg. [WWW] [www.lfu.bayern.de](http://www.lfu.bayern.de). (30. Oktober 2018).
- Beck, H., van Dijk, Albert I. J. M., Roo, A. de, Miralles, D., McVicar, T., Schellekens, J. und Bruijnzeel, L. 2016. Global-scale regionalization of hydrologic model parameters. *Water Resources Research*, **52** (5), S. 3599–3622.
- Bell, F. und Om Kar, S. 1969. Characteristic response times in design flood estimation. *Journal of Hydrology*, **8** (2), S. 173–196.
- Bergström, S. 1976. *Development and application of a conceptual model for scandinavian catchments*. SMHI, Hrsg. Norrköping: Swedish Meteorological and Hydrological Institute.

- Beven, K. 1989. Changing ideas in hydrology — The case of physically-based models. *Journal of Hydrology*, **105** (1-2), S. 157–172.
- Beven, K. und Binley, A. 1992. The future of distributed models. Model calibration and uncertainty prediction. *Hydrological Processes*, **6** (3), S. 279–298.
- Beven, K. und Kirkby, M. 1979. A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Bulletin*, **24** (1), S. 43–69.
- Beven, K., Wood, E. und Sivapalan, M. 1988. On hydrological heterogeneity — Catchment morphology and catchment response. *Journal of Hydrology*, **100** (1-3), S. 353–375.
- Blöschl, G. und Sivapalan, M. 1995. Scale issues in hydrological modelling. A review. *Hydrological Processes*, **9** (3-4), S. 251–290.
- Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A. und Savenije, H. 2013. *Runoff Prediction in Ungauged Basins. Synthesis across Processes, Places and Scales*. Cambridge: Cambridge University Press.
- Boorman, D., Hollis, J. und Lilly, A. 1995. *Hydrology of soil types. A hydrologically-based classification of the soils of the United Kingdom*. Wallingford, Oxfordshire, UK: Institute of Hydrology. (Report; 126).
- Bossard, M., Feranec, J. und Otahel, J. 2000. *CORINE land cover technical guide. Addendum 2000*. European Environmental Agency, Hrsg. Copenhagen: EEA. [WWW] <http://www.eea.eu>.
- Breiman, L. 2001. Random forests. *Machine Learning*, **45** (1), S. 5–32.
- Bremicker, M. 2000. *Das Wasserhaushaltsmodell LARSIM. Modellgrundlagen und Anwendungsbeispiele* [Freiburger Schriften zur Hydrologie]. Institut für Hydrologie. UNIVERSITÄT FREIBURG. Freiburg.
- Brunner, M., Furrer, R., Sikorska, A., Viviroli, D., Seibert, J. und Favre, A.-C. 2018. Synthetic design hydrographs for ungauged catchments: a comparison of regionalization methods. *Stochastic Environmental Research and Risk Assessment*, **32** (7), S. 1993–2023.
- Brunner, M., Viviroli, D., Sikorska, A., Vannier, O., Favre, A.-C. und Seibert, J. 2017. Flood type specific construction of synthetic design hydrographs. *Water Resources Research*, **53** (2), S. 1390–1406.
- Bundesanstalt für Geowissenschaften und Rohstoffe. 2006. *Bodenübersichtskarte. (BÜK1000)*. Hannover. [WWW] <https://www.bgr.bund.de>. (31. Oktober 2018).
- Casper, M., Gronz, O. und Gemmar, P. 2015. Prozesstreue Parametrisierung und Kalibrierung eines Wasserhaushaltsmodells. *Hydrologie und Wasserwirtschaft*, **59** (4), S. 136–

144.

- Chacon-Hurtado, J., Alfonso, L. und Solomatine, D. 2017. Rainfall and streamflow sensor network design. A review of applications, classification, and a proposed framework. *Hydrology and Earth System Sciences*, **21** (6), S. 3071–3091.
- Chapman, T. 1991. Comment on “Evaluation of automated techniques for base flow and recession analyses” by R. J. Nathan and T. A. McMahon. *Water Resources Research*, **27** (7), S. 1783–1784.
- Chen, B., Krajewski, W., Zhou, X. und Helmers, M. 2015. Organized variability of surface runoff responses across neighboring hillslopes in Iowa. *Journal of Hydrology*, **523**, S. 1–13.
- Cortes, C. und Vapnik, V. 1995. Support-vector networks. *Machine Learning*, **20** (3), S. 273–297.
- Cui, H. und Singh, V. 2015. Configurational entropy theory for streamflow forecasting. *Journal of Hydrology*, **521**, S. 1–17.
- Cunge, J. 1969. On The Subject Of A Flood Propagation Computation Method (Muskingum Method). *Journal of Hydraulic Research*, **7** (2), S. 205–230.
- Deutscher Wetterdienst. 2005. *KOSTRA-DWD-2000. Starkniederschlagshöhen für Deutschland (1951-2000), Grundlagenbericht*. Deutscher Wetterdienst, Hrsg. Offenbach am Main.
- Dooge, J. 1959. A general theory of the unit hydrograph. *Journal of Geophysical Research*, **64** (2), S. 241–256.
- Dooge, J. 1986. Looking for hydrologic laws. *Water Resources Research*, **22** (9S), S. 46S–58S.
- Drouge, G., Leviander, T., Pfister, L., Idrissi, A., Iffly, J.-F., Hoffmann, L., Guex, F., Hingray, B. und Humbert, J. 2002. The applicability of a parsimonious model for local and regional prediction of runoff. *Hydrological Sciences Journal*, **47** (6), S. 905–920.
- Dunn, S. und Lilly, A. 2001. Investigating the relationship between a soils classification and the spatial parameters of a conceptual catchment-scale hydrological model. *Journal of Hydrology*, **252** (1-4), S. 157–173.
- DVWK, Hrsg. 1990. *Arbeitsanleitung zur Anwendung von Niederschlag-Abfluß-Modellen in kleinen Einzugsgebieten*. 2., durchges. Aufl. DVWK-Regeln zur Wasserwirtschaft, **113**. Hamburg: Parey in Komm.
- DVWK, Hrsg. 1996. *Ermittlung der Verdunstung von Land- und Wasserflächen*. DVWK-Merkblätter zur Wasserwirtschaft, **238**. Bonn: Wirtschafts- und Verl.-Ges. Gas und Wasser.

- Elshorbagy, A., Corzo, G., Srinivasulu, S. und Solomatine, D. 2010a. Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1. Concepts and methodology. *Hydrology and Earth System Sciences*, **14** (10), S. 1931–1941.
- Elshorbagy, A., Corzo, G., Srinivasulu, S. und Solomatine, D. 2010b. Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 2. Application. *Hydrology and Earth System Sciences*, **14** (10), S. 1943–1961.
- Fang, X., Pomeroy, J., Westbrook, C., Guo, X., Minke, A. und Brown, T. 2010. Prediction of snowmelt derived streamflow in a wetland dominated prairie basin. *Hydrology and Earth System Sciences*, **14** (6), S. 991–1006.
- Flügel, W. 1995. Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany. *Hydrological Processes*, **9** (3-4), S. 423–436.
- Gaál, L., Szolgay, J., Kohnová, S., Hlavčová, K., Parajka, J., Viglione, A., Merz, R. und Blöschl, G. 2015. Dependence between flood peaks and volumes: a case study on climate and hydrological controls. *Hydrological Sciences Journal*, **60** (6), S. 968–984.
- Gaál, L., Szolgay, J., Kohnová, S., Parajka, J., Merz, R., Viglione, A. und Blöschl, G. 2012. Flood timescales: Understanding the interplay of climate and catchment processes through comparative hydrology. *Water Resources Research*, **48** (4), S. n/a.
- Gaborit, É., Ricard, S., Lachance-Cloutier, S., Anctil, F., Turcotte, R. und Polat, A. 2015. Comparing global and local calibration schemes from a differential split-sample test perspective. *Canadian Journal of Earth Sciences*, **52** (11), S. 990–999.
- Gharari, S., Hrachowitz, M., Fenicia, F. und Savenije, H. 2011. Hydrological landscape classification: investigating the performance of HAND based landscape classifications in a central European meso-scale catchment. *Hydrology and Earth System Sciences*, **15** (11), S. 3275–3291.
- Gharari, S., Shafiei, M., Hrachowitz, M., Kumar, R., Fenicia, F., Gupta, H. und Savenije, H. 2014. A constraint-based search algorithm for parameter identification of environmental models. *Hydrology and Earth System Sciences*, **18** (12), S. 4861–4870.
- Gong, W., Yang, D., Gupta, H. und Nearing, G. 2014. Estimating information entropy for hydrological data. One-dimensional case. *Water Resources Research*, **50** (6), S. 5003–5018.
- Goodfellow, I., Bengio, Y. und Courville, A. 2016. *Deep Learning*: MIT Press.
- Grayson, R. und Blöschl, G., Hrsg. 2001. *Spatial patterns in catchment hydrology. Observations and modelling*. Cambridge: Cambridge Univ. Press.
- Grimaldi, S., Petroselli, A., Alonso, G. und Nardi, F. 2010. Flow time estimation with spatially variable hillslope velocity in ungauged basins. *Advances in Water Resources*, **33**

- (10), S. 1216–1223.
- Grimaldi, S., Petroselli, A. und Nardi, F. 2011. A parsimonious geomorphological unit hydrograph for rainfall–runoff modelling in small ungauged basins. *Hydrological Sciences Journal*, **57** (1), S. 73–83.
- Gupta, H., Kling, H., Yilmaz, K. und Martinez, G. 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, **377** (1-2), S. 80–91.
- Gupta, H. und Nearing, G. 2014. Debates-the future of hydrological sciences. A (common) path forward? Using models and data to learn: A systems theoretic perspective on the future of hydrological science. *Water Resources Research*, **50** (6), S. 5351–5359.
- Gupta, V. und Mesa, O. 1988. Runoff generation and hydrologic response via channel network geomorphology — Recent progress and open problems. *Journal of Hydrology*, **102** (1-4), S. 3–28.
- Gupta, V., Waymire, E. und Wang, C. 1980. A representation of an instantaneous unit hydrograph from geomorphology. *Water Resources Research*, **16** (5), S. 855–862.
- Hall, D., Riggs, G. und Solomonson, V. 2015. *MODIS/Terra Snow Cover 8-Day L3 Global 500m SIN Grid*. NASA National Snow and Ice Data Center DAAC, Hrsg.: NASA MODAPS SIPS.
- Hall, M., Minns, A. und Ashrafuzzaman, A. 2002. The application of data mining techniques for the regionalisation of hydrological variables. *Hydrology and Earth System Sciences*, **6** (4), S. 685–694.
- Hallema, D. und Moussa, R. 2014. A model for distributed GIUH-based flow routing on natural and anthropogenic hillslopes. *Hydrological Processes*, **28** (18), S. 4877–4895.
- Han, J. und Kamber, M. 2010. *Data mining. Concepts and techniques*. 2. ed., [Nachdr.]. Amsterdam: Elsevier/Morgan Kaufmann. (The Morgan Kaufmann series in data management systems).
- Hattermann, F., Huang, S., Burghoff, O., Willems, W., Österle, H., Büchner, M. und Kundzewicz, Z. 2014. Modelling flood damages under climate change conditions – a case study for Germany. *Natural Hazards and Earth System Sciences*, **14** (12), S. 3151–3168.
- Hazenberg, P., Fang, Y., Broxton, P., Gochis, D., Niu, G.-Y., Pelletier, J., Troch, P. und Zeng, X. 2015. A hybrid-3D hillslope hydrological model for use in Earth system models. *Water Resources Research*, **51** (10), S. 8218–8239.
- He, Y., Bárdossy, A. und Zehe, E. 2011a. A catchment classification scheme using local variance reduction method. *Journal of Hydrology*, **411** (1-2), S. 140–154.
- He, Y., Bárdossy, A. und Zehe, E. 2011b. A review of regionalisation for continuous

- streamflow simulation. *Hydrology and Earth System Sciences*, **15** (11), S. 3539–3553.
- Herber Green, W. und Ampt, G. 1911. Studies on Soil Physics. *The Journal of Agricultural Science*, **4** (01), S. 1.
- Heřmanovský, M., Havlíček, V., Hanel, M. und Pech, P. 2017. Regionalization of runoff models derived by genetic programming. *Journal of Hydrology*, **547**, S. 544–556.
- Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M., Ehret, U., Fenicia, F., Freer, J., Gelfan, A., Gupta, H., Hughes, D., Hut, R., Montanari, A., Pande, S., Tetzlaff, D., Troch, P., Uhlenbrook, S., Wagener, T., Winsemius, H., Woods, R., Zehe, E. und Cudennec, C. 2013. A decade of Predictions in Ungauged Basins (PUB)—a review. *Hydrological Sciences Journal*, **58** (6), S. 1198–1255.
- Hughes, D. und Kapangaziwiri, E. 2007. The use of physical basin properties and runoff generation concepts as an aid to parameter quantification in conceptual type rainfall-runoff models. *IAHS PUBLICATION*, **313**, S. 311.
- Jarvis, A., Reuter, H., Nelson, A. und Guevara, E. 2008. *Hole-filled seamless SRTM data V4*. International Centre for Tropical Agriculture (CIAT), Hrsg. [WWW] <http://srtm.csi.cgiar.org>. (31. Oktober 2018).
- Jingyi, Z. und Hall, M. 2004. Regional flood frequency analysis for the Gan-Ming River basin in China. *Journal of Hydrology*, **296** (1-4), S. 98–117.
- Johnson, S. 2018. *The NLOpt nonlinear-optimization package* [Software].
- Kelleher, J., MacNamee, B. und D'Arcy, A. 2015. *Fundamentals of machine learning for predictive data analytics. Algorithms, worked examples, and case studies*. Cambridge, Massachusetts, London, England: The MIT Press.
- Keller, L., Rössler, O., Martius, O. und Weingartner, R. 2018. Delineation of flood generating processes and their hydrological response. *Hydrological Processes*, **32** (2), S. 228–240.
- Kirchner, J. 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), S. 2465.
- Kirkby, M. 1976. Tests of the random network model, and its application to basin hydrology. *Earth Surface Processes*, **1** (3), S. 197–212.
- Kirkby, M. 2005. Organization and Process. In: ANDERSON, Malcolm G., Hrsg. *Encyclopedia of hydrological sciences*. Chichester: Wiley, S. 42–58.
- Kisi, O., Shiri, J. und Tombul, M. 2013. Modeling rainfall-runoff process using soft computing techniques. *Computers & Geosciences*, **51**, S. 108–117.
- Klemes, V. 1983. Conceptualization and scale in hydrology. *Journal of Hydrology* (65), S.

1–23.

- Kumar, R., Chatterjee, C., Singh, R., Lohani, A. und Kumar, S. 2007. Runoff estimation for an ungauged catchment using geomorphological instantaneous unit hydrograph (GIUH) models. *Hydrological Processes*, **21** (14), S. 1829–1840.
- Li, C., Singh, V. und Mishra, A. 2012. Entropy theory-based criterion for hydrometric network evaluation and design. Maximum information minimum redundancy. *Water Resources Research*, **48** (5), S. 205.
- Liang, X., Lettenmaier, D., Wood, E. und Burges, S. 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, **99** (D7), S. 14415.
- Lindström, G., Johansson, B., Persson, M., Gardelin, M. und Bergström, S. 1997. Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology*, **201** (1-4), S. 272–288.
- Maidment, D., Hrsg. 1993. *Handbook of hydrology*. New York, NY: McGraw-Hill.
- Masih, I., Uhlenbrook, S., Maskey, S. und Ahmad, M. 2010. Regionalization of a conceptual rainfall–runoff model based on similarity of the flow duration curve. A case study from the semi-arid Karkheh basin, Iran. *Journal of Hydrology*, **391** (1-2), S. 188–201.
- May, R., Maier, H., Dandy, G. und Fernando, T. 2008. Non-linear variable selection for artificial neural networks using partial mutual information. *Environmental Modelling & Software*, **23** (10-11), S. 1312–1326.
- Merz, R. und Blöschl, G. 2003. A process typology of regional floods. *Water Resources Research*, **39** (12).
- Mesa, O. und Mifflin, E. 1986. On the relative role of Hillslope and Network Geometry in Hydrologic Response. In: GUPTA, V. K., RODRIGUEZ-ITURBE, I. & WOOD, E. F., Hrsg. *Scale Problems in hydrology: runoff generation and basin response*. Dordrecht: Reidel. (Scale Problems in hydrology; 2), S. 1–19.
- Mewes, B. und Opiel, H. 2019. Comparison of different machine learning algorithms for hydrograph separation. *Frontiers in Earth Science*, **(submitted)**.
- Mewes, B., Opiel, H. und Hartmann, A. 2019. Information based machine-learning for tracer signature prediction in karstic environments. *Water Resources Research* (submitted).
- Minns, A. und Hall, M. 2005. Artificial Neuronal Network Concepts in Hydrology. In: ANDERSON, Malcolm G., Hrsg. *Encyclopedia of hydrological sciences*. Chichester: Wiley, S. 307–319.
- Mishra, S. und Singh, V. 2013. *Soil conservation service curve number (SCS-CN) methodology*: Springer Science & Business Media. (42).

- Mount, N., Maier, H., Toth, E., Elshorbagy, A., Solomatine, D., Chang, F.-J. und Abrahart, R. 2016. Data-driven modelling approaches for socio-hydrology. Opportunities and challenges within the Panta Rhei Science Plan. *Hydrological Sciences Journal*, **8** (5), S. 1–17.
- Müller, C., Hellebrand, H., Seeger, M. und Schobel, S. 2009. Identification and regionalization of dominant runoff processes – a GIS-based and a statistical approach. *Hydrology and Earth System Sciences*, **13** (6), S. 779–792.
- Murawski, A., Zimmer, J. und Merz, B. 2016. High spatial and temporal organization of changes in precipitation over Germany for 1951–2006. *International Journal of Climatology*, **36** (6), S. 2582–2597.
- Nardi, F., Annis, A. und Biscarini, C. 2018. On the impact of urbanization on flood hydrology of small ungauged basins. The case study of the Tiber river tributary network within the city of Rome. *Journal of Flood Risk Management*, **11** (5), S. S594-S603.
- Nash, J. und Sutcliffe, J. 1970. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, **10** (3), S. 282–290.
- Nearing, G. und Gupta, H. 2015. The quantity and quality of information in hydrologic models. *Water Resources Research*, **51** (1), S. 524–538.
- Nied, M., Pardowitz, T., Nissen, K., Ulbrich, U., Hundecha, Y. und Merz, B. 2014. On the relationship between hydro-meteorological patterns and flood types. *Journal of Hydrology*, **519**, S. 3249–3262.
- Nobre, A., Cuartas, L., Hodnett, M., Rennó, C., Rodrigues, G., Silveira, A., Waterloo, M. und Saleska, S. 2011. Height Above the Nearest Drainage – a hydrologically relevant new terrain model. *Journal of Hydrology*, **404** (1-2), S. 13–29.
- Nourani, V., Hosseini Baghanam, A., Adamowski, J. und Kisi, O. 2014. Applications of hybrid wavelet–Artificial Intelligence models in hydrology. A review. *Journal of Hydrology*, **514**, S. 358–377.
- Oppel, H. und Schumann, A. 2017. A method to employ the spatial organization of catchments into semi-distributed rainfall–runoff models. *Hydrology and Earth System Sciences*, **21** (8), S. 4259–4282.
- Parajka, J., Blöschl, G. und Merz, R. 2007. Regional calibration of catchment models. Potential for ungauged catchments. *Water Resources Research*, **43** (6), S. 937.
- Patil, S. und Stieglitz, M. 2015. Comparing spatial and temporal transferability of hydrological model parameters. *Journal of Hydrology*, **525**, S. 409–417.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. und Duchesnay, E. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, **12**, S. 2825–2830.

- Perrin, C., Michel, C. und Andréassian, V. 2003. Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, **279** (1-4), S. 275–289.
- Piotrowski, A., Napiorkowski, M., Napiorkowski, J. und Osuch, M. 2015. Comparing various artificial neural network types for water temperature prediction in rivers. *Journal of Hydrology*, **529**, S. 302–315.
- Powell, M. 2009. The BOBYQA algorithm for bound constrained optimization without derivatives.
- Ramsay, J. und Silverman, B. 2005. *Functional Data Analysis*. Second Edition. New York, NY: Springer Science+Business Media Inc. (Springer Series in Statistics).
- Ratto, M., Young, P., Romanowicz, R., Pappenberger, F., Saltelli, A. und Pagano, A. 2007. Uncertainty, sensitivity analysis and the role of data based mechanistic modeling in hydrology. *Hydrology and Earth System Sciences*, **11** (4), S. 1249–1266.
- Rawls, W., Brakensiek, D. und Saxton, K. 1982. Estimation of soil water properties. *Transactions of the ASAE*, **25** (5), S. 1316–1320.
- Rice, J., Emanuel, R., Vose, J. und Nelson, S. 2015. Continental U.S. streamflow trends from 1940 to 2009 and their relationships with watershed spatial characteristics. *Water Resources Research*, **51** (8), S. 6262–6275.
- Rigon, R., Bancheri, M., Formetta, G. und de Lavenne, A. 2016. The geomorphological unit hydrograph from a historical-critical perspective. *Earth Surface Processes and Landforms*, **41** (1), S. 27–37.
- Rinaldo, A., Botter, G., Bertuzzo, E., Uccelli, A., Settin, T. und Marani, M. 2006. Transport at basin scales. 1. Theoretical framework. *Hydrology and Earth System Sciences*, **10** (1), S. 19–29.
- Rinaldo, A., Marani, A. und Rigon, R. 1991. Geomorphological dispersion. *Water Resources Research*, **27** (4), S. 513–525.
- Rodriguez, R., Singh, V., Pruski, F. und Calegario, A. 2016. Using entropy theory to improve the definition of homogeneous regions in the semi-arid region of Brazil. *Hydrological Sciences Journal*, **61** (11), S. 2096–2109.
- Rodríguez-Iturbe, I., Devoto, G. und Valdés, J. 1979. Discharge response analysis and hydrologic similarity: The interrelation between the geomorphologic IUH and the storm characteristics. *Water Resources Research*, **15** (6), S. 1435–1444.
- Rodríguez-Iturbe, I. und Valdés, J. 1979. The geomorphologic structure of hydrologic response. *Water Resources Research*, **15** (6), S. 1409–1420.
- Samaniego, L., Kumar, R. und Attinger, S. 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, **46** (5), S. n/a-n/a.

- Sartor, J. 1994. *Die Wahrscheinlichkeit des gleichzeitigen Auftretens von maßgebender Abflussereignisse in Kanalnetzen und natürlichen Gewässern*. Köhler, G., Hrsg. Kaiserslautern. (Berichte; 3).
- Savenije, H. 2010. HESS Opinions "Topography driven conceptual modelling (FLEX-Topo)". *Hydrology and Earth System Sciences*, **14** (12), S. 2681–2692.
- Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. und Carrillo, G. 2011. Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrology and Earth System Sciences*, **15** (9), S. 2895–2911.
- Shannon, C. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal*, **27** (3), S. 379–423.
- Sharma, A. und Mehrotra, R. 2014. An information theoretic alternative to model a natural system using observational information alone. *Water Resources Research*, **50** (1), S. 650–660.
- Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F.-J., Ganguly, S., Hsu, K.-l., Kifer, D., Fang, Z., Fang, K., Li, D., Li, X. und Tsai, W.-P. 2018. HESS Opinions. Incubating deep-learning-powered hydrologic science advances as a community. *Hydrology and Earth System Sciences*, **22** (11), S. 5639–5656.
- Sherman, L. 1932. Streamflow from rainfall by unit-hydrograph method. *Engineering News Record* (108), S. 501–505.
- Shortridge, J., Guikema, S. und Zaitchik, B. 2016. Machine learning methods for empirical streamflow simulation. A comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrology and Earth System Sciences*, **20** (7), S. 2611–2628.
- Sikorska, A., Viviroli, D. und Seibert, J. 2015. Flood-type classification in mountainous catchments using crisp and fuzzy decision trees. *Water Resources Research*, **51** (10), S. 7959–7976.
- Singh, P., Mishra, S. und Jain, M. 2014. A review of the synthetic unit hydrograph. From the empirical UH to advanced geomorphological methods. *Hydrological Sciences Journal*, **59** (2), S. 239–261.
- Sivapalan, M. 2005. Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale. In: ANDERSON, Malcolm G., Hrsg. *Encyclopedia of hydrological sciences*. Chichester: Wiley, S. 193–219.
- Sivapalan, M., Takeuchi, K., Franks, S., Gupta, V., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J., Mendiondo, E., O'Connell, P., Oki, T., Pomeroy, J., Schertzer, D., Uhlenbrook, S. und Zehe, E. 2003. IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012. Shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal*, **48** (6), S. 857–880.

- Skaugen, T., Peerebom, I. und Nilsson, A. 2015. Use of a parsimonious rainfall-run-off model for predicting hydrological response in ungauged basins. *Hydrological Processes*, **29** (8), S. 1999–2013.
- Skøien, J., Merz, R. und Blöschl, G. 2006. Top-kriging - geostatistics on stream networks. *Hydrology and Earth System Sciences*, **10** (2), S. 277–287.
- Snell, J. und Sivapalan, M. 1994. On geomorphological dispersion in natural catchments and the geomorphological unit hydrograph. *Water Resources Research*, **30** (7), S. 2311–2323.
- Solomatine, D. und Ostfeld, A. 2008. Data-driven modelling. Some past experiences and new approaches. *Journal of Hydroinformatics*, **10** (1), S. 3–22.
- Soulsby, C., Tetzlaff, D. und Hrachowitz, M. 2010. Spatial distribution of transit times in montane catchments. Conceptualization tools for management. *Hydrological Processes*, **24** (22), S. 3283–3288.
- Soulsby, C., Tetzlaff, D., Rodgers, P., Dunn, S. und Waldron, S. 2006. Runoff processes, stream water residence times and controlling landscape characteristics in a mesoscale catchment. An initial evaluation. *Journal of Hydrology*, **325** (1-4), S. 197–221.
- Sponagel, H., Hrsg. 2005. *Bodenkundliche Kartieranleitung. Mit 103 Tabellen und 31 Listen*. 5., verb. und erw. Aufl. Stuttgart: Schweizerbart.
- Steinschneider, S., Yang, Y.-C. und Brown, C. 2015. Combining regression and spatial proximity for catchment model regionalization. A comparative study. *Hydrological Sciences Journal*, **60** (6), S. 1026–1043.
- Thiessen, A. 1911. Precipitation averages for large areas. *Monthly Weather Review*, **39** (7), S. 1082–1089.
- Todini, E. 2007. A mass conservative and water storage consistent variable parameter Muskingum-Cunge approach. *Hydrology and Earth System Sciences*, **11** (5), S. 1645–1659.
- Toth, E. 2016. Estimation of flood warning runoff thresholds in ungauged basins with asymmetric error functions. *Hydrology and Earth System Sciences*, **20** (6), S. 2383–2394.
- Tyralla, C. und Schumann, A. 2013. *HydPy. - ein interaktiv nutzbares Framework zur Erstellung und Anwendung hydrologischer Modelle*. Bundesanstalt für Gewässergüte, Hrsg. Koblenz.
- Uhlemann, S., Thielen, A. und Merz, B. 2010. A consistent set of trans-basin floods in Germany between 1952–2002. *Hydrology and Earth System Sciences*, **14** (7), S. 1277–1295.
- Viglione, A., Chirico, G., Woods, R. und Blöschl, G. 2010. Generalised synthesis of

- space–time variability in flood response. An analytical framework. *Journal of Hydrology*, **394** (1-2), S. 198–212.
- Wagener, T., Sivapalan, M., Troch, P., McGlynn, B., Harman, C., Gupta, H., Kumar, P., Rao, P., Basu, N. und Wilson, J. 2010. The future of hydrology. An evolving science for a changing world. *Water Resources Research*, **46** (5), S. 1080.
- Wagener, T., Sivapalan, M., Troch, P. und Woods, R. 2007. Catchment Classification and Hydrologic Similarity. *Geography Compass*, **1** (4), S. 901–931.
- Wagener, T., Wheater, H. und Gupta, H. 2004. *Rainfall-runoff modelling in gauged and ungauged catchments*. London: Imperial College Press.
- Westerberg, I., Guerrero, J.-L., Younger, P., Beven, K., Seibert, J., Halldin, S., Freer, J. und Xu, C.-Y. 2011. Calibration of hydrological models using flow-duration curves. *Hydrology and Earth System Sciences*, **15** (7), S. 2205–2227.
- White, A., Kumar, P., Saco, P., Rhoads, B. und Yen, B. 2004. Hydrodynamic and geomorphologic dispersion. Scale effects in the Illinois River Basin. *Journal of Hydrology*, **288** (3-4), S. 237–257.
- Winter, T. 2001. The concept of hydrological landscapes. *Journal of the American Water Resources Association*, **37** (2), S. 335–349.
- Wolpert, D. und Macready, W. 1997. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, **1** (1), S. 67–82.
- Yadav, M., Wagener, T. und Gupta, H. 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, **30** (8), S. 1756–1774.
- Yao, C., Zhang, K., Yu, Z., Li, Z. und Li, Q. 2014. Improving the flood prediction capability of the Xinanjiang model in ungauged nested catchments by coupling it with the geomorphologic instantaneous unit hydrograph. *Journal of Hydrology*, **517**, S. 1035–1048.
- Yaseen, Z., El-shafie, A., Jaafar, O., Afan, H. und Sayl, K. 2015. Artificial intelligence based models for stream-flow forecasting. 2000–2015. *Journal of Hydrology*, **530**, S. 829–844.
- Young, P. 2003. Top-down and data-based mechanistic modelling of rainfall-flow dynamics at the catchment scale. *Hydrological Processes*, **17** (11), S. 2195–2217.
- Zhang, B. und Govindaraju, R. 2003. Geomorphology-based artificial neural networks (GANNs) for estimation of direct runoff over watersheds. *Journal of Hydrology*, **273** (1-4), S. 18–34.
- Zhang, Y., Chiew, F., Li, M. und Post, D. 2018. Predicting Runoff Signatures Using Regression and Hydrological Modeling Approaches. *Water Resources Research*, **54** (10), S. 7859–7878.

## Abbildungsverzeichnis

---

Abbildung 1-1: Einfluss (schematisch) der Geomorphologie und Transitzeit auf GIUH-Ganglinien .....	3
Abbildung 3-1: Übersicht über die ausgewählten Flussgebiete Iller, Regen und Main im Bundesland Bayern.....	15
Abbildung 3-2: DGM, ausgewählte Pegel und Teileinzugsgebiete im oberen Mains .....	17
Abbildung 3-3: Landnutzung und Gesamtporenvolumen im oberen Mains .....	17
Abbildung 3-4: DGM, ausgewählte Pegel und Teileinzugsgebiete im Regen .....	19
Abbildung 3-5: Landnutzung und Gesamtporenvolumen im Regen.....	19
Abbildung 3-6: DGM, ausgewählte Pegel und Teileinzugsgebiete der Iller.....	21
Abbildung 3-7: Landnutzung und nutzbare Feldkapazität der Iller .....	21
Abbildung 3-8: Lage und Umrisse der Untersuchungs- und Zusatzgebiete.....	22
Abbildung 3-9: Veränderung der Schneebedeckung im Untersuchungsgebiet Regen im Vorfeld eines Hochwassers am 16.03.2001 abgeleitet aus MODIS Daten .....	28
Abbildung 3-10: Entscheidungsbaum zur Identifizierung der Ereignistypen Regen auf Schnee (RoS), Schneeschmelze (SM), kurzes Regenereignis (SR) und langes Regenereignis (LR).....	30
Abbildung 3-11: Relative Häufigkeiten der Ereignistypen im Untersuchungsgebiet Regen .....	31
Abbildung 3-12: Füllen-Scheitel Verhältnisse für das Untersuchungsgebiet oberer Main, lokale (links) und regionale (rechts) Klassifikation .....	33
Abbildung 4-1: Geokennwerte im Einzugsgebiet der Mulde in Sachsen (v.r.n.l.): Höhe, GPV bis 2m Tiefe und Landnutzung.....	38
Abbildung 4-2: Geokennwerte im Einzugsgebiet der Salzach, Österreich (v.r.n.l.): Höhe, nFK bis 2m Tiefe und Landnutzung.....	38
Abbildung 4-3: Unterteilung des EZGs Marienthal in Distanzklassen mit $\Delta s = 20$ km .....	39
Abbildung 4-4: Distanz-Faktor-Funktion des gesamten Porenvolumens im EZG Mulde, dargestellt als Mittelwert und Standardabweichung der $x_S$ -Distanzklasse (links) und Mittelwerte differenziert nach $x_H$ (rechts) .....	40

Abbildung 4-5: Übersicht über die entwickelte Modellstruktur mit Kapitelnummern.....	44
Abbildung 4-6: Interpolierte Niederschlagsverteilung Ereignis 14.02.2002 im oberen Main .....	46
Abbildung 4-7: Umwandlung Niederschlagszeitreihe in räumlich differenzierte Reihe, Beispiel Niederschlagsereignis Februar 2002 im Flussgebiet oberer Main.....	46
Abbildung 4-8: Konzeptionelle Darstellung der Hangflächen eines Einzugsgebietes mit $N_s$ Distanzklassen.....	48
Abbildung 4-9: Empirische Häufigkeitsverteilung von Hangfließlängen in zwei exemplarischen Distanzklassen des Einzugsgebietes oberer Main (links: 5km vom Auslass; rechts 27 km); Darstellung einer Auswahl von angepassten Verteilungsfunktionen.....	49
Abbildung 4-10: Quantil-Quantil-Plot für die Verteilung Hangflächen dargestellt mit Johnson <sub>SB</sub> -Verteilung für 5 Distanzklassen (4km; 27km; 49km; 79km; 97km) im Gebiet des oberen Mains.....	50
Abbildung 4-11: Geomorphologische Einheitsganglinien des Einzugsgebietes Bad Berneck mit $\Delta s = 1\text{km}$ ; $v_{E,H} = 0.05\text{ m/s}$ .....	51
Abbildung 4-12: Konzeptioneller Aufbau des Gewässermodells für ein Gebiet mit $N_s$ Distanzklassen.....	52
Abbildung 4-13: Konzeptionelle Darstellung der Wellenfortpflanzung im Gewässer mit gleichförmiger Entwässerungsgeschwindigkeit.....	53
Abbildung 4-14: Konzept parallele Struktur Dynamikmodul für prozessorientierte Unterteilung der Daten/Ereignisse .....	54
Abbildung 4-15: Beispielhafte Regression mittels SVM (links), KNN (Mitte) und CART (rechts) für Prädiktor $x_1$ und Zielvariable $y$ .....	56
Abbildung 4-16: Darstellung eines mehrlagigen Perzeptrons mit Eingabeebene, einem Hidden Layer und einer Ausgabebene. Genutzte Verknüpfungen dargestellt in Rot, ungenutzte in Grau .....	57
Abbildung 4-17: Erweiterte Darstellung der Modellkonzeption mit internen Zuständen, Informationsaustausch und internen Flüssen .....	60
Abbildung 5-1: Höhenstruktur des Einzugsgebietes Bernstein a.W.....	63
Abbildung 5-2: Abflussreaktion auf einen synthetischen Blockregen des GIUH, ELS und der LSK mit unterschiedlichen Parametrisierungen.....	64
Abbildung 5-3: Simulierte Abflussreaktion auf Blockregen in Abhängigkeit von $E(x_H)$ (links) und $s(x_H)$ (rechts) .....	65
Abbildung 5-4: Abflussreaktion (links) synthetische EZG-Formen (rechts) mit Angabe der	

minimalen und maximalen $E(x_H)$ , Gewässerachse (Punktlinie) und schematischen Distanzklassen (rot) .....	66
Abbildung 5-5: Einfluss von $v_E$ (links) und $\Delta s$ (rechts) auf simulierte Abflussreaktion.....	67
Abbildung 5-6: Einfluss von $\Delta s$ auf Simulationsgüte beobachteter Ganglinien, Verhältnis Standardabweichung (links), Mittelwert (Mitte) und Korrelationskoeffizient simulierter und beobachteter Ganglinie (rechts).....	68
Abbildung 5-7: Güteparameter der Simulation mit GIUH-Modell bei Verwendung analytischer Parameter, räumlich differenziert im Untersuchungsgebiet Main .....	71
Abbildung 5-8: Güteparameter der Simulation mit GIUH-Modell bei Verwendung analytischer Parameter, räumlich differenziert im Untersuchungsgebiet Iller .....	72
Abbildung 5-9: Güteparameter der Simulation mit GIUH-Modell bei Verwendung analytischer Parameter, räumlich differenziert im Untersuchungsgebiet Regen .....	72
Abbildung 5-10: Vergleich beobachteter und simulierter Ganglinien, berechnet mit und ohne Variation von $v_{E,H}$ und $v_{E,S}$ , Ereignis vom 10.05.2003, Pegel Kemmern.....	73
Abbildung 5-11: Güteparameter der Simulation mit GIUH-Modell bei Verwendung analytischer Parameter, zeitlich differenziert im Untersuchungsgebiet Main.....	75
Abbildung 5-12: Mittlere Güteparameter pro Teilgebiet im Untersuchungsgebiet Main...	77
Abbildung 5-13: Mittlere Güteparameter pro Teilgebiet im Untersuchungsgebiet Iller.....	77
Abbildung 5-14: Mittlere Güteparameter pro Teilgebiet im Untersuchungsgebiet Regen .	77
Abbildung 5-15: Gegenüberstellung Güteparameter analytischer Parametrisierung, $\alpha$ gegen $\beta$ (links) und $\alpha$ gegen $r$ (rechts) .....	79
Abbildung 5-16: Gegenüberstellung Güteparameter kalibrierter Parametrisierung, $\alpha$ gegen $\beta$ (links) und $\alpha$ gegen $r$ (rechts) .....	79
Abbildung 5-17: Boxplot analytische und kalibrierte $v_E$ Parameter .....	80
Abbildung 5-18: Vergleich analytische und kalibrierte Parameter mit Veränderung des KGE .....	81
Abbildung 5-19: Quantil-Quantil-Plot $Q_0$ (links) und $Q_{Vol}$ (rechts) aller Daten und Ereignisse der mit sehr geringer Abflussbildung .....	83
Abbildung 6-1: Übersicht Auswahlmöglichkeiten bei der Regression von Parametern in unbeobachteten Gebieten.....	87
Abbildung 6-2: Entwicklung Vorhersagegüte Klassifikationen nach Algorithmen.....	89
Abbildung 6-3: Entwicklung Vorhersagegüte Selektion nach Algorithmen.....	89
Abbildung 6-4: Gegenüberstellung relativer MAE ( $v_E$ und $Q_{Vol}$ ) mit und ohne Prozessunterscheidung für PUB-Regression in allen Untersuchungsgebieten.....	90

Abbildung 6-5: MAE je Untersuchungsgebiet und Algorithmus, links: $Q_{Vol}$ , rechts: $v_E$ ....	95
Abbildung 6-6: Relativer MAE $Q_{Vol}$ und $v_E$ je Teilgebiet am oberen Main berechnet mit RF .....	96
Abbildung 6-7: Relativer MAE $Q_{Vol}$ und $v_E$ je Teilgebiet am Regen berechnet mit RF ....	97
Abbildung 6-8: Relativer MAE $Q_{Vol}$ und $v_E$ je Teilgebiet am Iller berechnet mit RF.....	97
Abbildung 6-9: Relativer MAE $Q_{Vol}$ und $v_E$ je Teilgebiet am Iller berechnet mit ADA ....	98
Abbildung 6-10: MAE je Untersuchungsgebiet und Algorithmus, kalibrierte Parameter links: $Q_{Vol}$ , rechts: $v_E$ .....	99
Abbildung 6-11: Darstellung des Dynamikmoduls mit ausgewählter Prädiktorkombination und prozessorientierter Differenzierung der Ereignistypen .....	101
Abbildung 7-1: Einfluss der Lernmenge auf MAE( $v_E$ ) (oben) und MAE( $Q_{Vol}$ ) (unten) für ANN2 im Untersuchungsgebiet Iller.....	105
Abbildung 7-2: Schematische Berechnung Semivarianz DFF-Ranking.....	108
Abbildung 7-3: Vergleich MAE( $v_E$ ) (oben) und ( $Q_{Vol}$ ) (unten) bei Verwendung aller Daten (Balken) und optimaler Datenkomposition (Markierungen) mit ADA im Main .....	109
Abbildung 7-4: Vergleich MAE( $v_E$ ) (oben) und ( $Q_{Vol}$ ) (unten) bei Verwendung aller Daten (Balken) und optimaler Datenkomposition (Markierungen) mit SVM im Main .....	111
Abbildung 7-5: Lernkurven CART im Untersuchungsgebiet Iller mit selektiver Datenkomposition .....	112
Abbildung 7-6: Lernkurven ANN1 im Untersuchungsgebiet Main mit selektiver Datenkomposition .....	112
Abbildung 8-1: Median und Unsicherheitsbereiche KGE-Komponenten der Fallstudie im Untersuchungsgebiet Main, differenziert nach Teilgebiet .....	117
Abbildung 8-2: Median und Unsicherheitsbereiche KGE-Komponenten der Fallstudie im Untersuchungsgebiet Regen, differenziert nach Teilgebiet .....	117
Abbildung 8-3: Median und Unsicherheitsbereiche KGE-Komponenten der Fallstudie im Untersuchungsgebiet Iller, differenziert nach Teilgebiet.....	118
Abbildung 8-4: Ganglinien Prognose Bad Berneck (links) 30.11.2002;(rechts) 14.12.2003 .....	119
Abbildung 8-5: Vergleich Median Gütekriterien mit und ohne räumlicher Variation der Dynamikparameter .....	120
Abbildung 8-6: Normierte Anzahl von Ereignissen für die mindestens eine simulierte Ganglinie max. 0.25/0.5 vom Optimum der Güte abweicht, Untersuchungsgebiet Iller .....	121

Abbildung 8-7: Relative Häufigkeit von Ereignissen für die eine (oder mehr Prognosen) vorlagen, deren Güteparameter maximal 0.5 vom Optimalwert abwichen (oben links: $\alpha$ ; rechts: $\beta$ ; unten links: $r$ ; rechts: KGE) .....	122
Abbildung 8-8: Ensemble Unsicherheitsbänder berechnet mit gesamtem Ensemble (ENS) und GLUE für Ereignisse vom 30.11.2002 Bad Berneck (links); 14.12.2003 Wirsberg (rechts).....	124
Abbildung 8-9: Güte Vorhersagen je ML-Algorithmus, Untersuchungsgebiet Regen.....	126
Abbildung 8-10: Güte Vorhersagen je ML-Algorithmus, Untersuchungsgebiet Iller .....	127
Abbildung 8-11: Relative Häufigkeit von Ereignissen für die eine (oder mehr Prognosen) vorlagen, mit $r \geq 0.5$ ; Iller (links), Regen (Mitte), Main (rechts).....	128
Abbildung 8-12: Vergl. Kalibrierung und Fallstudie: Normierte Anzahl von Ereignissen mit mind. eine Prognose deren Güte max. Opt. $\pm 0.5$ ., Untersuchungsgebiet Regen.....	128
Abbildung 8-13: Median und Standardabweichung KGE-Komponenten in Iller, Main und Regen bei Simulation mit und ohne Klassifikationsprediktor.....	130
Abbildung 8-14: Fehlerhafte Simulationen (links) Volumenfehler, Ereignis vom 13.02.2005 Bad Berneck; (rechts) $v_E$ Fehlprognose, Adlerhütte am 28.01.2002.....	131
Abbildung 8-15: Gegenüberstellung empirischer Quantile $Q_{Vol}$ und $Q_P$ , farbliche Markierung stellt Abstand zu optimaler Simulationsgüte dar, Ergebnisse des Untersuchungsgebiets Main .....	133
Abbildung 8-16: Mittlere Ereignisgüte je Cluster, Untersuchungsgebiet Regen .....	134
Abbildung 8-17: Relativer MAE( $Q_P$ ) gegenüber der empirischen Unterschreitungswahrscheinlichkeit des Scheitels $Q_P$ , Untersuchungsgebiet Iller...	136
Abbildung 8-18: MAE( $t_A$ ) normiert auf Ereignisdauer, gegenüber der empirischen Unterschreitungswahrscheinlichkeit des Scheitels $Q_P$ , Untersuchungsgebiet Main (links) & Iller (rechts).....	137
Abbildung 9-1: Gegenüberstellung der KGE-Komponenten nach der Kalibrierung des GIUH Modells und HBV.....	140
Abbildung 9-2: Gegenüberstellung der KGE-Komponenten regionales GIUH-RF Modell und HBV Kalibrierung .....	141
Abbildung 9-3: Schematische Darstellung HBV-Regionalisierung.....	142
Abbildung 9-4: Gegenüberstellung KGE-Komponenten GIUH-RF und HBV bei Anwendung in unbeobachteten Gebieten .....	142
Abbildung 9-5: Veränderung des MAE (Rot: $v_E$ ; Blau: $Q_{Vol}$ ) durch geordnete Hinzunahme von Prädiktoren. (links oben) RF Main, Cl. 1; (rechts oben) KNN Regen, Cl. 2; (unten) ANN2 Main, Cl. 1. ....	145

Abbildung 9-6: Unterschiedliches Erklärungspotential der Indikatorgruppen abhängig vom  
Volumen/Scheitel Verhältnis ..... 146

## Tabellenverzeichnis

---

Tabelle 3-1: Ausgewählte Pegel, Einzugsgebietsgrößen und Anzahl separierter N-A-Ereignisse im EZG des oberen Mains .....	18
Tabelle 3-2: Ausgewählte Pegel, Einzugsgebietsgrößen und Anzahl separierter N-A-Ereignisse im EZG des Regens .....	20
Tabelle 3-3: Ausgewählte Pegel, Einzugsgebietsgrößen und Anzahl separierter N-A-Ereignisse im EZG der Iller .....	22
Tabelle 6-1: Spearman's $\rho$ zwischen den Ereignisindikatoren.....	91
Tabelle 6-2: Ranganalyse der besten 9 Prädiktorkombinationen .....	93
Tabelle 6-3: Ergebnis Ranganalyse getrennt nach Algorithmus, Veränderung des mittleren Fehlers bei Verwendung der Kombination NQTS .....	94
Tabelle 6-4: Lineare Korrelationskoeffizienten der MAE(Q <sub>Vol</sub> )-Werte verschiedener Algorithmen im Untersuchungsgebiet Iller .....	100
Tabelle 7-1: Auswertung präferierter Metriken nach Algorithmen und Untersuchungsgebieten (I = Iller; M = Main; R = Regen) .....	113
Tabelle 7-2: Auswahl optimaler Anzahl Datensätze und Ranking.....	114
Tabelle 8-1: Anteil Ereignisse die innerhalb des Unsicherheitsband lagen, berechnet mit ENS und GLUE; Verhältnis der Bandbreiten GLUE/ENS und normierte Bandbreite GLUE .....	125
Tabelle 8-2: Rangkorrelation zwischen KGE-Komponenten der Fallstudien und Abflussscheitel differenziert nach Untersuchungsgebiet und ML-Algorithmus.....	132
Tabelle 9-1: Ausgewählte Parametergruppe auf Position 1-4 und Form der MAE-Kurven, differenziert nach Untersuchungsgebiet, Algorithmus und Cluster. Parametergruppen: N – Niederschlag, C – Vorfeuchte, S-Schnee.....	146

## Abkürzungs- und Symbolverzeichnis

---

### Allgemein

EZG	Einzugsgebiet
DGM	Digitales Gelände Modell, Höhenangaben [m ü.N.N.]
GPV	Gesamtes Porenvolumen [mm]
nFK	nutzbare Feldkapazität [mm]
SL	Gefälle [°]

### Indikatoren

$Q_{Vol}$	Abflussvolumen [mm] oder [m <sup>3</sup> ]
$v_E$	Entwässerungsgeschwindigkeit [m/s]
$I_{Max}$	Maximale Niederschlagsintensitäten (3 Werte Mittel) [mm/h]
$I_0$	Mittelwert Niederschlagsintensitäten > 0 mm/h [mm/h]
$D_{50}$	Minimales Zeitintervall in dem 50% des Ereignisniederschlags fallen, normiert mit DP [-]
DP	Dauer Niederschlag [h]
$Q_0$	(Basis-) Abfluss bei Ereignisbeginn [m <sup>3</sup> /h] oder [mm/h]
VP	Volumen Niederschlagsereignis [m <sup>3</sup> ] oder [mm]
VRI	Vorregenindex [-]
$\psi$	Verhältnis Niederschlag/Abfluss vor Ereignisbeginn [-]
$T_E$	Mittlere Temperatur während des Ereignisses [°C]
$T_{Pre}$	Mittlere Temperatur vor dem Ereignis [°C]
$WI_E$	Kennwert Vorfeuchte Ereignismitte [-]
$WI_{Pre}$	Kennwert Vorfeuchte bei Ereignisbeginn [-]
$\Delta WI$	Änderung Vorfeuchte über Ereignisdauer [-]
SC	prozentuale Schneeüberdeckung des EZG [-]
$\Delta SC$	Veränderung der prozentualen Schneeüberdeckung [-]

Indikatorgruppen

NC	Niederschlagscharakteristika
CS	Zustandsgrößen
SC	Schneecharakteristika
UN	Unkorreliert

Klassifikation

tq	(Klassifikation basierend auf) Füllen-Scheitel Verhältnis
_local	Klassifikation basierend auf Datensatz eines einzelnen Pegels
_regional	Klassifikation basierend auf allen Datensätzen eines Untersuchungsgebietes
FuzzyEB	(Klassifikation basierend auf) Fuzzy-Entscheidungsbaum
RoS	Regen auf Schnee
SM	Schneesmelze
LR	Langer Regen
SR	Kurzer Regen
D	Fuzzyzahl Dauer
v	Fuzzyzahl Entwässerungsgeschwindigkeit
rk	Fuzzyzahl Räumliche Kohärenz

Geomorphologie

DFF	Distanz-Faktor-Funktion
$x_S$	Gewässerfließstrecke
$x_H$	Hangfließstrecke
$N_S$	Anzahl Distanzklassen innerhalb eines Einzugsgebietes basierend auf $x_S$
$\Delta s; \Delta o$	Räumliche Auflösung für Gewässer - und Hangfließstrecke der DFF
$W(i\Delta s)$	Nicht normierte Flächenfunktion einer Distanzklasse
$E(C)_i, s(C)_i, z(C)_i$	Mittelwert, Standardabweichung und Schiefe eines Kennwertes in Distanzklasse $i$

Modell

GIUH	Geomorphologische momentane Einheitsganglinie (engl. <i>Geomorphologic instantaneous unit hydrograph</i> )
GIUH-ML	Gekoppeltes Modell aus geomorphologischer Einheitsganglinie und Machine Learning
ELS	Einzellinear Speicher
LSK	Lineare Speicher Kaskade
$\Delta s$	Schrittweite Gewässerstrecke definiert die räumliche Auflösung des Modells [m]
$\Delta t$	Zeitschrittweite des Modells [h]
$v_{E;H}$	Entwässerungsgeschwindigkeit des Hangs [m/s]
$v_{E;S}$	Entwässerungsgeschwindigkeit des Gewässers [m/s]
CR	Courant-Zahl [-]
$P_{\text{eff}}$	Effektiver Niederschlag [mm/h]
$q_H$	Hangabfluss [mm/h]

Machine Learning

DDM	Datengetriebene Modelle
ML	Machine Learning / Maschinelles Lernen
MLR	Multiple lineare Regression
SVM	Support Vector Machine
KNN	K-Nearest-Neighbour
CART	Classification and regression trees
ANN	Artificial neural network
RF	Random Forest
ADA	Random Forest mit adaptivem Boosting

Gütekriterien

KGE	Kling-Gupta Effizienz
MAE	Mittlerer absoluter Fehler

# Anhänge

---

Verzeichnis der Anhänge:

Anhang A	Ergänzende Darstellungen zu Kapitel 3.3 - Fuzzyzahlen und Fuzzy Entscheidungsbaum
Anhang B	Ergänzende Darstellungen zu Kapitel 4.4.4 - Parametrisierung ML-Algorithmen - Kerndichte-Schätzer
Digitaler Anhang	Ergänzende Darstellungen zu Kapitel 5 Ergänzende Darstellungen zu Kapitel 6 Ergänzende Darstellungen zu Kapitel 7 Ergänzende Darstellungen zu Kapitel 8

## Anhang A: Fuzzyzahlen & -funktionen

Zur Klassifikation der Ereignisse eines Untersuchungsgebietes wurde unter anderem der in Kapitel 3.3.1 vorgestellte Entscheidungsbaum verwendet. An den Knoten des Baumes wird basieren auf den *Membership*-Werten verschiedener Indikatoren entschieden. Diese Werte sind die fuzzyfizierten Indikatorwerte aus Kapitel 3.2. Dabei werden Indikatoren wie das Niederschlagsvolumen VP in die Fuzzyzahlen „hoch“ und „niedrig“ umgerechnet. Folgende Tabelle A-1 listet alle Fuzzyzahlen und zugehörige Indikatoren auf. Zusätzlich wird angegeben für welche Entscheidungen die Fuzzyzahlen verwendet wurden.

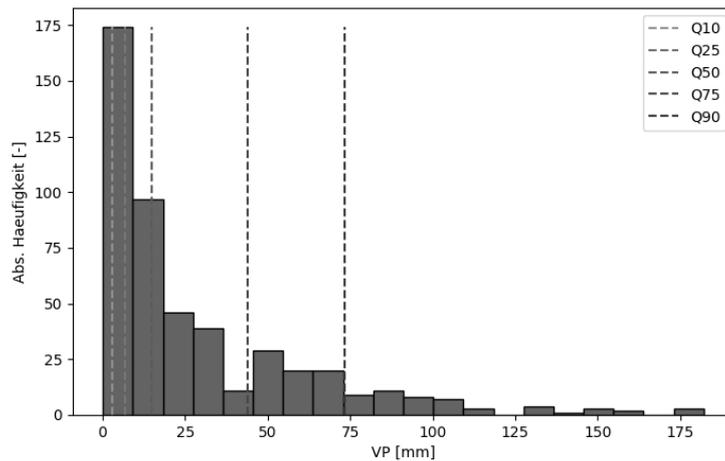
**Tabelle A1: Eingeführte Fuzzyzahlen mit zugehörigem Indikator, sortiert nach Entscheidung**

Indikator	Entscheidung		Fuzzyzahlen		
$\Delta WI$	Feuchte	fallend	stabil	steigend	
$T_{Pre}$	Schneesmelze	ja	nein		
DP	N-Dauer	lang	mittel	kurz	sehr kurz
VP	N-Menge	wenig	mittel	viel	
$v_E$	Geschwindigkeit	langsam	mittel	schnell	
RK	Ausdehnung	lokal	regional	überregional	

Die eingeführten Zahlen basieren weitestgehend auf den von Merz & Blöschl (2003) sowie Sikorska, Viviroli & Seibert (2015) verwendeten Zahlen. Als einzige Fuzzyzahlen mit physikalischer Begründung zu definieren waren in dieser Anwendung die Schneesmelzzahlen. Liegt die Temperatur unter 0°C, liegt keine Schneesmelze vor, liegt sie über 5°C liegt definitiv Schneesmelze vor. Dazwischen muss linear zwischen 0-1 interpoliert werden. Aus diesen Überlegungen leiten sich die Fuzzyzahlen  $\mu_{Schmelze}$  und  $\mu_{Keine\ Schmelze}$  ab:

$$\mu_{Schmelze} = \begin{cases} 0 & \text{wenn } T_{Pre} < 0 \\ 1 & \text{wenn } T_{Pre} > 5 \\ \frac{T_{Pre} - 0}{5 - 0} & \text{sonst} \end{cases}, \mu_{Keine\ Schmelze} = \begin{cases} 1 & \text{wenn } T_{Pre} < 0 \\ 0 & \text{wenn } T_{Pre} > 5 \\ \frac{T_{Pre} - 5}{0 - 5} & \text{sonst} \end{cases} \quad A-1$$

Die hier dargestellten Zahlen sind halb unendliche Fuzzyzahlen, da sie eine Membership von 1 für einen einseitig unbegrenzten Bereich definieren. Diese Form von Zahlen wurde für alle Fuzzyzahlen in der links äußeren und rechts äußeren Spalte von Tabelle A1 verwendet. Außer für die Schneesmelze berechnen sich diese Zahlen über die Quantile der jeweiligen Histogramme. In Abbildung A1 ist beispielhaft die Häufigkeitsverteilung von VP, samt empirischen Quantilen im Untersuchungsgebiet Iller dargestellt.



**Abbildung A1: Empirische Häufigkeitsverteilung Niederschlagsvolumen in der Iller mit Darstellung der Quantile**

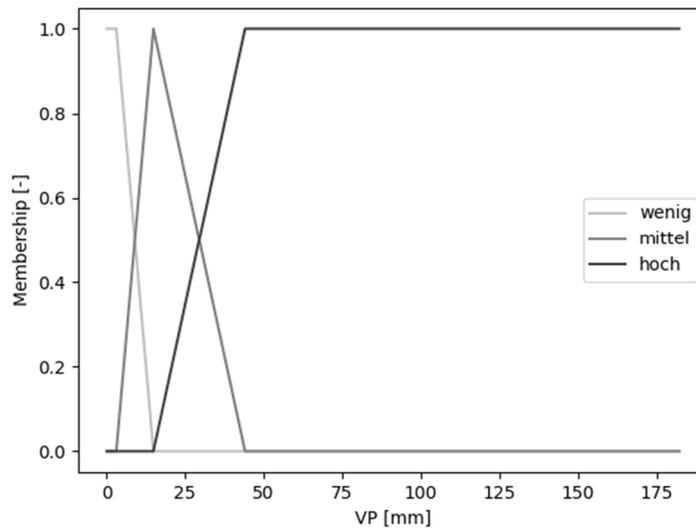
Zur Aufstellung der linksseitig unbegrenzten Fuzzyfunktionen wurden dann die 10% und 50% Quantile verwendet:

$$\mu_{\text{Niedrig}} = \begin{cases} 1 & \text{wenn } VP < Q_{10}(VP) \\ 0 & \text{wenn } VP < Q_{50}(VP) \\ \frac{VP - Q_{10}(VP)}{Q_{50}(VP) - Q_{10}(VP)} & \text{sonst.} \end{cases} \quad \text{A-2}$$

Gleichung A-2 lässt sich so mit anderen Indikatoren für alle linksseitigen Begrenzungen anwenden. Für die rechtsseitigen Begrenzungen wurde das 10% Quantil durch das 75%-Quantil ersetzt. Alle anderen Fuzzyzahlen sind Dreieckszahlen, mit links- und rechtsseitiger Begrenzung. Sie wurden über das 10%, 50% und 75% Quantil bestimmt. Folgende Fuzzyfunktionen wurden definiert:

$$\mu_{\text{Niedrig}} = \begin{cases} 0 & \text{wenn } VP < Q_{10}(VP) \\ \frac{VP - Q_{10}(VP)}{Q_{50}(VP) - Q_{10}(VP)} & \text{wenn } Q_{10}(VP) \leq VP < Q_{50}(VP) \\ \frac{VP - Q_{50}(VP)}{Q_{75}(VP) - Q_{50}(VP)} & \text{wenn } Q_{50}(VP) \leq VP < Q_{75}(VP) \\ 0 & \text{wenn } VP > Q_{75}(VP) \end{cases} \quad \text{A-3}$$

Analog zu Gleichung A-3 wurden alle anderen Dreieckszahlen bestimmt. Abbildung A-2 zeigt beispielhaft die Fuzzyfunktionen der hier dargestellten Niederschlagsdauerzahlen.



**Abbildung A-2: Fuzzyfunktionen zur Einschätzung der Niederschlagsmenge „wenig“, „mittel“ und „hoch“.**

Einzige Ausnahme des dargestellten Vorgehens bilden die Fuzzyzahlen der Niederschlagsdauer. Da in den genannten Veröffentlichungen eine Differenzierung zwischen kurzer und sehr kurzer Dauer vorgenommen wurde, wurden diese Zahlen in diese Analysen eingeschlossen. Die Fuzzyzahlen wurden über die folgenden Quantile definiert:

$$\begin{aligned} \mu_{lang} &= \{Q_{50}(DP); Q_{75}(DP)\} \\ \mu_{mittel} &= \{Q_{25}(DP); Q_{50}(DP); Q_{75}(DP)\} \\ \mu_{kurz} &= \{Q_{10}(DP); Q_{25}(DP); Q_{50}(DP)\} \\ \mu_{sehr\ kurz} &= \{Q_{10}(DP); Q_{25}(DP)\} \end{aligned} \tag{A-4}$$

Die Darstellung aller Fuzzyfunktionen und Histogramme der Indikatoren können dem digitalen Anhang entnommen werden.

### Räumliche Kohärenz

In den in Abbildung 3-10 dargestellten Fuzzy-Entscheidungsbaum sowie in die Typ-Wahrscheinlichkeiten fließt eine Fuzzyzahl der räumlichen Ausdehnung ein, die sich aus dem Kennwert der räumlichen Kohärenz RK berechnet. Dieser Indikator wurde nur in diesem Zusammenhang verwendet, daher wird seine Berechnung in diesem Anhang dargestellt. Seine Berechnung basiert auf den Definitionen von Merz & Blöschl (2003).

Im Gegensatz zu allen anderen Indikatoren basiert dieser auf dem Vergleich von Daten aller Teilgebiete aus allen Untersuchungsgebieten. Ausgangspunkt ist ein Ereignis an einem Teilgebiet, dessen räumliche Kohärenz bestimmt werden soll.

Basierend auf dem Eintrittsdatum des Abflussscheitels wird das gleichzeitige Auftreten des Ereignisses in allen anderen Gebieten ermittelt. „Gleichzeitig“ wurde im Rahmen dieser Anwendung als das identische Datum  $\pm 5$  Tage definiert.

Die Wahrscheinlichkeit, dass es sich um ein regionales Ereignis handelt, wird über die Anzahl der Teilgebiete mit einem gleichzeitig auftretenden Ereignis normiert über die Anzahl aller Teilgebiete im Untersuchungsgebiet berechnet. Beispiel: Am Pegel Birgsau (Iller) tritt am 22.11.1986 ein Hochwasserereignis auf. An 6 der 9 umliegenden Pegel der Iller traten zum selben Datum  $\pm 5$  Tage ebenfalls Hochwasser auf. Die Wahrscheinlichkeit, dass es sich um ein regionales Ereignis handelt ist  $6/9 = 66.7\%$ .

Um überregional auftretende Ereignisse zu identifizieren, wurde das gleichzeitige Auftreten eines Ereignisses in allen verfügbaren Teilgebieten untersucht. Die Wahrscheinlichkeit, dass es sich um ein überregionales Ereignis handelt, wird dann über die Anzahl gleichzeitiger Ereignisse geteilt durch die Anzahl aller verfügbaren Teilgebiete bestimmt. Beispiel: Am Pegel Birgsau (Iller) tritt am 22.11.1986 ein Hochwasserereignis auf. In 6 der 9 verbleibenden Gebieten der Iller, 10 von 11 Teilgebieten des Regens und in 12 von 22 Teilgebieten am oberen Main traten ebenfalls Ereignisse auf. Die Wahrscheinlichkeit, dass es sich um ein überregionales Ereignis handelt ist demnach  $(6+10+12)/(9+11+22) = 66.7\%$



## Anhang B: Ergänzende Darstellungen zu Kapitel 4

---

In der Anwendung des GIUH-ML Modells wird bei der Initialisierung des Modells für ein definiertes Zielgebiet dem Dynamikmodul ein Trainingsdatensatz  $X$  übergeben und der dazu gehörige Datensatz  $Y$  mit den Zielwerten. Datensatz  $X$  besteht aus  $K$  Variablen (Prädiktoren, nach Kapitel 6 sind dies stets 8) und  $N$  Stichproben (Ereignissen), Datensatz  $Y$  besteht aus 2 Variablen und ebenfalls aus  $N$  Stichproben.

Im Prognosefall wird dem Modul ein neuer Datensatz  $X_P$  übermittelt und dieses gibt als Ergebnis einen Datensatz  $Y_V$  mit 2 Variablen und 8 Prognosen zurück. Die zwei Variablen entsprechen den Variablen der Zielwerte während der Trainingsphase, und die 8 Prognosen werden mit unterschiedlichen ML-Algorithmen erstellt. In diesem Anhang soll auf die 8 Algorithmen und ihre Parametrisierung eingegangen werden.

Zu Beginn der Trainingsphase werden den Datensätzen  $X$  und  $Y$  20% der Daten entzogen, um die Anpassungsgüte der Algorithmen bestimmen zu können. Die Güte wird über den linearen Korrelationskoeffizienten zwischen Prognose der zurückgehaltenen Daten und den zurückgehaltenen Stichproben aus  $Y$  bestimmt.

### Multiple Lineare Regression

Den bereits reduzierten Datensätzen  $X_{Red}$  und  $Y_{Red}$  werden weitere 20% entzogen, um eine MLR-interne Modellauswahl durchzuführen. Basieren auf den verbleibenden Daten werden fünf MLR-Modell entwickelt. Die Modelle unterscheiden sich durch die Zielfunktion der Parameteranpassung (Darstellung nach Pedregosa et al., 2011):

$$\text{Fehlerquadrate:} \quad \min_{\omega} \|X\omega - Y\|_2^2 \quad \text{B-1}$$

$$\text{Ridge Regression:} \quad \min_{\omega} \|X\omega - Y\|_2^2 + \alpha \|\omega\|_2^2 \quad \text{B-2}$$

$$\text{Lasso:} \quad \min_{\omega} \frac{1}{2N} \|X\omega - Y\|_2^2 + \alpha \|\omega\|_1 \quad \text{B-3}$$

$$\text{Elastic net:} \quad \min_{\omega} \frac{1}{2N} \|X\omega - Y\|_2^2 + \alpha \rho \|\omega\|_1 + \frac{\alpha(1-\rho)}{2} \|\omega\|_2^2 \quad \text{B-4}$$

*LARS:* *Iteratives Vorgehen*

Dabei steht  $\omega$  jeweils für den gesuchten Vektor von Gewichten der MLR, und besteht aus  $K$  Elementen. Bei der Minimierung der Fehlerquadrate (Gleichung B-1) wird die Betragsnorm der quadrierten Fehler zwischen Prognose und Beobachtungen berechnet und als alleinige

Zielgröße verwendet. Bei der *Ridge Regression* (B-2) wird ein zweiter Term eingeführt, ein Strafterm, der die absolute Größe der Gewichte kontrolliert. Steuerparameter  $\alpha$  wurde gleich 1 gesetzt, hätte aber zur stärkeren Regulierung der Gewichte erhöht werden können. Durch den Strafterm wird einer Überanpassung vorgebeugt, die durch Multikolarität einzelner Prädiktoren erzeugt werden kann. Bei der *Lasso* (B-3) Anpassung wird der Strafterm der Parametergröße über die L1-Betragsnorm (Ramsay & Silverman, 2005) gebildet und die Fehlerquadrate über die Anzahl der Stichproben minimiert. Mit *Lasso* können so parameterarme MLR-Modelle entstehen, da Parametersätze mit möglichst vielen Null-Gewichten bevorzugt werden. Die Vorteile von *Lasso* und *Ridge Regression* werden mit dem *Elastic Net* (B-4) kombiniert. Ein L1-Strafterm wird hier mit einem L2-Strafterm zusammen verwendet. Parameter  $\rho$  (mit 0.5 verwendet) dient zur Kombination der Normen. *Elastic Net* wird angewendet, um die Anzahl an Parametern zu reduzieren und gleichzeitig Multikolarität der Prädiktoren zu berücksichtigen. *LARS* steht für *Least angle Regression* und steht für einen iterativen Algorithmus, der vor allem für Datensätze mit vielen Prädiktoren und wenig Stichproben entwickelt wurde. Wie bei einer stufenweisen Regression werden in jedem Schritt weitere Parameter hinzugenommen. Diese werden aber nicht einzelnen angepasst sondern in jedem Schritt gemeinsam in die Richtung der geteilten Korrelation mit dem Residuum der Regression angepasst.

### Support Vector Machines

Die verwendete Support Vector Machine (SVM) in der Implementierung von Pedregosa et al. (2011) kann mit vier unterschiedlichen Kernel-Funktionen betrieben werden. Die Kernel-Funktionen dienen zur impliziten Projektion der Eingangsdaten  $X$  vom  $K$ -dimensionalen Raum in den  $\Lambda$ -dimensionalen Raum. Durch diese Projektion soll das Problem linearisiert werden und somit eine Regression zu ermöglichen. Der sog. „Kernel Trick“ (Han & Kamber, 2010) besteht darin, die Daten nicht tatsächlich in den höher dimensionalen Raum zu verschieben sondern das innere Produkt des höher dimensionalen Vektorraums. Diese Berechnung ist numerisch effizienter, als die vollständige Berechnung aller Koordinaten des höherdimensionalen Raums. Für eine detailliertere Beschreibung des „Kernel Trick“ sei an dieser Stelle auf Cortes & Vapnik (1995) verwiesen.

Folgende Kernel standen zur Verfügung, dabei ist  $k$  die Abbildung des Eingaberaums  $(x,y)$  (also Raum der Eingangsdaten) in den höher dimensionalen Merkmalsraum:

$$\textit{linear} \quad k(x, y) = \langle x, y \rangle \quad \text{B-5}$$

$$\textit{polynomisch} \quad k(x, y) = (\gamma(x, y) + r)^d \quad \text{B-6}$$

$$\textit{rbf} \quad k(x, y) = \exp(-\gamma \|x, y\|^2) \quad \text{B-7}$$

$$\textit{sigmoid} \quad k(x, y) = \left\langle \tanh(-\gamma \|x, y\|^2 + r) \right\rangle \quad \text{B-8}$$

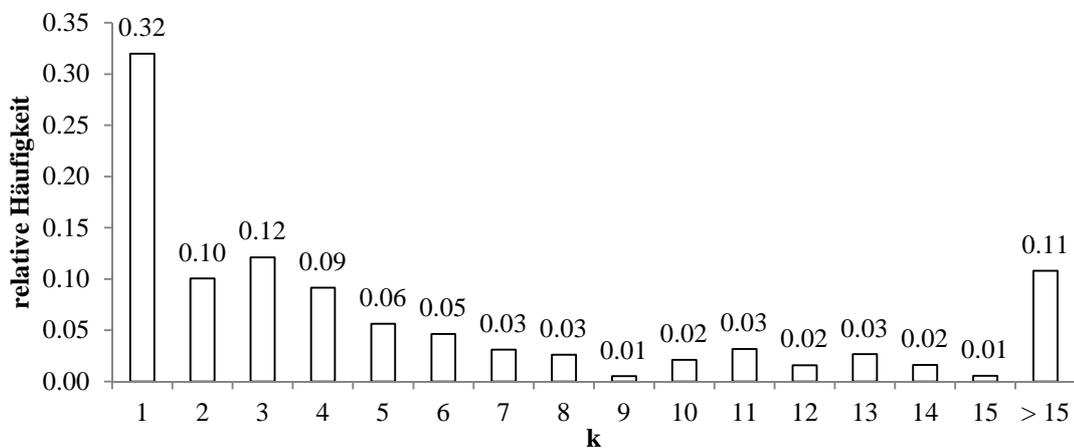
In den durchgeführten Tests zu Anpassung der SVR (*Support Vector Machine Regressor*)

wurden alle Kernelfunktionen getestet. Die dabei verwendeten Parameter wurden mit  $d = 3$ ,  $\gamma = 1/N$  und  $r = 0$  gewählt. Diese Parametrisierung der Kernel folgte den Empfehlungen von Pedregosa et al. (2011). Der in allen Untersuchungsgebieten durchgeführte Test beinhaltete eine Anpassung auf 80% der zur Verfügung stehenden Daten und eine Validierung auf den zurückgehaltenen 20%. Dieser Test wurde 50-Mal wiederholt und die Ergebnisse verglichen. Sigmoid und polynomischer Kernel erwiesen sich dabei als sehr rechenintensiv, der *rbf* kam in einigen Iterationen zu gar keinem Ergebnis. Der lineare Kernel erwies sich als beste Alternative, da er zum einen am effizientesten arbeitete und zuverlässig gleichbleibende Ergebnisse erzeugte. Daher wurde der lineare Kernel als Standard zur Anpassung der SVR verwendet.

### k-nächste Nachbarn (KNN)

Zur Parametrisierung des KNN wurden wie bei MLR weitere 20% der Trainingsdaten zur internen Anpassungsprüfung zurückgehalten. Der KNN Regressor definiert die Zielvariable als Mittelwert der  $k$  ähnlichsten Datensätze, die ihm zur Verfügung stehen. Die Auswahl von  $k$  ist daher entscheidend für die Ergebnisse des Schätzers.

Vorab wurde ein Test mit allen Daten des Untersuchungsgebietes durchgeführt, wieder mit 80%-Training und 20% Validierungsdaten mit 50 Iterationen. Dabei wurde  $k$  zwischen 1 und 100 variiert und für jeden Durchgang das optimale  $k$  gespeichert. Abbildung B-1 zeigt die Häufigkeitsverteilung der gewählten Parameter für die Zielvariable  $Q_{Vol}$ , Ergebnisse für  $v_E$  analog.

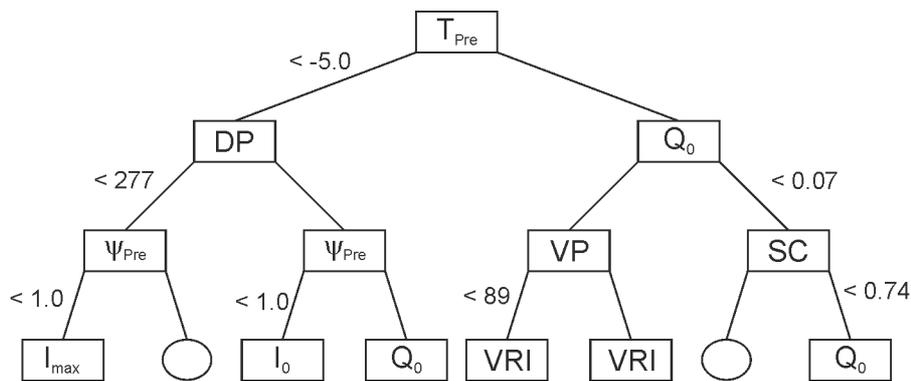


**Abbildung B-1: Häufigkeitsverteilung optimales  $k$  im Test des KNN in allen Untersuchungsgebieten**

Es zeigt sich, dass in 32% der Fälle die direkte Verwendung des Wertes des nächsten Nachbarn optimal war. Der Mittelwert über weitere 2-10 Nachbarn wurde in weiteren 50% der Fälle. Die kumulierte Häufigkeit bis 15 Nachbarn beträgt 89%. Daher wurde bei der individuellen Anpassung, um die numerische Effizienz zu erhöhen, zwischen 1 und 15 iteriert und das beste  $k$  gewählt.

**Entscheidungsäume (CART)**

Die verwendeten Regressionsäume richten sich in ihrer Parametrisierung nach den Empfehlungen von Pedregosa et al. (2011). Im Aufbau des Baumes werden Verästelungen, d.h. Aufteilungen der Lernmenge  $X$  in die Teilmengen  $X_1$  und  $X_2$ , vorgenommen. Der Mittelwert aller zugehörigen Teilmengen der Zielvariablen  $Y_1$  und  $Y_2$  werden dann als Schätzung des Regressors  $\hat{y}_1$  und  $\hat{y}_2$  verwendet. Die Teilung der Daten, die den mittleren quadratischen Fehler zwischen  $Y_1$  und  $\hat{y}_1$  sowie  $Y_2$  und  $\hat{y}_2$  minimiert, wird als erster „Split“ verwendet. Alle weiteren Verästelungen des Entscheidungsbaumes folgen dieser Logik. Es werden so lange Verästelungen durchgeführt, bis jede Stichprobe des Datensatzes alleine am Ende einer Verästelung steht, den sogenannten *Leafs*. Abbildung B-2 zeigt beispielhaft die obersten 4 Ebenen eines Entscheidungsbaums, dessen erster Split über der Temperatur vor Ereignisbeginn durchgeführt wird. Auf der linken Verästelung wird im weiteren Verlauf mit der Niederschlagsdauer und dem Abflussbeiwert der vorangegangenen 5 Tage differenziert. Auf der rechten Verästelung wird über den Basisabfluss, das Niederschlagsvolumen und die Schneebedeckung differenziert.



**Abbildung B-2: Oberste 4 Ebenen eines Entscheidungsbaumes, angepasst an Daten des Mains, *tq\_regional* Cluster 2; Rechteck = Verästelung; Kreis = Leaf**

Über die maximale Tiefe der Bäume kann die Anzahl der Ebenen und *Splits* minimiert werden. Bei der Parametrisierung der Regressionsäume werden stets drei Begrenzungen getestet (Training 80%, Validierung 20%): 15 (Anzahl aller verfügbaren Prädiktoren), K (Anzahl der verwendeten Prädiktoren) und unbegrenzt. In den meisten Fällen wird die restriktivste