Thesis for the Degree of Doctor of Philosophy

Towards Prediction in Ungauged Aquifers – Methods for Comparative Regional Analysis

Ezra M. Haaf

Department of Earth Sciences

Faculty of Science

UNIVERSITY OF GOTHENBURG

Gothenburg, Sweden 2020

Cover illustration: Marit Fahlander

Towards Prediction in Ungauged Aquifers – Methods for Comparative Regional Analysis

ezra.haaf@gmail.com

ISBN 979-91-7833-848-1 (PRINT)
ISBN 979-91-7833-849-8 (PDF)
Available at: http://hdl.handle.net/2077/63694

Printed in Borås, Sweden 2020
Printed by Stema Specialtryck AB, Borås

The nature of our intelligence is such that it is stimulated far less by the will to know than by the will to understand, and, from this, it results that the only sciences which it admits to be authentic are those which succeed in establishing explanatory relationship between phenomena.

Marc Bloch

Abstract

Hydrogeological investigations and in particular groundwater resource assessments are strongly reliant on understanding the factors controlling groundwater level dynamics. However, historical records of measured groundwater levels are often scarce and unevenly distributed in space and time. This irregularity of measurements, combined with hydrogeological systems with heterogeneous properties and unclear inputs and driving processes, leads to the need for systematic methods for prediction of groundwater in poorly-observed (ungauged) groundwater systems. In this thesis, methods of comparative regional analysis are presented to estimate groundwater level dynamics at ungauged sites based on similarity of groundwater system response and climatic and non-climatic characteristics. In order to carry out comparative regional analysis, methods were developed and compared for measuring similarity of groundwater system response based on entire (Paper I) and on features (Paper II) of groundwater levels time series. The relationship between similar groundwater response and groundwater system characteristics are evaluated further by identifying groups of similar sites using similarity-based classification (Paper I-III). Finally, climatic and physiographic system characteristics are identified that can be linked to groundwater dynamics aided by regression analysis and conceptual models (Paper IV). They can therefore serve as a basis for prediction in ungauged aquifers (Paper V).

The thesis presents novel methods for regional analysis of groundwater resources that can be used to link groundwater dynamics to groundwater system characteristics. It demonstrates the strong potential of the presented methods and ways forward for prediction of groundwater dynamics in ungauged aquifers.

Keywords: Groundwater, Prediction in ungauged aquifers, Comparative hydrogeology, Similarity, Groundwater hydrographs, Groundwater dynamics, Groundwater dynamics features, Time series clustering, Groundwater climate interaction, Classification, Regression, Groundwater storage, Groundwater resources management.

Sammanfattning

Att ha kunskap om fluktuationen av regionala grundvattennivåer och därmed vattenmagasinering är central för en hållbar hantering av grundvattenresurser. Grundvattensystem är dock komplexa och dess egenskaper uppvisar en stor rumslig variation, vilket leder till en betydande variabilitet i grundvattnets respons till klimatet. Responsen mäts främst som grundvattennivåer i observationsbrunnar i olika grundvattenmagasin. Men på många platser är grundvattennivåmätserier inte tillgängliga. Denna avhandling presenterar därför metoder och verktyg baserade på jämförande regionalanalys. Syftet är att sammanlänka grundvattennivåfluktuationer och egenskaper hos grundvattensystem. Detta gör det möjligt att prediktera grundvattnets kvantitativa status och dess variabilitet på platser där mätningar saknas. I avhandlingen utvecklas och jämförs metoder för att mäta likhet i grundvattensystemrespons baserat på historiska grundvattennivåmätserier (Paper I) och från mätserierna aggregerade statistiska mått (Paper II). Förhållandet mellan grundvattenrespons och -systemegenskaper utvärderas genom att liknande system identifieras med hjälp av likhetsbaserad klassifikation (Paper I-III). Vidare identifieras klimat- och miljövariabler som kan länkas till grundvattennivåfluktuationer genom regressionsanalys och konceptuella modeller (Paper IV). Dessa modeller ligger till grund för uppskattningen av grundvattennivåmätserier på platser utan observationer (Paper V).

Avhandlingen presenterar nya metoder för jämförande regionalanalys av grundvattenresurser. Den påvisar stor potential i metoderna, och banar vägen mot en sammanhållen strategi för prediktering i grundvattenmagasin där mätningar saknas.

List of papers

This thesis is based on the following studies, referred to in the text by their Roman numerals.

Appended to the thesis

I. **Haaf, E.**, Barthel, R. (2018). *An inter-comparison of similarity-based methods for organisation and classification of groundwater hydrographs*. Journal of Hydrology 2018; 559: 222-237.

II. Heudorfer, B.*, **Haaf, E.***, Stahl, K., Barthel R. (2019). *Index-Based Characterization and Quantification of Groundwater Dynamics* (*equal contribution). Water Resources Research 55(7): 5575-5592.

III. Giese, M., **Haaf, E.**, Heudorfer, B., Barthel, R. (2020). *Comparative hydrogeology – reference analysis of groundwater dynamics from neighbouring observation wells*. Hydrological Sciences Journal (accepted).

IV. **Haaf, E.**, Giese M., Heudorfer, B., Stahl, K., Barthel, R. (2020). *Physiographic and climatic controls on regional groundwater dynamics*. Revision submitted to Water Resources Research.

V. **Haaf, E.**, Giese M., Reimann, T., Barthel, R. (2020). *Estimation of daily groundwater levels in ungauged aquifers based on climatic and physiographic controls*. Manuscript.

Division of work between the authors

In Paper I, Haaf and Barthel conceived the study. Haaf prepared the data and performed the data analysis. Barthel carried out the visual classification. Haaf wrote the manuscript. All co-authors edited and revised the manuscript and approved the final version.

In Paper II, Haaf and Heudorfer conceived the study together with Barthel and Stahl. Heudorfer and Haaf prepared the data and performed the statistical analysis. Haaf and Heudorfer wrote the manuscript. All co-authors edited and revised the manuscript and approved the final version.

In Paper III, Giese initiated the study and designed it together with Haaf. Giese prepared and analyzed the site data, Haaf prepared the time series data and performed the statistical analysis. Giese and Haaf wrote the manuscript. All co-authors edited and revised the manuscript and approved the final version.

In Paper IV, Haaf conceived the study with input from all co-authors. Haaf, Giese and Heudorfer calculated indices and climatic, geologic and DEM-derived descriptors, respectively. Haaf performed the statistical analysis and wrote the manuscript. All co-authors edited and revised the manuscript and approved the final version.

In Paper V, Haaf conceived the study with input from all co-authors. Haaf performed the statistical analysis and wrote the manuscript with input from Giese. All co-authors edited and revised the manuscript and approved the final version.

Other peer-reviewed publications not included in this thesis

Sundell J, Haaf E, Norberg T, Alén C, Karlsson M, Rosén L. 2017. Risk Mapping of Groundwater-Drawdown-Induced Land Subsidence in Heterogeneous Soils on Large Areas. Risk Analysis, **39**: 105-124. DOI: 10.1111/risa.12890.

Sundell J, Haaf E, Tornborg J, Rosén L. 2019. Comprehensive risk assessment of groundwater drawdown induced subsidence. Stoch Environ Res Risk Assess, **33**: 427-449. DOI: 10.1007/s00477-018-01647-x.

Sundell J, Norberg T, Haaf E, Rosén L. 2019. Economic valuation of hydrogeological information when managing groundwater drawdown. Hydrogeology Journal. DOI: 10.1007/s10040-018-1906-z.

Sundell J, Rosén L, Norberg T, Haaf E. 2016. A probabilistic approach to soil layer and bedrock-level modeling for risk assessment of groundwater drawdown induced land subsidence. Engineering Geology, **203**: 126-139. DOI: https://doi.org/10.1016/j.enggeo.2015.11.006.

Table of Contents

Acknowledgements

Throughout the years working on the thesis, I have had the great privilege of learning from and working together with plenty of inspiring and knowledgeable people. Roland Barthel, of course, my main supervisor, was the one to conceive the initial idea that got me excited for this project and has been helpful and supportive throughout while keeping me on my toes. Markus Giese and Benedikt Heudorfer have contributed a lot towards probing and implementing research ideas, and have been invaluable sounding boards at various phases. I also want to thank my co-supervisor Thomas Reimann, who despite entering the project at a relatively late stage, has provided moral support and new perspectives, as well as a thorough review of this thesis (all remaining errors are mine). Highly appreciated are also perspectives and rigorous manuscript reviews by Kerstin Stahl. I am grateful to Deliang Chen for advice and for being a supportive examiner.

I want to thank my friend and colleague Jonas Sundell, an early reader of this thesis, for his good advice throughout these years. Thanks to Stefan Banzhaf and Michelle Nygren for their support. I also want to acknowledge my all PhD colleagues and especially Aifang Chen, Lorenzo Minola, Magdalena Wallman and Hui-Wen Lai for their midterm and kappa reviews. Thanks to David Rayner for his expertise and time, as well as trust in teaching matters. I am grateful of all present and former colleagues at the Department of Earth Sciences for providing great company, interesting discussions and an environment, where critical thinking and research can flourish. I also want to thank my former and present colleagues of the hydrogeology groups at COWI Stockholm and Gothenburg as well as Lars Rosén's lab at Chalmers for strongly contributing to making hydrogeology a field pleasurable to work in. I gratefully acknowledge the foundations that awarded me stipends and grants for courses, conferences, research secondments, and writing retreats throughout my PhD education: Sven Lindqvist Forskningsstiftelsen, Adlerbertska Hospitiestiftelse, Adlerbertska Forskningsstiftelse, Adlerbertska Stipendiestiftelsen, Donationsnämndens stipendier, Stiftelsen ÅForsk, Jonseredsstiftelsen, Kungliga och Hvitfeldtska stiftelsen.

I deeply appreciate all the support by the Haaf and Fahlander families throughout the years. Hilding, you have helped me to hurry up and finish writing this "boring book", so I can spend more time with you. Finally, Marit, thank you for all your patience, understanding, and encouragement.

Ezra Haaf, Gothenburg in May 2020

1 Introduction

1.1 Background

Estimating the dynamics of groundwater storage is essential for the assessment of groundwater availability and the design of sustainable groundwater management. Worldwide, groundwater is the only freshwater source for roughly 3 billion people and it also supplies a crucial portion of the world's agricultural (42%) and industrial (27%) freshwater demand (Döll *et al.*, 2012; UNESCO, 2015). Simultaneously, groundwater plays a central role in sustaining ecosystems, carbon storage, and buffering the effects of a changing hydrological cycle as a result of climate change (Taylor *et al.*, 2013; de Graaf *et al.*, 2019; Qiu *et al.*, 2019). Since storage of freshwater in glaciers and snowpack will continue to decrease, groundwater storage is also predicted to increase in importance for societies worldwide (Green *et al.*, 2011). Furthermore, while it is certain that climate change will also impact the quantity (and quality) of groundwater, the IPCC (Intergovernmental Panel on Climate Change) stated in their last assessment report (AR5) that the understanding of why and how is still limited (Jiménez Cisneros *et al.*, 2014). One of the underlying reasons is a lack of understanding of which factors control the response of groundwater storage to climate over time (Barthel, 2014; Boutt, 2017; Li *et al.*, 2019).

Without good knowledge of the factors controlling storage dynamics, water balance calculations - the bread and butter of water resources planning - can be significantly erroneous (Istanbulluoglu *et al.*, 2012). This is particularly true at the regional scale (Wang *et al.*, 2009; Billah and Goodall, 2011; Fan, 2015), where water managers assess the current and future status of groundwater resources (Lóaiciga and Leipnik, 2001). The regional scale (here understood as regions, e.g. catchments, covering areas of 10^3 to 10^5 km^2) plays an important role, since it connects global projections to local impacts (Wilbanks and Kates, 1999; VanRoosmalen *et al.*, 2007), allowing integrated analysis of feedback interactions between nature and society (Holman *et al.*, 2012). One of the dominant problems on the regional scale is that an enormous variety of hydrogeological conditions can be exhibited here. These conditions or settings dictate the variations of groundwater levels. Consequently, the variability in settings is also reflected in the large variability of patterns of groundwater level fluctuations. Figure 1 shows the variability of such patterns at three different sites. However, groundwater observations in both time and space are scarce and unevenly distributed. This patchiness of observations, combined with the fact that groundwater is hidden from view in the subsurface, makes comprehensive measurement of the groundwater status a difficult task. The satellite mission Gravity Recovery and Climate Experiment (GRACE) is an attempt to give estimates of groundwater storage over time with (nearly) global coverage (Strassberg *et al.*, 2007). However, resolution in time and space is coarse (Rodell *et al.*, 2007). Further, translating the satellite data into groundwater storage has proven to be difficult, resulting in large deviations from groundwater level measurements, which are direct observations of day-to-day change in groundwater storage in actual aquifers (Van Loon *et al.*, 2017). Consequently, methods based on extrapolating from or calibrating to groundwater level observations (groundwater hydrographs) in a hydrologically meaningful manner are still the main approach for groundwater status assessment in water resources management. However, in order to achieve a meaningful prediction of status at locations with none or few measurements (ungauged), methods need to integrate system understanding and, therefore, the factors controlling changes in groundwater levels.

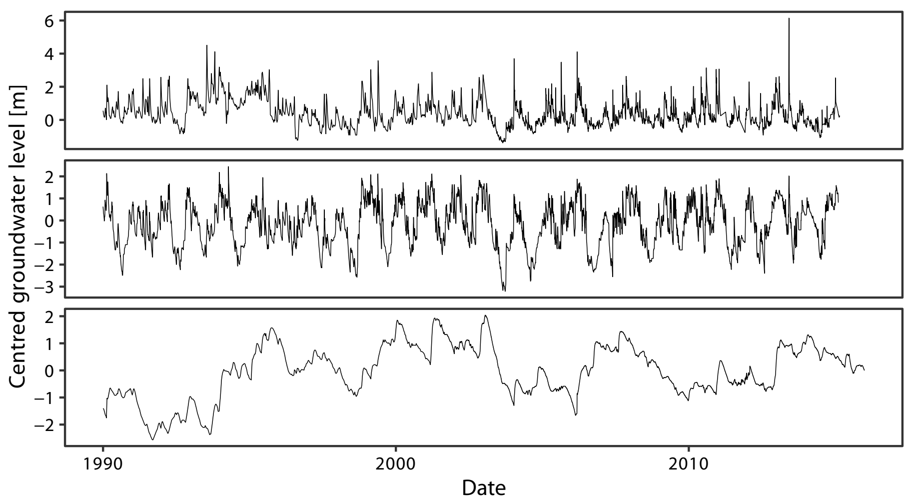

Figure 1. Variability in patterns of groundwater level dynamics in three daily groundwater level time series from southern Germany (from Paper II).

Today, the most common strategies to solve groundwater problems on the regional scale are either the application of process-based, numerical models of groundwater flow and transport, or with conceptual numerical hydrological models which include groundwater components (Barthel and Banzhaf, 2016). However, both approaches are problematic when applied at the regional scale. Using numerical groundwater models on the regional scale may provide elusive results as a consequence of lacking system understanding (Zektser and

Dzyuba, 2014). At the same time, data scarcity, and in particular uneven spatial distribution of data of the subsurface is seen as the greatest challenge to regional-scale groundwater modelling (Candela *et al.*, 2014). Conceptual (rainfall-runoff) hydrological models, on the other hand, are problematic as they usually disregard the foremost local characteristics of hydrogeological systems and their three-dimensional setup. Generally, conceptual models provide adequate descriptions of groundwater systems only for simple hydrogeological situations such as shallow, unconfined aquifers (Barthel, 2006; Götzinger *et al.*, 2008), but not for deeper and confined systems, which may be of much higher importance from a long-term and regional management perspective. Large scale integrated surface and subsurface models for meaningful local projections are still at an early stage (Berg and Sudicky, 2019).

A different approach is to use methods usually classified as data-driven methods, which are still uncommon for regional scale assessments. These methods use for example impulse-response functions (e.g. Gottschalk, 1977; Von Asmuth, 2012; Bakker and Schaars, 2019) or artificial neural networks (Wunsch *et al.*, 2018). Here, less data on system characteristics is required, while relatively long and measurement-dense series of groundwater measurements are needed to achieve good calibrations. No formal method is attached to these approaches to transfer information from gauged to ungauged aquifer. This means that these data-driven methods can only be used to make predictions given sufficient time series data at the local point of interest. In summary, neither numerical models nor the currently available data-driven tools provide directly applicable results with respect to status assessment and predictions for groundwater resources on the regional scale. Thus, new and complementary approaches are required to overcome the problem of scarcity and uneven data distribution. These approaches should be less data hungry than numerical models, while still accounting for local hydrogeological conditions and allowing data-driven prediction with little or no local data.

In surface-water-orientated hydrology, several authors have proposed and applied classification and similarity analysis as new concepts to cope with data scarcity (McDonnell and Woods, 2004; Wagener *et al.*, 2007; Sivakumar and Singh, 2012; Blöschl *et al.*, 2013; Hrachowitz *et al.*, 2013) based on the concept of comparative analysis, populated by Falkenmark *et al.* (1989). These concepts attempt to link the physical form and structure of surface-water systems to their functioning. This link can then be used to transfer information to similar systems. These concepts, however, have not yet been extended to groundwater hydrology. Hydrogeological classification has been applied in groundwater hydrology for a long time, e.g. Blank and Schroeder (1973), and tends to use very low levels of formalization, which are mainly verbally descriptive and are only made for and applied in specific geographical contexts. In the context of numerical groundwater models, arrays of cells (zones) with similar properties are commonly defined based on an underlying classification of mapped hydrogeological units. Yet, these zones linked to mapped units are hardly ever explicitly based on a systematic analysis of groundwater response, such that units are quantitatively linked to similarity of groundwater response. While such response-based classification schemes have been proposed for groundwater dependent ecosystems (Bertrand *et al.*, 2011; Stein *et al.*, 2012; Martens *et al.*, 2013), the general lack of systematic methods based on comparative analysis in groundwater research has been pointed out by various authors (e.g. de Marsily *et al.* (2005); Voss (2005); Green *et al.* (2011)).

Although principles of comparative analysis in hydrogeology (based on hydrogeological units) have been in existence within hydrogeology since at least the 1970s (Knutsson and Fagerlind, 1977), methods for informed and quantitative analysis as well as prediction of groundwater level dynamics have not. Such methods are based on the assumption that similar input to similar systems results in a similar response. This assumption is based on the idea that groundwater levels are the result of the temporal and spatial superposition of a multitude of processes, which in their turn are dependent on system properties. This means that information on the system properties is contained in the groundwater hydrograph (e.g. Law, 1974). Therefore, by finding methods to extract this information from hydrographs, hydrogeological response can be quantitatively linked to systems properties. This is an important step to transfer hydrogeological response from gauged to ungauged location, since linking the response to system elements has two main benefits (c.f. Gottschalk, 1985): (1) System properties are generally available at greater resolution than groundwater level observations and, therefore, can enable extrapolation to ungauged aquifers; and (2) system properties lie at the basis of explaining the changes in groundwater storage and therewith the formation of the water balance. Consequently, the present thesis provides different approaches to comparative regional analysis for quantitative prediction, i.e. estimation of the outcome of unseen data, in ungauged aquifers.

1.2 Objectives

This study aims to investigate and develop methods based on comparative regional analysis in order to make predictions of groundwater dynamics for regions with few observations, effectively transferring information from a gauged to an ungauged site. Specific objectives are listed below.

(1) Contrast visually-based and mathematical methods for measuring similarity in hydrogeological response by comparison of similarity-based classification methods of groundwater level time series.

(2) Develop a vocabulary for description coupled to mathematical indices for quantification of groundwater level dynamics that can be used for comparative analysis of groundwater level dynamics.

(3) Investigate the skillfulness of (groundwater dynamics) indices for differentiating (classified) local, intermediate and regional groundwater flow systems, and drivers of physical processes.

(4) Use regression analysis for exploratory regional analysis to understand which factors control groundwater dynamics.

(5) Investigate predictability of groundwater level time series at ungauged sites based on geological, topographical and climatic system properties with a nearest neighbor approach.

1.3 Scope and outline of thesis

The scope of the thesis is to explore, adapt and develop methods for predictions in ungauged aquifers by linking system properties to hydrogeological response through comparative regional analysis (Figure 2). The body of this work is documented in five appended papers:

Paper I. An inter-comparison of similarity-based methods for organisation and classification of groundwater hydrographs

Paper II. Index-based characterization and quantification of groundwater dynamics

Paper III. Comparative hydrogeology – reference analysis of groundwater dynamics from neighbouring observation wells

Paper IV. Physiographic and climatic controls on regional groundwater dynamics

Paper V. Estimation of daily groundwater level time series in ungauged aquifers based on climatic and physiographic controls

In Paper I, methods that can be used for similarity-based grouping of groundwater hydrographs and classification are contrasted. Paper II follows up on one of the main conclusions from Paper I, delivering a set of tools to quantify more precisely similarity between groundwater hydrographs. Here also an example case of similarity-based classification is shown. Paper III further explores the usefulness of the methods developed in Paper II to distinguish different groundwater flow systems and physical processes. The results are expanded on in Paper IV, where climatic and physiographic properties at each well location are used to establish regression-based relations to groundwater response behavior. Finally, Paper V shows an attempt of estimating groundwater level time series at ungauged sites based on system characteristics and nearest neighbors.

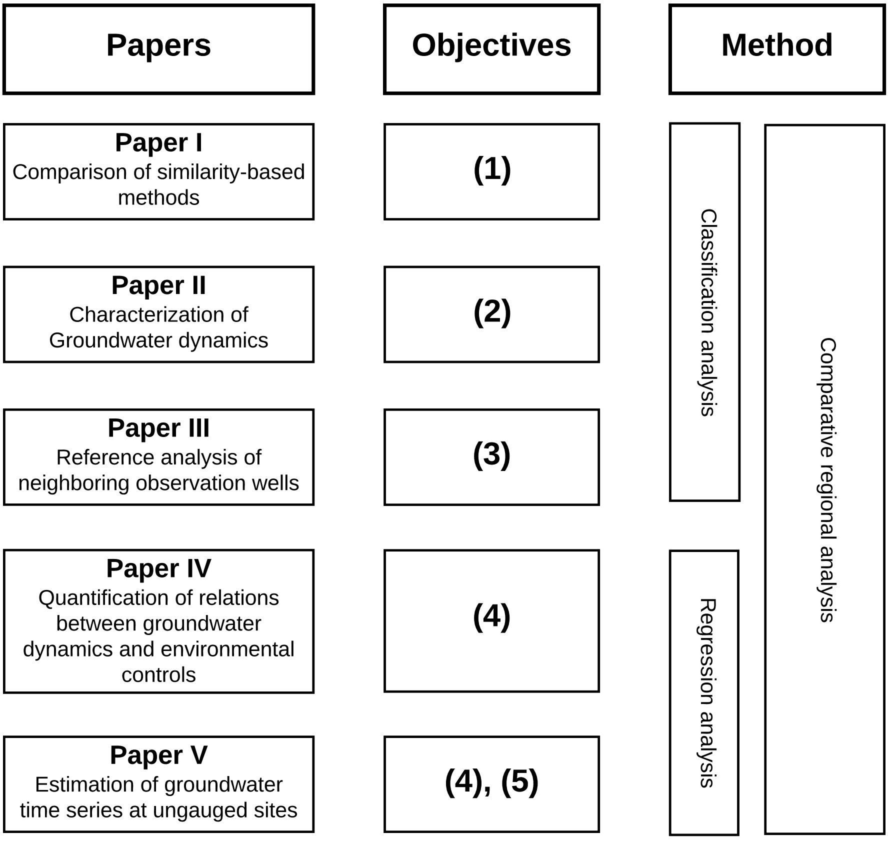

Figure 2. Relation between papers, method and objectives.

The thesis is structured in the following manner: After the introduction, a theoretical background of the approach taken in this thesis is given in chapter 2. Data and study domain are described in chapter 3. A review of the methods is given in chapter 4 and a summary of the appended papers can be found in chapter 5. The results of the thesis are summarized in chapter 6, which also includes a synthesizing discussion as well as suggestions for future work. The conclusions of this work are presented in chapter 7.

2 Prediction based on comparative regional analysis

The central idea that is explored in this thesis is the transfer of observations and understanding from gauged to ungauged groundwater systems using comparative regional analysis. Figure 3 demonstrates how three sites with groundwater level observations (gauged) at location close to rivers in alluvial aquifers could be used to predict groundwater observations at an ungauged site, As mentioned in chapter 1.1, two different modelling approaches are generally used for characterization of groundwater resources on the regional scale, process-based and data-driven modelling (Anderson *et al.*, 2015). The rationale of the process-based modelling approach is that hydrological behavior can be described by physical laws with exact mathematical representations. Data-driven modelling on the other hand means measuring of observable variables and building empirical, explicit relations between these and an unknown variable. This is the approach taken in this thesis, which means that empirical relationships are found between physical parameters involved in hydrologic events and physical variables. The relationships are identified at the regional scale and then generalized to ungauged sites.

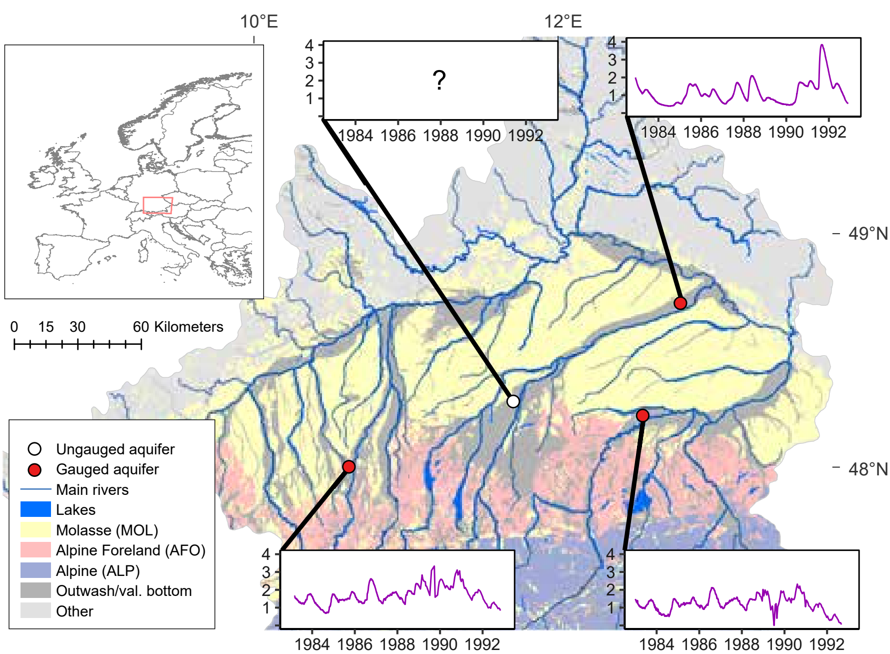

Figure 3. Comparative regional analysis for prediction in an ungauged aquifer.

For transfer of information from gauged to ungauged systems, similarity-based regionalization has been the dominant data-driven model within the surface water community. Regionalization relates to all methods that allow extension of records in space via transfer of hydrological information from gauged to ungauged locations (Riggs, 1973; Oudin *et al.*, 2010). The advances made in that field serve as inspiration for translating similarity-based methods into the groundwater field. Therefore, methods for regionalization were deemed a suitable starting point for this thesis. The idea of regionalization in hydrology is closely linked to the concept of comparative hydrology and the similarity of processes (Falkenmark *et al.*, 1989). As similarity of hydrological processes is difficult to observe and to describe in nature, regionalization uses the relationship of climate, hydrological response, and system properties from proxy systems to base extrapolation on. The principle hypothesis of how similarity in processes relate to physical system properties (which are the basis of classifications) is shown in Figure 4, here translated to an example in groundwater. In Figure 4, two groundwater systems Figure 4a) and Figure 4c) with differing system properties respond to a given, similar (climatic) forcing signal in dissimilar ways. In return, two similar systems respond to a similar system in similar ways (Figure 4a) and b)). This means in consequence, that system information is inherent in the groundwater response. This is the baseline hypothesis of similarity-based classification.

Regionalization within hydrology has two different main flavors: classification or distance-based, and regression-based regional analysis (He *et al.*, 2011). The basic idea of these two classes of methods are contrasted in Figure 5 (not real data). Classification is the practice of finding hydrologically homogenous regions (of contiguous or discontiguous nature) to transfer information based on geographical distance and/or hydrological similarity. The example of classification in Figure 5 (left) shows groundwater observation wells as points, displayed in terms of their inter-annual variability of groundwater levels and their distance to stream. The hydrologically homogenous regions here are displayed by the color of the points based on class membership: discharge-dominated (red) and recharge-dominated (blue). Within classification, two main approaches can be identified according to Olden *et al.* (2012), deductive and inductive reasoning. These opposing approaches can be summarized as the following: 1) Deductive-based classification, based on similarity of relevant structures of groundwater systems and forcing (e.g. precipitation, hydrogeological properties) and 2) inductive-based classification, based on similarity of time series of groundwater observations The disadvantage with the deductive approach (1) within groundwater is that relevant physical characteristics of groundwater systems that control groundwater dynamics are not easily derivable, often unknown or hidden in qualitative descriptions.

The inductive approach, (2), on the other hand, groups systems directly by the characteristic of interest, i.e., the hydrogeologic behavior and can be used to find system controls. In consequence, only inductive, response-based approaches were chosen in this thesis.

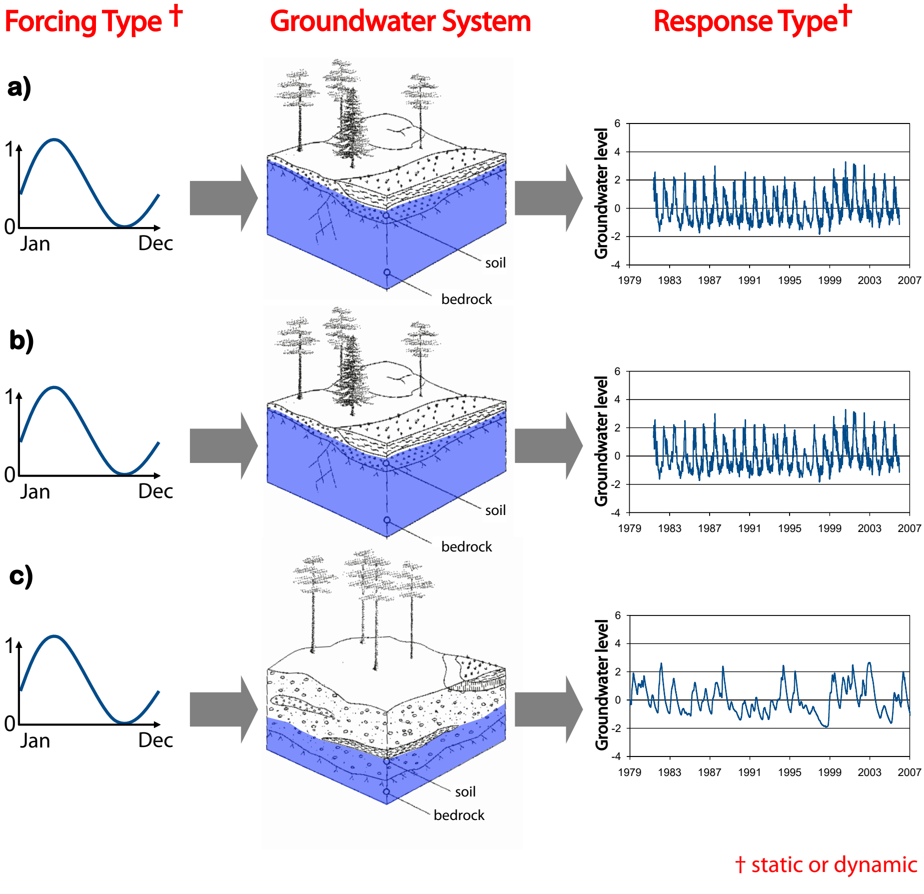

Figure 4. Principle of hypothesis: Similar inputs to different systems lead to different responses. While similar inputs to similar systems lead to similar responses.

While classification returns a discrete number of homogenous regions or classes, regression-based regional analysis yields continuous output. The example of classification in Figure 5 (left) can also be expressed as a regression problem, where inter-annual variability increases linearly with the groundwater wells' distance to stream (Figure 5, right). Regression can be used directly for prediction of hydrological response by descriptors of the system's structure and forcing. Further, regression can be used to understand the importance of individual system characteristics on groundwater dynamics, while classification yields a joint set of important characteristics for each class with similar groundwater dynamics. In consequence, similarity is implicit in regression-based methods by means of the continuous distribution of the output.

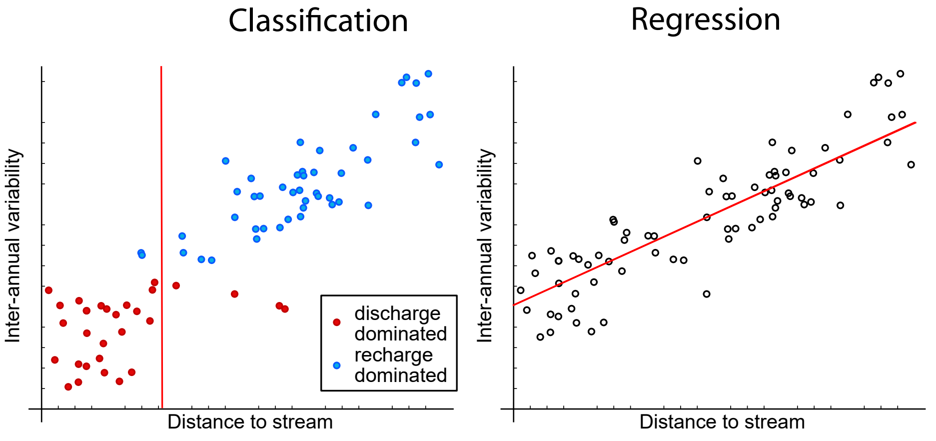

Figure 5: Example of classification and regression. Classification separates observation wells (dots) into two discrete groups. Blue dots to the left and red dots to the right of the line are misclassified. Regression describes a continuous relationship between two variables.

Both approaches, classification and regression, necessitate finding of measures that define similarity of groundwater response. In this thesis, similarity of groundwater response is measured on the sequential record of groundwater level observations. The hydrogeological response as exhibited in groundwater levels is the result of a complex spatial and temporal superposition of a multitude of static and dynamic factors. Relevant static factors are related to the structure of the system (e.g. hydrogeological units like aquifers, geomorphologic and pedologic properties). Dynamic factors are related to the forcing (e.g. hydro-climate, inter-aquifer exchange, regional-recharge). These structural system characteristics and their interactions need to be identified and quantified so that they can be linked to the hydrogeological response of the system. Differentiating static and dynamic in this context is of course dependent on the time scale of interest. At larger time scales, static factors can become dynamic as they co-evolve with climate. On the other hand, dynamic factors can also be averaged over a defined time scale and become static, e.g. long-term hydro-climatic characteristics, such as the average annual precipitation or evapotranspiration. This type of transferring of dynamic to static is necessary to turn a factor into an observable/derivable parameter, which can be used as a measure to define similarity. Transferring as well as the choice of relevant measure is aided by studies within the surface water community, but also from hydrogeological theory and hypothesis testing.

In the previous paragraphs, groundwater system characteristics and measures are discussed that can be used to transfer information from one groundwater system to another. However, what defines a groundwater system? While streamflow is usually the integrative output of a spatially well-definable river basin, groundwater time series are merely records of point measurements in a three dimensional aquifer matrix (compare discussion in Barthel, 2014). Defining the representativeness of these measurements in space and time and, therefore, boundaries and characteristics of groundwater systems is a difficult task. While Von Asmuth (2012) defines a groundwater system as its impulse response, here groundwater systems are defined heuristically as the sum of its system characteristics that control the groundwater response in a measurable way (i.e. in the groundwater level hydrograph).

3 Regional geology, climate and hydrology

The comparative regional analysis was carried out in a large study domain covering southern Germany and stretching from eastern France to northwestern Austria (Figure 6a). This area was initially chosen due to the availability of a large number of groundwater level time series with high measurement density (more in section 4.1). The time series were recorded in a number of observation wells, which are unevenly distributed throughout the study domain. The studies use observation wells from the northern alpine and peri-alpine regions, while the majority of wells is primarily located in the Molasse Basin (MB) in Southern Germany and secondly in the Upper Rhine Graben (URG) in northeastern France and southwestern Germany. The MB is the northern foreland basin of the Alps, containing up to 5,000 meter thick, partly coarse-grained Cenozoic tertiary and quaternary sediments. The URG is a Cenozoic rift structure filled with Cenozoic sediments up to 3,500 meter thick. The majority of observations are located in the hydrogeologically more relevant upper quaternary, shallow sediments for both the MB and the URG; observation wells in the deeper tertiary sediments are few. Gauges tend to cluster along the major river valleys and the fluvial gravel plains. Some remaining observations are drilled in Mesozoic sedimentary rock formations covering the region between the URG and the MB. Additional gauges can be found in: the quaternary sediments of the Inn river valley in Austria; a limited range of unconsolidated fluvial aquifers within the northern Alps; as well as on the northern fringe of a block of Paleozoic sedimentary, igneous and metamorphic rocks known as the Bavarian Forest.

Precipitation shows large contrasts over the study domain, with annual values ranging from 350 mm year^{-1} in Northern Bavaria to more than 2,000 mm year^{-1} in the low mountain areas (e.g. Black Forest) and just under 3,000 mm in the Austrian Alps (Barry, 2008; Thierion *et al.*, 2012), see (Figure 6b). During the summer season peak precipitation in the Alpine region can reach values of up to 1,400 mm, whereas the northern part of the study domain receives a maximum of 800 mm (northern meteorological seasons). Precipitation during the winter season is substantially lower, even in the Alpine region rarely exceeding 800 mm. Snowfall above 200 mm occurs in the Alpine mountain regions, the low mountain ranges and in the western part of the MB (Mauser and Reiter, 2016). Consequently, snow accumulation and melting are important processes for the dynamics of streamflow in the rivers of the study area. This can lead to extreme discharges from alpine snowmelts through tributaries into the two main river systems of the study area, Danube and

Rhine (e.g. Thierion *et al.*, 2012). Additionally, inflows from low mountain ranges and lowlands contribute to elevated discharges during long-lasting rainfalls (Willems *et al.*, 2016). Air temperature follows largely the topography (Figure 6c). The annual radiation balance is 0 – 300 kWh year^{-1} m^{-2} in the Alps to 700 kWh year^{-1} m^{-2} in the plains and river valleys. Translated to potential evapotranspiration this amounts to around 600 mm year^{-1} in the Alps and 800 mm year^{-1} in the plain. In consequence, water transported from mountainous catchments recharging into the basin aquifers play a crucial role in the study domain (Thierion *et al.*, 2012; Mauser and Reiter, 2016). Finally, changes in climate can be seen in large parts of the study domain with regard to temperature, which has increased on average 1.5°C between 1960 and 2006. Precipitation patterns have changed as well, showing slightly lower summer precipitations in the basins within this period, yet with low significance (Reiter *et al.*, 2011).

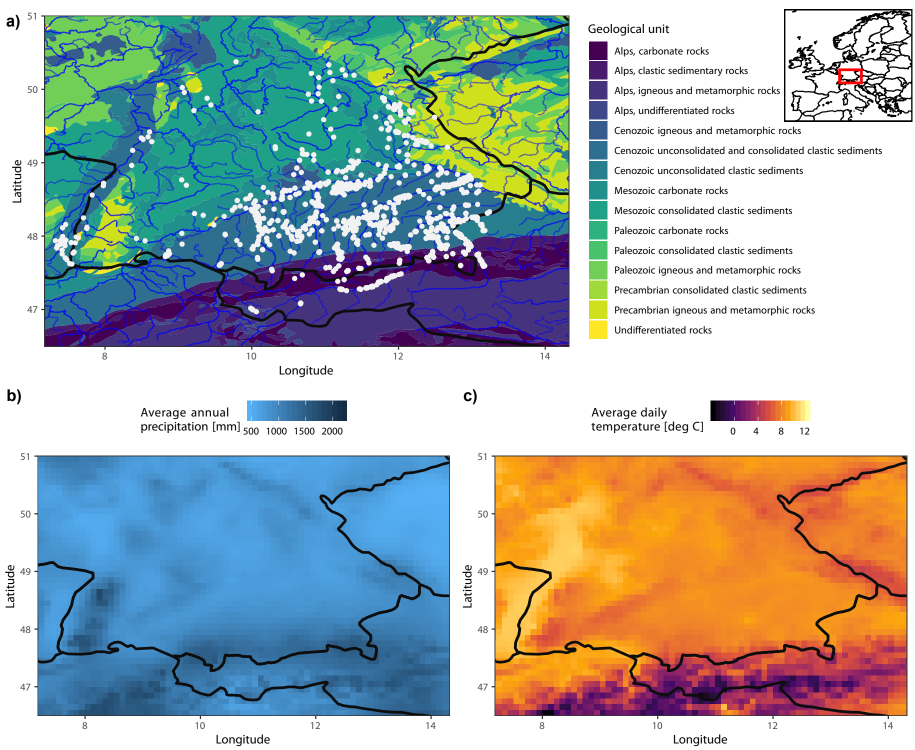

Figure 6. a) Study area with simplified geologic units based on the International Geological Map of Europe IGME5000 (Asch, 2007). White dots represent groundwater wells. b) Average annual precipitation. c) Average daily air temperature.

4 Data and methods

This chapter summarizes the data and methods that were used to develop a methodology for predictions in ungauged aquifers. Figure 7 describes the sequence of methods used, including the number of section in this chapter. The first step of each study is presented in section 4.1 explaining the selection of the groundwater level data, which is a function of quality of time series and site information. Section 4.2 describes the derivation of system characteristics, which can be used to link structure and hydrological behavior, while section 4.3 and 4.5 describe classification and regression methods to quantitatively link the two. In section 4.4 a system for characterizing and quantifying information from groundwater level time series is described, which can be used for improved classification and regression. Finally, section 4.6 describes the principle of regionalizing groundwater level time series.

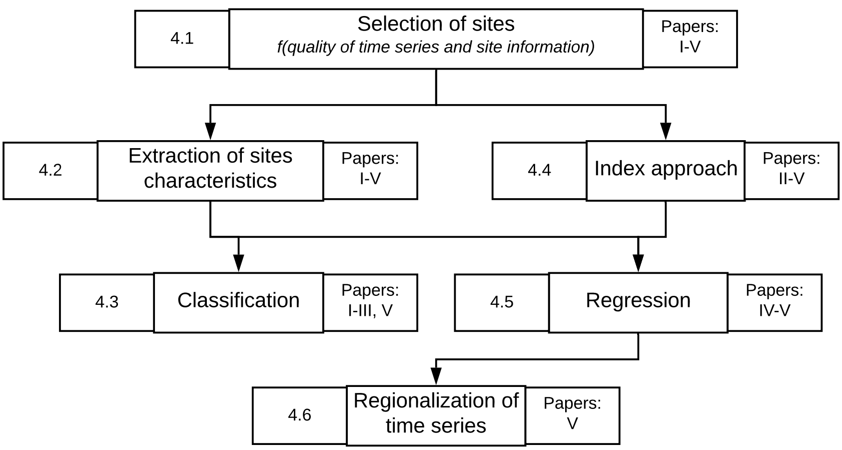

Figure 7. Methods used in each paper including sequence of application. Numbers refer to section number in chapter.

4.1 Selection and processing of groundwater level time series

For the task at hand, a data set was required, which had high measurement density in space and time, while covering an area which can be defined as regional scale with geologically diverse structures. An appropriate data set was identified within the GLOWA-Danube project, documented by Barthel (2011). An original data set was available, which was updated by assembling data retrieved from different national and federal state agencies in Austria Federal Ministry of Agriculture, Regions and Tourism (BMNT) (BMNT, 2015), in the German states Bavaria (LfU) and Baden-Württemberg (LUBW) (LfU, 2017; LUBW, 2017), as well as in France (BRGM, 2017). The data set was assembled for each study according to its requirements. Updates were made at various stages, through acquiring data via the agencies' respective online services or through direct delivery under license agreements. Combining all sources, a data set of around 5,000 groundwater level time series with varying amount of additional site-specific information was available. However, as seen in Table 1, each study used only a fraction of this data. This is due to data selection and preprocessing and explained in the following paragraph.

In Paper I, a selection was made in order to maximize the number of observation wells with weekly, concurrent time series of at least 10 years length. Paper II used an updated data set which included time series from the Rhine valley in Germany and France. Here, the selection was based on maximization of the number of observation wells with daily time series of at least 10 years length. However, time series weren't required to be concurrent. Three hydrologically similar sites were chosen from the data set in Paper III again with non-concurrent, daily time series of 10-year length. Paper IV used a subset of the time series data set from Paper II, for which a more detailed geological description was available. Finally, in Paper V, a subset from Paper IV was used that included the maximum number of concurrent time series from that data set.

Groundwater level dynamics and their records can be influenced by human activities such as groundwater extraction, irrigation, and operation of hydraulic structures. Although groundwater is the most important source of drinking water supply within the study area, only about 1 – 3 % of effective precipitation is used for human consumption (Nickel *et al.*, 2005). Groundwater extraction therefore generally shows only local impacts and irrigation does not take place on a large scale. However, to be sure, the time series of all data sets were visually screened to remove those with strong indications of anthropogenic influence (refer to Paper I and Paper II for more details). After visual selection, criteria for number of consecutive missing values and percentage of total missing values was set in each study (see papers for details). For daily time series, a series of missing values shouldn't be longer than 3 – 6 days. For weekly and monthly time series, only up to two consecutive missing values were allowed. Furthermore, daily groundwater hydrographs were smoothed with a low-pass frequency filter to reduce noise associated with pressure transducers.

When working with groundwater level time series different types of scaling are used for different purposes. Here, we use centering (C), standardization (S) and normalization (N). While centering is merely the subtraction of the mean value from the data set (leading to a time series mean of 0), standardization rescales the centered data by dividing by the standard deviation (leading to a time series with unit variance). Normalization rescales the values into a range from zero to one. This is useful where all values require the same non-negative scale.

Table 1. Characterization of data used for individual studies appended to thesis. *Frequency: M (monthly), W (weekly), D (daily). Concurrency: C (concurrent), NC (non-concurrent). Countries: AT (Austria), F (France), G (Germany). **Scaling: C (centered), S (standardized), N (normalized).

Dataset	Time Series [n]	Freque ncy*	Length [years]	Concur rency	Countries	Scaling**
Paper I	512	M,W	10	C	G, AT	C/S
Paper II	751	D,W	10,21	NC	G, AT, F	N
Paper III	50	D	10	C	G	C/S
Paper IV	341	D,W	10	NC	G	C/S, N
Paper V	198	D	10	C	G	C/S, N

4.2 Physiographic and climatic descriptors

As discussed in section 2, structural system characteristics and forcing have to be derived so that they can be effectively linked to hydrogeological system response. The physiographic information used in this thesis can be distinguished into local and regional geology, morphological and boundary descriptors. In the five studies, different types and sets of physiographic characteristics were used to describe the system structure at each well location.

Two sets of geological information were used, local and regional. Local geological characteristics, such as thickness of saturated and unsaturated zone, aquifer pressure state and aquifer material were defined based on analysed stratigraphy given in borehole logs and used in studies III-V (for a complete list, see 0). In paper I, a combined approach using generalized regional hydrogeological classes (Barthel *et al.*, 2016) was used together with borehole data.

The morphologic characteristics Slope, Topographic Wetness Index, Curvature Classification, Convergence index and Catchment area divided by flow accumulation were calculated from the 3 arc second USGS HydroSheds digital elevation model (DEM) (Lehner *et al.*, 2008) using SAGA-GIS (Conrad *et al.*, 2015). Subsequently, the mean and skewness of these descriptors were estimated for DEM cells within a 300 m and 3 km buffer surrounding the wells (local and regional radius). These buffer areas are visualized in Figure 8. The two averaged areas are interpreted to represent the surface conditions for recharge and groundwater flow at the local and regional-scale. In Paper V, the 3 km buffer was also used to estimate average percentage of land cover radius of each site, derived from the CORINE land cover data set (Bossard *et al.*, 2000). Figure 8 also shows hydrologic boundary descriptors that were introduced based on topographically indicated locations of flow divides and drainage boundaries. Estimating these boundaries was necessary as reliable information on aquifer extents, hydraulic gradients and boundary conditions are rarely available. These measures were also derived from the HydroSheds DEM. A more detailed description can be found in Paper IV.

Finally, climatic measures were derived for each well from the 0.25º E-OBS grid (Version 17.0) covering the period 1950-01-01 to 2017-12-31 (Haylock et al., 2008). Assigning the gridded time series of precipitation, temperature and radiation to each well location was done by finding the grid cell that covered the area of the specific well location. Then, the reference period to calculate climatic descriptors was selected for each well, based on the selected measurement period of the well. Within this reference period twelve long term average climatic descriptors were calculated from the gridded time series of precipitation, temperature and radiation as described in 0.

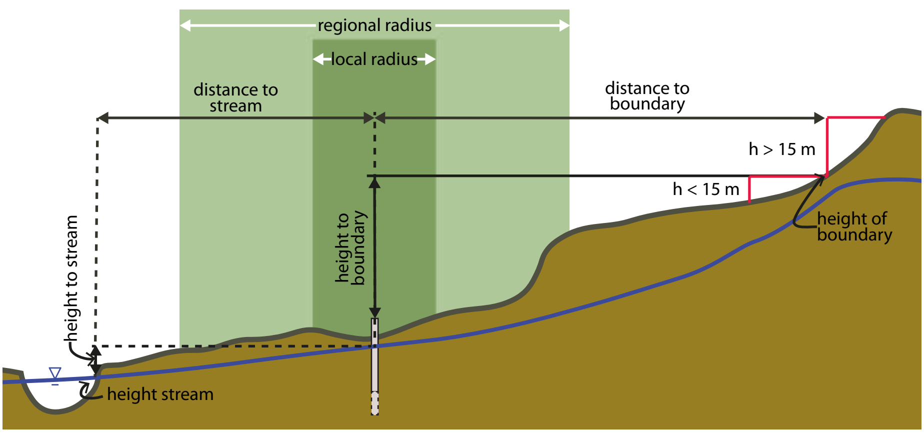

Figure 8. Principle of estimating topography derived hydrologic boundary and morphology descriptors for each well location. Distance and height to as well as height of boundary are based on a recharge boundary, wheras distance height of and height to stream are based on the nearest stream segment. Shaded areas show the local and regional radius for morphology descriptors.

4.3 Classification-based regional analysis

Hydrological classification as described in section 2 requires three steps: definition of groundwater hydrograph similarity using distance measures (4.3.1), grouping of a groundwater hydrographs based on these distance measures (4.3.2) and assignment of hydrologically meaningful labels to clusters (4.3.3). Finally, methods for comparison and evaluation of classification methods are described in section 4.3.4.

4.3.1 Measuring similarity from groundwater hydrographs

The central idea in this thesis is transferring observations and understanding from a gauged to an ungauged groundwater system using similarity-based methods. All methods for similarity search used in this study are based on entire groundwater level time series. Additionally, aspects that do not emerge in the time domain were mined through transformation or extraction of features from time series. These three different paths resulting in six different aspects (domains) from which groundwater hydrographs are analyzed. These are shown in Figure 9, juxtaposed with visual classification. Domains refer to the space in which groundwater hydrographs are analyzed (explained in the next section).

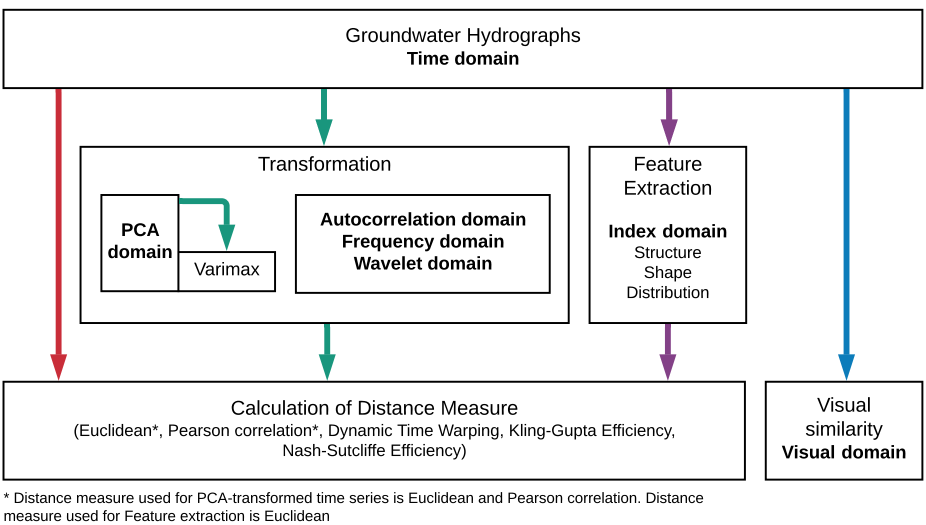

Figure 9. Pathways of similarity-based grouping of groundwater hydrographs carried out within this study.

The method following the red path, e.g. used in papers I, III and V, compares concurrent pairwise groundwater hydrographs in the data set in the original form, or time domain (1), by distance measures (more on distance measures below). Figure 10a) illustrates three groups of groundwater level hydrographs that are similar in the time domain. The green path includes a prior step, where the hydrograph is removed from the time domain and transformed. Transformation of time series is carried out using four different methods: (2) Principal Component Analysis (PCA)/PCA in combination with Varimax rotation, (3) auto-correlation function (Figure 10b), (4) Fourier-transform of the auto-correlation function (frequency domain) (Figure 10c) and (5) wavelet transform. PCA describes the major part of the total variation in a data set of correlated variables, by a set of uncorrelated principal components (PCs). This is done to reduce noise and redundancy in the time domain, effectively transforming the variables into a lower dimensional space. Autocorrelation is the Pearson correlation of the time series at different times as a function of the time lag. Transformation into autocorrelation space is carried out to find similarity between observations within a hydrograph, reduce noise and mine a potential periodic signal. By transforming the hydrograph into the frequency domain, the energy of the signal with respect to different frequencies can be compared, identifying strength of different periodic signals in the hydrograph. Wavelet transform is carried out using orthogonal wavelets, comparing wavelet coefficients that describe a function of frequency and timing of the time series.

The third path (purple) in Figure 9, bases similarity on hydrogeological indices. These indices are metrics that quantify aspects of groundwater dynamics. An example of such metrics are the mean annual maxima and inter-annual fluctuations, which will vary depending on the fluctuation pattern of a hydrograph, see Figure 10d). Indices take a prominent role in papers II, III and IV and are therefore described in a separate section 4.4. Finally, to quantify similarity in hydrogeological behavior groundwater hydrographs in the original time domain or transformed into one of the other domains are compared by their pairwise distance. In many applications the most common distance measures are the Pearson's distance (d_P) and the Euclidean distance (d_E), given two hydrographs with n sequential values i, $U = \{u_i, \dots, u_n\}$ and $V = \{v_i, \dots, v_n\}$ in the same domain they take the form in the discrete case of measurements as follows.

$$d_P(U,V) = 1 - \rho_{U,V}, \text{ where} \tag{1}$$

$\rho_{U,V}$ is the Pearson correlation coefficient

$$d_E(U,V) = \sqrt{\sum_{i=1}^{n}(u_i - v_i)^2} \tag{2}$$

For d_E to work adequately in the time domain, pairs of sequences must be well-aligned in terms of timing, which makes it not optimal for hydrological time series because of the sensitivity to small offsets in timing. Therefore, alternative distance measures were investigated in the study that take these offsets into account. More details can be found in the appendix of Paper I.

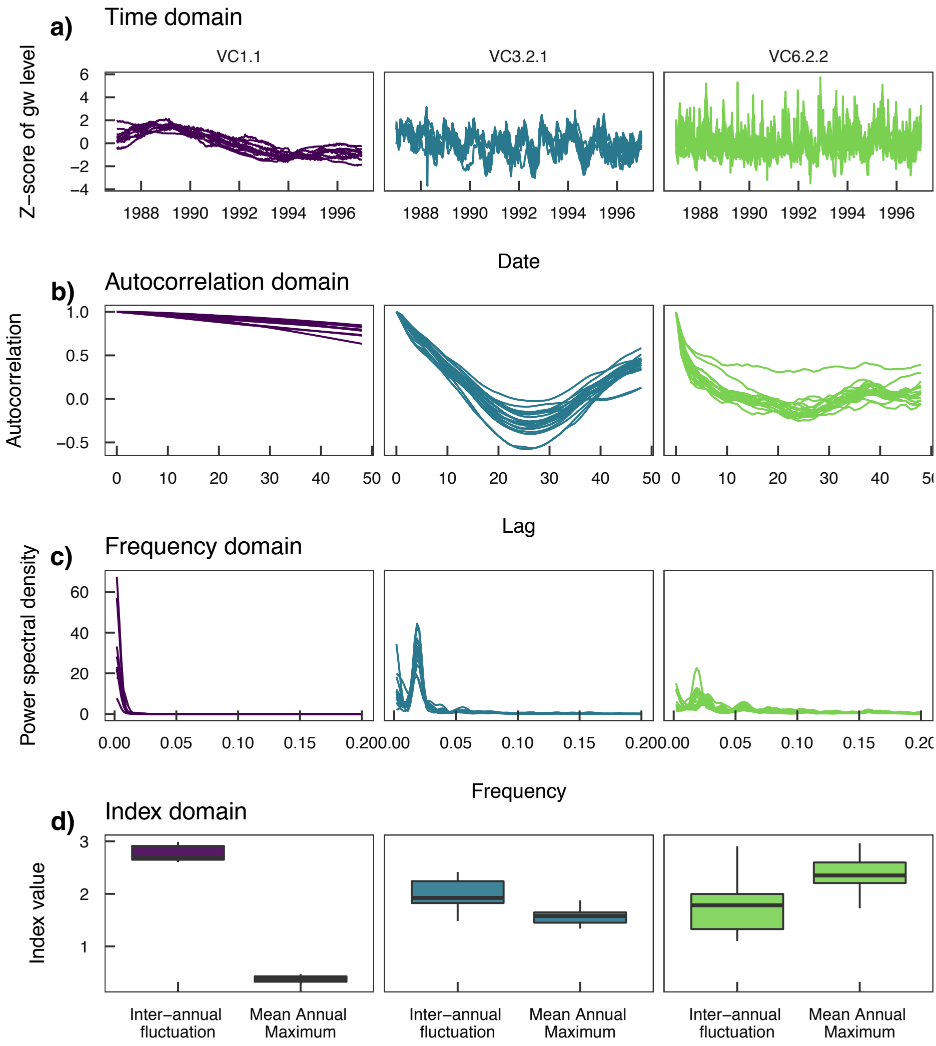

Figure 10. Illustration of transformation of time series from a) time domain to b) autocorrelation, c) frequency and d) index domain. The boxplots of indices within the index domain represent the distribution of index values, i.e. aggregated statistical measures of groundwater level dynamics.

The blue path shown in Figure 9 is visual classification, which was used as a baseline and evaluation method in paper I. To best knowledge, there is no prior study on the relationship between visual perception and numerical measures of similarity of groundwater level time series. Therefore, no standardized technique for visual perception-based similarity is available, so an iterative heuristic approach was used. Grouping of similar hydrographs was achieved by displaying the plots as icons on a computer screen and simply sorting such with similar patterns into new sub-folders. This was done in several iterations, i.e. by refining and revising the groups several times until further reorganization was not perceived as an improvement. The visual grouping was aided by plots on different scales and ensemble empirical mode decomposition, which is described in more detail in Paper I.

4.3.2 Cluster analysis

For grouping a set of objects into similar groups (clusters), a large number of methods are available in statistical data analysis. This group of methods is generally referred to as cluster analysis. The grouping is done in such a manner that each cluster contains objects with the highest joint similarity compared to other clusters. Here, *objects* refers to groundwater hydrographs. In Paper I, two different methods for finding clusters were employed, portioning-based (k-means) and agglomerative hierarchical cluster analysis (HCA) (e.g. Kaufman and Rousseeuw, 2005). In HCA, each object starts out as a cluster of its own and is then merged into larger clusters until the desired number is obtained. In both methods, clusters are split or merged according to the distance between objects (see section 4.3), as well as based on what provides the largest between-group dissimilarity. Three different algorithms for merging objects into clusters were used in Paper I: (1) complete-linkage uses the pair of observations with maximum distance between them as the cluster-cluster distance, also called furthest-neighbor technique; (2) average linkage calculates the averages of the distances at each step of the joined cluster and the remaining objects; and, (3) Ward’s linkage selects clusters by minimizing the error sum of squares between objects in a cluster at each step. Ward's linkage was also used in study III and V. One of the advantages of HCA compared with k-means is that the final number of clusters does not have to be predefined, but can be derived from the clustering result. K-means cluster analysis is a partitioning method that minimizes the sum of squared errors within each cluster and is therefore implicitly based on pairwise Euclidean distance.

4.3.3 Labelling of clusters

The step of labelling each cluster was done in a mixed qualitative and quantitative approach. In Paper I dominant land cover and hydrogeological units combined with depth to water table and thickness of aquifer among cluster members were used to label each cluster. In Paper III, clusters were labelled as belonging to a certain groundwater flow system, which was decided based on joint cluster membership with 15 reference members. These reference members' hydrogeological conditions were characterized prior to cluster analysis including local geology, distance to stream, hydrogeology and response pattern.

4.3.4 Comparison of classification methods

No general strategy is available to measure the performance of classification strategies. Evaluation needs to be approached from different aspects, depending on the objective and nature of the information to be classified (Brun *et al.*, 2007). When no true classification is known (ground truth), the commonly available methods use minimization of numerical criteria. However, numerical criteria to measure homogeneity with regard to similarity of patterns of groundwater fluctuations within a group, has not yet been universally formalized. As a main interest of this thesis lies in measuring the similarity of hydrogeological behavior as seen in groundwater hydrographs, any single optimization criteria is insufficient for the cause. Therefore, in Paper I, an ensemble-based approach was chosen, combining qualitative and quantitate types of evaluation: evaluation of classification methods was carried out based on similarity to visual classification (1), on cluster homogeneity with regard to hydrograph patterns (2.1) as well as numerical criteria (2.2) and (3) homogeneity of physiographic properties. Finally (4), an inter-comparison of cluster methods with regard to cluster composition was performed (not part of Paper I).

(1) Cluster analysis-based classifications were compared with regard to group composition to visual classification using the adjusted Rand index (ARI) (Hubert and Arabie, 1985). The ARI is based on the Rand index (Rand, 1971), which can be used to calculate the overlap or similarity between two partitions. It can be interpreted as a ratio between number of agreements and number of disagreements between the partitions compared. The ARI is the Rand index adjusted for chance grouping of elements with the generalized hypergeometric distribution. ARI yields values between -1 and 1, where 1 means that compared groupings are identical, while 0 equals to complete randomness of class membership and is defined below (Equation 3). Negative values express disagreement below what is expected from a random clustering. Here, n refers to number of elements in a contingency table composed of two clustering results, where i is row and j column index, a refers to row means and b to column means of the contingency table.

$$ARI = \frac{\sum_{ij}\binom{n_{ij}}{2} - \left[\sum_i \binom{a_{ij}}{2} \sum_j \binom{b_{ij}}{2}\right] \div \binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_{ij}}{2} + \sum_j \binom{b_{ij}}{2}\right] - \left[\sum_i \binom{a_{ij}}{2} \sum_j \binom{b_{ij}}{2}\right] \div \binom{n}{2}} \quad (3)$$

(2.1) Cluster homogeneity was first analyzed with regard to groundwater level dynamics using visual perception. The number of clusters is set equal to the classes found in the visual classification. Then, hydrographs and their spectral densities within a class are plotted on top of each other. Using the spectral densities, dominant patterns were identified of inter-annual, annual and intra-annual fluctuation calculating as fraction of clusters that are perceived as homogenous.

(2.2) In hydrology, mathematical criteria to summarize dynamics features of hydrological time series are commonly used. Six such criteria, or indices, were implemented to assess within-group homogeneity. This is done by first calculating each index for each groundwater hydrograph, then calculating mean within cluster standard deviation of each index for the various classification methods. Indices express contingency (Colwell 1974), i.e. dependency of fluctuation on season, flashiness, the frequency of peaks occurring, as described by the Richards-Baker index (Baker *et al.*, 2004) as well as inter-annual fluctuation (Martens *et al.*, 2013). Additionally, the Pardé-coefficient (Pardé, 1933) was calculated as well as the year-of-day maxima and minima (Castellarin *et al.*, 2001).

(3) Homogeneity of physiographic properties is tested by evaluating whether the members of groups of similar hydrographs originate from observation wells with similar hydrogeological characteristics. Again, each mean standard deviation within clusters for each property is calculated for the various classification methods.

(4) In order to compare classifications with regard to group composition, normalized mutual information (NMI) was used (Strehl and Ghosh, 2002). An information theoretic measure can be used to describe the similarity between two groupings Υ, Λ and for comparing groupings independent of set of classes $\Upsilon = (\lambda_1, \dots, \lambda_k)$ and $\Lambda = (\upsilon_1, \dots, \upsilon_j)$ due to normalization. The probabilities of objects in classes $p(\lambda_k)$ and $p(\upsilon_j)$ are estimated, as well as the joint probability distributions of objects in classes of Υ, Λ. Mutual information $I(\Upsilon; \Lambda)$ is determined with:

$$I(\Upsilon;\Lambda) = \sum_{\upsilon \in \Upsilon} \sum_{\lambda \in \Lambda} p(\lambda_k, \upsilon_j) \log \frac{p(\lambda_k, \upsilon_j)}{p(\lambda_k) p(\upsilon_j)} \quad (4)$$

NMI is then calculated by normalizing the entropy H of each grouping:

$$\mathrm{NMI}(\Lambda, \Upsilon) = \frac{\mathrm{I}(\Lambda;\Upsilon)}{\frac{\mathrm{H}(\Lambda)+\mathrm{H}(\Upsilon)}{2}} \quad (5)$$

Normalized mutual information gives values between 0 and 1, where 1 means that ϒ and Λ are identical and 0 randomness of class membership.

4.4 Groundwater dynamics typology and index assignment

A typology of groundwater dynamics was developed in Paper II analogous to the ecological flow regime classification, widely used in hydrology introduced by Poff (1997). The rationale for bringing a typology of dynamics into groundwater was twofold. Firstly, results from Paper I showed the necessity of feature extraction for measurement of similarity of hydrogeological response. Secondly, it is common for groundwater to use descriptive adjectives such as fast, slow, smooth, flashy, seasonal, without an objective definition. In consequence, to be able to describe similarity through features of groundwater dynamics, these had to be selected based on a systematic description of perceived groundwater dynamics. Such perceived features of dynamics may be summarized through descriptive components such as the regularity of annual seasonality, the dominance of inter-annual variability or flashiness of the hydrograph and so on. By combining components that describe independent information within a hydrograph, a typology could be designed, which links perceived groundwater dynamics features to actual quantifiable statistical aggregates (indices) of these features (c.f. Olden and Poff, 2003).

To arrive at a typology that matches many uses, different existing descriptive time series classifications from literature were combined with the domain knowledge of the authors. Subsequently, a large number of indices was derived from literature, adapted to fit the specifics of groundwater hydrographs. Alternatively, indices were designed from scratch to describe specific features, deemed unique to groundwater hydrographs. According to the purpose of the index, it was then assigned to a component of the typology. This means that for an index that e.g. yields a low value for very smooth and a high value for very flashy, time series will be assigned to the typology's component *Flashiness* (fast and slow). An example of this is the Richards-Baker flashiness index (Baker *et al.*, 2004), where l is mean daily groundwater level of n observations, and i is the index of values in the groundwater hydrograph (an example of high and low flashiness can be seen in Figure 15).

$$I_{RB} = \frac{\sum_{i=1}^{n} |l_i - l_{i-1}|}{\sum_{i=1}^{n} l_i} \quad (6)$$

The suitability of assignment was verified with a visual skill test that resulted either in verification, reassignment or dismissal of the index. Finally, the assumption was tested that the typology's components describe independent information. This was done by calculating index values for hydrographs of the data set of Paper II and comparing inter-correlations based on assignment to components. The test was passed when indices assigned to the same component were correlated – and uncorrelated to indices of other components.

In Paper III, indices were further tested on how well they could distinguish groundwater hydrographs by local, intermediate and regional groundwater flow systems. This was done calculating an index skill measure for groups of hydrographs classified into one of the three flow systems. The index skill measure is called the average-normalized pairwise dissimilarity (ANPD). ANPD is the average distance between the index values of a pair of hydrographs standardized by all pairwise distances. If ANPD = 1, the distance is equal to the average distance between index values and serves as a threshold for skillfulness of an index. To separate groundwater wells that are driven by different dominant processes, index values are differentiated using the overlapping coefficient (Gini and Livada, 1943) implemented in the statistical programming language R (Pastore and Calcagnì, 2019). The overlapping coefficient approximates the joint area of two sample distributions, by estimating empirical density and then integrating over the joint density.

4.5 Regression-based regional analysis

Stepwise multiple linear regression analysis was carried out to identify (select) physical and climatic system characteristics that controls groundwater dynamics. Regression analysis was performed on two different response variables, groundwater indices (Paper IV, section 4.5.1) and segments of the duration curve of groundwater levels (Paper V, section 4.5.2). Prior to variable selection, all variables were centred (for interpretability) and scaled due to the descriptors different units.

4.5.1 Linking indices to system characteristics

For describing single relations between groundwater dynamics indices and system characteristics, correlation analysis was carried out. Correlations between dynamics indices and descriptors were estimated using correlation methods for linear (Pearson correlation), monotonous (Spearman rank correlation), and non-linear relationships (distance correlation). In order to structure the resulting correlation matrices, cluster analysis was performed using Ward clustering with Euclidean distance (see sections 4.3.1 and 4.3.2). Correlation matrices and clusters are displayed in heatmaps to identify groups with similar relations. Then, multiple linear regression with a forward selection procedure (see section 4.5.3) was used to find statistical models describing relationships between multiple physiographic and climatic characteristics and groundwater dynamics. The statistical models were based on the entire data set stratified by aquifer pressure states, confined and unconfined.

4.5.2 Linking duration curve segments to system characteristics

In Paper V, daily groundwater level time series are estimated at ungauged sites using regionalization of groundwater level duration curves. Duration curves of groundwater levels are constructed at a gauged site by first ranking all n observed, groundwater levels $l_i, i = 1,2, \dots, n$ in descending order and i is the rank of an observation. The duration curve is then constructed based on the Weibull plotting positions (Sugiyama *et al.*, 2003):

$$q_i = P(L \geq l_i) = \frac{i}{n+1}, \tag{7}$$

where q_i is the quantile of time that a given level l_i is equaled or exceeded. Groundwater level duration curves are the relation between quantile q_i and the corresponding groundwater level l_i. Multiple linear regression with a forward selection procedure (see section 4.5.3) was used to find statistical models describing relationships between physiographic and climatic characteristics and 15 fixed quantile levels (0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99).

4.5.3 Variable selection using selective inference

A regression relationship from n observations with $i = 1,2 \dots n$ (each site contributes one observation point) is built with the response variable y_i and x_i the predictors. In Paper IV the response variable y_i is the set of index values of all sites for an index Y. In Paper V the response variable y_i is the set of groundwater levels l_i at all sites corresponding to a fixed quantile level q_i. Standard linear regression has the form:

$$y_i = \beta_0 + \sum_j x_{ij}\beta_j + \epsilon_i \tag{8}$$

with errors ϵ_i being independent and normally distributed and where β is a vector of unknown coefficients that are estimated. Here, the selective inference method proposed by Taylor and Tibshirani (2015) was employed, since it adjusts p-values for the effect of sequentional selection of variables. This is necessary since in standard stepwise regression, computation of p-values based on the t-test leads to an overestimation of the strength of apparent relations and thus yields non-significant predictors. Forward regression starts from an empty model, iteratively building a regression model from the set of descriptors. Descriptors become predictors x to optimize the fit to the response vector y. New predictors are added at each step from X by selecting the variable that yields the maximum absolute correlation with the residual (after orthogonalization of the variables with respect to the current model). Selection of predictors for each model is stopped when the p-value of the latest added predictor has a 1 in 10 chance of being a false-positive (Taylor and Tibshirani, 2015; G'Sell *et al.*, 2016). Stepwise regression and selective inference was carried out using the R package "selectiveInference" (Tibshirani *et al.*, 2017).

4.5.4 Evaluation of regression models

When empirical relationships can be found between dynamics and climatic and/or physiographic descriptors, the coefficient of determination (R^2) is used to express predictive power. Regression models are furthermore controlled for plausibility with regard to physical process understanding. This is done by describing a conceptual model based on the significant predictors, signs of coefficients and leveraging with knowledge on regional aquifer dynamics compiled in literature. Additionally, in Paper IV relative variable importance is computed based on standardized regression coefficients. This variable importance gives indication of average importance of a descriptor for groundwater dynamics.

4.6 Statistical regionalization of groundwater levels at an ungauged site

In Paper V, to estimate groundwater levels at an ungauged site (compare Figure 3), three steps are required. Reconstruction of duration curves through quantile model estimates, inter and extrapolation to a continous duration curve, and spatial interpolation of the time series. First, a duration curve is reconstructed from predicted exceedence levels at 15 quantile levels (black dots Figure 11a). After this, levels between the 15 fixed quantile levels are inter- and extrapolated with a logarithmic base (solid line Figure 11a).

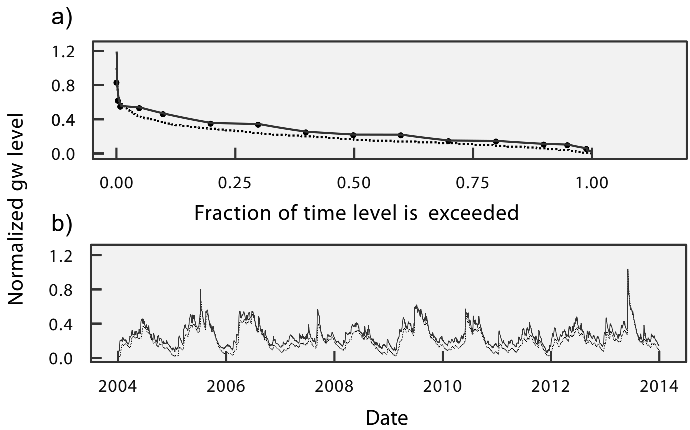

Figure 11. a) Original groundwater level duration curve (dashed line), estimated fixed quantiles from regional regression model groundwater level duration curve (black points) and estimated interpolated groundwater level duration curve (solid line). b) Estimated (solid) and observed (dashed) daily groundwater levels.

Then, using inversed (geographical) distance weighting, time series are reconstructed from neighboring sites at an ungauged location (red line Figure 11b). The estimated time series at the ungauged site L_u is calculated from groundwater level time series at gauged sites L_{uj} (where $j = 1,2, \dots, n$ is the number of gauged sites) and weighted with weights w_j and then normalized by the sum of all weights, as follows:

$$L_u = \sum_{j=1}^{n} w_j L_{uj} \,/\, \sum_{j=1}^{n} w_j \qquad (9)$$

The weights are calculated based on a dissimilarity measure, d_u:

$$w_j = \frac{1/d_u}{\sum_{j=1}^{n} 1/d_u} \qquad (10)$$

Similarity of groundwater level time series is related to spatial proximity (based on findings in Paper III). Therefore, the geographical distance between gauged (X, Y) und the ungauged site (X_u, Y_u) was selected as the similarity measure d_u, as follows:

$$d_u = \sqrt{(X_u - X)^2 + (Y_u - Y)^2} \qquad (11)$$

Evaluation is carried out using leave-one-out cross-validation, such that estimated and observed values can be compared (observed values shown as black line in Figure 11a and b). The methodology of time series reconstruction is an adaption from Shu and Ouarda (2012) and described in detail in Paper V.

5 The papers

This chapter provides a brief overview of the studies appended to the thesis as well as a summary of each paper.

5.1 Similarity-based hydrograph classification (Paper I)

The central aim of the first study was to determine effective and appropriate strategies to classify and group original time series of groundwater levels according to similarity in groundwater fluctuation patterns. A crucial element of the strategy is to determine the optimal approaches to group groundwater hydrographs, which means that prior knowledge of the respective groundwater systems was not included. The decision to exclude information other than the hydrograph values as such in the classification strategies was taken with the aim to achieve independency and consequently transferability of the strategies between different regional settings.

One of the challenges with this undertaking is to define similarity in response. Although a number of similarity measures and grouping algorithms have been used for hydrogeological time series, no single agreed-upon numerical measure of similarity exists. This gap aims to be filled by an inter-comparison of a number of numerical grouping methods, as well as a visual classification (VC) of groundwater hydrographs. The VC is based on visual similarity of rising and recession limbs, shape and magnitude of peaks as well as periodicity, yet does not involve a systematic description of these features. Based on these considerations, 33 candidate grouping methods, including a visually based grouping, are applied to a large regional data set of concurrent groundwater level time series. Subsequently, the homogeneity and separation of the resulting classes of time series are evaluated, both visually, numerically and with regard to regional and local hydrogeological descriptors. The granularity of the candidate methods is explored, i.e. how well the target number of class is identified by internal validation indices.

The study shows that visual similarity cannot be directly reproduced by statistical, similarity-based grouping methods. No universal similarity measures perform substantially better than others across all evaluation methods. Statistical methods for classification based on groundwater level time series is promising, but the transferability is hampered by the need for concurrent time series.

5.2 Index-based classification (Paper II)

The objective of Paper II is to provide a typology of groundwater dynamics. This typology should enable two steps simultaneously: a systematic description of perceived groundwater dynamics features and a quantification of these features with statistical aggregates (indices).

The approach is inspired by a methodology used in surface water hydrology and described by Olden and Poff (2003) and is adapted to groundwater time series as well as enhanced where needed. The typology was devised through the following three steps: (1) a set of different "dynamics-aspects" was formulated (hereafter called dynamics components), which have been conceptualized and synthesized iteratively, based on what the authors perceive as constituting the dynamics in a large and diverse set of groundwater time series. Then (2), within a broad literature review, a large number of quantitative indices were assigned to the dynamics components. The assignment was based on an understanding of which indices were descriptive of individual dynamics components. Finally (3), the indices were calculated on a large data set and subsequently verified if the index is skillful in expressing the dynamics components. This was done using a visual skill test and resulted in either acceptance, reassignment or dismissal of indices. Subsequently, a principal component analysis, PCA, was applied to the indices to identify if the dynamics components and the indices were expressing different information within the time series. Indices were also tested for sensitivity towards different time series lengths and record periods.

The study provides a typology, where statistical aggregates are classified according to what feature of groundwater level dynamics they describe. The indices show promise for linking physiographic system properties to groundwater dynamics.

5.3 Enabling process understanding with indices (Paper III)

In Paper III, an analysis of the capability of indices of groundwater level observations was carried out to improve hydrogeological system understanding. Instead of using a large sample approach to investigate general understanding (as in the remaining papers), a smaller number of wells at three sites were selected in order to ensure a detailed prior understanding of the hydrogeological situation at each well. The study gives a more detailed understanding of what different components of the groundwater dynamics typology express and what they can be used for.

The study includes a regional and a local hydrogeological analysis of observation wells through a joint analysis of individual wells local characteristics and similarity of groundwater level fluctuations. To transfer this characterization to a larger set, an expert-guided cluster analysis was performed, resulting in five distinct clusters. The dominant flow system within the five clusters was then assigned and groundwater dynamics indices (developed in Paper II) were calculated. Similarity of index values was then compared for pairs of groundwater wells with common system characteristics and groundwater flow systems. This was done to test the skill of indices to classify groundwater flow systems in general. Finally, indices were tested in how well they distinguish groundwater hydrographs that show groundwater-surface water interaction, with a strong local signal, i.e. local river stage and such that only show a common pattern in connectivity to surface water.

It can be demonstrated that a large number of indices is skillful for separating time series of local, intermediate and regional flow systems. Some of these skillful indices overlap with those that are powerful in distinguishing hydrographs with local or common groundwater-surface water patterns.

5.4 System controls of groundwater dynamics (Paper IV)

Paper IV presents an empirical study of relations between temporal dynamics of local groundwater levels and climatic as well as physiographic controls. The overarching question is how well groundwater dynamics can be quantified with borehole and map-derivable descriptors.

The analysis is based on observational data from 340 well locations in Southern Germany. The wells are installed in confined and unconfined sand and gravel aquifers from mountain valleys to more extensive lowland alluvial aquifers and groundwater hydrographs exist with ten-year daily measurements. These groundwater hydrographs are decomposed into indices. The following system descriptors are derived at groundwater well locations: geologic descriptors from borehole logs, local and regional morphologic features, topography-based boundary conditions as well as long-term climatic aggregates. Correlation analysis was carried out between the indices and the 54 candidate controls (geologic, morphologic, boundary, and climatic). These correlations allow identifying indices and controls with the strongest relationship. Using only non-collinear and significant relationships, global regression relationships are established by mining the data for associations between dynamics and controls with forward stepwise regression.

The study shows that groundwater dynamics are most strongly linked to geology and hydrogeologic boundary conditions and secondarily to climate, but also to some morphologic features. Simple statistical models are in agreement with general process understanding linked to groundwater dynamics. This study suggests that statistical regionalization of groundwater dynamics in ungauged aquifers based on map-derived physiographic and climatic controls can be feasible.

5.5 Estimation of daily groundwater level time series (Paper V)

In Paper V, the feasibility is studied of predicting daily groundwater level time series at ungauged sites by transferring time series based on system characteristics and proximity from gauged locations. Two methods for transferring information from gauged to ungauged locations are compared: a purely proximity-based method and a combined method based on regionalization of system characteristics and proximity.

Groundwater level time series at 198 sites in Southern Germany with ten-year daily measurements were selected from aquifers situated in diverse landscapes from mountain valleys to more extensive lowlands aquifers. At each site the percent of time that a particular groundwater level is equaled or exceeded (groundwater level duration curve) was calculated. Using regression analysis and logarithmic interpolation, a regional duration curve model was constructed. From this model, a local duration curve (dc) at an ungauged site was estimated based on system characteristics, such as aquifer thickness and distance to nearest stream. Groundwater level time series were then reconstructed from the dc at an ungauged sites based on spatial interpolation from neighboring sites. The estimation method based on a regional duration curve was compared with an estimation method using inverse distance-weighted groundwater levels from neighboring sites at the ungauged site. The optimal number of neighboring sites was evaluated for different groundwater systems. Evaluation is carried out using leave-one-out cross-validation (time series at "ungauged sites" are temporarily left out for estimation and then compared to the estimate for evaluation).

Performance measures show generally high R^2 for estimation at ungauged sites. However, the study indicates that regionalization is only in a third of cases improved by including system characteristics (use of dc method). Further, groundwater dynamics are controlled by different factors depending on aquifer state, which has implications for modelling hydrogeological extremes.

6 Results and discussion

6.1 Comparison of similarity-based classification approaches

In Paper I, hydrograph classification based on entire time series (time domain), spectral densities (frequency domain), autocorrelation, wavelet domain and PCA loadings was carried out on 512 time series and 33 candidate methods – foremost with the aim to detect the optimum combination of distance metrics/clustering approach. As a baseline to contrast different clustering approaches, two visual perception-based approaches were developed: visual classification (VC), a grouping of time series based on including visual perception and systematic visual inspection (VI) of cluster analysis results. It could be shown that VC is useful since it combines the eye's pattern recognition with domain-knowledge, which improves understanding of data and grouping results. However, VC has some deficits as it is subjective, inefficient and hard to reproduce. Also, it is shown that it doesn't guarantee homogeneity of groups with regard to evaluation with mathematical measures. This becomes clear, when comparing similarity of VC with cluster analysis methods. In Figure 12 the group composition from three hierarchical cluster analysis algorithms in combination with different distance measures (see legend), k-means and PCA/Varimax is compared with the visual classification. Similarity (measured as adjusted Rand index, ARI) between VC and clustering approaches generally increases with the number of clusters until it reaches an inflection point at ten to 30 clusters. The highest similarity value between formalized clustering approach and visual classification is around 40% (adjusted Rand index, ARI: 0.4) and many methods are far below this score.

Average homogeneity with regard to hydrological indices was also calculated in Paper I, which showed that the VC in most cases performs about average compared to clustering methods. This can be compared with the result by Crochemore *et al.* (2015), who showed that the perception of similarity between time series of runoff as identified by the human eye is often incongruent with the available numerical measures. From this it may be concluded that the worth of visual classification can only be assessed based on similarity of hydrogeological properties of cluster members. Visual inspection, however, is shown to be a powerful tool to assess the skill of cluster analysis in grouping visually similar hydrographs together. Figure 13 shows the ranking according to VI, where Discrete Wavelet transform (DWT) yields the highest summarized skill, across all tested algorithms with regard to inter-annual,

annual and intra-annual homogeneity. Besides DWT, the similarity measures Euclidean (EUCL) and Nash-Sutcliffe Efficiency (NSE) rank highest, and relatively consistently across all clustering algorithms. In general, methods comparing hydrographs with distance measures in the time domain are more skillful (DWT included, which uses the wavelet or time-frequency domain), while autocorrelation-, frequency-, and PCA-based methods have lower scores. The latter is surprising, since many studies dealing with groundwater hydrograph grouping use PCA in combination with CA (Upton and Jackson, 2011; Triki *et al.*, 2014; Machiwal and Singh, 2015). Here, it was shown that PCA transformation before cluster analysis delivers lower cluster homogeneity than groupings performed on the original time series. The reduction of information resulting from PCA generally leads to few clusters with many members. In consequence, when granularity should be high, such as in this regional study with hundreds of time series and different patterns, PCA-transformation is less useful. However, its use might be preferable for studies on smaller scales or less diversity, with fewer time series and fewer principal components, e.g. Upton and Jackson (2011).

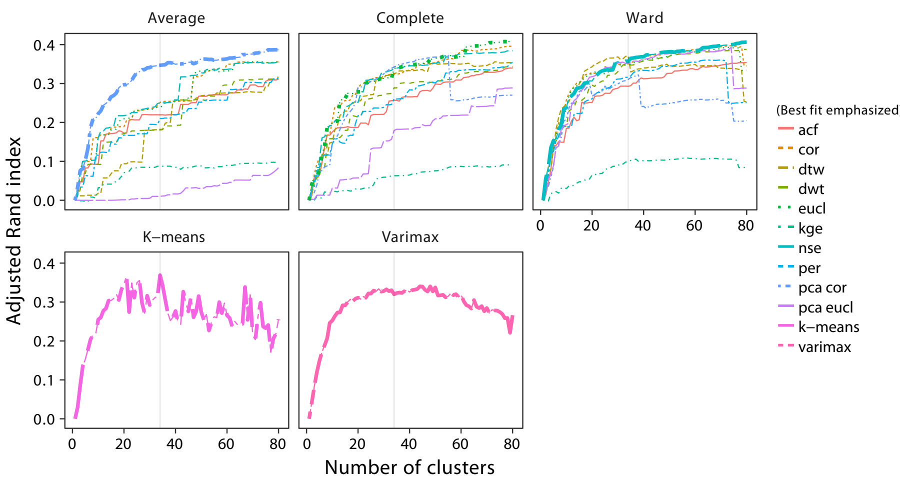

Figure 12. Similarity between visual classification and candidate grouping methods as expressed by the adjusted Rand index. Vertical straight line at 34 clusters represents the number of groups found by visual classification (from Paper I).

Despite large differences in visual homogeneity and low similarity with the visual classification, different clustering results are often similar to each other, when it comes to group composition (i.e. which hydrographs end up in the same cluster). In Figure 14, similarity between classification methods with regard to cluster composition is shown, expressed as normalized mutual information between different classification methods. Mutual information is generally highest when comparing classification methods that are based on the same similarity measure. While correlation (COR), EUCL, and NSE have very high mutual information across different algorithms, cluster compositions based on DWT and dynamic time warping (DTW) are less similar to other classifications. Finally, autocorrelation- (ACF) and periodicity-based (PER) groupings are similar to each other, but together with Kling-Gupta efficiency- (KGE) and PCA-based groupings very distinct from other classifications, having less than 50% mutual information. When comparing the results of the visual inspection among linkage algorithms, the results do not reject the conclusions made by Machiwal and Singh (2015), that linkage has a major effect on the resulting grouping. However, judging from the visual inspection and the grouping composition analysis shown in Figure 14, transformation of time series from time domain into other domains (PCA, frequency, etc) results in the largest difference between groups.

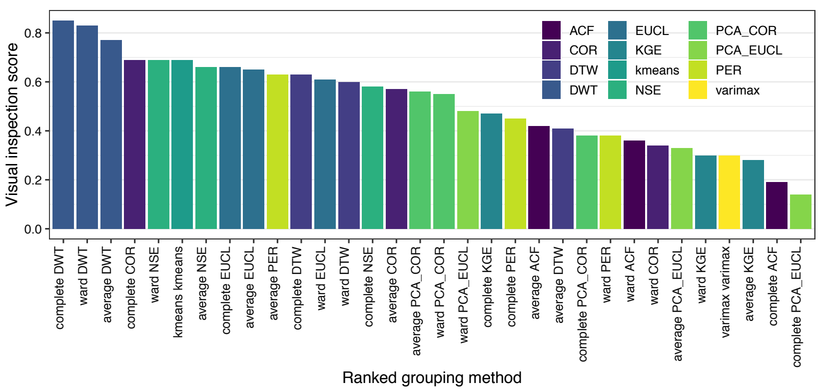

Figure 13. Ranked scores based on visual homogeneity of cluster results with regard to inter-annual/annual/intra-annual periodicity (figure based on results from Figure 10 in Paper I). A score of homogeneity of inter-annual, annual and intra-annual periodicity is given to clusters of each method. The visual inspection score is the sum of the three individual scores.

Finally, interpretation of cluster analysis-based groups of groundwater hydrographs can be difficult to interpret. Often, the question why hydrographs have been clustered into the same or different groups arises. This is especially true, when higher granularity is required. Only relative information (distance between objects) is available and no quantitative summary of the dynamics as the underlying reason for group membership can be made. The reason for this is that in all methods (but wavelet-based distance), single values of each object's data series do not carry much information individually, but give information on similarity when compared to other objects. The semi-quantitative visual inspection using three classes (inter-annual/annual/intra-annual periodicities) has shown to be useful to characterize the dynamics of groups, despite its relatively inefficient implementation.

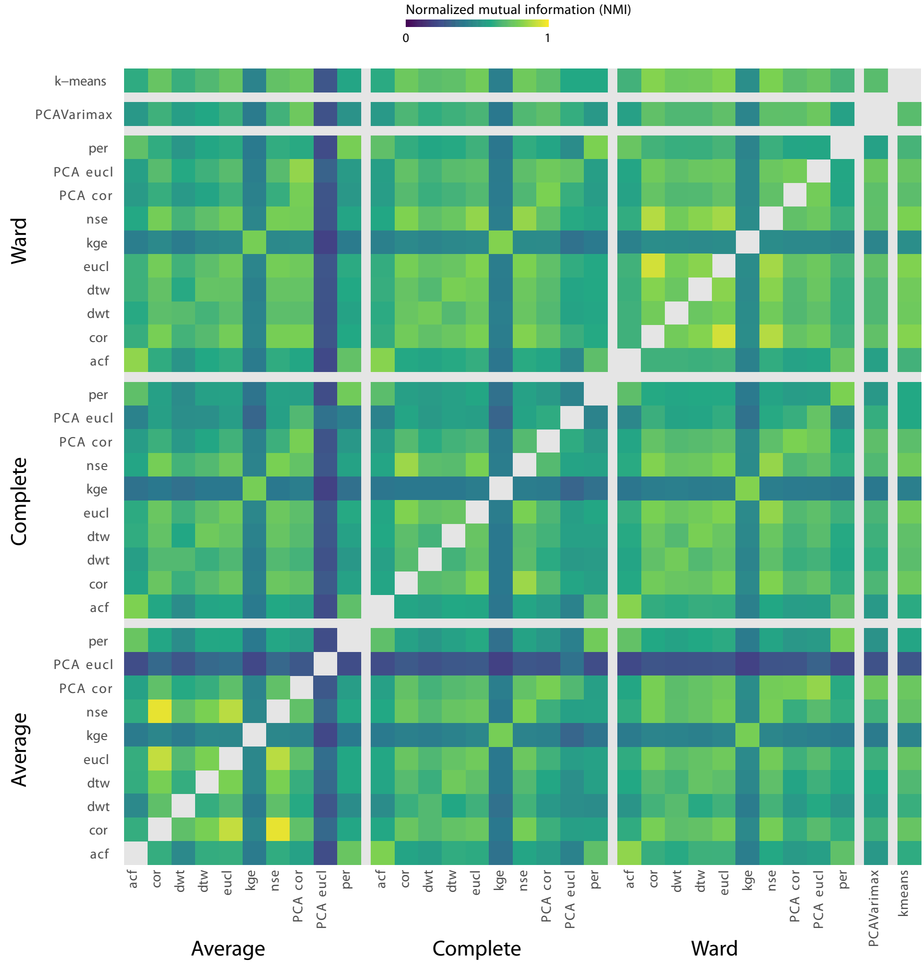

Figure 14. Similarity of hydrograph classification methods expressed as normalized mutual information (NMI) of pairwise group composition. Lighter colors have higher mutual information thus similarity (graph based on results from Paper I).

6.2 Index-based description of groundwater hydrographs

In Paper II, a typology for descriptive characterization of groundwater dynamics was developed. Figure 15 shows the result of the typology with three main aspects: Structure, Shape, and Distribution, which capture characteristics of groundwater dynamics at three different perceptual levels. Structure (purple) describes global features in the time domain, while Distribution (green) describes global features as seen in the form of the empirical distribution of groundwater levels. Shape (orange) characterizes the short-term recharge and recession event- or peak-based features of a time series. These three aspects are separated into ten components that further decompose the aspects. Structure is divided into five components: seasonal magnitude and timing, inter-annual variability, flashiness, and amplitude (not shown in Figure 15). Distribution is decomposed into three components: Boundedness, density, and modality. Shape contains the components scale and slope. Figure 15 also shows illustrative minimum and maximum end members of each component. Using this typology, each groundwater level time series in the data set can be classified in terms of the ten components as described above. Within groundwater hydrology specifically, few attempts have been made to explicitly label aspects of groundwater dynamics in a systematic way, based on perceived patterns in time series. von Asmuth and Knotters (2004), for example, use the terms "fast" and "slow", which here are considered by flashiness. As described in chapter 6.1 within the descriptive visual inspection, hydrographs are classified into three groups of dominant inter-annual, annual, and intra-annual variability based on spectral densities. Within the proposed typology, these classes are captured by four components: magnitude, timing, inter-annual variability, and flashiness in the aspect structure (Figure 15).

The critical point mentioned in chapter 6.1 is quantification of dynamics within a group. Therefore, 63 indices were assigned to the components that are capable of translating the descriptive terms into numbers (see Appendix B). The number of indices varied between studies: all 63 were used in Paper III, 45 and 46 indices in Paper II and Paper IV, respectively. This number of components fitted with indices is large compared to other studies (Schürch *et al.*, 2010; Martens *et al.*, 2013) and express different information as seen in the PCA in Figure 16. The PCA also shows that the indices assigned to the typology's components generally are orthogonal to each other. This means that the components (via indices as proxies) express independent information despite their derivation based in the perceived features of groundwater hydrographs. The typology can therefore be used to objectively describe dynamics of individual or group-wise groundwater level time series. An example of the combination of descriptive and quantitative power of the typology can be seen in Figure 10 (section 4.3.1, p.19), where three distinct clusters of time series, are compared in time, frequency, autocorrelation, and finally index domain. The index domain shows two indices: inter-annual fluctuations and mean annual maximum groundwater level, which have been assigned to the components inter-annual variability and seasonality-magnitude. Now the components can be used to describe that the group of hydrographs in Figure 10a) have a dominant inter-annual variability, but a low regularity of the seasonal (annual) magnitude. The hydrographs in Figure 10b) have a much weaker inter-annual variability, but more regular magnitude, information that can be quantified using the relative differences of index values.

In Paper III, a joint procedure was used for classification. Accordingly, dynamics were used to classify 50 wells into local, intermediate, and regional flow systems according to dynamics and hydrogeological setting (Figure 18). The skill of 63 indices was assessed for each of these classes using average-normalized pairwise dissimilarity (ANPD) and overlapping coefficient. Hereby, it can be concluded that a large fraction of the indices is skillful in separating time series of local, intermediate, and regional flow system individually. Across all aspects and components of the typology, the best performing indices are shown in Figure 17. Combined with the overlapping analysis carried out in Paper III, it was shown that indices assigned to inter-annual variation (autocorrelation, inter-annual fluctuation, etc.) and generally associated with the hydraulic memory of the system, show the best performance at all levels. Bloomfield and Marchant (2013) could also show that autocorrelation (correlated with other indices within inter-annual variation) is related to aquifer diffusivity and, therefore, scales with aquifer transmissivity and storage coefficient. Accordingly, indices describing long-term pattern of dynamics should be considered for every kind of regional scale analysis, supplemented by specific indices describing visible hydrograph dynamics or representing expected driving forces. Further, indices within the slope and boundedness components have high skill. This means that all three aspects, Structure, Shape, and Distribution add value to the analysis of groundwater hydrographs.

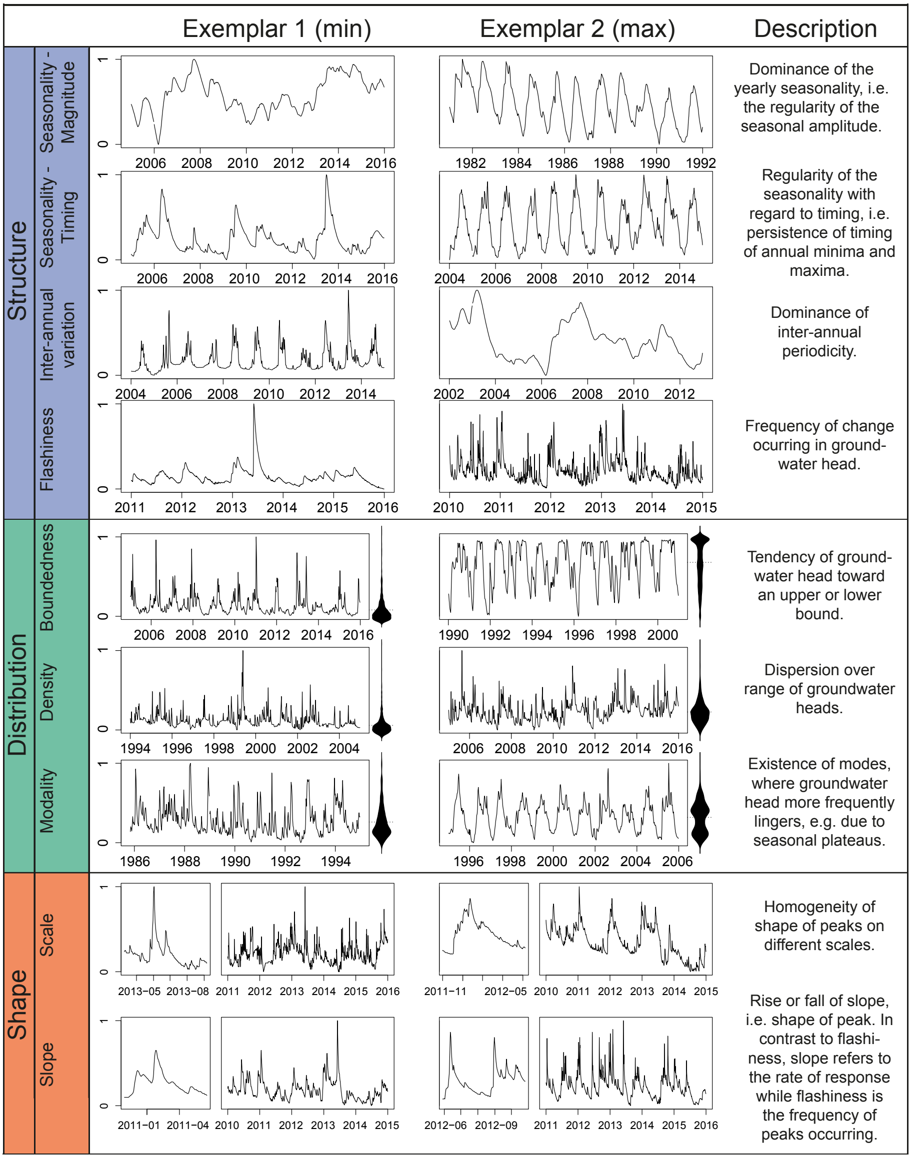

Figure 15. Typology of groundwater dynamics, divided into aspects and components. For each component an explanation is given and end-members are shown: where the respective component is weak and strong, for example, for Flashiness one with low and high flashiness. For the aspect Distribution, violin plots indicate the frequency distribution. For Shape, a recharge-event-based zoom window is added (from Paper II).

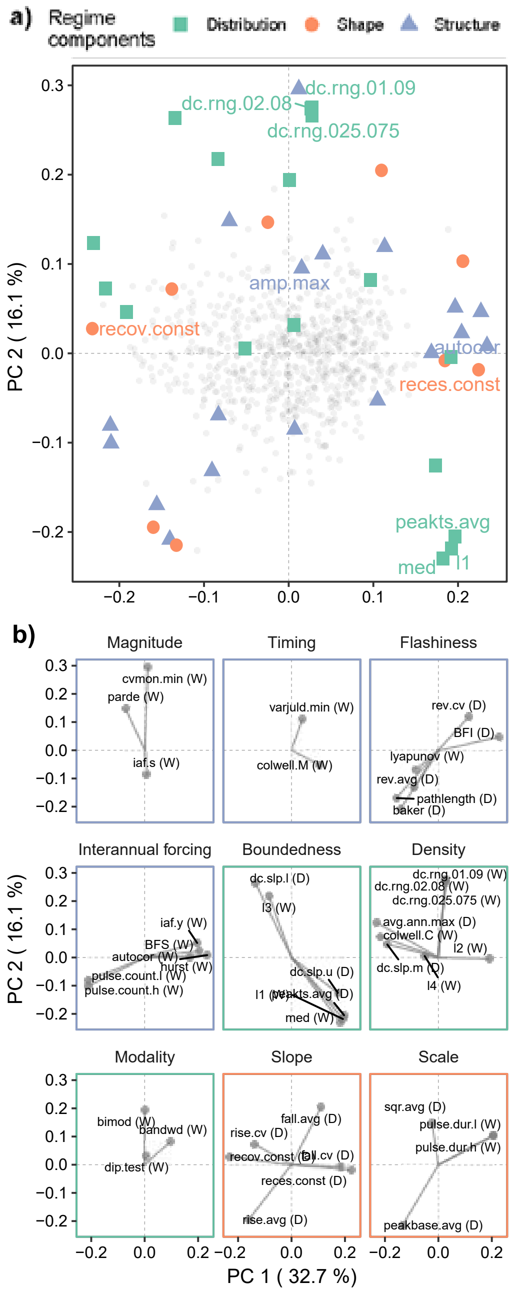

Figure 16. a) Biplot of index loadings (coloured) and scores of the individual groundwater time series (opaque background) on the first two principal components, PC1 and PC2. b) Index loadings on PC1 and PC2, for each component of the groundwater dynamics, showing the similarity of indices (explanation of names in Appendix B) assigned to a component (based on direction of loadings) and difference between components (adapted from Paper II).

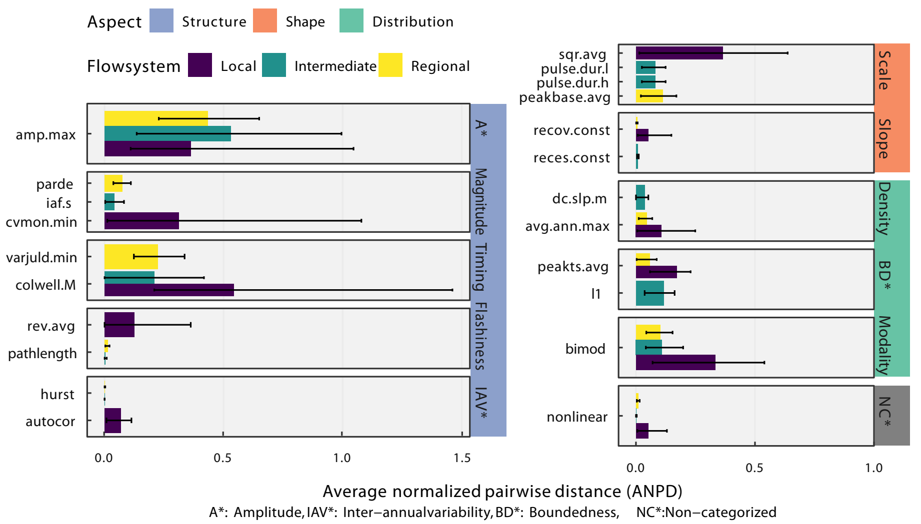

Figure 17. Average normalized pairwise distance of the most skillful indices (names in Appendix B) across dynamics components and flow systems (adapted from Paper III).

6.3 Linking groundwater dynamics to climatic and non-climatic controls

In this chapter the two main approaches of regional analysis, classification and regression, are compared for linking groundwater dynamics to physiographic and climatic controls.

6.3.1 Classification-based approaches

In Paper I, classification was carried out to test the possibility to transfer information from a data-rich to a data-poor region, based on the baseline visual classification (compare chapter 6.1). For 80 wells in an extensive and diverse geographical region, a classification of hydrogeological settings was performed. Groups were found with members originating from similar settings or regional geology and aquifer extent. Three of these groups are shown in Figure 10 on page 21, where a) for example represents deep sandstone aquifers of large horizontal and vertical spatial extent, covered by rocks of lower permeability. Figure 10b) contains mainly wells tapping thick fluvial deposits with shallow water tables in wetlands within a small geographical region, while the class of hydrographs in Figure 10c) are located in gravel and sand fluvial aquifers. Among the members of the classes, some were geographically dispersed over larger areas and some concentrated to a certain locality. This means that the local climate signal does not necessarily make the response locally unique. The fact that groundwater dynamics can be similar despite large distances (hundreds of kilometers) has also been identified by studies in other hydrogeological settings less (Wunsch *et al.*, 2020) and similarly diverse (Bloomfield *et al.*, 2015). The study carried out in Paper I shows that similar groundwater dynamics are also linked to similar hydrogeological settings. The semi-quantitative classification, based on grouping of hydrographs with similar dynamics (VC) was carried out on several groups of hydrographs, all showing a homogeneity with regard to setting. This means that the fundamental idea that information can be transferred from similar systems is generally valid.

In Paper III, a joint procedure was used for classification, were dynamics were used to classify 50 wells into local, intermediate, and regional flow systems according to dynamics and hydrogeological setting. Cluster analysis yielded five clusters of dynamics, with two clusters each, classified as local and intermediate flow and one cluster as regional flow (Figure 18). The local flow clusters were located in shallow gravel aquifers with a saturated thickness < 10 m and mean groundwater depth close to the surface < 5 m. All class members are connected to the adjacent river, with differences in dynamics within the clusters correlated with distance to stream or anthropogenic influence. The wells classed as regional flow systems tap deep aquifers with mean groundwater depth > 20 m and high mean saturated thicknesses of 14 - 28 m. Classification of the remaining clusters was less straightforward. Since aquifer characteristics were in between the values found for local and regional flow systems, they were defined as representative of intermediate flow systems. Here the steps of classification are clearly linked, which allows a straightforward prediction of groundwater regimes in similar settings within the study domain.

Generalizing such a classification to a larger data set, in a diverse study domain using statistical methodology, however, is more difficult. In Paper I for example, an attempt was made to understand, which clustering method at which granularity (number of clusters) yields the most homogenous clusters with regard to depth to groundwater table and aquifer thickness (shown for the best performing methods, average and Ward linkage with a dynamic time warping distance Figure 19). Here, the mean of within cluster standard deviation (a measure of heterogeneity of clusters) decreases with higher granularity until reaching ten clusters. After that, homogeneity increases only slightly. Average linkage performs slightly better than the second-best Ward linkage and substantially better than the visual classification. This result is surprising, since neither of the clustering methods were among the results with highest visual homogeneity (compare Figure 13). The main difference between the two linkage methods is that average linkage generally yields fewer, large clusters with many outlier clusters. This uneven cluster size distribution of average linkage explains the higher heterogeneity as seen in the shaded bands in Figure 19a) and b). Although an optimal number of clusters can be found with regard to granularity, heterogeneity remains high throughout. Conversely, an average error of 5 ± 5 m for depth to groundwater (Figure 19a) and 13.4 ± 20 m for aquifer thickness (Figure 19b) will not provide accurate prediction of groundwater dynamics. Adding supplementary descriptors should improve such a prediction since the variety of hydrogeological settings could be better captured and, therefore, different systems can be better distinguished. Such system characteristics should be inspired by the qualitative descriptors used in the section above, for example the location in valley, distance to stream or land cover.

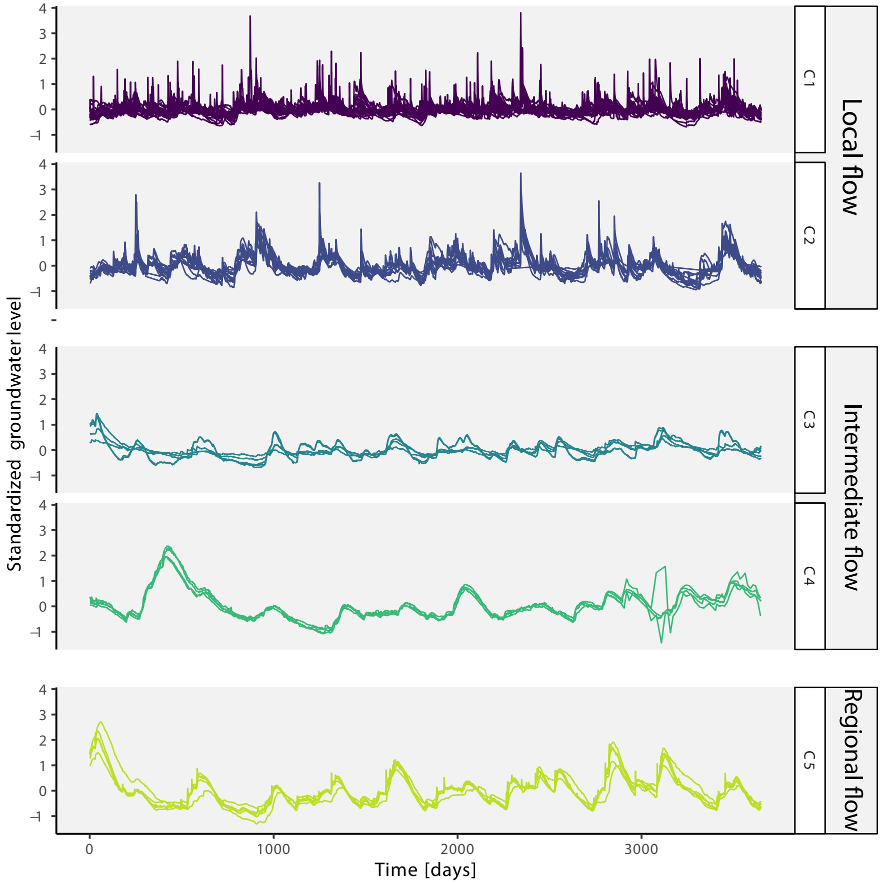

Figure 18. Groundwater hydrographs classified by dynamics and hydrogeological settings into local, intermediate and regional flow systems (adapted from Paper III).

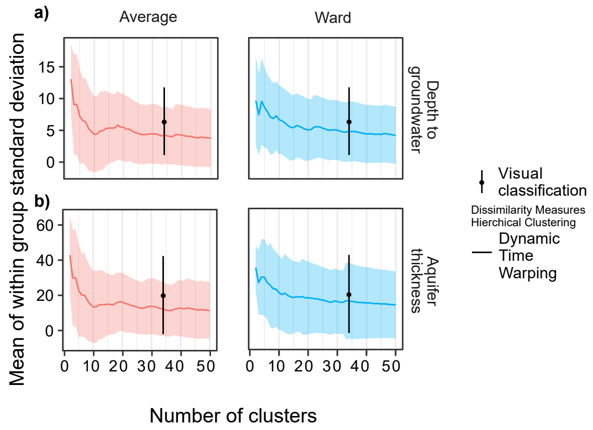

Figure 19. Within group heterogeneity with regard to two selected hydrogeological descriptors: a.) depth to groundwater table [m] and b.) Aquifer thickness [m]. The shaded area shows the range of the standard deviation for each method, i.e. the minimum and maximum within group standard variation for each cluster (from Paper I)

6.3.2 Regression-based approaches

While classification based on cluster analysis of groundwater level hydrographs has been carried out by other scholars (Moon *et al.*, 2004; Triki *et al.*, 2014; e.g. Bloomfield *et al.*, 2015; Machiwal and Singh, 2015), regression analysis is uncommon. A successful prediction using multiple regression analysis, as argued in the section above, is dependent on system characteristics that are surrogates for processes influencing groundwater dynamics. A stepwise regression strategy was adapted to find these dependent system characteristics that can be linked to groundwater dynamics from a large number of quantifiable geologic, boundary, topographic, and climatic system descriptors (4.2), a strategy that has previously been taken in surface water, (e.g. Bloomfield *et al.*, 2009; Ali *et al.*, 2012; Van Loon and Laaha, 2015; Kuentz *et al.*, 2017).

In Paper IV, prior to regression analysis, Pearson, Spearman, and distance correlation analysis, stratified by aquifer pressure state were estimated. Significant correlations vary between 0.11 and 0.64 ($p > 0.05$) where boundary and geologic descriptors show the highest correlations and the highest number of significant pairs between indices and descriptors. Both, climatic and topographic descriptors have significantly lower correlation coefficients and fraction of significant pairs. Further, the bulk of index-descriptor relations are relatively weak while a few strong dependencies exist. Comparing Pearson correlation to monotonous Spearman's and non-linear distance correlation similar patterns can be seen. However, the latter two show a larger number of significant pairs and higher absolute correlation coefficients. A few descriptors show large increases of more than r = 0.2 when estimating distance and Spearman correlation compared to Pearson correlation (e.g. distance and height to stream, depth to groundwater, median and skewness of slope on the regional scale, thickness of saturated zone, as well as height position of observation well). As will be demonstrated in the subsequent paragraphs, all of these descriptors are important controls of groundwater dynamics. Their non-linear behavior should therefore be explored more deeply and considered in the statistical (regression) model design.

Forward stepwise regression was used to build explanatory models for each index individually by selection of significant predictors in Paper IV. Despite the hydrologically diverse landscape and extensive spatial scale, clear relationships could be found, while using a relatively simple methodology. Modelling linear relations only, groundwater dynamics components had an average R^2 of 0.34 for unconfined aquifers and of 0.47 for confined aquifers. The frequency of predictors in models for unconfined and confined aquifers can be seen in Figure 20. As seen from correlation analysis in the section above, boundary descriptors are most important, especially distance and height difference to stream, the former also being the most important predictor of groundwater dynamics. Only three aggregated climate descriptors were significant: average temperature, precipitation and the seasonality index. Within geology, the most important predictors are depth to groundwater head and aquifer depth - variables that have previously been used for classification. Among topography, the most frequent predictors are mean of regional slope and regional catchment area divided by flow accumulation. In summary among the four classes of descriptors, all show relevance and most significant predictors are clearly coupled to groundwater flow processes: Distance to stream as a drainage boundary, average annual precipitation as system forcing, depth to groundwater level, indicating time lag of recharge and finally topographic features, such as slope characteristics of the surrounding environment, which is coupled to surface runoff and the hydraulic gradient.

Topographic controls, however, show realtively little importance as predictors. This is surprising, since the Topgraphic Wetness index (TWI), curvature and local slope that have shown good explanatory power of groundwater dynamics at catchment and hillslope scales (e.g. Seibert *et al.*, 2003; Rinderer *et al.*, 2016; Rinderer *et al.*, 2017; Rinderer *et al.*, 2019) have little or no relevance at the scales studied here. Except for the question of scale, a reason for this inconsistency can be the choice of representative area, which was taken using averaged information within two different radii on local and regional scale respectively. This was a compromise taken due to the unknown representative aquifer volume. Nonetheless, with little guidance given in published literature, the adapted approach appears successful, as some predictors such as slope and convergence are important at the regional scale and result in plausible models as seen in the paragraph below. Another class of predictors derived from topgraphy showing lower relevance is the group of drainage divide boundaries. Cuthbert (2014), e.g. shows that distance to boundary is an important control of groundwater recession behaviour for unconfined aquifers. In conclusion, while derivation of many predictors is straightforward, the unexpected weakness of some topographic predictors that are relevant with regard to factors affecting groundwater recharge processes (cf. Tóth, 1963; Haitjema and Mitchell-Bruker, 2005) calls for a specific sensitivity study for choosing threshold of drainage divide definition as well as radius and DEM-resolution for local and regional scale aggregation.

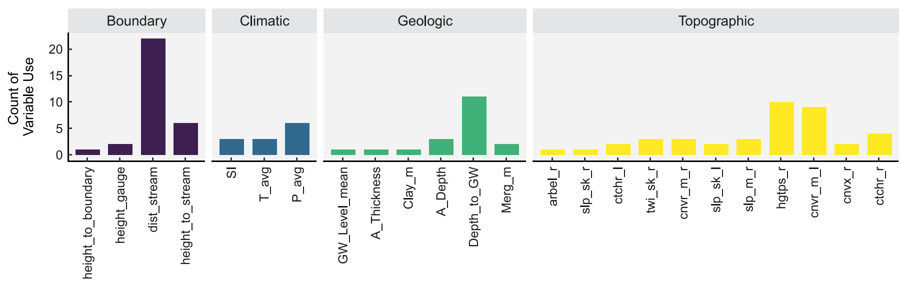

Figure 20. Absolute count of predictors in exploratory regression models across all models. Explanations of predictors can be found in 0 (from Paper IV).

In Paper IV, a number of models were explored in depth by relating significant predictors to literature and conceptual thinking. Figure 21 shows one out of five models in Paper IV for indices in unconfined aquifers. Figure 21a) shows upper and lower end members of groundwater hydrogaphs based on the index value within the data set. As shown in Figure 21b) (coefficients in **Error! Reference source not found.**), the amplitude of a groundwater time series increases when the *mean slope on regional scale* (slp_m_r), *average precipitation* (P_avg) increase and *skewness of slope at the regional scale* (slp_sk_r) is negative. The setup in Figure 21b) shows a typical hydrogeological setting, where mountain block recharge occurs (Wilson and Guan, 2004). This interpretation is supported by mapped amplitude values in Figure 21c), where the highest values are found in the South of the study domain, which is a peri-alpine area. Another example studied in-depth is magnitude, which is particularly stable in topographical depressions, such as incised river valleys, where winter and summer floods occur, driven by snow melt and summer precipitation. Timing is strongly correlated with amount and seasonality of precipitation and consequently the vicinity to the Alps, with a clear north-south gradient. High flashiness generally occurs near streams and in shallow aquifers, but also the height of the observation well within the surrounding environment plays a role. Explaining this latter factor is more difficult and published literature is partially conflicting due to different observational scales and physical environments. In conclusion, while in some cases models and process understanding cannot be unambigously reconciled at the regional scale, most models link climatic and physiographic properties to groundwater dynamics in a plausible manner with regard to physical process understanding. Confined aquifers on the other hand are comparitively understudied with regard to observational studies, which makes description of empirical models more difficult. These models were also small with two significant predictors at most, which indicates a lack of available, relevant descriptors. Local information related to recharge mechanisms are generally not available and borehole data only describe the condition at the location of the observation well, not larger scale structure can be considered. However, models for confined aquifers have higher effect sizes than unconfined, indicating that the dynamics of unconfined aquifers might be easier to describe with empirical models.

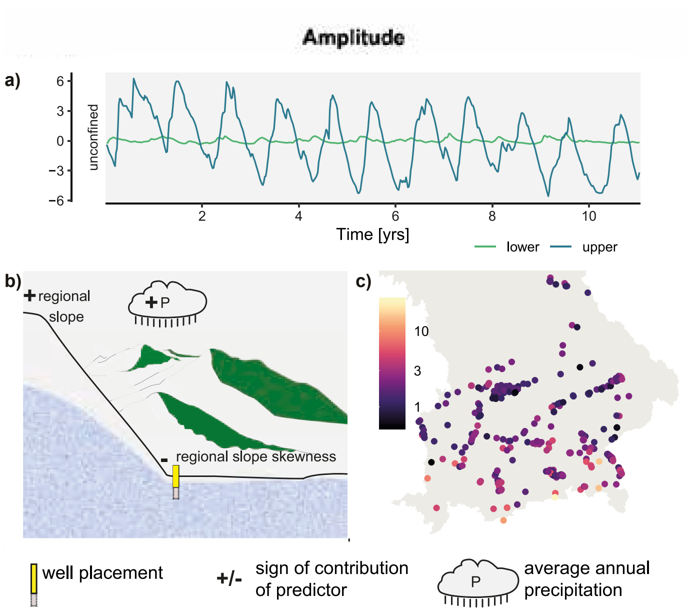

Figure 21. a) Example time series showing the end member of amplitude within the data set (centred time series). b) Landscape with explanatory factors indicating and well placement indicates the position in the landscape where high groundwater level amplitudes are expected. c) Spatial distribution of amplitude values, with a South-North gradient (adapted from Paper IV).

Regression analysis performed in Paper V, linking individual quantiles of the groundwater level duration curve to climatic and non-climatic characteristics also result in plausible models with coefficients of determination between 0.3 – 0.5. Based on quantiles of groundwater level, four aquifer states can be distinguished: Flooding, wet state, baseline state and drought state. When the aquifer is in the flooding state, groundwater dynamics are controlled by aquifer thickness, summer precipitation and percentage of agriculture at the site. Within the wet state, dynamics are controlled by aquifer thickness and distance to stream, the baseline state is dependent on thickness of unsaturated zone and distance to stream. When the aquifer is in a drought state, apart from thickness of unsaturated zone and distance to stream, seasonality of precipitation, and silt layer thickness control the groundwater level.

6.4 Prediction of groundwater hydrographs at ungauged sites

In Paper V, groundwater hydrographs were estimated at ungauged sites using statistical interpolation. Results are analyzed for classified groups of hydrographs found by cluster analysis and labelled according to hydrogeological function. Figure 22a) shows three clusters (C1 (n=27), C4 (n=38), C7 (n=6) are only three out of seven in Paper V, which were chosen here for sake of brevity) with differing response and hydrogeological features. C1 shows high flashiness, thus a duration curve that is steep at the low quantiles and flat above (Figure 22b). The flashiness is related to the C1 member's proximity to streams and distance to drainage divides and shallow depth to groundwater. This together with the aquifers high hydraulic conductivity suggests connectivity to streams. Groundwater dynamics in C7 are dominated by inter-annual variability and therefore show a flatter duration curve (Figure 22a) and b)). Hydrogeologically, these wells tap deeper systems and have lower hydraulic conductivity and low connectivity to streams. C4 has an intermediate setting.

Prediction of time series at ungauged locations was carried out using neighbors only and dc-based techniques. The mean (median) performance, R^2, when comparing observed and estimated groundwater level time series is 0.73 (0.76) for dc-based and 0.74 (0.78) and neighbors only. This means that using system characteristics for predicting daily time series at ungauged sites actually performs slightly lower than if just averaging the nearest neighbor's time series. However, when studying Figure 23a), it becomes clear that the dc-based method performs better, when few similar neighbors are available. Performance of the dc-based method is for example higher for C7, where only five similar neighbors are available, whereas the performance is higher for the neighbor-only method for C1 and C4 with 26 and 37 members. This effect can also be seen in Figure 23b), where sites are colored according to which method gives higher performance. The sites with better performance of dc-based generally show higher mean distance to nearest neighbor. Figure 23b) also shows that performance generally decreases logarithmically with mean distance to nearest neighbor.

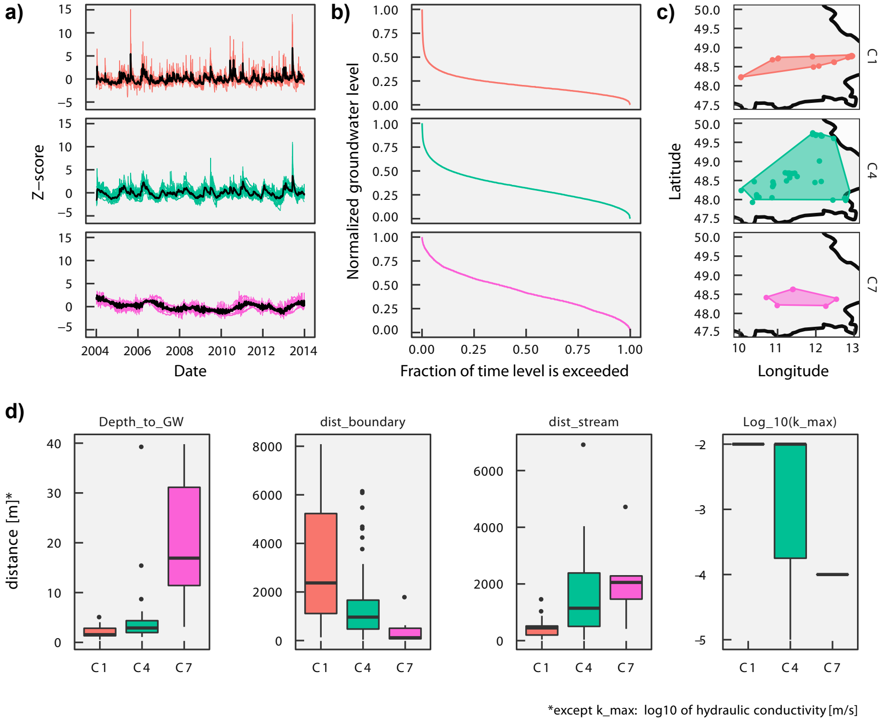

Figure 22. a) Time series within cluster, black line represents cluster mean. b) Groundwater level duration curve. c) Spatial distribution of cluster members. d) Hydrogeological properties of site, stratified by cluster.

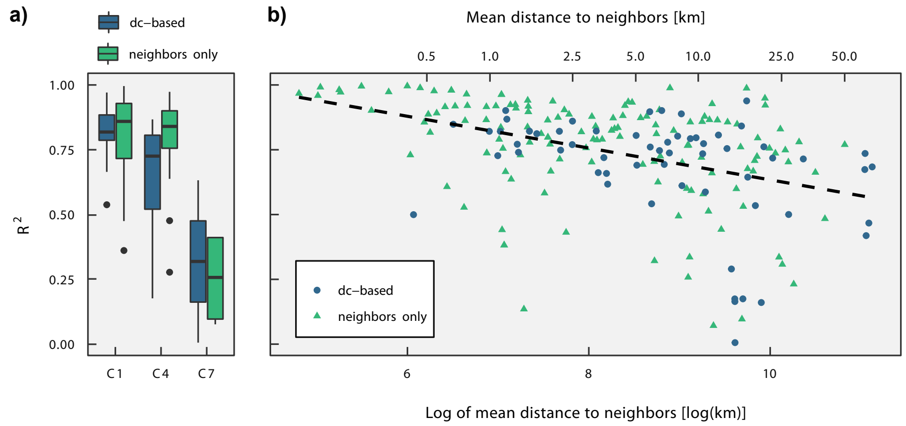

Figure 23. a) Performance of estimation (R^2) decreases logarithmically with mean distance to neighbors. b) Performance of models stratified by cluster and regionalization type, based on site characteristics (dc-based) or only on nearest neighbors.

6.5 Outlook for regionalization and prediction of groundwater dynamics

In the sections above, classification and regression-based approaches for regional analysis of groundwater systems were explored. The main aim was to prepare a starting point for predictive modelling of groundwater dynamics, by a) finding appropriate strategies to derive information from groundwater level time series, and b) investigating links between dynamics and controls. Below, some theoretical limitations of groundwater regionalization are discussed (6.5.1), contrasted with the results of attempts of prediction in Paper V, and general strategies for predicting groundwater response at ungauged sites (6.5.2).

6.5.1 Representative scale and volume

The methods discussed in the chapters above showed that groundwater dynamics can be clearly linked to system characteristics. However, transfer to an ungauged system is not straightforward since in the presented studies, definition of system volume is diffuse and representative scales are mixed. Papers I-III, e.g. show that both, similar but also dissimilar responses, are found at very close locations, which has also been identified within surface water as flow regimes can show large variation within catchments (Poff *et al.*, 2006). While in surface hydrology this can be attributed to contrasts in, e.g. soils (Bloschl and Sivapalan, 1995), these local dissimilarities in groundwater response are dependent on the hydrogeological unit, boundary conditions and properties of the observation well, which may show large contrasts of multiple orders of magnitude on a small scale in three dimensions.

These contrasts are due to different geological structures, described through e.g. geomorphology, geological genesis and therefore varying hydrogeological properties as well as differing recharge mechanisms at different scales. Since these local and regional aspects play a role, a hydrogeologist's intuition is that, as a first step, a classification should be based on regions delineated by known geological structures and hydrogeological type settings, a common exercise in hydrogeological research (Blank and Schroeder, 1973; Anderson, 1989; Klingbeil *et al.*, 1999; Stejmar Eklund, 2002) (inductive reasoning approach). However, as mentioned in chapter 2, this type of classification is often linked to a particular geological environment, qualitative, and therefore not easily derivable. Further, it has other objectives than relating the classified areas to groundwater level dynamics. In consequence, a new type of response-based classification is needed, using classification of geological structures but based on system controls that can be quantifiably linked to groundwater level dynamics as a basis for definition of representative scale and system.

Apart from geological structures, emergent behavior can be expected at a well, where processes from regional and local scale are combined. Disentangling the resulting complexity presented at different scales, calls for a multi-scale approach for predictions in ungauged aquifers. A first step for this type of system description was taken in Paper IV and V, which combines system characteristics at regional, local and point scale. At these scales, physiographic and climatic descriptors were derived with the aim to improve understanding of factors that control groundwater dynamics and their scale-dependency. With no prior response-based classification available, system characteristics were linked globally (within the domain) to groundwater dynamics, meaning that dynamics measured in observation are treated as a point in a continuous space (one environmental class). Although this assumption is only valid in cases where aquifer conditions are homogenous , the study showed that clear central tendencies exist. When using information from all three scales, this means that groundwater level dynamics generally behave consistently despite varying geological structures across multiple discontiguous and discontinuous aquifers.

While these findings might appear as an argument against a prior response-based classification, Paper II showed that even a simple prior classification compared with global models, results in models with a better fit and strongly differing or even inverse relations between dynamics and system descriptors. In Figure 24, wells were stratified according to a simplified response-based classification, based on depth of system, aquifer pressure state, and aquifer material. Flashiness (calculated as the Richards-Baker index), e.g. shows positive correlation for deep wells with depth to groundwater and for shallow wells to distance to stream. Flashiness decreases for shallow wells with depth of groundwater table and for deep wells with distance to stream. Opposing correlations based on depth and pressure state can be seen also for Inter-annual variability and Seasonality-Timing. These dynamics components also indicate further stratification of shallow wells with regard to their relation to distance to stream, as indicated in Figure 24b) by dotted regression lines and ellipses. The dotted elipses envircles a sub-group of the class of shallow wells that show deviating behaviour, which can attributed to their great vicinity to streams (< 150 m). Correlations between indices and system characteristics with prior stratification are higher than for global models. In effect, prior functional classification appears necessary, to achieve better model fits than were achieved in papers IV and V.

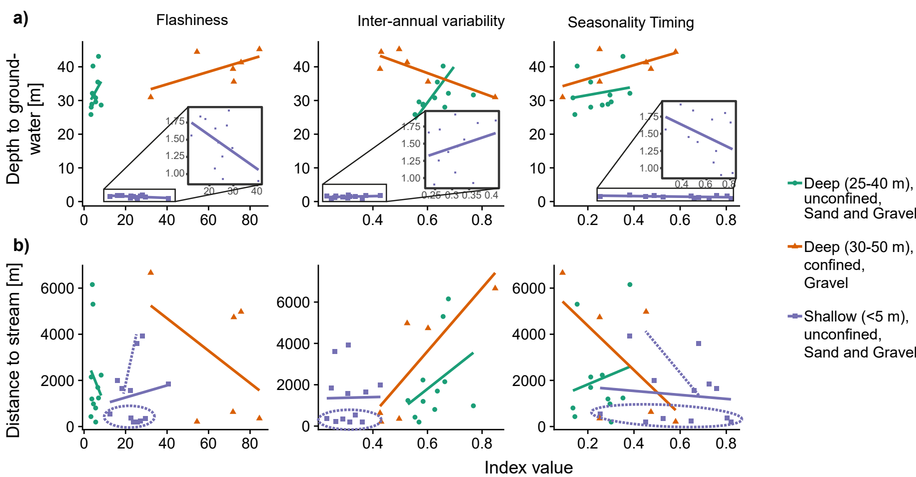

Figure 24. a) Relationship between indices (Index values stratified by class for components Flashiness (baker2), Inter-annual variation (iaf.y), Seasonality/magnitude (parde)), and distance to groundwater level fitted with a regression line. b) The relationship between indices and distance to closest stream level fitted with a regression line. Blue points in dotted ellipse are wells deviating from the average behavior with regard to pattern (adapted from Paper II)

In summary, combining the two approaches shows promise for reducing variability by deriving functions that link groundwater level dynamics and system characteristics within each group of response classes. Through this, a framework for classification of groundwater systems can be achieved with the advantages that both approaches offer. This will provide a tool for organization of groundwater data sets to improve insight into expected hydrogeological behavior and can be a first step for a holistic framework for similarity-based groundwater management as suggested by Barthel (2014).

6.5.2 Towards prediction

As shown within this chapter and the sections above, regression and classification are useful methods to advance the thesis overarching aim of prediction in ungauged aquifers. However, so far uncertainty is either unquantified (classification methods) or relatively large as expressed by coefficients of variations of 0.3 – 0.5 (regression methods). This means that using the methods in operational groundwater resource assessment, requires careful evaluation and expert knowledge. Below, necessary steps for improving methods to reduce uncertainty are discussed.

As shown above in section 6.5.1, model improvements are expected by combining a functional response classification. However, for substantial reduction in uncertainty, advances are needed in design and type of system descriptors. For example, Paper IV shows that the design of topographic and boundary descriptors still requires a comprehensive sensitivity analysis and that there is a lack of system descriptors for unconfined aquifers. As shown in Paper III, local, intermediate and regional flow systems according to Tóth (1963) show distinct patterns of groundwater level dynamics. Therefore, a descriptor that characterizes which groundwater flow system is being tapped by the well, would potentially add to explanatory power. Such a descriptor could be derived from a DEM using the methodology proposed by e.g. Wörman *et al.* (2006). This could further improve the performance of the regional groundwater level duration curve method, presented in Paper V not only by reducing the uncertainty of the fixed quantile models, but also improving the neighbor selection. Further, the properties of the unsaturated zone are underrepresented. However, in their current form (thickness of local silt, peat and clay), they haven't shown strong links to groundwater dynamics.

The steps mentioned in the paragraph above are important, but the main focus going forward, should be to use classifications that systematize knowledge of groundwater systems response to optimize prediction of groundwater levels in ungauged aquifers. Nevertheless, populating such predictive models can and should be explored simultaneously with development of a response-based classification, to achieve a tight coupling of the two. Only when classification is based on process understanding and benefits prediction, will it retain its raison d'être. This thesis focuses on classification-based or regression-based methods as a basis to explore the relation between groundwater dynamics and system characteristics. These steps were performed to enable prediction through transfer of information. Such prediction from cluster analysis-based classification as carried out in Paper I is limited (see discussion in section 6.3.1) but could be improved by making use of the physiographic and climatic descriptors developed in Paper IV. When distinguishable, homogenous classes are found and time series at ungauged similar locations can be transferred (either directly or through back-transformation to time domain) based on the results of classification. Regression models, do not require homogenous classes, but take into account the more continuous nature of groundwater response. Paper IV shows how regression models can be used to predict features of groundwater dynamics. This information can be used for characterization of expected groundwater regime at ungauged locations. In Paper V, entire time series are estimated at ungauged sites through regression-based statistical-interpolation.

The regression-based method such as the regionalization of duration curves shown in Paper V and based on Shu and Ouarda (2012) has shown merit and potential for further improvement through prior classification and empirical models with higher explanatory power. Such models should take into account interaction between predictors and nonlinearity between groundwater dynamics and system characteristics, which is present in a number of such relations as shown in Paper IV (e.g. nonlinear regression, self-organizing maps, …). Within statistical streamflow time series prediction other methods using neighborhood selection (Jayawardena *et al.*, 2002; Bárdossy *et al.*, 2005) or geostatistical methods (e.g. Skøien *et al.*, 2006) are conceivable. An alternative approach also used in PUB is to regress model parameters of transfer functions or lumped groundwater models on system characteristics to ungauged locations. These models are driven by local forcing data (rainfall, evapotranspiration, etc.) and can thus reflect the specific meteorological conditions at a site. Examples of such transfer functions are impulse response functions (Bakker and Schaars, 2019) and artificial neural networks (Wunsch *et al.*, 2018) which have also shown merit for estimating the magnitude of anthropogenic impacts on groundwater dynamics. Ecrepont *et al.* (2019) shows a strategy of transferring transfer functions to ungauged catchment based on site characteristics for streamflow. Similar to these methods are empirical transfer functions, where groundwater dynamics indices are used to predict groundwater level time series, however, so far only at annual temporal resolution (Chen *et al.*, 2002). Conceptual or lumped numerical models are routinely parametrized at ungauged sites through regionalization in surface water hydrology (He *et al.*, 2011). Such models have also been developed for particular groundwater settings (e.g. Jackson *et al.*, 2016) These methods can be used for prediction and their application and comparison in various case studies will be necessary to clarify which method is appropriate for specific settings and research questions.

Finally, the methods presented in this thesis are in principle transferable to other regions where data is available. The methods require time series and descriptors that can be generalized from widely available mapped data and digital elevation models to establish a multiple descriptor-space. However, so far, the methods have only been demonstrated on sites located within the same study domain. Despite the area's hydrogeological diversity and landscapes from mountain valleys, riverine valleys to lowlands, other hydrogeological settings and climate zones should be probed to deepen understanding of relevant process and data requirements, as well as ensure transferability of the methods.

7 Conclusions

The overall aim of the present thesis was to enable quantitative prediction of groundwater dynamics in ungauged aquifers using comparative regional analysis. The scientific contributions are mainly related to the introduction of new quantitative measures of groundwater dynamics and system properties, as well as their application within classification and regression approaches for regional analysis. Based on the objectives in section 1.2 the contributions to the field can be summarized as follows.

(1) Methods to group sites with similar groundwater dynamics are useful in aiding response-based classification. The visual perception of similarity is only weakly reproduced by the tested mathematical tools of similarity analysis, but is valuable as a tool for understanding data and evaluating the results of classification (Paper I).

(2) The developed typology for characterization and quantification of groundwater dynamics is a useful tool to compare and characterize groundwater level time series on the regional scale as demonstrated in papers II-IV. The quantitative indices can be used to perform comparative regional analysis and link features of groundwater hydrographs to site characteristics.

(3) Groundwater dynamics measured in wells tapping local, intermediate or regional flow system can be distinguished using indices. The skillfulness of the indices however differs, such that certain indices perform best for particular flow systems and groundwater regime (Paper III).

(4) Descriptors of system characteristics can be linked to groundwater dynamics through indices with multiple linear regression models with R^2 0.3 – 0.6. The most important factors controlling groundwater dynamics are hydrogeological boundaries and local geology, regional slope and preciptation patterns (Paper IV). Further, groundwater dynamics are controlled by different factors depending on aquifer state (Paper V).

(5) Groundwater level time series can be estimated at ungauged sites based on spatial interpolation supplemented by physiographic site characteristics (average R^2 = 0.74). Predictability depends on nearest neighbor and decreases logarithmically with distance. Including site characteristics improve prediction in 30% of cases (Paper V).

Jointly, the studies show that despite a heterogeneous landscape and a large study domain, relationships between groundwater dynamics and system characteristics can be modelled and are consistent with process understanding. Through these relationships, groundwater dynamics can in principle be estimated at ungauged sites through classification as averaged hydrographs (Paper I and III) and locally through regression, as a set of statistically averaged features of groundwater dynamics (Paper IV) or as estimated daily groundwater hydrographs (Paper V). However, at this stage, all approaches require expert knowledge in addition to a dataset that covers a large multiple-descriptor space. Significant work has to be invested to reduce uncertainty by improving descriptors of groundwater systems and explanatory models. Moreover, there are strong indications that predictions will be further improved, when supplementing the regression approaches with a prior functional response-based classification. Through predictions in ungauged aquifers, the methods presented in this thesis show promise to improve assessment of groundwater storage dynamics and aid water balance calculations for operational regional scale groundwater resource assessment.

References

Ali G, Tetzlaff D, Soulsby C, McDonnell JJ, Capell R. 2012. A comparison of similarity indices for catchment classification using a cross-regional dataset. Advances in Water Resources, **40**: 11-22. DOI: DOI 10.1016/j.advwatres.2012.01.008.

Anderson MP. 1989. Hydrogeologic facies models to delineate large-scale spatial trends in glacial and glaciofluvial sediments. Geological Society of America Bulletin, **101**: 501-511. DOI: 10.1130/0016-7606(1989)101<0501:HFMTDL>2.3.CO;2.

Anderson MP, Woessner WW, Hunt RJ. 2015. Applied Groundwater Modeling - Simulation of Flow and Advective Transport (2nd Edition). Elsevier.

Asch K. 2007. Europe's Geology on-line: The IGME 5000.

Baker DB, Richards P, Loftus TT, Kramer JW. 2004. A new flashiness index: characteristics and applications to midwestern rivers and streams. Journal of the American Water Resources Association, **40**: 503-522.

Bakker M, Schaars F. 2019. Solving Groundwater Flow Problems with Time Series Analysis: You May Not Even Need Another Model. Ground Water. DOI: 10.1111/gwat.12927.

Bárdossy A, Pegram GGS, Samaniego L. 2005. Modeling data relationships with a local variance reducing technique: Applications in hydrology. Water Resources Research, **41**: n/a-n/a. DOI: 10.1029/2004wr003851.

Barry RG. 2008. Mountain Weather and Climate. 3 Edn., Cambridge University Press.

Barthel R. 2006. Common problematic aspects of coupling hydrological models with groundwater flow models on the river catchment scale. Advances in Geosciences, **9**: 63-71.

Barthel R. 2011. An indicator approach to assessing and predicting the quantitative state of groundwater bodies on the regional scale with a special focus on the impacts of climate change. Hydrogeology Journal, **19**: 525-546. DOI: 10.1007/s10040-010-0693-y.

Barthel R. 2014. A call for more fundamental science in regional hydrogeology. Hydrogeology Journal, **22**: 507-510. DOI: DOI 10.1007/s10040-014-1101-9.

Barthel R. 2014. HESS Opinions "Integration of groundwater and surface water research: an interdisciplinary problem?". Hydrology and Earth System Sciences, **18**: 2615-2628. DOI: 10.5194/hess-18-2615-2014.

Barthel R, Banzhaf S. 2016. Groundwater and Surface Water Interaction at the Regional-scale – A Review with Focus on Regional Integrated Models. Water Resources Management, **30**: 1-32. DOI: 10.1007/s11269-015-1163-z.

Barthel R, Braun J, Rojanschi V, Wolf J. 2016. Hydrogeology – A Consistent Basin-Wide Representation of the Major Aquifers in the Upper Danube Basin. In: Regional Assessment of Global Change Impacts: The Project GLOWA-Danube, Mauser W, Prasch M (eds.) Springer International Publishing, pp: 125-131.

Berg SJ, Sudicky EA. 2019. Toward Large-Scale Integrated Surface and Subsurface Modeling. Ground Water, **57**: 1-2. DOI: 10.1111/gwat.12844.

Bertrand G, Goldscheider N, Gobat J-M, Hunkeler D. 2011. Review: From multi-scale conceptualization to a classification system for inland groundwater-dependent ecosystems. Hydrogeology Journal, **20**: 5-25. DOI: 10.1007/s10040-011-0791-5.

Billah MM, Goodall JL. 2011. Annual and interannual variations in terrestrial water storage during and following a period of drought in South Carolina, USA. Journal of hydrology, **409**: 472-482.

Blank HR, Schroeder MC. 1973. Geologic Classification of Aquifers. Groundwater, **11**: 3-5. DOI: 10.1111/j.1745-6584.1973.tb02959.x.

Blank HR, Schroeder MC. 1973. Geologic Classification of Aquifers. Ground water, **11**: 3-5. DOI: 10.1111/j.1745-6584.1973.tb02959.x.

Bloomfield JP, Allen DJ, Griffiths KJ. 2009. Examining geological controls on baseflow index (BFI) using regression analysis: An illustration from the Thames Basin, UK. Journal of Hydrology, **373**: 164-176. DOI: https://doi.org/10.1016/j.jhydrol.2009.04.025.

Bloomfield JP, Marchant BP. 2013. Analysis of groundwater drought building on the standardised precipitation index approach. Hydrol. Earth Syst. Sci., **17**: 4769-4787. DOI: 10.5194/hess-17-4769-2013.

Bloomfield JP, Marchant BP, Bricker SH, Morgan RB. 2015. Regional analysis of groundwater droughts using hydrograph classification. Hydrology and Earth System Sciences, **19**: 4327-4344. DOI: 10.5194/hess-19-4327-2015.

Bloschl G, Sivapalan M. 1995. Scale issues in hydrological modelling: a review. Hydrological Processes, **9**: 251-290.

Blöschl G, Sivapalan M, Wagener T, Viglione A, Savenije H. 2013. Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales. Cambridge University Press.

BMNT. 2015. eHYD [Data Web Service]. Retrieved February 12, 2015, from https://ehyd.gv.at/.

Bossard M, Feranec J, Otahel J. 2000. CORINE land cover technical guide: Addendum 2000.

Boutt DF. 2017. Assessing hydrogeologic controls on dynamic groundwater storage using long-term instrumental records of water table levels. Hydrological Processes, **31**: 1479-1497. DOI: 10.1002/hyp.11119.

BRGM. 2017. BDLISA [Data Web Service]. Retrieved May 15, 2017, from https://bdlisa.eaufrance.fr/.

Brun M, Sima C, Hua J, Lowey J, Carroll B, Suh E, Dougherty ER. 2007. Model-based evaluation of clustering validation measures. Pattern Recognition, **40**: 807-824. DOI: https://doi.org/10.1016/j.patcog.2006.06.026.

Candela L, Elorza FJ, Tamoh K, Jiménez-Martínez J, Aureli A. 2014. Groundwater modelling with limited data sets: the Chari-Logone area (Lake Chad Basin, Chad). Hydrological Processes, **28**: 3714-3727. DOI: 10.1002/hyp.9901.

Castellarin A, Burn DH, Brath A. 2001. Assessing the effectiveness of hydrological similarity measures for flood frequency analysis. Journal of Hydrology, **241**: 270-285.

Chen Z, Grasby SE, Osadetz KG. 2002. Predicting average annual groundwater levels from climatic variables: an empirical model. Journal of Hydrology, **260**: 102-117. DOI: https://doi.org/10.1016/S0022-1694(01)00606-0.

Colwell RK. 1974. Predictability Constancy and Contingency of periodic phenomena. Ecology, **55**: 1148-1153.

Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, Wehberg J, Wichmann V, Böhner J. 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geoscientific Model Development, **8**: 1991-2007. DOI: 10.5194/gmd-8-1991-2015.

Crochemore L, Perrin C, Andréassian V, Ehret U, Seibert SP, Grimaldi S, Gupta H, Paturel J-E. 2015. Comparing expert judgement and numerical criteria for hydrograph evaluation. Hydrological Sciences Journal, **60**: 402-423. DOI: 10.1080/02626667.2014.903331.

Cuthbert MO. 2014. Straight thinking about groundwater recession. Water Resources Research, **50**: 2407-2424. DOI: 10.1002/2013wr014060.

de Graaf IEM, Gleeson T, Rens van Beek LPH, Sutanudjaja EH, Bierkens MFP. 2019. Environmental flow limits to global groundwater pumping. Nature, **574**: 90-94. DOI: 10.1038/s41586-019-1594-4.

de Marsily G, Delay F, Gonçalvès J, Renard P, Teles V, Violette S. 2005. Dealing with spatial heterogeneity. Hydrogeology Journal, **13**: 161-183. DOI: 10.1007/s10040-004-0432-3.

Döll P, Hoffmann-Dobrev H, Portmann FT, Siebert S, Eicker A, Rodell M, Strassberg G, Scanlon BR. 2012. Impact of water withdrawals from groundwater and surface water on continental water storage variations. Journal of Geodynamics, **59-60**: 143-156. DOI: 10.1016/j.jog.2011.05.001.

Ecrepont S, Cudennec C, Anctil F, Jaffrézic A. 2019. PUB in Québec: A robust geomorphology-based deconvolution-reconvolution framework for the spatial transposition of hydrographs. Journal of Hydrology, **570**: 378-392. DOI: https://doi.org/10.1016/j.jhydrol.2018.12.052.
Falkenmark M, Chapman TG, Unesco. 1989. Comparative Hydrology: An Ecological Approach to Land and Water Resources. Unesco.
Fan Y. 2015. Groundwater in the Earth's critical zone: Relevance to large-scale patterns and processes. Water Resources Research, **51**: 3052-3069. DOI: 10.1002/2015wr017037.
G'Sell MG, Wager S, Chouldechova A, Tibshirani R. 2016. Sequential selection procedures and false discovery rate control. J R Stat Soc B, **78**: 423-444. DOI: 10.1111/rssb.12122.
Gini C, Livada G. 1943. Transvariazione a più dimensioni. Paneto & Petrelli.
Gottschalk L. 1977. Mathematical modelling of groundwater level response in different geological environments.
Gottschalk L. 1985. Hydrological regionalization of Sweden. Hydrological Sciences Journal, **30**: 65-83. DOI: 10.1080/02626668509490972.
Green TR, Taniguchi M, Kooi H, Gurdak JJ, Allen DM, Hiscock KM, Treidel H, Aureli A. 2011. Beneath the surface of global change: Impacts of climate change on groundwater. Journal of Hydrology, **405**: 532-560. DOI: 10.1016/j.jhydrol.2011.05.002.
Götzinger J, Barthel R, Jagelke J, Bárdossy A. 2008. The role of groundwater recharge and baseflow in integrated models. IAHS-AISH Publication, **321**: 103-109.
Haitjema HM, Mitchell-Bruker S. 2005. Are Water Tables a Subdued Replica of the Topography? Groundwater, **43**: 781-786. DOI: doi:10.1111/j.1745-6584.2005.00090.x.
He Y, Bárdossy A, Zehe E. 2011. A review of regionalisation for continuous streamflow simulation. Hydrology and Earth System Sciences, **15**: 3539-3553. DOI: 10.5194/hess-15-3539-2011.
Holman IP, Allen DM, Cuthbert MO, Goderniaux P. 2012. Towards best practice for assessing the impacts of climate change on groundwater. Hydrogeology Journal, **20**: 1-4. DOI: DOI 10.1007/s10040-011-0805-3.
Hrachowitz M, Savenije HHG, Blöschl G, McDonnell JJ, Sivapalan M, Pomeroy JW, Arheimer B, Blume T, Clark MP, Ehret U, Fenicia F, Freer JE, Gelfan A, Gupta HV, Hughes DA, Hut RW, Montanari A, Pande S, Tetzlaff D, Troch PA, Uhlenbrook S, Wagener T, Winsemius HC, Woods RA, Zehe E, Cudennec C. 2013. A decade of Predictions in Ungauged Basins (PUB)—a review. Hydrological Sciences Journal, **58**: 1198-1255. DOI: 10.1080/02626667.2013.803183.

Istanbulluoglu E, Wang T, Wright OM, Lenters JD. 2012. Interpretation of hydrologic trends from a water balance perspective: The role of groundwater storage in the Budyko hypothesis. Water Resources Research, **48**. DOI: 10.1029/2010wr010100.
Jackson CR, Wang L, Pachocka M, Mackay JD, Bloomfield JP. 2016. Reconstruction of multi-decadal groundwater level time-series using a lumped conceptual model. Hydrological Processes: n/a-n/a. DOI: 10.1002/hyp.10850.
Jayawardena AW, Li WK, Xu P. 2002. Neighbourhood selection for local modelling and prediction of hydrological time series. Journal of Hydrology, **258**: 40-57. DOI: https://doi.org/10.1016/S0022-1694(01)00557-1.
Jiménez Cisneros BE, T. O, Arnell NW, Benito G, Cogley JG, Döll P, T. Jiang T, Mwakalila SS. 2014. Freshwater resources. . In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global andSectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Field CB, V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir,M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken,P.R. Mastrandrea, and L.L. White (ed.) Cambridge University Press, pp: 229-269.
Kaufman L, Rousseeuw PJ. 2005. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley.
Klingbeil R, Kleineidam S, Asprion U, Aigner T, Teutsch G. 1999. Relating lithofacies to hydrofacies: outcrop-based hydrogeological characterisation of Quaternary gravel deposits. Sedimentary Geology, **129**: 299-310. DOI: https://doi.org/10.1016/S0037-0738(99)00067-6.
Knutsson G, Fagerlind T. 1977. Grundvattentillgångar i Sverige. In: Rapporter och meddelanden Nr 9, SGU.
Kuentz A, Arheimer B, Hundecha Y, Wagener T. 2017. Understanding hydrologic variability across Europe through catchment classification. Hydrol. Earth Syst. Sci., **21**: 2863-2879. DOI: 10.5194/hess-21-2863-2017.
Law AG. 1974. Stochastic analysis of groundwater level time series. Colorado State University.
Lehner B, Verdin K, Jarvis A. 2008. New Global Hydrography Derived From Spaceborne Elevation Data. **89**: 93-94. DOI: doi:10.1029/2008EO100001.
LfU. 2017. Gewässerkundlicher Dienst Bayern. Umwelt BLf (ed.) Available online at http://www.gkd.bayern.de/index.php.

Li B, Rodell M, Sheffield J, Wood E, Sutanudjaja E. 2019. Long-term, non-anthropogenic groundwater storage changes simulated by three global-scale hydrological models. Sci Rep, **9**: 10746. DOI: 10.1038/s41598-019-47219-z.

Lóaiciga HA, Leipnik RB. 2001. Theory of sustainable groundwater management: an urban case study. Urban Water, **3**: 217-228. DOI: https://doi.org/10.1016/S1462-0758(01)00040-1.

LUBW. 2017. Daten- und Kartendienst der LUBW. Landesanstalt für Umwelt MuNB-W (ed.) Available online at http://udo.lubw.baden-wuerttemberg.de/public/index.xhtml.

Machiwal D, Singh PK. 2015. Understanding factors influencing groundwater levels in hard-rock aquifer systems by using multivariate statistical techniques. Environmental Earth Sciences, **74**: 5639-5652. DOI: 10.1007/s12665-015-4578-1.

Martens K, Van Camp M, Van Damme D, Walraevens K. 2013. Groundwater dynamics converted to a groundwater classification as a tool for nature development programs in the dunes. Journal of Hydrology, **499**: 236-246. DOI: https://doi.org/10.1016/j.jhydrol.2013.06.045.

Mauser W, Reiter A. 2016. Spatial and Temporal Interpolation of the Meteorological Data: Precipitation, Temperature and Radiation. In: Regional Assessment of Global Change Impacts: The Project GLOWA-Danube, Mauser W, Prasch M (eds.) Springer International Publishing, pp: 99-107.

McDonnell JJ, Woods R. 2004. On the Need for Catchment Classification. Journal of Hydrology, **299**: 2-3. DOI: 10.1016/j.jhydrol.2004.09.003.

Moon S-K, Woo NC, Lee KS. 2004. Statistical analysis of hydrographs and water-table fluctuation to estimate groundwater recharge. Journal of Hydrology, **292**: 198-209. DOI: http://dx.doi.org/10.1016/j.jhydrol.2003.12.030.

Nickel D, Barthel R, Braun J. 2005. Large-scale water resources management within the framework of GLOWA-Danube—The water supply model. Physics and Chemistry of the Earth, Parts A/B/C, **30**: 383-388. DOI: 10.1016/j.pce.2005.06.004.

Olden JD, Kennard MJ, Pusey BJ. 2012. A framework for hydrologic classification with a review of methodologies and applications in ecohydrology. Ecohydrology, **5**: 503-518.

Olden JD, Poff NL. 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Research and Applications, **19**: 101-121. DOI: 10.1002/rra.700.

Oudin L, Kay A, Andréassian V, Perrin C. 2010. Are seemingly physically similar catchments truly hydrologically similar? Water Resources Research, **46**. DOI: 10.1029/2009wr008887.

Pardé M. 1933. Fleuves et rivières. Librairie Armand Colin.
Pastore M, Calcagnì A. 2019. Measuring Distribution Similarities Between Samples: A Distribution-Free Overlapping Index. Frontiers in Psychology, **10**. DOI: 10.3389/fpsyg.2019.01089.
Poff NL. 1997. The Natural Flow Regime. BioScience, **47**: 769-784.
Poff NL, Olden JD, Pepin DM, Bledsoe BP. 2006. Placing global stream flow variability in geographic and geomorphic contexts. River Research and Applications, **22**: 149-166. DOI: 10.1002/rra.902.
Qiu J, Zipper SC, Motew M, Booth EG, Kucharik CJ, Loheide SP. 2019. Nonlinear groundwater influence on biophysical indicators of ecosystem services. Nature Sustainability, **2**: 475-483. DOI: 10.1038/s41893-019-0278-2.
Reiter A, Weidinger R, Mauser W. 2011. Recent Climate Change at the Upper Danube—A temporal and spatial analysis of temperature and precipitation time series. Climatic Change, **111**: 665-696. DOI: 10.1007/s10584-011-0173-y.
Riggs HC. 1973. Regional analyses of streamflow characteristics. In: Techniques of Water-Resources Investigations.
Rinderer M, McGlynn BL, van Meerveld HJ. 2017. Groundwater similarity across a watershed derived from time-warped and flow-corrected time series. Water Resources Research, **53**: 3921–3940. DOI: 10.1002/2016wr019856.
Rinderer M, Meerveld HJ, McGlynn BL. 2019. From Points to Patterns: Using Groundwater Time Series Clustering to Investigate Subsurface Hydrological Connectivity and Runoff Source Area Dynamics. Water Resources Research, **55**: 5784-5806. DOI: 10.1029/2018wr023886.
Rinderer M, van Meerveld I, Stähli M, Seibert J. 2016. Is groundwater response timing in a pre-alpine catchment controlled more by topography or by rainfall? Hydrological Processes, **30**: 1036-1051. DOI: 10.1002/hyp.10634.
Rodell M, Chen J, Kato H, Famiglietti JS, Nigro J, Wilson CR. 2007. Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE. Hydrogeology Journal, **15**: 159-166. DOI: 10.1007/s10040-006-0103-7.
Schürch M, Kozel R, Biaggi D, Weingartner R. 2010. Typisierung von Grundwasserregimen in der Schweiz - Konzept und Fallbeispiele. Schweizerischer Verein des Gas- und Wasserfaches (SVGW).
Seibert J, Bishop K, Rodhe A, McDonnell JJ. 2003. Groundwater dynamics along a hillslope: A test of the steady state hypothesis. Water Resources Research, **39**. DOI: 10.1029/2002wr001404.

Shu C, Ouarda TBMJ. 2012. Improved methods for daily streamflow estimates at ungauged sites. Water Resources Research, **48**. DOI: 10.1029/2011wr011501.

Sivakumar B, Singh VP. 2012. Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework. Hydrology and Earth System Sciences, **16**: 4119-4131. DOI: 10.5194/hess-16-4119-2012.

Skøien JO, Merz R, Blöschl G. 2006. Top-kriging - geostatistics on stream networks. Hydrol. Earth Syst. Sci., **10**: 277-287. DOI: 10.5194/hess-10-277-2006.

Stein H, Griebler C, Berkhoff S, Matzke D, Fuchs A, Hahn HJ. 2012. Stygoregions--a promising approach to a bioregional classification of groundwater systems. Sci Rep, **2**: 673. DOI: 10.1038/srep00673.

Stejmar Eklund H. 2002. Hydrogeological type settings: A tool to assess groundwater quality, to identify groundwater resources and for risk-decision analysis along roads. In: Geologiska Institutionen, Chalmers University of Technology.

Strassberg G, Scanlon BR, Rodell M. 2007. Comparison of seasonal terrestrial water storage variations from GRACE with groundwater-level measurements from the High Plains Aquifer (USA). Geophysical Research Letters, **34**. DOI: 10.1029/2007gl030139.

Strehl A, Ghosh J. 2002. Cluster Ensembles - A Knowledge Reuse Framework for Combining Multiple Partitions. Journal of Machine Learning Research, **3**: 583-617.

Sugiyama H, Vudhivanich V, Whitaker AC, Lorsirirat K. 2003. STOCHASTIC FLOW DURATION CURVES FOR EVALUATION OF FLOW REGIMES IN RIVERS. JAWRA Journal of the American Water Resources Association, **39**: 47-58. DOI: 10.1111/j.1752-1688.2003.tb01560.x.

Taylor J, Tibshirani RJ. 2015. Statistical learning and selective inference. Proc Natl Acad Sci U S A, **112**: 7629-7634. DOI: 10.1073/pnas.1507583112.

Taylor RG, Scanlon B, Döll P, Rodell M, van Beek R, Wada Y, Longuevergne L, Leblanc M, Famiglietti JS, Edmunds M, Konikow L, Green TR, Chen J, Taniguchi M, Bierkens MFP, MacDonald A, Fan Y, Maxwell RM, Yechieli Y, Gurdak JJ, Allen DM, Shamsudduha M, Hiscock K, Yeh PJF, Holman I, Treidel H. 2013. Ground water and climate change. Nature Climate Change, **3**: 322-329. DOI: 10.1038/nclimate1744.

Thierion C, Longuevergne L, Habets F, Ledoux E, Ackerer P, Majdalani S, Leblois E, Lecluse S, Martin E, Queguiner S, Viennot P. 2012. Assessing the water balance of the Upper Rhine Graben

hydrosystem. Journal of Hydrology, **424-425**: 68-83. DOI: 10.1016/j.jhydrol.2011.12.028.
Tibshirani R, Tibshirani R, Taylor J, Loftus J, Reid S. 2017. selectiveInference: Tools for Post-Selection Inference,.
Tóth J. 1963. A theoretical analysis of groundwater flow in small drainage basins. Journal of Geophysical Research, **68**: 4795-4812. DOI: doi:10.1029/JZ068i016p04795.
Triki I, Trabelsi N, Hentati I, Zairi M. 2014. Groundwater levels time series sensitivity to pluviometry and air temperature: a geostatistical approach to Sfax region, Tunisia. Environ Monit Assess, **186**: 1593-1608. DOI: 10.1007/s10661-013-3477-8.
UNESCO. 2015. The United Nations world water development report 2015: water for a sustainable world. pp: 122.
Upton KA, Jackson CR. 2011. Simulation of the spatio-temporal extent of groundwater flooding using statistical methods of hydrograph classification and lumped parameter models. Hydrological Processes, **25**: 1949-1963. DOI: 10.1002/hyp.7951.
Wagener T, Sivapalan M, Troch P, Woods R. 2007. Catchment Classification and Hydrologic Similarity. Geography Compass, **1**: 901-931. DOI: 10.1111/j.1749-8198.2007.00039.x.
Van Loon AF, Kumar R, Mishra V. 2017. Testing the use of standardised indices and GRACE satellite data to estimate the European 2015 groundwater drought in near-real time. Hydrology and Earth System Sciences, **21**: 1947-1971. DOI: 10.5194/hess-21-1947-2017.
Van Loon AF, Laaha G. 2015. Hydrological drought severity explained by climate and catchment characteristics. Journal of Hydrology, **526**: 3-14. DOI: https://doi.org/10.1016/j.jhydrol.2014.10.059.
Wang T, Istanbulluoglu E, Lenters J, Scott D. 2009. On the role of groundwater and soil texture in the regional water balance: An investigation of the Nebraska Sand Hills, USA. Water Resources Research, **45**. DOI: 10.1029/2009wr007733.
VanRoosmalen L, Christensen BSB, Sonnenborg TO. 2007. Regional Differences in Climate Change Impacts on Groundwater and Stream Discharge in Denmark. Vadose Zone Journal, **6**: 554-571.
Wilbanks TJ, Kates RW. 1999. Global Change in Local Places: How Scale Matters. Climatic Change, **43**: 601-628. DOI: 10.1023/A:1005418924748.
Willems W, Kasper G, Klotz P, Stricker K, Zimmermann A. 2016. Mean Daily Discharge and Discharge Variability. In: Regional Assessment of Global Change Impacts, pp: 133-137.
Wilson JL, Guan H. 2004. Mountain block hydrology and mountain-front recharge. In: Groundwater Recharge in a Desert Environment: The

Southwestern United States, Hogan JF, Philips FM, Scanlon BR (eds.) American Geophysical Union, pp: 113-117.
Von Asmuth JR. 2012. Groundwater System Identification Through Time Series Analysis.
von Asmuth JR, Knotters M. 2004. Characterising groundwater dynamics based on a system identification approach. Journal of Hydrology, **296**: 118-134. DOI: 10.1016/j.jhydrol.2004.03.015.
Voss CI. 2005. The future of hydrogeology. Hydrogeology Journal, **13**: 1-6. DOI: 10.1007/s10040-005-0435-8.
Wunsch A, Liesch T, Broda S. 2018. Forecasting groundwater levels using nonlinear autoregressive networks with exogenous input (NARX). Journal of Hydrology, **567**: 743-758. DOI: 10.1016/j.jhydrol.2018.01.045.
Wunsch A, Liesch T, Broda S. 2020. Uncover Similarities of Groundwater Dynamics with Machine Learning based Hydrograph Clustering. in review.
Wörman A, Packman AI, Marklund L, Harvey JW, Stone SH. 2006. Exact three-dimensional spectral solution to surface-groundwater interactions with arbitrary surface topography. Geophysical Research Letters, **33**. DOI: 10.1029/2006gl025747.
Zektser IS, Dzyuba AV. 2014. Uncertainties in the assessment of climate change impacts on groundwater. Episodes, **38**: 49-53.

Appendices

Appendix A. Climatic and physiographic descriptors

System descriptors, source and description. Type of descriptor is indicated in bracket after variable name. All topographic descriptors are calculated with regard to the radial surrounding at local scale (r = 300 m) and regional scale (r = 3 000 m). Class of variable in parenthesis: (G) Geology, (M) Morphology, (L) Land cover, (B) Boundaries and (C) Climate. Data Sources: Geology (Bore profiles supplied by LfU, Bavaria), Land cover (Copernicus CORINE CLC 2000), Morphology and Boundary (HydroSheds 3 arc sec), Climate (E-OBS (Europe, 0.25° grid)). Variables marked with (^{V}) are only used in Paper V.

Variable	Description	Range		Unit
		Minimum	Maximum	
A_thickness (G)	Aquifer thickness	0	182	m
Aquifer_Depth (G)	Thickness of unsaturated zone	3	246	m
Depth_to_GW (G)	Depth to groundwater head	0	52	m
K_max ^{V}	Maximum hydraulic conductivity	1E-7	1E-2	m s^{-1}
Screen_length (G) ‡	Length of well screen	1	172	m
Screen_Low (G) ‡	Lower level of well screen	1	239	m
Screen_Upp (G)	Upper level of well screen	1	167	m
GW_Level_mean (G)	Elevation of mean groundwater level	309	932	masl.
Silt_m (G)	Cumulative thickness of Silt within profile	0	8	m
Peat_m (G)	Cumulative thickness of Peat within profile	0	2	m
Clay_m (G)	Cumulative thickness of Clay within profile	0	14	m
Marl_m (G)	Cumulative thickness of Marl within profile	0	26	m
arbel (M)	Area below obersvation well (OW): Percentage of area below OW height	0/0	100/100	%

hgthps (M)	Position of OW relative to the height range	0/0	100/100	%
cncv (M)	Concavity: Percentage of area with concave structure	0/0	100/43	%
cnvx (M)	Convexity: Percentage of area with convex structure	0/0	100/44	%
cnvr (M)†	Mean Convergence index	-11/-1 (-3/-.5)	9/1 (3.6/.3)	-
slp (M) †	Mean slope	0/1 (0/-.5)	13/27 (3.8/2.6)	-
twi (M) †‡	Mean value of Topographic Wetness index	5/6 (-.3/-.1)	10/9 (.5/.3)	-
ctchr (M)	Catchment Area/Flow accumulation: Maximum value of catchment area/flow accumulation	30234/ 120934	1070269/ 3120105	m^2/nr max grid cells contributing to each grid cell in radius
Agriculture ^{V} (L)	% of 3 km buffer occupied by agriculture	1.34	97.78	%
Broadleaved_forest ^{V} (L)	% of 3 km buffer occupied by broadleaved forest	0	44.51	%
Coniferous_forest ^{V} (L)	% of 3 km buffer occupied by coniferous forest	0	93.5	%
Mixed_forest ^{V} (L)	% of 3 km buffer occupied by mixed forest	0	40.81	%
Other ^{V} (L)	% of 3 km buffer occupied by other types of land cover	0	27.7	%
Urban ^{V} (L)	% of 3 km buffer occupied by urban fabric	0	74.87	%
Water ^{V} (L)	% of 3 km buffer occupied by surface water bodies	1.34	97.8	%
dist_boundary (B)	Estimated distance from GW OW to closest segment of the outer aquifer boundary	z	9928	M
dist_stream (B) ‡	Estimated distance from GW OW to nearest stream (main rivers)	6	10958	m
height_gauge (B)	Estimated Elevation of GW OW	310	950	masl.
height_stream (B)	Estimated Elevation of closest stream segment	309	944	masl.
height_to_stream (B)	Meters that the stream is below/above the GW OW (gradient to stream)	-95	82	m
height_boundary (B) ‡	Estimated Elevation of closest segment of outer aquifer boundary	313	967	masl.
height_to_boundary (B)	Meters that the aquifer boundary is below/above the OW (gradient to upper boundary)	-24	55	m

P_avg (C)	Mean annual precipitation	663	1649	mm
T_avg (C)	Mean annual temperature	3.5	9.3	°C
PE_avg (C)	Mean annual Potential Evaporation transpiration. Calculated from temperature and latitude with Thorntwaite equation	503	623	mm
AI (C) ‡	Aridity Index	1.1	3.0	-
meanPcoldest (C) ‡	Mean precipitation in coldest quarter of the year	41	114	mm
meanPwarmest (C) ‡	Mean precipitation in warmest quarter of the year	72	182	mm
meanPdriest (C) ‡	Mean precipitation in driest quarter of the year	30	80	mm
meanPwettest (C) ‡	Mean precipitation in wettest quarter of the year	83	198	mm
ratioPcoldwarm (C) ‡	meanPcoldest divided by meanPwarmest	.38	.7	-
ratioPdrywett (C) ‡	meanPdriest divided by meanPwettest	.28	.41	-
tdiffdrywett (C)	Time difference (in months between driest and wettest quarter	4.3	5.8	Months
SI (C)	Seasonality index of precipitation	.11	.35	-

† skewness was calculated apart from the area mean. For these, the range values are given in parenthesis. ‡ removed after pre-screening for multicollinearity

Appendix B. Indices

Descriptive list of indices. The column C1 is the dynamics components, C2 is the analogue sub-component. The index abbreviation is used in text, figures and tables of this paper are also found in the table. Time indicates on which subset the index was calculated (W: weekly, D: daily) [after Heudorfer et al., 2019]

C1	C2	Index Name	Index Abb.	Time	Principle	Ref	Paper
Structure	Seasonality-Magnitude	CV of Mean Minimum Monthly Head	cvmon.min	W	Coefficient of variation of mean minimum monthly groundwater heads.	[11]	II-IV
		Pardé Seasonality	parde	W	Pardé seasonality is the difference between the maximum and minimum Pardé coefficient. A Pardé series consists of 12 Pardé coefficients, corresponding to 12 months. Pardé coefficient for e.g. January is its long-term monthly mean groundwater head divided by the overall mean groundwater head.	[21]	
		Average Seasonal Fluctuation	iaf.s	W	Mean annual difference between the averaged 3 highest monthly groundwater heads per year and the averaged 3 lowest monthly groundwater heads per year.	[13]	
	Seasonality-Timing	Colwell's Contingency	colwell.M	W	The difference between the sum of entropy for each time step and possible state of the seasonal cycle, and the overall entropy across all states and time steps, divided by the logarithm of the absolute number of possible states. Entropy according to definition in information theory, see reference for details.	[3]	

		CV of Date of Annual Minimum Head	vardoy.min	W	Coefficient of variation of the date of annual minimum groundwater head.	[16]	
	Interannual variation	Base Flow Stability	BFS	W	Originally developed for streamflow, here the Base Flow Index [19] algorithm is analogously adapted to groundwater time series as a filter to separate the slow component ('baseflow') of the time series. Then, the mean annual 'baseflow' is calculated. Base Flow Stability is the difference of maximum and minimum annual 'baseflow'.	[21]	
		Inter-yearly Variation	iaf.y	W	The average between the range in annually averaged 3 highest monthly groundwater heads and the range in annually averaged 3 lowest monthly groundwater heads.	[13]	
		Low/high* Pulse Count	pulse.count.l/ pulse/count.h	W	Number of times during which the groundwater head drops below/exceeds a certain threshold. The threshold is defined as the 20th/80th percentile of non-exceedance.	[16]	
		Hurst Exponent	hurst	W	The slope of a linear model fitted to the relationship between the sample size and the logarithmized sample range of k contiguous subsamples from the time series.	[18]	
		Autocorrelation Frequency	autocor	W	Location (lag) where the first peak in the autocorrelation function occurs.	[18]	
	Flashiness	Base Flow Index	BFI	D	Adapted analogously to its application in streamflow. Here, a 'baseflow' time series is separated from a five-day minimum	[19]	

					groundwater head in a moving window. BFI equals the total sum of heads of original time series divided by the total sum of heads from the 'baseflow' type of time series.		
		Richards Pathlength	pathlength2	D	The pathlength of the time series, standardized by time series length. Original calculations also involve standardization by median head, which is not implemented here due to the prior scaling of the time series.	[1]	
		Richards-Baker Index	baker	D	Sum of absolute values of day-to-day changes in head divided by the sum of scaled daily head. Equivalent the Richards Pathlength without the time component.	[1]	
		Reversals in Time Series	rev.avg	D	Average annual number of rises and falls (i.e. change of sign) in daily head.	[16]	
		CV in Reversals	rev.cv	D	Coefficient of Variation in annual number of rises and falls in daily head.	[16]	
		Lyapunov Exponent	lyapunoc	W	The exponential rate of divergence of nearby data points when moving away in time from a certain data point in the series. Iteratively estimated for every point in the time series, and then averaged.	[9]	
	Amplitude	Range of Amplitude	amp.max	W	Range of unscaled groundwater head.	**	
Distribution	Boundness	Slopes in DC 0-0.1 Slope in DC 0.8-1	dc.slp.l/ dc.slp.u	D	Slope of the duration curve (DC; analogue flow duration curve for	[14]	

					streamflow) between percentile 0 and 0.1 or 0.8 and 1.		
		1st / 3rd L-moment	l1/ l2	W	First linear moment (equals mean) of the probability distribution of the groundwater time series or third linear moment (equals skewness).	[10]	
		Median in 0-1 Scale	med	W	Median in the groundwater time series after rescaling to fit the 0-1 scale.	**	
		Peak Timescale	peakts.avg	D	Area under peak divided by difference of peak head to peak base, averaged over all peaks.	[6]	
	Density	Slopes in DC 0.3-0.7	dc.slp.m	D	Slope of the duration curve (DC; analogue flow duration curve for streamflow) between percentile 0.3 and 0.7.	[14]	
		Mean of Annual Maximum	avg.ann.max	D	Mean of annual maximum.	[2]	
		Range in DC 0.1-0.9/ 0.2-0.8/ 0.25-0.75	dc.rng.01.09/ dc.rng.02.08/ dc.rng.025.075	W	Range of the duration curve (DC) between the percentile 0.1 and 0.9 (Variants: between 0.2 and 0.8, 0.25 and 0.75).	[15]	
		2nd / 4th L-moment	l2/ l4	W	Second linear moment (equals standard deviation) of the probability distribution of the groundwater time series. Variant: fourth linear moment (equals kurtosis).	[10]	
		Colwell's Constancy	colwell.C	W	One minus the sum of entropy with respect to state, divided by the logarithm of the absolute number of possible states.	[3]	
	Modality	Bimodality Coefficient	bimod	10	Squared product moment skewness plus one, divided by product moment kurtosis.	[4,5]	

		Excess Mass	dip	10	Test statistic of the dip test; maximum distance between the empirical distribution and the best fitting unimodal distribution. By default the best fitting distribution is the uniform.	[8]	
		Critical Bandwidth	bandwd	10	Test statistic of the Silverman test; minimum kernel bandwidth required to create a unimodal distribution estimated by fitting a Kernel Density Estimation.	[17]	
Shape	Scale	Peak Base Time	peakbase.avg	D	Difference between peak and base head, standardized by duration of peak.	[21]	
		Hydrograph Magnitude	sqr.avg	D	Difference of peak head to base head, divided by base head.	[7]	
		Low/high Pulse Duration	pulse.dur.l/ pulse.dur.h	W	Average duration of pulses where the groundwater head drops below/exceeds a certain threshold. The threshold is defined as the 20th/80th percentile of non-exceedance.	[16]	
	Slope	Recession/Recovery Constant	reces.const/ recov.const	D	Slope of the linear model fitted to percentile-wise binned means in a log-log plot of negative/ positive head versus negative/positive head one time step ahead.	[12]	
		Rise Rate	rise.avg	D	Mean rate of positive changes in flow from one day to the next.	[16]	
		CV in Rise Rate	rise.cv	D	Coefficient of Variation in S.SL_3.	[16]	
		Fall Rate	fall.avg	D	Mean rate of negative changes in flow from one day to the next.	[16]	
		CV in Fall Rate	fall.cv	D	Coefficient of Variation in S.SL_5.	[16]	
Not categorize		Colwell's Predictability	colwell.P	W	The sum of Colwell's Contingency and Colwell's Constancy.	[3]	III

		CV of Mean Maximum Monthly Head	cvmon.max	W	Coefficient of variation of mean maximum monthly groundwater heads.	[11]	
		Slopes in DC 0.1-0.8	dc.slp.m	D	Slope of the duration curve (DC; analogue flow duration curve for streamflow) between percentile 0.1 and 0.8	[14]	
		Flow Variability Index	fvi.cvdry	D	Coefficient of Variation of groundwater heads in the three consecutive "driest" months (i.e. months with on average lowest groundwater heads).	[20]	
		Flow Variability Index	fvi.cvwet	D	Coefficient of Variation of groundwater heads in the three consecutive "wettest" months (i.e. months with on average lowest groundwater heads).	[20]	
		Duration of peak climax	gwpc.tdur	D	Timespan between the time when the groundwater level reach 95% of the maximum peak level on the rising limb of the hydrograph and the corresponding point on the falling limb (95% recession). Averaged over all peaks.	[20]	
		CV in duration of peak climax	gwpc.tdur.cv	D	Timespan between the time when the groundwater level reach 95% of the maximum peak level on the rising limb of the hydrograph, and the corresponding point on the falling limb (95% recession). Coefficient of variation taken over all peaks.	[20]	
		Duration of peak recession	gwpc.trec	D	Timelag between the time after the peak when the groundwater level is still on 95% of the maximum peak level, and the time when it reaches 20% of the	[20]	

					maximum peak level. Averaged over all peaks.		
		CV in duration of peak recession	gwpc.trec.cv	D	Timelag between the time after the peak when the groundwater level is still on 95% of the maximum peak level, and the time when it reaches 20% of the maximum peak level. Coefficient of variation taken over all peaks.	[20]	
		Duration of peak rise	gwpc.tris	D	Timelag between the time when groundwater levels are still on 20% of the maximum peak level prior to the peak, and the time when they reach 95% of the maximum peak level. Averaged over all peaks.	[20]	
		CV in duration of peak rise	gwpc.tris.cv	D	Timelag between the time when groundwater levels are still on 20% of the maximum peak level prior to the peak, and the time when they reach 95% of the maximum peak level. Coefficient of variation taken over all peaks.	[20]	
		Mean of Annual Minimum	meanann.min	D	Mean of annual minimum groundwater head.	[2]	
		Non-linear Autoregressive Structure	nonlinear	W	Test statistic of the Teräesvirta neural network linearity test.	[18]	
		CV in Peak Base Time	peakbase.cv	D	Coefficient of Variation in annual Peak Base Time.	[21]	
		CV in Peak Timescale	peakts.cv	D	Coefficient of Variation in annual Peak Timescale.	[6]	
		CV in High Pulse Count	pulse.cv.h	W	Coefficient of Variation in annual High Pulse Count.	[16]	
		CV in Low Pulse Count	pulse.cv.l	W	Coefficient of Variation in annual Low Pulse Count.	[16]	

		CV in Hydrograph Magnitude	sqr.cv	D	Coefficient of Variation in annual Hydrograph Magnitude.	[7]	
		CV of Date in Annual Maximum Head	varjuld.m	W	Coefficient of Variation in Date of Annual Maximum Head.	[16]	

* Value inverted for easy interpretability, ** Standard statistic,

[1] Baker et al. (2004), [2] Clausen and Biggs (2000), [3] Colwell (1974), [4] Ellison (1987), [5] Deevi and 4D Strategies (2016), [6] Gaál et al. (2012), [7] Hannah et al. (2000), [8] Hartigan and Hartigan (1985), [9] Hilborn (1994), [10] Hosking (1990), [11] Hughes and James (1989), [12] Kirchner (2009), [13] Martens et al. (2013), [14] Oudin et al. (2010), [15] Richards (1990), [16] Richter (1996), [17] Silverman (1981), [18] Wang et al. (2006), [19] WMO (2008), [20] Hughes and Hannart (2003), [21] Heudorfer et al. (2019)

Publications I-V

I. **Haaf, E.**, Barthel, R. (2018). *An inter-comparison of similarity-based methods for organisation and classification of groundwater hydrographs*. Journal of Hydrology 2018; 559: 222-237.

II. Heudorfer, B.*, **Haaf, E.***, Stahl, K., Barthel R. (2019). *Index-Based Characterization and Quantification of Groundwater Dynamics* (*equal contribution). Water Resources Research 55(7): 5575-5592.

III. Giese, M., **Haaf, E.**, Heudorfer, B., Barthel, R. (2020). *Comparative hydrogeology – reference analysis of groundwater dynamics from neighbouring observation wells*. Hydrological Sciences Journal (accepted).

IV. **Haaf, E.**, Giese M., Heudorfer, B., Stahl, K., Barthel, R. (2020). *Physiographic and climatic controls on regional groundwater dynamics*. Revision submitted to Water Resources Research.

V. **Haaf, E.**, Giese M., Reimann, T., Barthel, R. (2020). *Estimation of daily groundwater levels in ungauged aquifers based on climatic and physiographic controls*. Manuscript.