Optimization Strategies for Spatio-temporal Groundwater Dynamics Monitoring

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN
(Dr. rer. nat.)

von der KIT-Fakultät für
Bauingenieur-, Geo- und Umweltwissenschaften
des Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von

M.Sc. Marc Ohmer

aus Herxheim

Tag der mündlichen Prüfung:
27. Juli 2022

Referent: Prof. Dr. Nico Goldscheider

Korreferent: Prof. Dr. Ty Ferré

Karlsruhe 2022

Abstract

Spatially continuous data play an essential role in hydrogeology for scientific research, risk assessment, and water management decision-making. However, most of this information is collected only selectively through measurements at groundwater monitoring wells and then regionalized. The predictive accuracy of these regionalized data, which usually provide the foundation for subsequent analysis and decision-making processes, depends on the design of the groundwater monitoring network (GMN), i.e., the spatial distribution, the monitoring density, the sampling frequency, as well as the interpolation technique and the interaction of these factors. This leads to a considerable optimization potential for the GMN and the regionalization technique.

Appropriate GMNs are therefore important tools for the sustainable management and protection of groundwater resources. They provide access points for monitoring groundwater levels and sampling, thereby providing insight into groundwater conditions. Due to the combination of high initial investment and relatively low spatial representativeness of wells due to hydrogeologic heterogeneity, planning a suitable GMN is a major challenge. This work addresses techniques for better understanding groundwater dynamics through (i) spatial and (ii) spatiotemporal optimization of groundwater level monitoring networks (GLMNs) and (iii) more accurate spatial prediction of data obtained at these monitoring wells, employing interpolation techniques.

For this purpose, the first part of this work comprises a comprehensive assessment of the most frequently applied deterministic and geostatistical, uni- and multivariate interpolation techniques for groundwater contour mapping in a study area characterized by a complex interaction between karst and an alluvial aquifer and the low-permeability strata of the Subalpine Molasse sediments. The studied methods were evaluated by global cross-validation and eco-hydrogeologic information at karst springs, wetlands, surface waters, and profile sections. The potential effect of the choice of interpolation method on subsequent calculations was made by estimating the exchange processes between karst and alluvial aquifers based on the calculated potential differences. The results show that the choice of method, especially in the case of inadequate monitoring design, can have drastic effects on further analyses. This study showed that geostatistical or kriging interpolation

methods were superior to deterministic interpolation methods. When groundwater data are sparse, co-kriging with spatially cross-correlated secondary variables (e.g., elevation, river level) that are sampled more frequently can provide meaningful information about the primary variable and thus reduce the variance of the estimation error.

In the second part of this work, different GLMN designs with varying monitoring densities were investigated using numerically modeled groundwater surfaces with a range of scales and dynamics. The objective was to provide insight into appropriate monitoring approaches for reliable spatial estimation of groundwater levels (GWLs) and derive the monitoring density at which a proper information/cost ratio is achieved. The interpolation results were evaluated using global cross-validation and the actual spatial error, computed from the numerical model surfaces as "a priori" references. Monitoring networks with regular grid arrangements provided the most accurate spatial predictions for the density spectrum studied but are not suitable due to their drawbacks, such as a lack of extensibility. Comparably good performance was obtained when the maximum kriging prediction standard error was used as a selection criterion for the selection of new well locations to existing networks. In addition, a novel optimization strategy for GMN based on mathematical quasi-random sequences was applied in this study. The approach also produced good results and offers several advantages. No prior knowledge of the aquifer through existing wells is required, and reproducible designs are obtained independently of the expansion steps.

In part three, a data-driven sparse sensing algorithm approach for selecting sparse sensor placements utilizing dimensionality reduction techniques was explored and adjusted for the temporal and spatial optimization of an existing GLMN in the Upper Rhine Graben. The optimization was performed by using a greedy search (QR) algorithm that selects and ranks monitoring wells according to their information content about aquifer dynamics. First, long-term hydrograph records were used as input data to indicate representative monitoring wells or wells with redundant or low information content. Second, optimization based on regionalized, weekly gridded GWL contour maps was performed to identify eligible locations for additional wells. This search was controlled with a spatial cost function that penalized less suitable sites. The approach studied has proven to be a potentially valuable tool for optimizing the number of wells and their locations, for network downsizing and expansion, or both combined.

Kurzfassung

Räumlich kontinuierliche Daten sind in der Hydrogeologie für die wissenschaftliche Forschung, die Risikobewertung und wasserwirtschaftliche Entscheidungsprozesse von wesentlicher Bedeutung. Die meisten dieser Informationen werden allerdings nur punktuell durch Messungen an Grundwassermessstellen erhoben und anschließend regionalisiert. Die Vorhersagegenauigkeit dieser räumlich interpolierten Daten, die in der Regel die Grundlage für weitere Berechnungen und Entscheidungen bilden, ist stark abhängig von der Konzipierung des Grundwassermessnetzes, d.h. von der räumlichen Verteilung und Dichte der Grundwassermessstellen, der Beprobungshäufigkeit, dem Interpolationsverfahren sowie dem Wechselspiel zwischen diesen Faktoren. Daraus ergibt sich ein erhebliches Optimierungspotenzial hinsichtlich des Grundwassermessnetzes und der Regionalisierungstechnik.

Geeignetes Grundwassermessnetze sind daher wichtige Instrumente für die nachhaltige Bewirtschaftung und für den Schutz der Grundwasserressourcen. Sie bieten Zugangspunkte für die Überwachung von Grundwasserständen und -proben und ermöglichen so einen Einblick in die Grundwasserverhältnisse. Die Kombination aus hohen Erschließungskosten und einer verhältnismäßig geringen räumlichen Repräsentativität der Brunnen aufgrund der hydrogeologischen Heterogenität machen die Konzeption eines geeigneten Überwachungsnetzes zu einer großen Herausforderung. Diese Arbeit beschäftigt sich mit Techniken zum verbesserten Verständnis der Grundwasserdynamik durch (i) räumliche und (ii) räumliche-zeitliche Optimierung von Grundwasserstandsmessnetzen und (iii) verbesserter räumlicher Vorhersage der an diesen Überwachungsbrunnen gewonnenen Daten unter Verwendung von Interpolationstechniken.

Zu diesem Zweck wurde im ersten Teil dieser Arbeit eine umfassende Untersuchung der meistgenutzten deterministischen und geostatistischen, uni- und multivariaten Interpolationstechniken für die Erstellung von Grundwassergleichenplänen in einem Untersuchungsgebiet durchgeführt, das durch eine komplexe Interaktion zwischen Karst, einem alluvialen Grundwasserleiter und gering durchlässigen Schichten der alpinen Molasse gekennzeichnet ist. Die untersuchten Methoden wurden durch globale Kreuzvalidierung und öko-hydrogeologische Informationen an Karstquellen, Feuchtgebieten,

Oberflächengewässern und Profilschnitten bewertet. Der mögliche Effekt der Methodenwahl auf die weitere Berechnung wurde durch Abschätzung der Austauschprozesse zwischen Karst- und Alluvialgrundwasserleiter auf Basis der geschätzten Potentialunterschiede durchgeführt. Die Ergebnisse zeigen, dass die Verfahrenswahl, insbesondere bei unzureichendem Überwachungskonzept, drastische Auswirkungen auf die nachfolgenden Berechnungen haben kann. Die Studie hat ergeben, dass geostatistische oder Kriging-Interpolationsmethoden den deterministischen Interpolationsmethoden überlegen sind. Bei dürftiger Grundwasserdatenlage kann das Co-Kriging mit räumlich kreuzkorrelierten Sekundärvariablen (z. B. Höhenlage, Flusspegel), die häufiger erfasst werden, wertvolle Informationen über die Primärvariable bereitstellen und so die Varianz des Schätzfehlers verringern.

Im zweiten Teil dieser Arbeit wurden räumliche Monitoringkonzepte mit unterschiedlichen Messdichten an numerisch modellierter Grundwasseroberflächen mit verschiedenen Skalen und Dynamiken untersucht. Ziel war es, Einblicke in geeignete Monitoringansätze für eine verlässliche räumliche Abschätzung des Grundwasserspiegels zu gewinnen und eine Überwachungsdichte abzuleiten, bei der ein angemessenes Informations/Kosten-Verhältnis erreicht wird. Die Interpolationsergebnisse wurden mit globaler Kreuzvalidierung und dem tatsächlichen räumlichen Fehler evaluiert, der anhand der numerischen Modellflächen als A-priori-Referenz errechnet wurde. Überwachungsnetze mit einer regelmäßigen Gitteranordnung boten zwar genaueste räumliche Vorhersagen für das betrachtete Dichtespektrum, sind jedoch aufgrund ihrer Nachteile, wie der mangelnden Erweiterungsfähigkeit, tendenziell ungeeignet. Eine vergleichbar gute Leistung wurde erzielt, wenn der maximale Vorhersage-Standardfehler als Auswahlkriterium für zusätzliche Brunnen für bestehende Messnetze verwendet wurde. In dieser Studie wurde außerdem eine neuartige Optimierungsstrategie für Überwachungsnetze angewandt, die auf mathematischen Quasi-Zufallsfolgen basiert. Der Ansatz liefert ebenfalls überzeugende Ergebnisse und bietet mehrere Vorteile. Er bedarf keinerlei Vorkenntnisse über den Grundwasserleiter durch vorhandene Brunnen und es werden unabhängig von den Ausbaustufen reproduzierbare räumliche Anordnungen erzielt.

Im dritten Teil wurde ein datengesteuerter Sparse-Sensing-Algorithmus-Ansatz zur Auswahl von spärlichen Sensorpositionen unter Nutzung von Techniken zur Dimensionsreduktion untersucht und für die zeitliche und räumliche Optimierung eines bestehenden Grundwasserstandsmessnetzes im Oberrheingraben adaptiert. Die Optimierung erfolgt mit einem greedy search (QR)-Algorithmus, der die Überwachungsbrunnen nach ihrem Informationsgehalt über Aquifer-Dynamik selektiert und einordnet. Als Eingangsdaten wurden langjährige Ganglinien-Aufzeichnungen verwendet, um repräsentative Messstellen oder Messstellen mit redundantem oder niedrigem Informationsgehalt zu

bestimmen. Des Weiteren wurde eine Optimierung auf der Grundlage regionalisierter, wöchentlicher Grundwassergleichenkarten durchgeführt, um mögliche geeignete Standorte für zusätzliche Messstellen zu identifizieren. Die Suche wurde durch eine räumliche Kostenfunktion gelenkt, bei der weniger geeignete Standorte abgewertet wurden. Der untersuchte Ansatz hat sich als potenziell wertvolles Instrument für die Optimierung der Brunnenanzahl und deren Standorte, für die Reduzierung und den Ausbau des Netzes aber auch für eine kombinierte Nutzung beider Möglichkeiten erwiesen.

Table of Contents

List of Figures X

List of Tables XII

Abbreviations XIII

1 Introduction 1

1.1 Motivation . . . 1
1.2 Groundwater Monitoring . . . 3
1.2.1 A Brief History of Groundwater Monitoring . . . 3
1.2.2 Objectives and Types of Groundwater Monitoring Networks . . . 4
1.2.3 Groundwater Monitoring in the International Context . . . 7
1.3 Groundwater Monitoring Network Optimization . . . 10
1.4 Regionalization of Groundwater Data . . . 12
1.5 Outline . . . 14

2 Optimal Selection of Interpolation Methods for Groundwater Contouring 19

2.1 Introduction . . . 20
2.2 Methods . . . 23
2.2.1 General Overview . . . 23
2.2.2 Deterministic Methods . . . 24
2.2.3 Geostatistical Methods . . . 26
2.2.4 Validation Methods . . . 28
2.2.5 Study Area . . . 30
2.2.6 Available Data . . . 32
2.2.7 Data Processing . . . 33
2.2.8 Calculation of Groundwater Exchange . . . 33
2.3 Results and Discussions . . . 34
2.3.1 Cross-Validation Results . . . 34
2.3.2 Groundwater Level Contour Maps . . . 35
2.4 Conclusions . . . 41

3 Optimal Spatial Design for Groundwater Level Monitoring Network 46
3.1 Introduction . . . 47
3.2 Methods . . . 50
3.2.1 Interpolation Methods . . . 50
3.2.2 Data and Data Processing . . . 52
3.2.3 Sampling Designs . . . 54
3.2.4 Validation Methods . . . 60
3.3 Results and Discussion . . . 62
3.3.1 Resulting Spatial Distribution . . . 62
3.3.2 Absolute Prediction Errors . . . 64
3.3.3 Cross-Validation Results . . . 66
3.4 Conclusions . . . 67

4 Spatiotemporal Optimization of Groundwater Monitoring Networks 74
4.1 Introduction . . . 75
4.2 Methodology . . . 79
4.2.1 Mathematical Background . . . 79
4.2.2 Application Cases . . . 83
4.2.3 Error metrics . . . 83
4.2.4 Data and Study Area . . . 85
4.3 Results and Discussions . . . 87
4.3.1 Grid Search Results . . . 87
4.3.2 Ranking of Wells and Network Reduction with Hydrograph Data . 89
4.3.3 Network Reduction and Extension Based on Gridded GWL Maps . 93
4.4 Conclusions . . . 98

5 Synthesis and Outlook 104
5.1 Summary and Synthesis . . . 104
5.2 Perspective and Outlook . . . 110

Acknowledgements 114

Declaration of Authorship 115

References 118

List of Figures

1.1 Concept of Pareto optimality 12
1.2 Graphical abstract of the study presented in Chapter 2 16
1.3 Graphical abstract of the study presented in Chapter 3 17
1.4 Graphical abstract of the study presented in Chapter 4 18

2.1 Geological map of the study area. 31
2.2 Observed vs. predicted GWLs 36
2.3 Estimated GWL contours and potential differences I/II 38
2.4 Cross section with topography and GWL contours 39
2.5 Maps of groundwater depth in the Donaumoos area 40
2.6 Maps of groundwater depth near karst springs 41
S2a Overview of the interpolation parameters used 44
S2b Estimated GWL contours and potential differences II/II 45

3.1 GWL contour maps of the MODFLOW models 53
3.2 Overview of the GMN designs approaches 55
3.3 Resulting spatial distribution of the monitoring wells for ACF A 63
3.4 APE of all methods for ACF A 64
3.5 MSAPE for all methods and surfaces 65
3.6 Error statistics for all methods and surfaces 68
3.7 Method ranking for all surfaces and CV error estimates 69
S3c APE and SAPE for all methods and surfaces 73

4.1 Study area and GMN 86
4.2 Grid search cross-validation (CV) results 88
4.3 QR based ranking of the monitoring wells 90
4.4 Z-transformed hydrographs of important/redundant monitoring wells . . . 91
4.5 Reconstruction error metrics 92
4.6 Hydrograph reconstruction at the five reduction stages 94
4.7 Summary of the results of the 3D GMN optimization 96
S4a Comparison of QR-based ranking 1D and 2D input data 101

S4b Stacked z-transformed hydrographs of monitoring wells. Rank 1 -240 . . . 102
S4c Stacked z-transformed hydrographs of monitoring wells. Rank 241 -480 . . 103

5.1 Overview of the studies carried out within the scope of this thesis 104

List of Tables

2.1 Interpolation methods applied in comparative studies 22
2.2 Evaluation statistics used in comparative studies 23
2.3 Cross-validation results 35
2.4 Estimated vertical GW exchange rates 42

S3a Overview of important parameters of the GW model surface 71
S3b Result comparison of the methods used 72

Abbreviations

ACF	Apalachicola-Chattahoochee-Flint river basin, Georgia, potentiometric surface of the *MODFLOW* model
APE	absolute prediction error
ASE	average standard error
CoK	co-kriging
CoOK	co-ordinary kriging
CV	cross-validation
DEM	digital elevation model
DSN	densify sampling network
EBK	empirical Bayesian kriging
EEA	European Environment Agency
GA	genetic algorithm
GLMN	groundwater level monitoring network
GMN	groundwater monitoring network
GPI	global polynomial interpolation
GQMN	groundwater quality monitoring network
GWL	groundwater level
GWQ	groundwater quality
IDW	inverse distance weighting
KGE	Kling-Gupta efficiency
LOO	leave-one-out
LPI	local polynomial interpolation
MAE	mean absolute error
MAPE	mean absolute percentage error
maxAE	maximum absolute error
MCA	multicriteria decision-making analysis
ME	mean error
MSAPE	mean standardized absolute prediction error
MSE	mean square error
NSE	Nash–Sutcliffe model efficiency coefficient

$\mathbf{P_1}$	Point at which the numerical derivatives of the mean standardized absolute prediction error (SAPE) of all methods, except reference method (REF), is 5×10^{-3}
$\mathbf{P_2}$	Point at which the numerical derivatives of the mean SAPE of all methods, except REF, is 1×10^{-3}
$\mathbf{P_{max}}$	Point with maximum number of monitoring wells
PCA	principle component analysis
POD	proper orthogonal decomposition
r	Pearson correlation coefficient
$\mathbf{R^2}$	coefficient of determination here: squared Pearson r
$\mathbf{R_2}$	2-dimensional additive recurrence sequences
RBF	radial basis function
rBias	relative Bias
REF	reference method
RMSE	root mean square error
RMSSE	root mean square standardized error
rRMSE	relative root mean square error
SAPE	standardized absolute prediction error
SK	simple kriging
SRS	simple random sampling
SVD	singular value decomposition
OK	ordinary kriging
UK	universal kriging
URG	Upper Rhine Graben
WFD	Water Framework Directive

Chapter 1

Introduction

1.1 Motivation

Groundwater is one of our most vital resources, providing about half of the world's drinking water and about 40 % of the water needed for irrigated agriculture (Siebert et al., 2010; World Water Assessment Programme, 2009). It is essential to many ecosystems, maintaining groundwater flow for surface waters and preventing land subsidence and seawater intrusion. This makes groundwater a local resource with a global impact. Despite these facts, groundwater, due to being hidden in the subsurface, is literally out of sight for most of us, and is often perceived as an inexhaustible and quality-assured natural resource. Increasing overexploitation due to population and economic growth and changing climatic conditions expand the pressure on this resource. The majority of groundwater is fossil groundwater (Jasechko et al., 2017). More than non-renewable energy sources, the finite resource of safe and clean freshwater will limit global population growth unless changes are made to how we utilize and conserve this resource (Kinzelbach and Kunstmann, 1998).

Groundwater resources are monitored worldwide by surveying GWLs, groundwater qualitys (GWQs) parameters, abstraction and spring discharge. Monitoring wells provide access for measuring groundwater levels and collecting groundwater samples that represent groundwater conditions at the particular well. This makes monitoring networks essential instruments to make the "invisible" groundwater "visible". So far, however, there are few uniform regulations and standards concerning the number, spatial distribution and density of the monitoring locations, their monitoring frequency, and the parameters measured. Groundwater's lack of visibility and the resulting low awareness of it has led to the neglect of systematic monitoring in many countries, despite the great importance of groundwater resources for drinking water supply and their crucial role in conserving some aquatic ecosystems. The planning and expansion (of mostly historically grown)

monitoring networks are therefore mainly done on an ad hoc basis, which means there is usually a high degree of subjectivity. Moreover, in many emerging and developing countries, aquifers are often insufficiently or completely unmonitored.

Besides the importance of groundwater for a drinking water and ecosystems, and its "invisibility", the monitoring of groundwater is of great importance due to its mostly low flow velocities and thus long residence times. If contamination is not detected and appropriate action taken in time, it may take decades or longer to regain the requisite quality. Consequences of overexploitation due to misestimation of the groundwater supply are manifold and can take decades to manifest. Impacts include ecosystem collapse, salinization of coastal aquifers, loss of aquifer storage capacity due to compaction and subsidence, and even desertification (Graaf et al., 2019), most of which are partly or totally irreversible.

Groundwater monitoring is generally more challenging than monitoring surface waters. Reasons for this are higher initial investments (e.g., drilling costs, well development, especially for deep aquifers), comparatively lower spatial representation of monitoring wells due to hydrogeological heterogeneity, and the inaccessibility of the hidden groundwater resource (IGRAC, 2020). Unlike surface water, groundwater offers hardly any visual evidence since accessible sampling is generally only possible at spring locations (e.g., Frank et al., 2019). Therefore, groundwater-related issues and crises often go unnoticed, especially on a smaller scale. The assistance of modern remote sensing methods is limited, in contrast to surface waters, to coarse assessments of storage changes on a large spatial scale (Liesch and Ohmer, 2016). This underlines the great importance of groundwater monitoring in general, particularly the need for groundwater monitoring networks with appropriate spatial coverage and monitoring frequencies (together often referred to as monitoring optimization) and reliable techniques for regionalizing information obtained at monitoring wells into spatially continuous information such as groundwater contour maps, including uncertainty estimates. Especially since the quality of the monitoring network dictates the quality of most subsequent investigations, analyses, and analytical or numerical groundwater models (e.g., Ohmer et al., 2022a,b).

The following chapter is structured as follows: Section 1.2 provides a general overview of the concept of groundwater monitoring, starting with the historical development from the first systematic measurements on wells to the present day. The differences between GMNs in their systematics, monitoring parameters, and objectives are explained, and examples of national and international groundwater monitoring networks are given to illustrate the range of current groundwater monitoring. Section 1.3 introduces network optimization concepts with a focus on Pareto optimality, a cornerstone in the field of optimization. Section 1.4 illustrates the most common methods to estimate spatially

continuous GWL and groundwater quality parameters from point observations. The focus is on kriging techniques since this is the most commonly used interpolation technique for groundwater contour maps and was applied in all three studies presented here. Section 1.5 provides a brief outline of the three studies conducted in this thesis in connection with the research questions that motivated this work.

1.2 Groundwater Monitoring

1.2.1 A Brief History of Groundwater Monitoring

According to Richter (1943) and Muzikar (2013), the first systematic recording of groundwater levels in a well was carried out between 1865 and 1880 by the abbot Johan Georg Mendel (Anonymous, 1881). He later became famous for his genetics research. Applied scientific groundwater monitoring dates back to the 19th century in England. This period is characterized by rapid technical and scientific progress in the context of industrialization and urbanization. The rapid increase in the urban population due to the high demand for factory workers and increasing industrialization led to shortages in the supply of sufficient clean drinking water in rapidly growing cities. The deteriorating quality of urban water supplies is often cited as the leading cause of the increase in mortality during this period (Davenport et al., 2019). To overcome this challenge, an interdisciplinary collaboration between geologists and engineers was established for the first time (Dassargues et al., 2021). At the same time, technical inventions such as mechanical drilling and steam-driven water pumps made it possible to increase extraction rates and access deeper groundwater resources. The expansive exploitation of groundwater began.

In addition, scientific disciplines became increasingly specialized during this period. Thus, the year 1856, when Henry Darcy published his famous law, is often cited as the starting point of hydrogeology. In the following years, Dupuit (1863) derived a formula for calculating the groundwater flow to a well based on Darcy's equation. In 1870, Dupuit's formula was extended by Thiem to determine the hydraulic properties of the aquifer with pumping tests.

In 1822, Conybeare and Phillips measured the GWL at four wells and drew the course in a diagram over 25 km from the River Thames to the town of Epping. This is probably the earliest hydrogeologic cross-section (Mather and London, 2004). The methodology of hydrogeological profile sections was taken up and extended in several studies in the following years (Clutterbuck et al., 1850; Lucas, 1876). Gustave Dumont produced the

first hand-interpolated potentiometric groundwater map in Belgium in 1856 as part of a feasibility study for large-scale drinking water production. Dumont used river level measurements and 204 wells in an unconfined Cretaceous chalk aquifer (Dassargues et al., 2021). The study was particularly innovative in that it examined the feasibility and efficiency of the project and its potential impacts (e.g., on the groundwater levels of private wells).

John Snow probably wrote the first study of groundwater contamination in 1854 during the second cholera epidemic in London. Through a map showing the cluster of deaths and the drinking water wells, Snow proved that the deaths were concentrated near a water pump and that this was, therefore, probably the source. Shutting off this well led to a quick end to the epidemic (Price, 2004).

In the late 19th to early 20th century, other European countries simultaneously followed the first regular groundwater level measurements at several wells (Blum et al., 2013). Muzikar (2013) describes that from 1899 onwards, quantitative and qualitative monitoring (both chemical and microbiological analysis) was carried out at selected wells in the Czech Republic. This would make it the oldest qualitative monitoring network. In most European countries, systematic groundwater quality monitoring was not undertaken before the 1970s to 1980s (Jousma and Roelofsen, 2004). In the USA, groundwater levels were systematically recorded beginning in 1885 as part of Kansas cooperative water resources study. In 1903, the preliminary division of the Office of Ground Water was established to collect well and spring data. The network of groundwater monitoring wells was expanded nationwide by the end of 1960 (Holmes, 1985). In China, monitoring wells have been operated since the 1970s, especially in large cities, to investigate groundwater pollution and seawater intrusion. The network was extended to other cities and some rural regions in the following years (Zaisheng and Mengxiong, 2013). In many developing countries, the first steps in monitoring were taken through international groundwater assessment projects. Unfortunately, many of these projects have resulted in only short-term records discontinued shortly after the project ended (Jousma and Roelofsen, 2004).

1.2.2 Objectives and Types of Groundwater Monitoring Networks

The factors that determine the design of a monitoring network are manifold. These are primarily the monitoring objectives and parameters, the operator (usually water supplier or public authority) and its financial capabilities, the scale of the monitoring network (local, regional, national to international), the monitoring regions (concerning the difference in climate conditions, land use, infrastructure, etc.) and the complexity of the hydrogeological system. A categorization of groundwater monitoring networks

can be made under numerous aspects. Except for a few clearly defined terms (e.g., GQMN/GLMN), there is no unified terminology describing and classifying the different monitoring strategies. For reasons of uniformity, the terminology used in the following section is aligned with that of Jousma and Roelofsen (2004).

Groundwater monitoring networks are divided according to their main tasks into groundwater level monitoring network for the monitoring of groundwater level and its temporal change, and groundwater quality monitoring network, for the monitoring of physicochemical parameters (e.g. temperature, pH, and electrical conductivity), natural water constituents and pollutants.

Groundwater monitoring network operator groups are usually either public authorities or water suppliers. Their GMNs differ in that the GMNs of water suppliers are designed more to pursue **operational objectives** in terms of water management, while the GMNs of authorities are designed to pursue **technical objectives** of oversight. This includes, for instance, ambient monitoring, where background data on water quality and quantity is collected in specific aquifers to detect and assess changes. However, there is no sharp line, so authorities can also operate operational GMNs and vice versa.

Monitoring networks vary further in terms of their spatial coverage. **Primary groundwater monitoring networks** (also referred to as background or reference networks) are large-scale monitoring networks that usually cover aquifers of a large regional extent. **Secondary groundwater monitoring networks**, on the other hand, are local and designed for specific purposes.

Whereas GMNs for management objectives usually focuses on observing areas of special interest with secondary networks, e.g., catchment areas of extraction wells, GMNs for the technical objective of a holistic description of the groundwater system requires a large-scale primary network of distributed monitoring wells covering the groundwater system under study.

Groundwater Level Monitoring

Operational objectives of GLMNs may include developing groundwater resources for drinking water supply, supplying agricultural production, ensuring protection and preservation of conservation areas, including wetland restoration, documenting compensation claims, and monitoring for ground subsidence.

Technical objectives of GLMN are the investigation of characteristics and parameters of the aquifer e.g., determination of the horizontal and vertical flow directions, aquifer delineation, assessment of the dynamics of the groundwater system and possible influ-

ence of climatic changes, the quantification of the impact of groundwater abstraction or surface water management activities.

While water suppliers usually solely operate secondary monitoring networks in the catchment area of wells, authorities usually maintain primary and secondary networks. Monitoring wells are selected to be relatively evenly distributed over the area to be monitored, thus providing an unbiased overall survey. Primary GLMN coverage ranges from small valley aquifers to international coverage. For special purposes (e.g. development of a groundwater level contour map) primary and secondary networks are often combined into a wide-spaced regional network with denser coverage in areas of special interest.

In the past decades, automatic data acquisition with groundwater level probes and real-time data transmission has steadily increased or has become the standard, especially in industrialized countries. In many parts of the world, however, manual measurement is still a common practice (IGRAC, 2020).

Groundwater Quality Monitoring

Operational objectives of groundwater quality monitoring networks (GQMNs) may include the spatial collection of reference background values, detection and monitoring of diffuse contamination from agriculture (pesticides, fertilizers, herbicides), industry, and point contamination from industry, landfills, etc., as well as the assessment of the effect of remedial actions. Categorization can be made between hazard-based monitoring, where known contaminants and substance releases are monitored, and flow-based monitoring, where unknown contaminants in the flow of extraction wells are monitored. The standard parameters measured are mostly TDS, EC, and major constituents. In addition, depending on the hazard potential on site, there is investigation-specific sampling for possible pollutants from agriculture (fertilizers, herbicides, pesticides), industry and domestic sewage (e.g., heavy metals, bacteria, pharmaceuticals, volatile organic compounds), or pollutants of geogenic origin (e,g, arsenic, boron, etc.).

Technical objectives of GQMN are used to characterize the groundwater quality in the observed region. Analogous to the GLMNs, the objective of primary GQMNs is to provide an overall view of the investigated area and, on this basis, to monitor general GWQ and trends. Secondary GLMNs are used locally for operational objectives.

1.2.3 Groundwater Monitoring in the International Context

The following section is intended to provide a brief overview of the status of monitoring networks on an international level. The aim is not to give a comprehensive overview but rather to show that there are no uniform rules, objective standards, and a high degree of subjectivity in developing the networks at this level.

Monitoring Networks at the National Level

As described, monitoring of groundwater levels has a long tradition in Europe. Groundwater quality monitoring, on the other hand, has been systematically undertaken in most European countries only since the 1970s and 1980s. A 1996 European Environment Agency (EEA) survey of all member countries on their current GLMNs/ and GQMNs revealed that the groundwater quality monitoring objectives of individual countries, particularly the spatial and temporal coverage of monitoring wells and measurements, varied significantly and reflect the broad spectrum of national regulations, institutional frameworks, and economic environments (Koreimann et al., 1996). Thus, GQMNs have developed based on national objectives and demands and the respective hydrogeological settings. Monitoring is commonly undertaken by water suppliers or public authorities to assess GWQ in relation to drinking water standards, industrial process requirements and irrigation needs. The GQMNs density of the individual EU countries varied at the time of the survey in 1996 between 0.003 wells/km^2 and 0.57 wells/km^2. The sampling frequency varied from weekly to perennial, depending on the parameters studied, the country, and the survey objective. The majority of respondents indicated that GLMNs are primarily used to collect primary groundwater data, record temporal and spatial changes, and manage groundwater resources and water supplies. While most GQMNs are located primarily near or directly sampled from extraction wells, most GLMNs are widely distributed across groundwater basins and aquifer types. In almost all countries, a continuous recording of the GWL is undertaken. The frequency of recording usually varies from weekly to twice a year. The monitoring network density of GLMNs varied from 0.004 wells/km^2 to 7.3 wells/km^2 in 1996 (Koreimann et al., 1996).

The following countries as exemplary cases illustrate general disparities in national monitoring networks. In **Germany**, the Federal Mnistry of the Enviroment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV) is responsible for regulating water resources through federal legislation. GMNs are the competence of the federal states. Thus, each federal state operates its own quantitative and qualitative monitoring network. In addition, there are numerous small-scale monitoring networks of local

water suppliers, primarily in groundwater protection zones and well catchments. For instance, the State Office for the Environment (LUBW) operates a network for monitoring GWQ on behalf of the state of Baden-Württemberg with 1,866 observation wells (status 2020), subdivided into subnetworks depending on anthropogenic influence and groundwater utilization in the catchment areas. Sampling of all monitoring wells takes place every four years. This full investigation includes geogenic and anthropogenic parameters and substances. 1,450 monitoring wells in- and outside water protection areas are monitored for nitrogen parameters at annual intervals. This analysis is carried out semi-annually at 230 monitoring wells in water protection areas. It is carried out quarterly at around 70 monitoring wells subject to special protection regulations (problem and remediation areas). The LUBW operates a GLMN with 412 monitoring wells (status 2020), subdivided into three groups according to the type of the well and monitoring interval. In cooperation with water suppliers, there is also a cooperative monitoring network with about 1,300 to 1,900 monitoring wells (status 2018/2019) (LUBW, 2021a). **France's** groundwater quality monitoring network consists of about 1,400 monitoring wells. The GMN is managed as a collaboration of more than 200 regional and local organizations. As many of these are local water suppliers, most monitoring wells are located near groundwater abstraction wells. The sampling procedures are not standardized and vary in sampling intervals and monitored parameters. The French GLMN operates on a national scale and covers all groundwater bodies of national importance. It consists of approximately 1,500 monitoring wells that monitor weekly in unconfined aquifers and monthly in confined aquifers. In **Italy**, groundwater monitoring is not centralized. It consists of a loose group of regional to local groundwater monitoring programs, operated by Local Health Units and Local Water Supply Services. Data are collected in so many ways that comparison does not seem helpful (IGRAC, 2020). The United States Geological Survey (USGS) is responsible for collecting and providing groundwater data in the **USA**. It operates several national monitoring networks (National Groundwater Monitoring Network (NGWMN), Active Groundwater Level Network, and Climate Response Network) and regional monitoring networks (High Plains Aquifer Monitoring Program; state-based and local networks). NGWMN comprises about 16,500 GWL and 4,000 GWQ monitoring wells selected from 10 federal, state, and local subnetworks across the nation. The monitoring frequency depends on the monitoring category, aquifer type, and groundwater extraction. It ranges from daily to quarterly for trend monitoring and quarterly to triennial for surveillance monitoring. Since 2020, **China's** Ministry of Natural Resources has operated nationwide (previously, only around eleven percent of the nation was monitored). The network operates approximately 20,500 GWL monitoring wells covering an area of 3.5 million km². It monitors groundwater temperatures, GWL, and GWQ parameters in major plains, basins, karst, and ecologically vulnerable

areas (Liu and Zheng, 2016). The Federal Office for the Environment operates a national groundwater monitoring network (NAQUA) with about 600 monitoring sites for **Switzerland**. These meet several selection criteria including adequate coverage of the spectrum of geological and hydrogeological conditions, altitudes, and land uses within the country. The distribution of monitoring sites is concentrated toward population centers, intensive agriculture, and industry (BAFU, 2019; Chilton, 2009).

Multinational Networks

The EU Water Framework Directive

The Water Framework Directive (WFD) (EC, 2000) is a legislative framework developed by the European Union to protect European waters. The Directive requires member states to establish GMN to assess the chemical and quantitative status of groundwater at the scale of groundwater bodies. The Groundwater Directive (EC, 2006) is the respective daughter directive. It defines quality standards for evaluating the chemical situation and establishes a framework for setting threshold values for nitrates and pesticides. It is linked to other regulations, such as the Pesticide Regulation and the Nitrates Directive.

The GMNs are composed of three subnetworks: surveillance monitoring, operational monitoring, and quantity monitoring. The surveillance monitoring program aims to obtain background trends and baseline information on groundwater conditions. The results are used to establish an operational monitoring program in a 6-year cycle. The measured core determinants are dissolved oxygen, pH, electrical conductivity, nitrate, and ammonium. In addition, selected determinants (temperature and several major and trace ions) are recommended. The development of monitoring network designs is based on a site-by-site basis. In principle, however, regional conceptual models of the groundwater body, its spatial distribution, land-use distribution, and risk assessments should be considered in selecting sites. The monitoring networks must provide a coherent and representative overview of the chemical status of the groundwater body. Monitoring frequency varies from semi-annual to quarterly, depending on aquifer type and expected groundwater system dynamics. Depending on the results, the aquifer is classified as "at risk" or "not at risk" for the next 6-year management cycle (EC, 2007). The operational monitoring is conducted for those groundwater bodies identified as "at-risk" of not achieving the The Water Framework Directive good status objective. In most cases, both the core and selected determinants are required at each monitoring site. The design of the GMN and the sampling frequency depend on the particular pollution situation and should capture the specific causes. Sampling is carried out at least once a year. The quantitative monitoring network is designed to supplement the risk assessment proce-

dure of the quantitative status of groundwater in groundwater bodies. The quantitative status of groundwater is considered "good" if the long-term average annual abstraction does not exceed the available groundwater resource. The selection of monitoring sites is based on conceptual models of the water bodies to be observed. Measurement density is based on the previous assessment of groundwater body status abstraction rates. It can be reduced at low risk or must be increased in density at high risk. However, the frequency of the monitoring also depends on the risk assessment and the annual fluctuation within the water body. In Germany, for instance (as of 2017), 4,892 surveillance monitoring sites, 2,273 operational monitoring sites, and nearly 6,000 monitoring sites for quantitative status are operated by the federal states. In practice, member states construct few if any new wells. Instead, they select a subset of existing wells that they report for the WFD network. As a result, the differences in the national networks trace through to the WFD network.

The EU Nitrates Directive

To protect groundwater in regions with intensive agricultural use, the European Commision adopted the Nitrates Directive (EC, 1991) in 1991. The Nitrates Directive requires the Member States to develop and implement appropriate monitoring programs to evaluate control programs' effectiveness and identify groundwaters with nitrate levels greater than 50 mg/l. The directive specifies basic principles and guidelines. Thus, each state must operate a general monitoring network in all aquifers of national importance. Article 5 requires that a connection between the nitrate contamination of the waters and agricultural use be established in selecting the monitoring sites. However, aspects such as sampling site density, site selection, and sampling frequency are the responsibility of the member states. The way sampling sites are chosen can strongly influence GWQ, making a direct comparison of groundwater conditions in countries complicated. Several countries (for example, Sweden, Hungary, Belgium, Lithuania, Ireland, Bulgaria) are dominated by sampling sites used for drinking water supply and have reasonably good GWQ (Lindinger and Scheidleder, 2004). In 2015, the German nitrate monitoring network was redesigned and expanded from 198 monitoring sites to 697. In contrast to the countries mentioned above, these monitoring wells are located in or downstream of agricultural areas (BMU, 2016).

1.3 Groundwater Monitoring Network Optimization

Optimization of a GMN attempts to identify the temporal and/or spatial redundancy or inadequacy of groundwater monitoring in an existing network in order to improve the

cost-effectiveness while meeting regulatory and other requirements or adding new wells in locations where data are insufficient for a cost-effective management plan (Reed et al., 2000). In recent studies, two main approaches have emerged to address this optimization problem: mathematical and geostatistical approaches. Geostatistical methods typically use local or global kriging variance to identify spatially or temporally over- and under-represented sites. Mathematical approaches employ different optimization algorithms, from random search to heuristic algorithms, to maximize or minimize a given objective function subject to constraints. Studies often combine these two approaches. A more detailed overview of previous studies can be found in chapters 3.1 and 4.1.

Pareto optimality, named after the economist and sociologist Vilfredo Pareto (1848-1923), has its origins in the economic equilibrium and welfare theories of the 19th century. According to this theory, a social situation is economically efficient or Pareto-optimal if it is impossible to make one person better-off without making someone else worse-off. The concept of Pareto optimality was increasingly adapted and further developed from the 1970s and 1980s in countless scientific fields (Chinchuluun et al., 2008). A Pareto-efficient groundwater monitoring network provides maximum information about the aquifer for a given cost budget. All Pareto-optimal solutions form the Pareto frontier. Additional information gain can only be accomplished, for example, by adding monitoring wells and thus increasing costs (Fig. 1.1). Conversely, in a Pareto-inefficient GMN, information gain can be improved while costs remain unchanged, or expenses can be decreased without loss of information. Identifying monitoring wells containing redundant information allows these monitoring wells to be omitted from the subsequent monitoring campaigns. The monitoring expenses are therefore decreased without any significant loss of information. Alternatively, the capacity thus gained (e.g., GWL logger or qualitative sampling) can be used on a more representative well that has not been considered so far. This increases the information gain at a constant monitoring cost. The principal objective of a GMN optimization is to find the best balance of required system information at a reasonable cost. For an initial groundwater monitoring network (consisting of very few monitoring wells), each additional well provides valuable information. The information gain per new well (information per cost unit) decreases with increasing numbers of wells in the network. A balanced, Pareto-optimal monitoring network is located on the sweet spot of this Pareto frontier. It is complex enough to detect relevant dynamics and provides only slightly less information than much more complex GMNs.

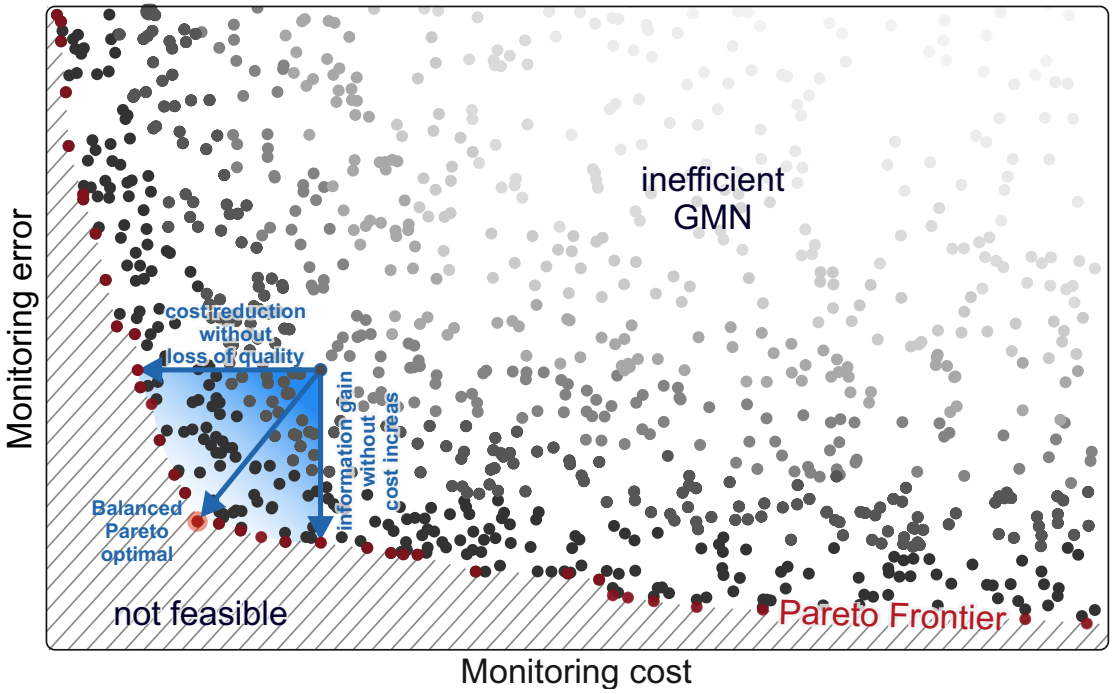


Figure 1.1: Concept of Pareto optimality for groundwater monitoring networks.

1.4 Regionalization of Groundwater Data

Spatially continuous data (or spatially continuous surfaces) are relevant for groundwater management planning, risk assessment, and decision making. Since, as mentioned above, the attributes of groundwater (level, chemical composition, etc.) are generally unknown except at monitored locations, they must be estimated at unmonitored sites, so regionalization from point to continuous spatial data is required. The most commonly used technique for regionalization is spatial interpolation, but other options exist. The following section discusses regionalization using interpolation of groundwater contour maps as an example. However, the techniques described are also applicable to regionalization of GWQ parameters.

Groundwater contour maps (unconfined aqui-fers) or piezometric surfaces (confined aquifers) are essential instruments in hydrogeologic investigations. In addition to determining groundwater flow directions, groundwater contour maps combined with additional data enable the analysis and estimation of hydraulic gradients, flow velocities, flow rates, particle travel time, hydraulic conductivity, transmissivity, groundwater quantification, and impact of groundwater extraction, and the interaction with surface waters (Krešic, 2007). Apart from the interpolation procedure itself, the results primarily depend on spatial and temporal data quality (the respective system under consideration).

Since the assumptions made by interpolation are usually applied as a foundation for further calculations, errors are propagated to all subsequent analyses. Groundwater contour maps and flow networks are commonly produced using one of the following methods::

1. Manually constructed based on the triangular linear interpretation of neighboring monitoring wells (Delaunay triangulation). Although the method is outdated, it allows expert knowledge to be incorporated. Since this is a very time-consuming method, it applies only to specific applications.

2. Computed using deterministic, geostatistical, or combined interpolation methods.

3. Computed through a numerical or analytical groundwater flow model.

Spatial interpolation and/or spatial prediction is a process of estimating values of the target variable over the entire area of interest using some input training point data, an algorithm, and the values of the covariates at new locations (Mitas and Mitasova, 2011). Numerous spatial interpolation methods have been developed and applied for various disciplines. General statements regarding selecting an appropriate interpolation method are somewhat limited because many factors, such as sample size, sampling design, and the variability and nature of the data influence the spatial interpolator. There are no consistent findings on how these factors affect the performance of spatial interpolators (Li and Heap, 2008).

Spatial interpolation methods can be separated into deterministic, geostatistical, and combined methods. Deterministic methods use empirical model parameters. These methods do not allow estimation of the uncertainty of the model, and there are usually no strict assumptions about the variability of the attribute. For groundwater contouring, commonly applied deterministic methods are, e.g., Thiessen's polygons, inverse distance weighting, or spline interpolation.

Although other techniques exist, kriging has been a synonym for geostatistical methods. It is one of the most robust and widely used methods for interpolation and contouring in different fields (Krešic, 2007). Kriging is, in fact, a collective term for a variety of interpolators, all based on the same idea and basic equation. It was named by G. Matheron after D. G. Krige, a South African mining engineer who first formulated the basic idea in 1951. Matheron took these ideas and developed them into his theory of regionalized variables.

Kriging assumes that the parameter being investigated is subject to a stationary random function described as $Z(\mathbf{x}) = \mu + \varepsilon(\mathbf{x})$. This means that the value Z at the location $\mathbf{x}$ is the mean of the process plus random component resulting from a distribution with zero mean and covariance function. The basic idea is that the sample points close to x_0

should get a larger weighting to improve the prediction. For this purpose, the technique relies on a certain spatial structure of the data, modeled by the second-order properties, i.e., variogram or covariance, of the underlying random function $Z(\mathbf{x})$. The goal is to determine the weights in the linear estimator (taking into account the structure) such that the final estimate is unbiased and has minimal error variance among all linear unbiased predictors. The resulting weights depend on the assumptions about the mean value $\mu(\mathbf{x})$ and the variogram or covariance function of $Z(\mathbf{x})$.

The variogram concept is founded on the premise that the spatial correlation between two sampling points does not depend on their absolute geographic location but only on their relative location (Wackernagel, 1995). For this purpose, the squared differences of the measured points are plotted in pairs based on their relative distances or in defined lags. The squared differences of the pairs are summed and divided by the number of points. This gives the semivariance (half variance), which is plotted in a two-dimensional graph as a function of distance from the reference point. In the ideal case, this results in processes that mathematical functions can approximate. A semivariogram approaches a limit (sill) in general (exception: linear variogram). The distance to the point where the sill is attained is called range. If $y(\mathbf{x} = 0) > 0$, then $\mathbf{y}$ is called a nugget, a measure of noise. Calculating the standard deviation of the manipulated values estimates the interpolation error. Such error estimates can be used as plausible parameter values for model calibration and forecast uncertainty analysis (Anderson et al., 2015).

Since collecting additional (independent) measurements is often impractical, validation of predictive models is usually done by CV, i.e., by subsetting the original set of points into two datasets - calibration and validation without the subset - and then repeating the analysis. The most common types of CV methods are k-fold CV (splitting into k equal parts), leave-one-out (LOO) CV (each measurement is used for CV), and jackknifing (similar to LOO, but aims to estimate the bias of the statistical analysis rather than the predictions). It should be noted that CV is not necessarily independent because the items used for CV are a subset of the original measurement. For example, if they are biased or unrepresentative, the CV will not show the actual accuracy of the CV.

1.5 Outline

The focus of this cumulative thesis is, first, on the investigation of techniques to improve the design of groundwater level monitoring networks in terms of a reasonable balance of monitoring quality and monitoring costs in a temporal and spatial context and, second, on the investigation of techniques for the spatial prediction of GWL based

on observations at these networks. This was accomplished by three studies, submitted to peer-reviewed (ISI-listed) journals. Two of them have already been published, and one is currently under review. In the process of the studies, the following research questions were formulated and addressed in chapters 2, 3, and 4. The synthesis in Chapter 5 establishes a connection between the results and findings of the presented studies.

RQ 1: Which interpolation technique provides the best results for the studied alluvial and karst aquifer?

RQ 2: How do the different interpolation methods deal with the

i spatially inhomogeneous distribution/patterns of the existing groundwater monitoring network?

ii different surface types change from a pronounced hilly topography to a flat riverine landscape?

ii different hydraulic pressure conditions within the aquifer which fluctuate from unconfined to artesian?

RQ 3: What are the possible influences of the chosen methods on further computations, namely the calculation of the estimated vertical groundwater exchange between different aquifer systems?

RQ 4: Which are the most suitable error statistics to compare the performance of the methods?

RQ 5: How can the results be validated with additional eco-hydrogeological data? This includes, e.g., a comparison between calculated groundwater depth and geographic locations of karst springs and wetlands and surface waters and a comparison of computed flow accumulation with the areas of receiving waters.

RQ 6: Is there an extensible and transferable GLMN design that allows reliable spatial estimates of GWL with a minimum number of monitoring wells?

RQ 7: What are the quality differences resulting from the use of various GLMN design approaches?

RQ 8: At what monitoring well density does a reasonable information/cost ratio result?

RQ 9: Which is the most suitable CV error statistic (MAE, RMSE, RMSSE, ASE, or NSE) to evaluate the quality of interpolated groundwater surfaces?

RQ 10: What is the ranking of monitoring wells in an existing network in terms of their information content/reconstruction performance, i.e., in which order should the wells be removed if a network reduction is desired?

RQ 11: How does a reconstruction/interpolation error develop when a given number of monitoring wells are reduced? How does the error of reducing wells according to information content compare to a random reduction?

RQ 12: When the goal is network extension, where should new wells be placed for maximum information gain? How much is the increase in information, i.e., how much will the spatial reconstruction error be reduced?

RQ 13: How well does a combined reduction/extension (i.e., replacement) of a certain number of wells perform compared to a straightforward extension?

The objective of the study, presented in Chapter 2, was to investigate uncertainties resulting from the choice of the interpolation techniques for predicting a continuous GWL from point measurements at monitoring wells and the resulting error propagation in downstream calculations, e.g., aquifer exchange processes (RQ 1, RQ 2, RQ 3). For this purpose, nine frequently used deterministic and geostatistical interpolation techniques were comprehensively compared. The evaluation was performed using leave-one-out CV with six commonly applied error metrics (RQ 4) as well as detailed plausibility checks at selected locations, e.g., the groundwater depths were calculated in wetlands, along cross-sections in valley regions, and the course of the calculated major groundwater streamlines was compared with the locations of large karst springs (RQ 5). In the co-kriging method, additional variables associated with the GWL (e.g., elevation-, river-level data) were added to improve the prediction accuracy.

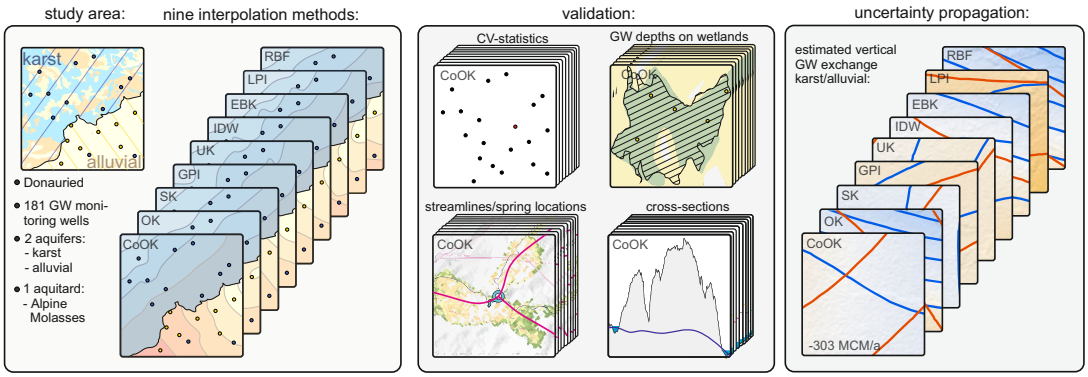


Figure 1.2: Graphical abstract of the study presented in Chapter 2

The region covered by the GLMN is characterized by a complex interaction of the regional highly productive aquifers of the Upper Jurassic karst and the Quaternary alluvial aquifer in the Danube valley with several wetlands, the Danube River, and the low permeability Molasse Formation controlling the aquifer exchange. In short, a wide range of hydrogeologic conditions is present here on a small scale. Although several regional water suppliers actively use the local groundwater, little is known about the exchange processes of the aquifers due to the hydrogeological complexity. The resulting question of the necessary complexity of a monitoring network to adequately represent the spatial GWL was the motivation for the study presented in Chapter 3.

The study presented in Chapter 3 addresses the requirements for an optimal GLMN to accurately predict spatial GWL with a reasonable information/cost ratio. For this purpose, a total of 6 monitoring design approaches were comprehensively evaluated in terms of monitoring density and spatial dynamics of the hydrogeological system (RQ 6). These approaches consist of random sampling (which corresponds most to the spatial distribution of most real existing measurement networks), two spatial coverage sampling approaches, one geostatistical approach, and two regular grid sampling approaches. Since the spatial and temporal variability of GWL beyond points such as wells and springs is generally unknown, the interpolation of spatial GWL, as well as the evaluation of the interpolation itself, must be based on information available at these points.

As the results of the first study indicated that the performance of CV assessment depends on the number and distribution of monitoring wells and the variation in data, the study in Chapter 3 instead used nine generic potentiometric groundwater surfaces extracted from three large-scale numerical groundwater models as an "a priori" reference. The model surfaces were chosen to reflect a wide range of hydrogeologic system characteristics at different scales so that the overall performance of the monitoring approach studied can be assessed. From an initial monitoring network of 10 wells, the network was extended stepwise to up to 500 wells for each method investigated. For each extension step, the GWL was interpolated, and the interpolation accuracy was calculated with the

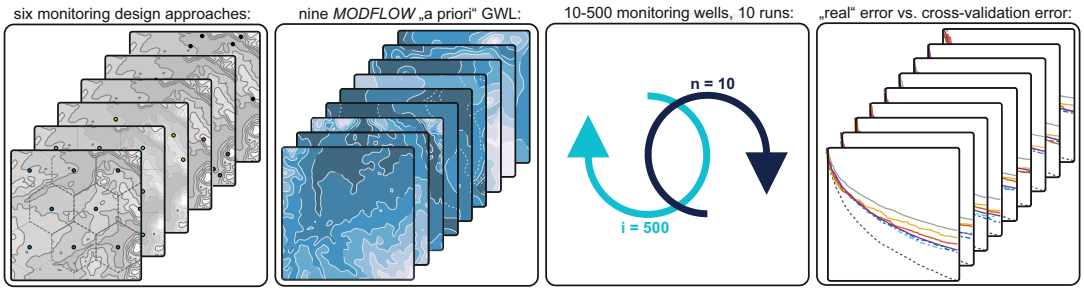


Figure 1.3: Graphical abstract of the study presented in Chapter 3

"real" groundwater surface and global CV. This procedure was repeated ten times for design strategies with random components, and the results were averaged. The use of the "a priori" GWL, assuming that it reflects actual groundwater conditions, allows the investigation of the development of the prediction accuracy as a function of the measurement density (RQ 7, RQ 8). In addition, conclusions can be made about the suitability of global CV results to compare monitoring network designs (RQ 9).

The study presented in Chapter 4 exploits data-driven learning to optimize an existing GLMN in the Upper Rhine Graben (URG). The applied algorithms, which are applied here for the first time to groundwater monitoring wells, identify dominant low-dimensional structures (e.g., principle component analysis (PCA)) from high-dimensional signals (e.g., hydrographs, gridded data) that enable low-loss reconstruction with a drastically reduced subset. The input signals are ranked according to their information content in this process. Both well records (i.e., hydrographs, 1D case) and spatially continuous GWL as gridded data (i.e. interpolation, numerical modelling, 2D case) were used as input data. The regionalization was performed using the automated ordinary kriging (OK) procedure, which was already used in the study in Chapter 3 and proved to be appropriate in it. For the 1D case, this results in a ranking of monitoring wells according to their importance to reconstruct the aquifer signal (RQ 10). This ranking thus allows the removal of unimportant wells with less or redundant information according to Chapter 1.3 (RQ 11) or to select a reasonable sub-network from the existing monitoring network. For the 2D case, the ranking refers to any pixel on the GWL contour maps as a potential monitoring site. The algorithm was controlled by a non-uniform spatial step function, set to zero at existing wells and to a large value at locations distant from infrastructure or on steep terrain. In addition to reducing the monitoring network, this also allows an extension or combination of both based on the spatially estimated groundwater dynamics (RQ 12, RQ 13).

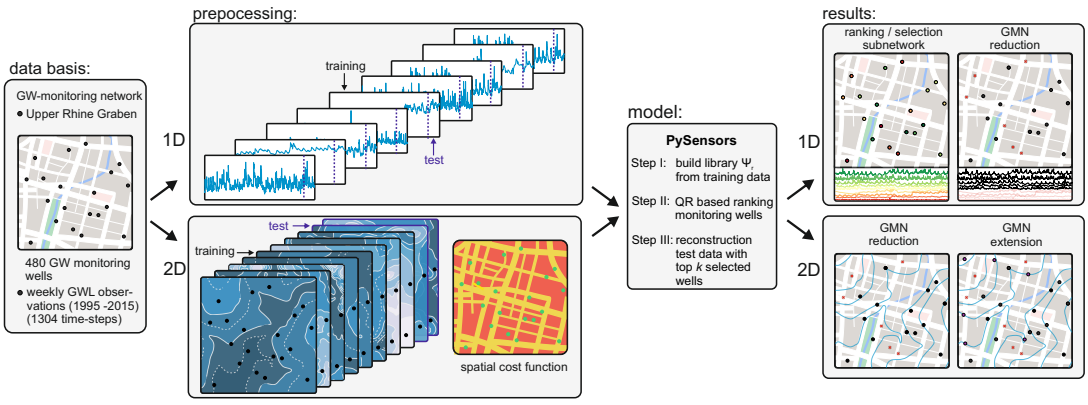


Figure 1.4: Graphical abstract of the study presented in Chapter 4

Chapter 2

On the Optimal Selection of Interpolation Methods for Groundwater Level Contouring

An example of propagation of uncertainty regarding inter-aquifer exchange

This chapter is based on a study published in the Elsevier journal *Advances in water resources*. The remainder of this part is an edited reprint of:

Ohmer, M., Liesch, T., Goeppert, N., Goldscheider, N., (2017). On the optimal selection of interpolation methods for groundwater level contouring: An example of propagation of uncertainty regarding inter-aquifer exchange. *Advances in Water Resources Volume 109*, 121-132, doi: 10.1016/j.advwatres.2017.08.016.

Highlights

- Nine interpolation methods for a Jurassic karst and a Quaternary alluvial aquifer are compared.
- Calculated inter-aquifer exchange rates vary greatly depending on the chosen method.
- Plausibility was additionally validated with eco-hydrogeological data (e.g., wetlands).
- Best results achieved with co-kriging incorporating additional data, e.g., topography.

Abstract

The selection of the best possible method to interpolate a continuous groundwater level (GWL) from point data of monitoring wells is a controversial issue. In the present study, four deterministic and five geostatistical interpolation methods: global polynomial interpolation (GPI), local polynomial interpolation (LPI), inverse distance weighting (IDW), radial basis function (RBF), simple kriging (SK), ordinary kriging (OK), universal kriging (UK), empirical Bayesian kriging (EBK), and co-ordinary kriging (CoOK)) and seven error statistics: mean error (ME), mean absolute error (MAE), mean square error (MSE), mean absolute percentage error (MAPE), root mean square error (RMSE), root mean square standardized error (RMSSE), Pearson r were examined for a Jurassic karst aquifer and a Quaternary alluvial aquifer. We investigated the possible propagation of uncertainty of the chosen interpolation method to calculate the estimated vertical groundwater exchange between the aquifers. Furthermore, we validated the results with eco-hydrogeological data, including comparing calculated groundwater depth and geographic locations of karst springs, wetlands, and surface waters. These results show that calculated inter-aquifer exchange rates based on different interpolations of groundwater potentials may vary greatly depending on the chosen interpolation method (by factor > 10). Therefore, the choice of an interpolation method should be made with care, taking different error measures and additional data for plausibility control into account. The most accurate results have been obtained with co-kriging incorporating secondary data (e.g., topography, river levels).

2.1 Introduction

Reliable GWL contour maps provide insight into manifold hydrogeological questions, e.g., determination of regional hydraulic gradients, flow directions, groundwater depth, flow velocities, recharge and discharge zones, hydraulic conductivities, aquifer susceptibility, and catchment sizes to delineate protection areas. The comparison of GWL contour maps of different points in time provides information about the temporal change and, therefore, the recharge and discharge of the area of interest. Differences in the hydraulic potential of two or more aquifers in the same area, separated by an aquitard, allow conclusions about possible vertical groundwater exchange.

With conventional methods, the GWL can only be measured at distinct observation points, such as monitoring wells, springs, and perennial surface water. By applying geostatistical and deterministic interpolation methods, the GWL can also be estimated

between those observation points. However, due to economic considerations, the number and spatial and temporal distribution of the hydraulic head observations are often insufficient to reliably represent the GWL (Delbari, 2014; Varouchakis and Hristopulos, 2013). In hilly terrain, the estimation of the GWL is often problematic because the data set is always sparse in relation to the topographic relief, and monitoring wells are almost exclusively located in valley regions, while the GWL is usually a subdued replica of the ground surface elevation (Hoeksema et al., 1989).

Spatial interpolation methods including geostatistic techniques, have been applied to various disciplines. A broad overview of comparative studies of interpolation methods in environmental science can be found in Li and Heap (2008). Zimmerman et al. (1999) compared the spatial interpolation accuracy of OK and UK and IDW as well as the influence of surface types, sampling patterns, noise level, and strength of small-scale spatial correlation on those methods by creating mathematical surfaces. They pointed out that the kriging methods outperformed the deterministic methods over all levels and factors. During past decades, different types of univariate kriging methods, e.g., OK and SK, have been used to interpolate the GWL (e.g., Ahmadi and Sedghamiz, 2007; Guekie et al., 2016; Möhler et al., 2014; Sadat Noori et al., 2013) and were compared with one to several deterministic methods, e.g., IDW and RBF (Arslan, 2014; Chung and Rogers, 2012; Cooper et al., 2015; Delbari, 2014; Hua et al., 2009; Sun et al., 2009; Varouchakis and Hristopulos, 2013; Xiao et al., 2016; Yao et al., 2014).

In addition, some studies used multivariate kriging methods like UK and co-kriging (CoK) to incorporate the influence of the topography on the interpolated GWL (Ahmadi and Sedghamiz, 2007; Arslan, 2014; Chung and Rogers, 2012; Cooper et al., 2015; Delbari, 2014; Guekie et al., 2016; Hoeksema et al., 1989; Möhler et al., 2014; Sadat Noori et al., 2013; Sun et al., 2009; Varouchakis and Hristopulos, 2013; Xiao et al., 2016; Yao et al., 2014). An overview of the methods and evaluation statistics used in those studies can be found in Tab. 2.1 and 2.2.

The majority of the previous studies compared only some of the methods and used selected error statistics for evaluation, and often did not investigate possible consequences of the chosen methods on further calculations or conclusions. In this study, almost all current deterministic and geostatistical interpolation methods, namely GPI, LPI, IDW, RBF, SK, OK, UK, EBK and CoOK, and error statistics, namely ME, MAE, MSE, MAPE, RMSE, RMSSE, and Pearson correlation coefficient (r) are systematically examined to answer the following research questions:

- Which interpolation technique provides the best results for the studied alluvial and karst aquifer?

- How do the different methods deal with the
  (i) spatially inhomogeneous distribution/patterns of the existing groundwater monitoring network (GMN)?
  (ii) different surface types which change from a pronounced hilly topography to a flat riverine landscape?
  (iii) different hydraulic pressure conditions within the aquifer which fluctuate from unconfined to artesian?
- Which are the most suitable error statistics to compare the performance of the methods?
- How can the results be validated with additional eco-hydrogeological data? This includes, e.g., a comparison between calculated groundwater depth and geographic locations of karst springs and wetlands and surface waters and a comparison of computed flow accumulation with the locations of receiving waters.
- What are the possible influences of the chosen methods on further computations, namely the calculation of the estimated vertical groundwater exchange between different aquifer systems?

Table 2.1: Methods for interpolating groundwater contour lines applied in comparative studies. GPI: global polynomial interpolation; LPI: local polynomial interpolation; IDW: inverse distance weighting; BF: radial basis function; OK: ordinary kriging; EBK: empirical Bayesian kriging; SK:simple kriging; UK: universal kriging; CoK: co-kriging; MLR: multiple linear regression; GWR: geographic weighted regression; MC: minimum curvature; DeK: Delaunay triangulation; ANN: artificial neural networks; ✓: applied method; ✓ best method.

Authors	GPI	LPI	IDW	RBF	OK	EBK	SK	UK	CoOK	MLR	GWR	MC	DeK	ANN
Ahmadi and Sedghamiz (2007)					✓			✓						
Arslan (2014)			✓	✓	✓				✓					
Chung and Rogers (2012)					✓				✓	✓				
Cooper et al. (2015)	✓		✓		✓	✓				✓	✓			
Delbari (2014)			✓		✓			✓	✓					
Guekie et al. (2016)					✓			✓	✓					
Hua et al. (2009)			✓		✓									
Möhler et al. (2014)					✓			✓						
Sadat Noori et al. (2013)			✓	✓	✓				✓					
Sun et al. (2009)					✓		✓	✓						
Tapoglou et al. (2014)					✓									✓
Varouchakis and Hristopulos (2013)			✓		✓			✓				✓	✓	
Xiao et al. (2016)	✓	✓	✓	✓	✓		✓	✓						
Yao et al. (2014)	✓	✓	✓	✓	✓		✓	✓						

Table 2.2: Evaluation statistics used in comparative studies. CV: cross-validation; OV: orthogonal-validation; ME: mean error; MAE: mean absolute error; MAPE: mean absolute percentage error; r: Pearson correlation coefficient; R^2: coefficient of determination; VSE: variance of standardized error; CVUD: cumulative vertical uncertainty standard deviation; MSE: mean standardized error; RMSSE: root mean square standardized error; τ-Test: Kendall rank correlation; 95 PPI: 95 percent prediction interval; ✓: applied statistics.

Authors	CV	OV	ME	MAE	MAPE	RMSE	R	R2	VSE	CVUD	MSE	RMSSE	τ-Test	95 PPI
Ahmadi and Sedghamiz (2007)	✓							✓						
Arslan (2014)	✓			✓		✓		✓						
Chung and Rogers (2012)	✓		✓			✓					✓	✓		
Cooper et al. (2015)	✓					✓		✓		✓				
Delbari (2014)	✓		✓	✓	✓	✓								
Guekie et al. (2016)	✓		✓	✓			✓		✓					
Hua et al. (2009)	✓		✓			✓								
Möhler et al. (2014)	✓							✓						
Sadat Noori et al. (2013)	✓			✓		✓		✓						
Sun et al. (2009)	✓					✓		✓					✓	✓
Tapoglou et al. (2014)	✓			✓		✓						✓		
Varouchakis and Hristopulos (2013)	✓		✓	✓	✓	✓	✓							
Xiao et al. (2016)	✓		✓			✓		✓						
Yao et al. (2014)	✓	✓				✓		✓						

2.2 Methods

2.2.1 General Overview

We employed nine spatial interpolation methods for this study on a Jurassic karst aquifer and a Quaternary alluvial aquifer. Of these were four deterministic methods (GPI, LPI, IDW, and RBF) and five geostatistical methods (OK, SK, UK, EBK, and CoOK). For the CoOK method, additional surface-elevation data, river levels, and long-term GWL of monitoring wells not observed at the reference date were included (see also 2.2.6). All model parameters were optimized by iterative cross-validation (CV). For clarity, the parameters are given in the supporting information of this Chapter (Fig. S2a).

Spatial interpolation methods can be classified either as global or local methods. Global methods use all available data of the study area for the estimation and show a general trend. In contrast, local methods operate within a smaller area of variable size around the point to be estimated and offer a more local variation. Furthermore, the interpolation methods can be divided into exact and inexact methods. Exact methods generate a prediction equal to the observation at the sample point, while inexact (smooth) methods generate a prediction that usually differs from the observed values at the sample point (Li and Heap, 2008), which may be appropriate for data that include considerable measurement errors compared to the total variations. Deterministic methods use closed-form mathematical formulas or the solution of a linear system of equations to estimate the value at a given location as a weighted sum of data values at surrounding locations, while geostatistical or stochastic methods include an assessment of a statistical spatial

autocorrelation of the data by variography. In contrast to the geostatistical methods, which allow a probabilistic estimate of the interpolation quality, deterministic methods cannot generate measures of uncertainty in addition to the estimations. Almost all interpolation methods share the same formula to estimate the unknown value of $\hat{Z}$ at the point $\mathbf{x}_0$:

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i) \tag{2.1}$$

where $Z(\mathbf{x}_i)$ is the observed value at the data point $\mathbf{x}_i$. The number of the existing data points is represented by n, and λ_i represents the weight function assigned to each data point (Li and Heap, 2008; Oliver et al., 1996).

2.2.2 Deterministic Methods

Global Polynomial Interpolation (GPI)

GPI is a deterministic, global and inexact (smooth) trend surface analysis. It fits a smooth two-dimensional polynomial function of first, second, or higher degree representing the surface through a set number of data points (Cooper et al., 2015). There are a few decisions to make regarding the model parameters. The order of the function sets the shape of the interpolated surface. We achieved the best results for both aquifers with a first-order Polynomial that fits a flat plane through the dataset. The interpolated GWL represents a gradual trend over the area of interest. The method is mainly recommended for GWL that change slowly and gradually. Otherwise, interpolated surfaces are highly susceptible to outliers, especially in the edge region (Johnston et al., 2004).

Local Polynomial Interpolation (LPI)

Unlike GPI, which adjusts a polynomial over the entire area, LPI adjusts several partially overlapping polynomials within defined neighborhoods. It is a moderately quick, inexact, and local, but compared to GPI, more flexible interpolator (Johnston et al., 2004). The neighborhood shape, the minimum and maximum number of neighbouring points to be included, and the sector configuration can be specified. The surface value at the center of the neighborhood is estimated as the predicted value (Wang et al., 2014). It provides a prediction, prediction standard error, and a condition number surface that are comparable to OK measurement errors. As with GPI, the order of the function determines the shape of the interpolated surface. As a constraint, the method depends on the data values being normally distributed within the search neighborhood (Johnston

et al., 2004). We used a first-order exponential Kernel function for both aquifers with six neighbors for the Quaternary alluvial aquifer and 29 for the Jurassic karst aquifer.

Inverse Distance Weighting (IDW)

IDW is an extensively used, quick, deterministic, exact, and local interpolator. It estimates the value of an unknown point by using a linear combination of values at sampled points, weighted by an inverse function of the distance between these points. The method assumes that closer points are more similar to predicted points than more distant points (Li and Heap, 2008; Tobler, 1970). The weights are computed according to the equation:

$$\lambda_i^{IDW} = \frac{1/d_i^p}{\sum_{i=1}^{n} 1/d_i^p} \tag{2.2}$$

where d_i is the distance between the predicted point $\mathbf{x}_0$ and the observed point $\mathbf{x}_i$, n is the total number of observation points used in the interpolation, and p is a power parameter which decides how the weight decreases as the distance increases (Xie et al., 2011). IDW cannot predict values above or below the maximum and minimum observed values (Johnston et al., 2004). We determined a p of 4.88 for the alluvial and p of 10.26 for the karst aquifer with CV. The shape of the searching neighborhood was chosen that all monitoring wells will be used for the prediction.

Radial Basis Function (RBF)

RBF covers a large series of exact, moderately quick interpolators (e.g., thin-plate spline, spline with tension, multiquadric function, completely regularized spline) that use a basic equation dependent on the distance between the predicted and the observed point (Aguilar et al., 2005). There is no assessment of prediction errors, and they also do not allow investigating the autocorrelation of the data. The name comes from the fact that the function is radial-symmetric by definition. The approximates have the general form:

$$\hat{Z}_{RBF}(\mathbf{x}_0) \approx \sum_{i=1}^{n} \lambda_i \varphi \left(|\mathbf{x}_i - \mathbf{c}_i| \right) \tag{2.3}$$

where $\hat{Z}_{RBF}(\mathbf{x}_0)$ is approximated by a sum of n radial basis function $\varphi(r)$ which have different centers $\mathbf{c}_i$ and are weighted by the coefficients λ_i. We achieved the smallest RMSE with completely regularized spline functions for both aquifers.

2.2.3 Geostatistical Methods

Kriging-Estimators

Kriging is a very flexible interpolator that can be exact or smooth. It allows a variety of output surfaces including predictions, prediction standard errors, and probability (Johnston et al., 2004). Each of the different kriging methods is based on the following basic equation, which is a slight modification of Eq. 2.1:

$$\hat{Z}(\mathbf{x}_0) - \mu(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i \left[Z(\mathbf{x}_i - \mu(\mathbf{x}_i)) \right] \tag{2.4}$$

where λ_i is the kriging weight derived from a covariance function or semivariogram, μ is a known stationary mean (trend component), assumed to be constant over the whole area of interest and calculated as the average of the data (Li and Heap, 2011).

Semivariance and Variogram

Before the actual prediction, the spatial correlation of the data is assessed by variography. The semivariance γ of Z between the observation point $\mathbf{x}_i$ and the prediction point $\mathbf{x}_0$ is defined as:

$$\gamma(\mathbf{x}_i\mathbf{x}_0) = \gamma(h) = \frac{1}{2} var \left[Z(\mathbf{x}_i) - Z(\mathbf{x}_0) \right] \tag{2.5}$$

where h is the distance between $\mathbf{x}_0$ and $\mathbf{x}_i$ and $\gamma(h)$ the semivariance. The plot of $\gamma(h)$ vs. h is called empirical semivariogram and represents the spatial autocorrelation of the observed points. It quantifies the assumption that nearby data points are more similar than distant data points (Tobler, 1970). Some important features of the $\gamma(h)$-plot are the nugget, the sill/partial sill, and the range. In theory, the semivariance at $h = 0$ should be zero. The nugget is a positive value of $\gamma(h)$ at h very close to 0. It represents variability at distances smaller than the typical sample spacing and includes measurement errors. The sill is the semivariance value at which the semivariogram levels off. The partial sill results from sill minus nugget. The range is the distance at which the sill is reached. Points that are further apart from each other than the range are considered spatially independent (Li and Heap, 2011). There is a great variety of semivariogram models, such as spherical-, exponential-, and Gaussian models, which significantly influence the prediction of the unknown values. The semivariogram model and the associated parameters nugget, sill, and range are optimized in this study by using CV with a focus on the estimation of the range parameter.

Simple Kriging (SK)

The estimation of SK is based on (2.4) and (2.5). It is assumed that the trend component is an exactly known constant over the whole area of interest and estimated by the mean value of observed data, $\mu(\mathbf{x}_0) = \mu$, so that:

$$\hat{Z}_{SK}(\mathbf{x}_0) = \mu \sum_{i=1}^{n} \lambda_i^{SK}(\mathbf{x}_0 + [Z(\mathbf{x}_i - \mu)] \tag{2.6}$$

We used a multiplicative skewing approximation method for both aquifers with a gamma distribution for the alluvial aquifer and a Student's t-distribution for the karst aquifer.

Ordinary Kriging (OK)

OK is similar to SK with the difference, that μ is an unknown trend constant that has to be estimated. The most important consideration in OK is the assumption that the mean value remains constant over the whole are to be interpolated:

$$\hat{Z}_{OK} = \sum_{i=n}^{n} \lambda_i^{OK}(\mathbf{x}_0) Z(\mathbf{x}_i) \qquad \text{with} \qquad \sum_{i=1}^{n} \lambda_i^{OK}(\mathbf{x}_0) = 1 \tag{2.7}$$

OK allows trend removal and data transformation. We interpolated the GWL with a local smoothing ($\mathrm{OK_{sm}}$) (Gribov and Krivoruchko, 2004) as well as a standard ($\mathrm{OK_{st}}$) neighborhood type.

Universal Kriging (UK)

UK, also known as kriging with a trend, kriging with an external drift, and regression kriging (Hengl, 2009). It is a multivariate extension of OK. Instead of a constant trend μ, it uses a linear or higher deterministic trend function $\mu(\mathbf{x}_i)$. As a trend function, we achieved the most satisfactory results with an exponential kernel function for the Quaternary alluvial aquifer and a Gaussian kernel function for the Jurassic karst aquifer.

Co-Kriging (CoK, CoOK)

CoK uses information from one or more correlated secondary variables. The variables do not necessarily have to be measured at the same location but should be in the same range (Johnston et al., 2004). CoK exists for all kriging methods mentioned. In this

study, we used co-ordinary kriging (CoOK). The primary variable of interest Z_1 and both autocorrelation for Z_1 and CV between Z_1 and the secondary variables are used to improve the prediction. In CV, individual observation points are systematically removed from the data set one by one, and their values are re-estimated with the model used. (see also Section 2.2.4). The prediction of Z_1 cannot impair thereby because if there is no cross-correlation, it falls back to autocorrelation for Z_1 (Johnston et al., 2004; Rivoirard, 1990). The secondary variables (support points) used for CoOK A-D are given in 2.2.6.

Empirical Bayesian Kriging (EBK)

EBK provides a straightforward and robust interpolation method that automates the most difficult aspect of building a solid kriging model by automatically calculating parameters through subsetting and simulation. The semivariogram parameters in EBK are estimated using intrinsic random functions (restricted maximum likelihood method) (Johnston et al., 2004).

2.2.4 Validation Methods

Cross-Validation (CV)

The performance of a spatial interpolation method is affected by several factors, such as sampling density and distribution and data variation (Li and Heap, 2011). Therefore, the performance must be carefully evaluated in each case. The CV method is a statistical method to assess interpolation accuracy. In CV, each observed point is sequentially omitted, and the value is predicted using the residual data. The difference between each observed and the respective predicted value is the CV error. CV can also be used to select the best possible modeling settings for the respective method (e.g., search radius, power option, kernel parameter). Based on the results of the CV, the following evaluation statistics or error measures were used to compare the accuracy of the different interpolation methods, where m_i is the observed value and p_i is the predicted value at position i.

Mean Error (ME)

The ME is the average (arithmetic mean) of the errors. It indicates the average direction of the errors. An overestimation is characterized by positive bias, and an underestimation is marked by negative bias. ME is only conditionally suitable as an indicator of accuracy

because negative and positive estimates counteract each other, and the resultant ME tends to be lower than the actual error (Li and Heap, 2011):

$$ME = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i) \tag{2.8}$$

Mean Absolute Error (MAE)

The MAE is the arithmetic mean of the absolute error values. It indicates the magnitude of the error, which shows the accuracy of the method:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|p_i - o_i| \tag{2.9}$$

Mean Square Error (MSE)

The MSE measures the magnitude of the error (accuracy), weighted on the squares of the errors. Therefore, it is sensitive to outliers as it attaches major weight on larger errors:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2 \tag{2.10}$$

Root Mean Square Error (RMSE)

The RMSE is the square root of MSE. It has similar properties as MSE but has the advantage that it possesses the same unit as the required value (e.g., meter):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2} \tag{2.11}$$

Root Mean Square Standardized Error (RMSSE)

The RMSSE should be close to 1. An RMSSE greater than 1 means a general underestimation in the variability of the predictions, a RMSSE smaller than 1 means a general overestimation in the variability:

$$RRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{p_i - o_i}{\sigma_i}\right]^2} \tag{2.12}$$

Mean Absolute Percentage Error (MAPE)

The MAPE expresses the accuracy as a percentage of the error:

$$MAPE = \frac{100}{n} \left| \frac{p_i - o_i}{p_i} \right| \tag{2.13}$$

Pearson r

The Pearson r measures the linear correlation between the predicted and the observed rescaled covariance. Generally, it is in the range of -1 to 1:

$$Pearson\ r = \frac{\sum_{i=1}^{n}(o_i - \bar{o})(p_i - \bar{p})}{\sqrt{\sum_{i=1}^{n}(o_i - \bar{o})^2}\sqrt{\sum_{i=1}^{n}(p_i - \bar{p})^2}} \tag{2.14}$$

We use the squared Pearson r as a general coefficient R^2.

2.2.5 Study Area

The study area is located in southern Germany, on the border region of the federal states of Baden-Wuerttemberg and Bavaria (Fig. 2.1). The area comprises one of the most important freshwater sources in Germany, which holds about one billion cubic meters and provides over 3 million people with high-quality drinking water (Flinspach et al., 1997). The region is hydrogeologically characterized by a complex interaction between the karst aquifer of the up to 550 m thick Upper Jurassic Limestones of the Swabian Alb and the alluvial aquifer situated within the Danube valley with its numerous ecologically important wetlands and the Danube River itself. The riverine landscape of the Donauried covers nearly half of the study area, which is 2800 km^2. North of the Danube River, the rocks of the Upper Jurassic make up the high plateau of the Swabian Alb, which submerges below the Tertiary sediments of the pre-alpine Molasse with an average of 1–2° in SSE-direction. The low-permeable heterogeneous layers of the wedge-shaped Molasse might act as a hydraulic barrier preventing significant exchange between the karst and the alluvial aquifer. The north-eastern part of the study area is covered by large parts of a Miocene meteorite impact crater, the Noerdlinger Ries. The impact destroyed extensive areas of exposed rocks down to the crystalline bedrocks, leaving only small, economically insignificant groundwater reservoirs (Winkler, 1972). Hence, it has been excluded for the interpolation of the GWL.

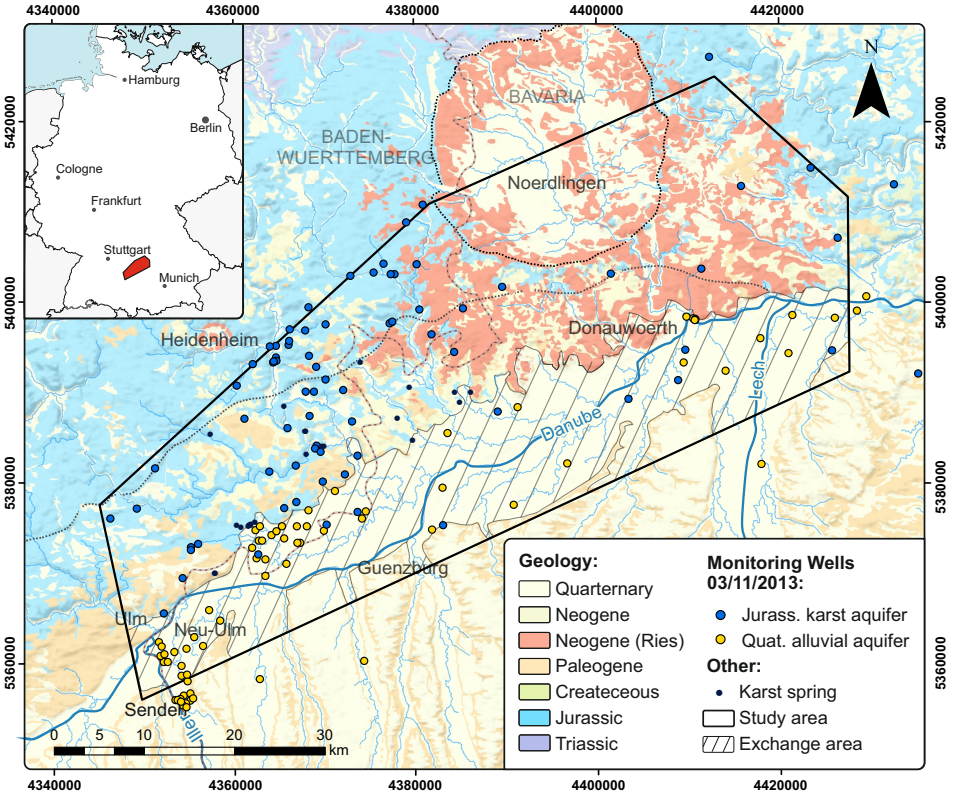


Figure 2.1: Geological map of the study area.

Upper Jurassic Karst Aquifer

Due to its thickness and heterogeneity, the Upper Jurassic karst consists of several individual karst aquifers. The groundwater flowing from the north to the Donauried belongs to the zone of deep karst. This deep karst zone is further divided (Villinger, 1977) into an open deep karst zone and a zone of covered karst below the Molasse. In the transitional area between these two zones, the covering layers are either locally present or thin so that the water level is unconfined in this northern edge region of the Molasse. The main karst aquifer is situated in the 200m – 450 thick sequence above the low permeable Lacunosa-Marls (ki 1), starting with the advanced karstified Upper Kimmeridge Limestones (Felsen- and Bankkalke, ki 2–3) to the Hangende Bankkalk Formation (ti H). Regional layers of the impermeable Cement-Marl can lead to local areas with perched groundwater.

Alpine Molasse

The Paleogene/Neogene Molasse is characterized by alternating bedding from sandy to clayey layers. Accordingly, a distinction must be drawn between strongly anisotropic horizontal and vertical hydraulic permeabilities. The horizontal permeabilities of the individual layers are in the range of 5×10^{-10} m/s for silty clays and 4×10^{-4} m/s for silty sands. A large-scale effective vertical hydraulic conductivity should be at the lower end of this range.

Quaternary Alluvial Aquifer

The homogeneous fluviatile Quaternary gravels with high to very high permeabilities represent valuable groundwater reservoirs,that completely cover the Molasse sediments within the Danube valley. They are found locally at the northern edge of the Molasse directly above the Upper Jurassic Limestones, where a large-scale groundwater exchange between the two aquifers can be assumed. Hydraulic conductivities between 5×10^{-4} to 1×10^{-2} m/s were determined in pumping tests (Bierer, 1987).

2.2.6 Available Data

Water level observation

Although a large number of monitoring wells are available, there is a large discrepancy between the period and the interval of the observations of the individual monitoring wells. GWL observations from 03/11/2013 were used as the basis for this study since this date ensured the best compromise between quantity and qualitative spatial coverage of the monitoring wells. Data from a total of 104 monitoring wells within the Quaternary alluvial aquifer and 77 within the Jurassic karst aquifer (65 unconfined, 12 confined to artesian) were included for all methods.

Support Points

Besides the reference date observations, additional variables were used for co-kriging interpolation to improve the prediction in areas with low spatial coverage of data points. For sites with no available monitoring wells with observations at 03/11/2013, long-term values of monitoring wells observed in other years have been used as long they have a coherent measuring interval of at least five years and no observable trend. This results in additional data from 40 monitoring wells in the karst aquifer and 57 monitoring wells

in the alluvial aquifer. A digital elevation model (DEM) based on SRTM1-data with a resolution of 1″ (about 25 m × 25 m) was used to define the topography (USGS, 2014). The surface elevation within the study area varies from 376 to 689 m asl. In addition, 16 river levels (daily average March 11, 2013, monitoring wells: 13 in the area of the Quaternary alluvial aquifer and 3 in the area of the Jurassic karst aquifer) were used, where direct contact of groundwater and surface water was expected. The following support points were used for the respective interpolation methods:

- CoOK A: monitoring wells with long-term observation values (not observed at the reference day),
- CoOK B: DEM,
- CoOK C: DEM, river levels,
- CoOK D: monitoring wells long-term values, DEM, river levels.

2.2.7 Data Processing

The GWL data were evaluated using *ArcGIS (10.4)*, *Spatial Analyst* and *Geostatistical Analyst tools*. The model parameters for each interpolation method, e.g., nugget, partial sill, and others, were optimized using CV, focusing on estimating the range parameter. A complete overview of all used model parameters is available in the support information of this chapter (Fig. S2a).

2.2.8 Calculation of Groundwater Exchange

The large-scale hydraulic effects of the Molasse on water exchange between the two aquifers have been poorly studied to date. The karst water level underlying the Molasse is in any part confined to artesian. A higher karst aquifer groundwater potential indicates areas of potential upward infiltration into the alluvial aquifer, while a lower potential indicates possible downward infiltration into the karst aquifer. The exchange of groundwater between the two aquifers in the area of the entire contact zone was quantified according to Darcy's law by using the following equation:

$$Q = \frac{K \cdot A \cdot \Delta h}{M} \tag{2.15}$$

K is the mean hydraulic conductivity for the leaky confining Molasse with 10^{-7} m/s, Δh the potential difference between the two aquifers, A the size of the contact area, and M

the thickness of the Molasse which was determined by interpolation of data from 126 borehole logs.

2.3 Results and Discussions

2.3.1 Cross-Validation Results

Tab. 2.3 shows the CV results for all methods used in this study including error ranking. By combining the results of all validation methods in an average error ranking the order of suitability of the methods are for the Jurassic karst aquifer: CoOK D > LPI> EBK > CoOK A > CoOK C > CoOK B > OK_{sm} > UK > RBF > OK_{st} > GPI > IDW > SK and the Quaternary alluvial aquifer: CoOK D > CoOK C > CoOK A > OK_{st} > OK_{sm} > EBK > LPI > CoOK B > IDW > RBF > SK > UK > GPI.

Therefore, CoOK D, which uses additional long-term values of monitoring wells not observed at the reference date and DEM data and river levels, provides the best results for both aquifers. CoOK B (additional DEM values only) has the largest error parameters within the co-kriging methods. EBK also offers promising results for both aquifers. LPI achieves good results for karst while underperformed for the alluvial aquifer. The negative bias of the ME of LPI, CoOK A, and CoOK D for the Jurassic karst aquifer and LPI, SK, UK, and CoOK D for the Quaternary alluvial aquifer indicates an underestimation while the other methods tend to overestimate the GWL. SK generates the greatest errors for the Jurassic karst aquifer and GPI for the Quaternary alluvial aquifer. The error estimations for the most frequently used methods, IDW and OK, show a substandard position in the ranking. The ranking of the methods shows no significant deviations due to the use of different error statistics. Only Pearson r shows deviations for some methods, which can lead to a different interpretation as long the method is compared with only one or a few other methods.

Fig. 2.2 shows an error histogram grouped in 1 m (Quaternary alluvial aquifer, yellow) and 2.5 m groups (Jurassic karst aquifer, blue) as well as a scatter-plot of observed and predicted values. For Jurassic karst aquifer, all methods overestimate the low GWL within the deep karst and underestimate the high water levels. This effect is most evident in SK and would result in a much lower gradient than in reality. Underestimated levels are primarily located in the area of the Swabian Alb. Those miscalculations could result from the variable topography or possible local areas with perched groundwater. Overestimated levels are located in the deep karst and the western part of the study area. Both are areas with a low density of monitoring wells.

Table 2.3: Cross-validation results. IDW: inverse distance weighting; GPI: global polynomial interpolation; RBF: radial basis function; LPI: local polynomial interpolation; OK: ordinary kriging; SK: simple kriging; UK: universal kriging; CoOK: co-ordinary kriging; EBK: empirical Bayesian kriging; ME: mean error; MAE: mean absolute error; MSE: mean standard error; RMSE: root mean standard error; RMSSE: root mean square standardized error; MAPE: mean absolute percentage error; Pearson r: Pearson correlation coefficient. Error ranking: 1 = "best method"...13 = "worst method"

Karst aquifer	Interpolation method												
Validation method	IDW	GPI	RBF	LPI	OK_{sm}	OK_{st}	SK	UK	CoOK A	CoOK B	CoOK C	CoOK D	EBK
ME	1.55	0.21	0.86	-0.95	1.14	1.06	2.57	1.34	-1.10	0.74	0.26	-0.47	0.85
MAE	7.51	12.48	8.89	6.72	6.52	6.93	14.06	7.81	7.07	6.62	6.65	6.38	6.38
MSE	206.7	291	243	97	196.6	198.3	492.2	196.6	187.1	188.7	167.8	149.8	160
RMSE	14.38	17.06	15.59	9.85	14.02	14.08	22.18	14.02	13.68	13.74	12.95	12.24	12.65
RMSSE	-	-	-	1.09	2.84	2.13	0.66	1.52	0.59	0.6	0.73	1.49	0.97
MAPE	1.61	2.6	1.91	1.46	1.39	1.47	3.12	1.67	1.29	1.42	1.44	1.38	1.37
Pearson r	0.914	0.999	0.985	0.961	0.918	0.918	0.867	0.919	0.922	0.922	0.932	0.938	0.934
	Interpolation method												
Error ranking	IDW	GPI	RBF	LPI	OK_{sm}	OK_{st}	SK	UK	CoOK A	CoOK B	CoOK C	CoOK D	EBK
MAE	9	12	11	6	3	7	13	10	8	4	5	1	2
MSE	10	12	11	1	8	9	13	7	5	6	4	2	3
RMSE	10	12	11	1	7	9	13	8	5	6	4	2	3
MAPE	9	12	11	7	4	8	13	10	1	5	6	3	2
Pearson r	12	1	2	3	11	10	13	9	8	7	6	4	5
Average	10.25	9.25	8.75	3	7.5	9	13	8.5	4.75	6	5	2.75	3.25
Alluvial aquifer	Interpolation method												
Validation method	IDW	GPI	RBF	LPI	OK_{sm}	OK_{st}	SK	UK	CoOK A	CoOK B	CoOK C	CoOK D	EBK
ME	0.3	0.04	0.27	-0.24	0.27	0.25	-0.14	-0.16	0.26	0.34	0.01	-0.09	0.45
MAE	2.04	3.26	2.19	1.89	1.46	1.37	2.33	3.12	1.41	1.53	1.09	1.02	1.5
MSE	13.59	18.71	15.55	7.31	4.86	4.75	16.1	18.79	4.62	9.08	3.15	2.28	7.29
RMSE	3.69	4.33	3.94	2.7	2.2	2.18	4.01	4.33	2.15	3.01	1.77	1.51	2.7
RMSSE	-	-	-	1.06	0.98	0.95	0.25	3.57	1.28	0.68	0.68	1.25	0.74
MAPE	0.47	0.74	0.5	0.43	0.33	0.31	0.54	0.7	0.32	0.35	0.25	0.23	0.34
Pearson r	0.993	0.99	0.991	0.996	0.997	0.997	0.992	0.99	0.997	0.995	0.998	0.999	0.996
	Interpolation method												
Error ranking	IDW	GPI	RBF	LPI	OK_{sm}	OK_{st}	SK	UK	CoOK A	CoOK B	CoOK C	CoOK D	EBK
MAE	9	13	10	8	5	3	11	12	4	7	2	1	6
MSE	9	12	10	7	5	4	11	13	3	8	2	1	6
RMSE	9	12	10	7	5	4	11	13	3	8	2	1	6
MAPE	9	13	10	8	5	3	11	12	4	7	2	1	6
Pearson r	9	13	11	7	5	4	10	12	3	8	2	1	6
Average	9	12.6	10.2	7.4	5	3.6	10.8	12.4	3.4	7.6	2	1	6

2.3.2 Groundwater Level Contour Maps

The comparison of the error statistics shows only minor differences between the results of the different interpolation methods. For example, a $R^2 > 0{,}9$ for almost all techniques shows that they all provide sufficient interpolation results. Therefore, additional methods for plausibility control are required, e.g., optical validation of the resulting GWL contour maps. Due to the pronounced anisotropy and heterogeneity of the karst aquifer caused by the local change in hydraulic permeabilities, the interpolated GWL contour maps can only give a general large-scale picture of the groundwater. This is also reflected in

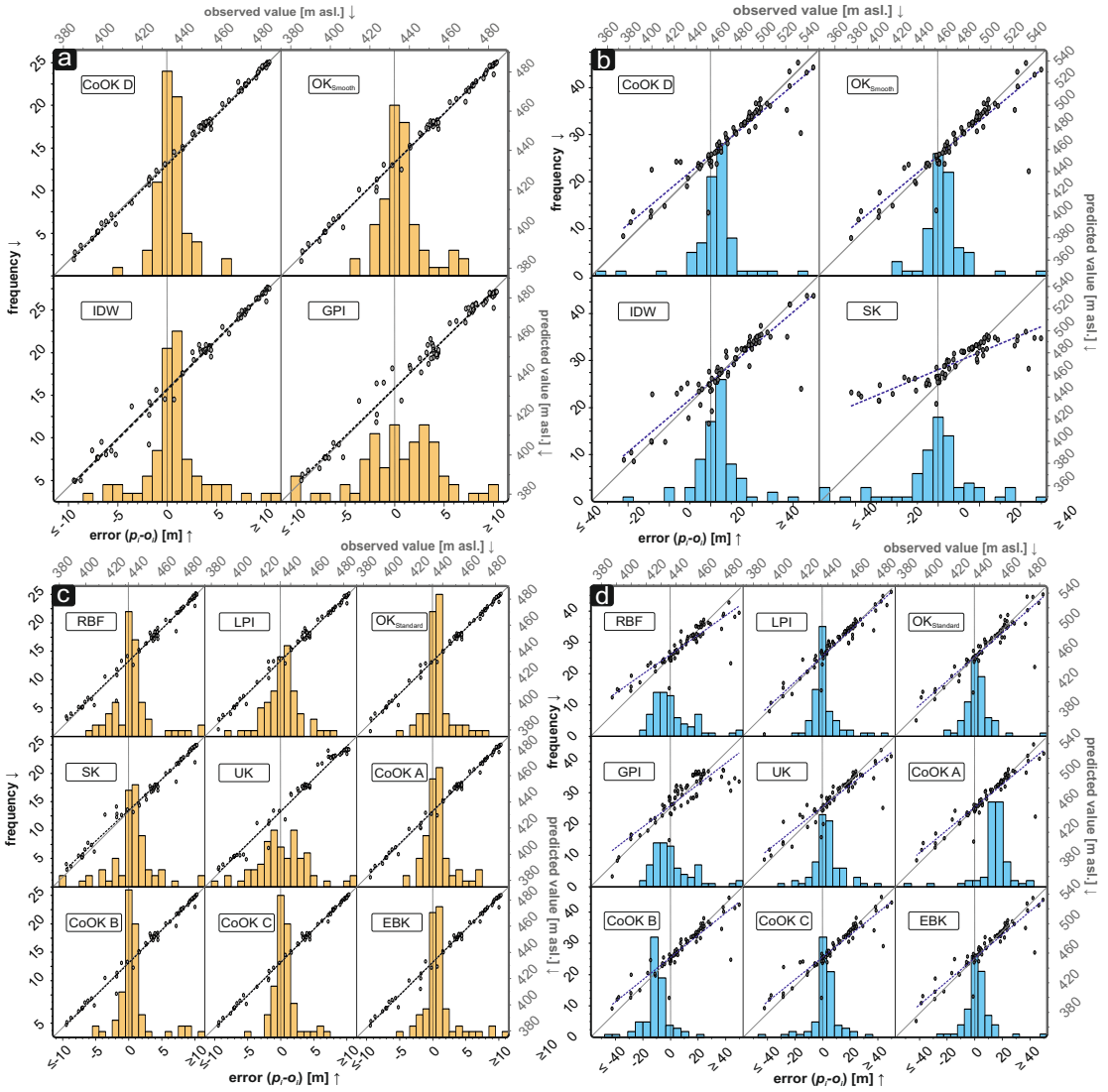


Figure 2.2: Histograms for ME and plots of observed vs. predicted GWL. CoOK; OK; IDW; GPI; SK; RBF; LPI; UK; EBK; [a]: best (CoOK D), worst (GPI) and most popular methods (IDW, OK) for Quaternary alluvial aquifer; [b]: best (CoOK D), worst (SK) and most popular methods (IDW, OK) for Jurassic karst aquifer; [c]: other methods for Quaternary alluvial aquifer and [d] for Jurassic karst aquifer. Gray line: observed values; purple dashed line: R^2 predicted values.

the higher errors and lower R^2 values compared to the Quaternary alluvial aquifer. A primary flow direction of karst water flowing southeast from the Swabian Alb is shown in all GWL contour maps. As the Jurassic strata dip below the molasse, the groundwater gradient decreases sharply in all interpolated maps, especially in the western areas where many karst springs are located at the alluvial boundary. Most GWL contour maps of the Quaternary alluvial aquifer show that flow conditions are determined by the main course of the Danube and the rivers Iller and Lech. Along the northern boundary of the

Quaternary alluvial aquifer near Langenau, a large-scale rotation of the flow direction in a southerly direction indicates significant amounts of infiltrated karst water from the Swabian Alb area.

Fig. 2.3 shows the generated GWL contour maps from the best [a] and worst [d] methods according to the error ranking in Tab. 2.3, as well as from OK [b] and IDW [c] as the most commonly used methods for both aquifers (Jurassic karst aquifer: blue lines; Quaternary alluvial aquifer: orange lines). In the area where the low permeable Molasse separates the two aquifers, the potential differences of the two aquifers were calculated to identify possible groundwater exchange zones. Within the blue areas, the Jurassic karst aquifer has a higher potential (likely rise). The Quaternary alluvial aquifer within the orange areas has a higher potential (potential descent). White areas mean an equilibrium. The results show an increase of the karst water potential around Langenau and southwest of Donauwoerth (see also Fig. 2.3). This confirms the results of previous investigations (Bierer, 1987; Udluft, 2000; Villinger, 1977). The different distances between the monitoring wells and the inversely proportional weight to the distance on IDW create vast, so-called "bull's eye" artifacts, which are circular regions of equal values around the known data points. The high gradient thus implied is limited to a small-scale region halfway between the monitoring wells. For the Jurassic karst aquifer, nearly all methods show an eastward change of direction in the area south of the Danube. The intensity and location of this bending vary significantly within the methods. The remaining GWL contour maps interpolated with the other methods can be found in the supporting information of this chapter (Fig. S2b).

Fig. 2.4 shows an SW-NE cross-section through the unconfined karst of the Swabian Alb. Within the green areas, the distance between the monitoring wells is less than 2.5 km, yellow less than 5 km, orange less than 10 km. In cross-section, only the areas with a higher density in monitoring wells show plausible results, independent of the method used, while artifacts can be seen in other regions. The best method can not replace a sufficient GMN. Again, the bullseye effect of IDW is evident. The GWL contours produced by OK and CoOK D show a very similar course in most regions. The GWL contours of CoOK D are more perturbed due to the incorporation of the secondary variable, especially the terrain surface. This leads to probable overfitting, especially in areas with high depths to groundwater. In the southwestern part of the cross section, the GWL contours of SK are roughly the same as those of the other methods. However, in the northeast direction, the method tends to over- and underestimate the GWL. In Kessel valley and Woernitz valley, all methods fail because of the low density of the measuring network. This results in an overestimation of the interpolated GWL within the valley. We validated the results additionally with hydrogeological expectations. This

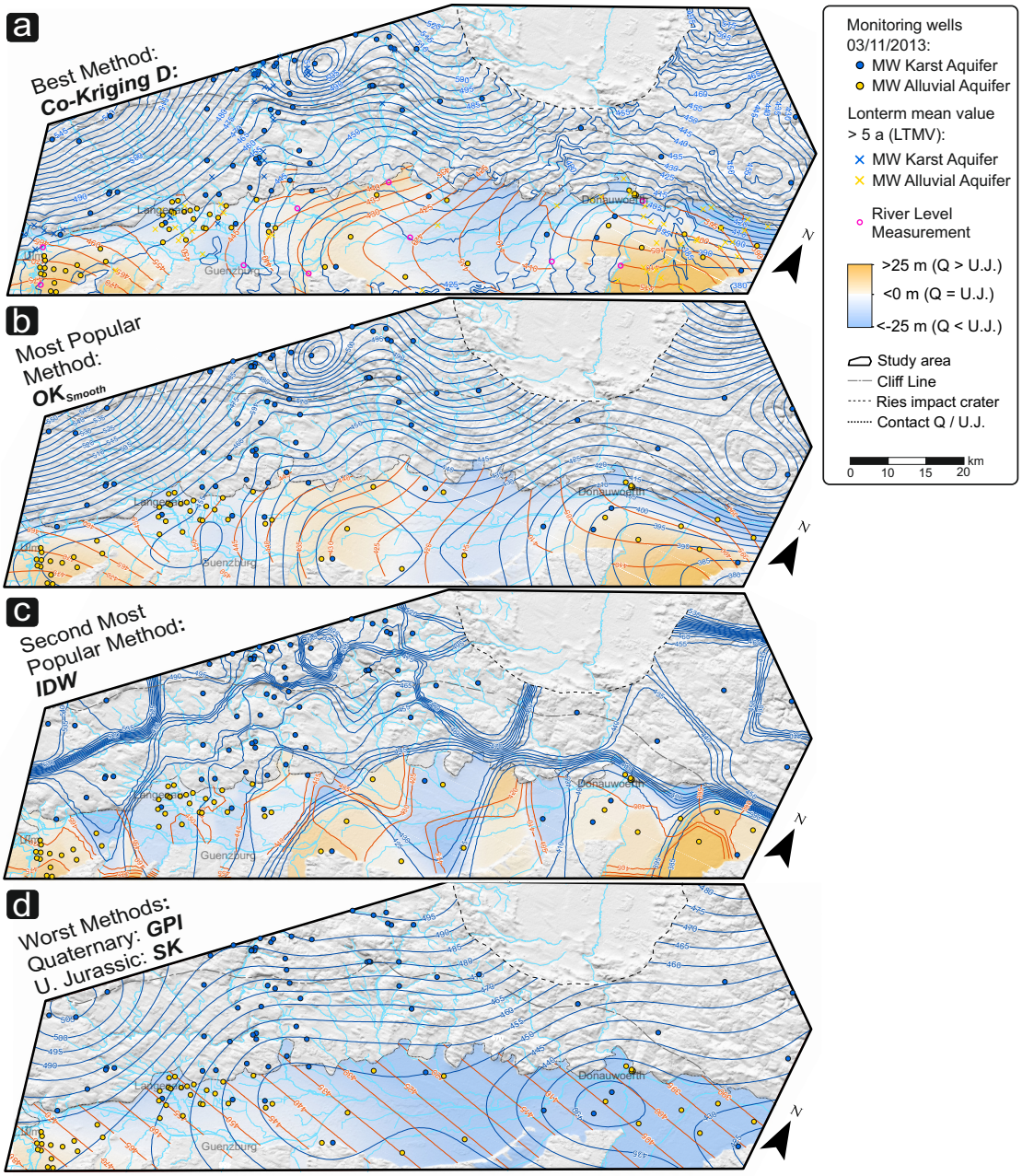


Figure 2.3: Estimated GWL contours and potential differences Jurassic karst aquifer vs Quaternary alluvial aquifer, based on a: CoOK D, b: OK, c: IDW, d: GPI and SK.

includes a comparison between calculated groundwater depth and geographic locations of karst springs and wetlands and surface waters as well as a comparison of the calculated flow accumulation resulting from the modeled GWL and the locations of real receiving waters.

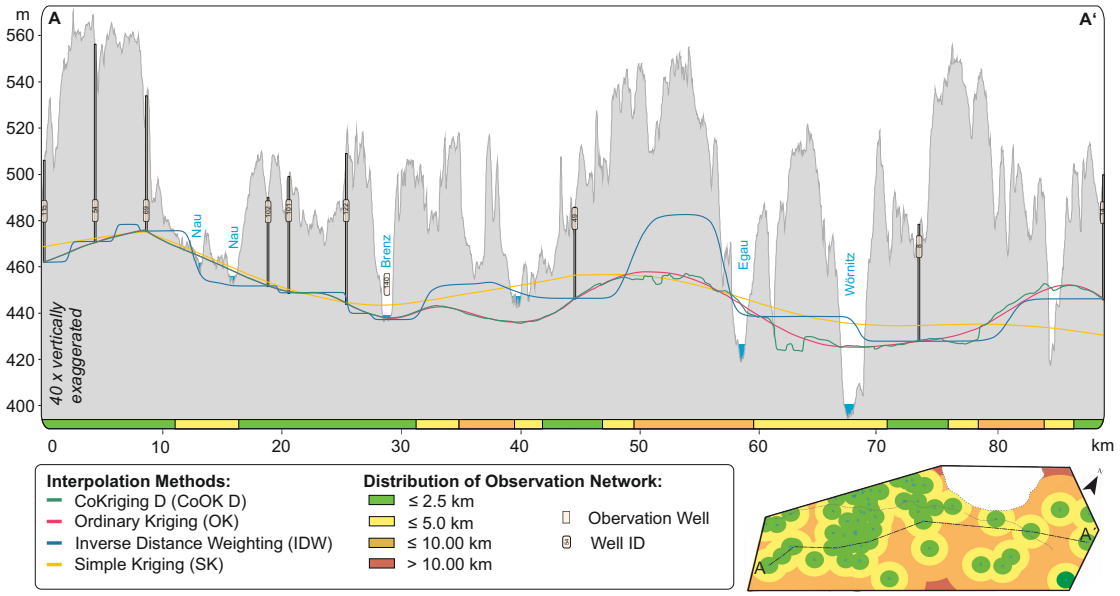


Figure 2.4: Cross section with topography and GWL contours generated by CoOK D, OK, IDW, and SK.

Fig. 2.5 shows the pattern of groundwater depth for the Quaternary alluvial aquifer in the lowland fens area Swabian Donaumoos (fen area marked black). Since lowland fens are permanently water-saturated wetlands fed by groundwater and rainfall, the difference between the modeled surface and the DEM surface should ideally be close to zero. The yellow areas show groundwater depth between 3 and 0 m (GWL < DEM), the green areas show negative groundwater depth between 0 and -3 m (GWL > DEM). All areas with groundwater depth higher than 3 m and -3 m were displayed colorless. The soil (Letten) in this area of the Donaumoos was accumulated from clayey lake deposits. These soils form a hydraulic barrier for the groundwater and can lead to confined conditions within the aquifer. Accordingly, negative groundwater depth is not unexpected. CoOK D and OK may best reflectthe edge of the fen best. For both methods, confined conditions within most fen areas were calculated. CoOK D seems to overestimate the GWL in the western part of the fen. This is shown by the artesian conditions with groundwater depth above 3 m (blank area). These overestimations are also apparent with IDW and, to a lesser extent, in OK. GPI overestimates the entire southern flank while the northeast flank is underestimated.

Fig 2.6 shows the results for the karst aquifer following the same procedure as for the alluvial aquifer. At a perennial spring, the GWL has at least the potential of the terrain. Therefore, the springs can also be used to validate the quality of the interpolated surfaces. Furthermore, the resulting flow direction of the respective methods was calculated with

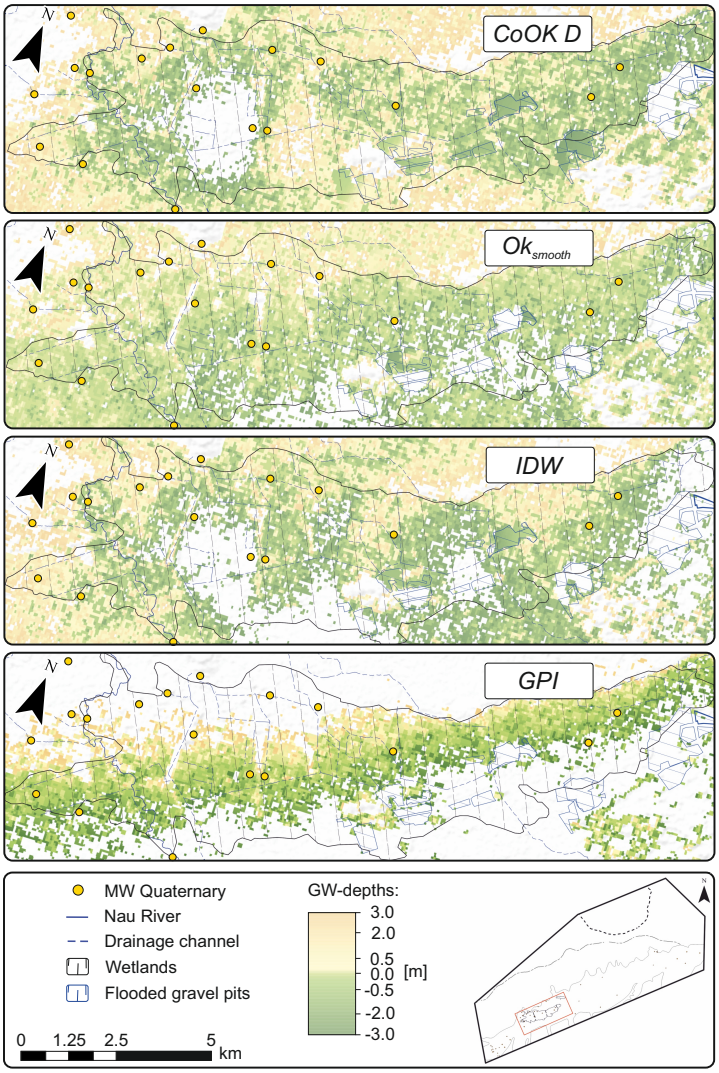


Figure 2.5: Maps of groundwater depth for the Quaternary aquifer in the Donaumoos area generated by CoOK D, OK, IDW and GPI

the D8 single-flow algorithm (O'Callaghan and Mark, 1983), which is implemented in *ArcGIS*. For an accurately modeled GWL, the course of the calculated receiving water (shown as purple lines) should be similar to the course of the real receiving waters (shown as light blue lines). Though none of the methods could reproduce the receiving waters in detail, the results of CoOK D and OK match the general course, whereas IDW and SK underperform.

Tab. 2.4 shows the calculated water exchange of both aquifers, calculated according to section 2.2.8 The estimated exchange rates within the methods fluctuate by up to a factor of more than 10. Depending on the methods, an exchange range from 273.96 million m^3/yr (Quaternary alluvial aquifer → Jurassic karst aquifer) to 2,772.63 million

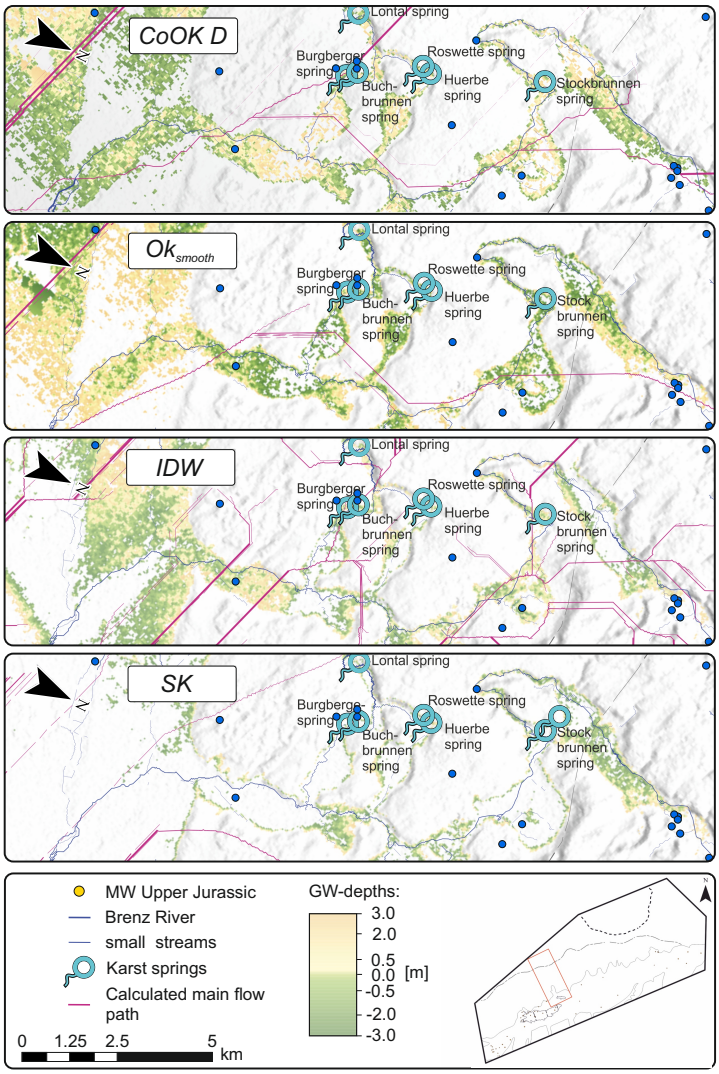


Figure 2.6: Maps of groundwater depth for the Jurassic aquifer and location of perennial karst springs as well as a comparison between rivers and flow directions resulting from the interpolations generated by CoOK D, OK, IDW and SK

m^3/yr (Quaternary alluvial aquifer $\leftarrow$ Jurassic karst aquifer) could be the consequence. However, the methods in the upper end of the error ranking are in a range between 300 million m^3/yr and 500 million m^3/yr.

2.4 Conclusions

A total of nine deterministic and geostatistical (multivariate and univariate) methods were compared with seven error statistic methods as well as with hydrogeological expectations, including the comparison between calculated groundwater depth and geographic

Table 2.4: Estimated vertical groundwater exchange rates between the Quaternary alluvial aquifer and the Jurassic karst aquifer. Pos. exchange rate: Quaternary → Jurassic. Neg. exchange rate: Quaternary ← Jurassic. ^{Q}: Quaternary , ^{J}: Jurassic

	IDW	GPI	RBF	LPI	OK_{smth}
$Quaternary_{mean}$ GWL [m asl.]	426.67	428.8	427.29	427.6	427.02
$Jurassic_{mean}$ GWL [m a.sl]	427.29	425.24	432.69	423.14	427.08
Exchange [mio m^3/yr]	-247.12	-484.45	-974.19	-196.97	-332.24
	OK_{std}	CoOK A	CoOK B	CoOK C	CoOK D
$Quaternary_{mean}$ GWL [m asl.]	426.78	426.97	426.97	426.46	426.57
$Jurassic_{mean}$c GWL [m a.sl]	425.72	428.61	426.87	428.28	427.41
Exchange [mio m^3/yr]	-227.06	-526.12	-429.64	-485.76	-303.24
	UK	EBK	SK	CoOK C^{Q}/SK^{J}	SK^{Q}/LPI^{J}
$Quaternary_{mean}$ GWL [m asl.]	427.25	426.78	429.02	426.57	429.02
$Jurassic_{mean}$ GWL [m a.sl]	429.61	425.55	447.66	447.66	423.14
Exchange [mio m^3/yr]	-640.34	-411.84	-2,451.13	-2,772.63	273.96

locations of karst springs as well as wetlands and surface waters. In the multivariate approaches, additional surface elevation data, river levels, and long-term GWL of wells not observed at the reference date were included to improve the predictions. The quality of the results for each method has been estimated by qualitative (maps and cross-section) and quantitative (CV) tools.

Several important points emerge from the results of the previous section. There is not a "universal superior method" for GWL interpolations. Which method performs best depends on the number and spatial distribution of the available monitoring well and the characteristics of the area (type of aquifer, topography, etc.). IDW is often used but is nearly never the "best" method. Particularly between remote monitoring wells, IDW tends to create "bull's eye" artifacts. Geostatistical methods mostly perform better than deterministic methods. In hilly terrain, the estimation of GWL is often problematic because the data set is always sparse with respect to topographic relief, and monitoring wells are almost exclusively located in valley regions, while the GWL is usually a subdued replica of the ground surface elevation (Hoeksema et al., 1989). This often leads to underestimating the GWL below mountains ranges and overestimating the GWL in valleys. If available, additional data (e.g. topography) can help to improve the results here. The consequences of the selection of the interpolation methods can be severe inaccurate or even false computations of recharge, discharge, flow directions, delineation of protection zones, etc. Therefore, it is advisable to compare the results of different methods, using various error statistics, to determine their plausibility and to indicate possible ranges for calculations based on interpolation results.

Of all tested geostatistical methods, it appears that CoOK D (incl. long-term values of monitoring wells not observed at the reference date, DEM data, and river levels) appears

to be the most accurate method for both aquifers studied. Our results show that the calculated exchange rates between the two aquifers, based on different interpolations of groundwater potentials, can vary greatly depending on the interpolation method chosen. In our example, by a factor of more than ten or from 273.96 million m^3/yr to -2,772.63 million m^3/yr when comparing the worst-case scenario (highest mean level Quaternary alluvial aquifer with lowest mean level Jurassic karst aquifer and vice versa). Therefore, the choice of an interpolation method should be made judiciously, taking into account the different error measures and additional data for plausibility check. If a clear choice is not possible, further calculations based on interpolated GWL contours should include an error estimate.

An inadequate GMN cannot be replaced by a suitable interpolation method. When the data variation is high, the monitoring density should be increased to capture the spatial change and, therefore, improve the respective method's performance.

Acknowledgments

We want to thank the environmental agencies of Baden-Wuerttemberg and Bavaria (LUBW and LfU) and the regional water management council of Donauwoerth (WWA-DON), particularly Dr. Rüdiger Zischak and Cornelius Jakob, for providing GWL data.

Supporting Information

This supporting information provides an overview of all used model parameters (S2a) and resulting GWL contour maps of the investigated interpolation methods (S2b).

GPI	Order			QUATERNARY			
	1			JURASSIC			
	1						
IDW	Pow er	Smoothing Factor	Angle	Major Semiaxis	Minor Semiaxis	Weights	
	4.88	1.00	0.00	35941.81	35941.81	104 Neighbors	
	10.26	1.00	0.00	27556.81	27556.81	77 Neighbors	
RBF	Kernel Function	Kernel Parameter	Sector Type	Max./Min. Neighbors	Major Semiaxis	Minor Semiaxis	Semiaxis
	Completely Regularized Spline	1.60E-03	1 Sector	8/8	35941.81	35941.81	
	Completely Regularized Spline	1.63E-03	4 Sectors	8/8	27556.81	27556.81	
LPI	Exploratory Trend Surface Analysis	Order of polynomial	Kernel Function	Goodness of Fit	Advanced mode	Neighborhood type	
	41	1	Exponential	2.629999201	False	Standard	
	45	1	Exponential	9.85124788	False	Standard	
	Maximum neighbors	Minimum neighbors	Sector type	Angle	Major semiaxis	Minor semiaxis	
	1000	0	1 Sector	0	23814.09	23814.09	
	1000	0	1 Sector	0	23658.76	23658.76	
	Anisotropy factor	Bandw idth	Use Spatial Condition Number Threshold	Spatial Condition Number Threshold	Weights		
	1	19051.27	True	27.92	6 neighbors		
	1	18927.01	True	21.23	29 neighbors		
OK_{sm}	Transformation type	Order of trend removal		Variable	Enable	Calculate Nugget	
	None	None		Semivariogram	True	False	
	None	None		Semivariogram	True	False	
	Nugget	Measurement Error	Type	Parameter	Major Range	Anisotropy	
	2.357589605	100	Stable	2	61181.43	False	
	1.118006499	100	Stable	2	24494.91	False	
	Calculate Partial Sill	Partial Sill	Lag Size	Number of Lags	Neighborhood type	Smoothing factor	
	False	2357.59	7647.68	12	Smooth	1	
	False	1118.01	3061.86	12	Smooth	1	
	Copy from Variogram	Angle	Major semiaxis	Minor semiaxis	Anisotropy factor	X	
	True	0	61181.43	61181.43	1	4411600.50	
	True	0	24494.91	24494.91	1	4391983.30	
	Y	Value	Weights				
	5363191.00	455.69	104 neighbors				
	5396312.60	458.54	75 neighbors				
OK_{st}	Transformation type	Order of trend removal		Variable	Enable	Calculate Nugget	
	None	None		Semivariogram	True	False	
	None	None		Semivariogram	True	False	
	Nugget	Measurement Error	Type	Parameter	Major Range	Anisotropy	
	2.357589605	0	Stable	2	61181.43	False	
	1.118006499	0	Stable	2	24494.91	False	
	Calculate Partial Sill	Partial Sill	Lag Size	Number of Lags	Neighborhood type	Maximum neighbors	
	False	2357.59	7647.68	12	Standard	15	
	False	1118.01	3061.86	12	Standard	5	
	Minimum neighbors	Sector type	Copy from Variogram	Angle	Major semiaxis	Minor semiaxis	
	10	1 Sector	True	0	61181.43	61181.43	
	2	1 Sector	True	0	24494.91	24494.91	
	Anisotropy factor	Weights					
	1	15 neighbors					
	1	5 neighbors					
SK	Transformation type	Decluster before transformation		Order of trend removal	Number of bins	Type	
	Normal Score	False		None	10	Multiplicative Skew ing	
	Normal Score	False		None	8	Multiplicative Skew ing	
	Number of modifiers	Base distribution	Variable	Model Nugget	Enable	Calculate Nugget	
	7	Gamma	Covariance		True	False	
	1	Student's t	Covariance	0.680021642	True	False	
	Nugget	Measurement Error	Type	Parameter	Major Range	Anisotropy	
	0.41	100	Stable	1.45	41409.38	False	
	0.68	100	Stable	1.50	28476.57	False	
	Calculate Partial Sill	Partial Sill	Lag Size	Number of Lags	Neighborhood type	Smoothing factor	
	False	0.66	5176.17	12	Smooth	1	
	False	0.32	3559.57	12	Smooth	1	
CoOK A	Semivariogram	0,38; 0,54	Stable	2	17503.32	2187.91	
CoOK A	Semivariogram	0; 0	Stable	1.87	15315.97	1914.50	
CoOK B	Semivariogram	0; 647,93	Stable	1.59E+00	29989.64	3748.7	
CoOK B	Semivariogram	0; 874,12	Stable	1.16328125	69628.17	8703.52	
CoOK C	Semivariogram	0; 828,29	Stable	1.59	45253.22	5656.65	
CoOK C	Semivariogram	0; 436,860; 214,77	Stable	1.35	34085.98	4260.75	
CoOK D	Semivariogram	0; 0; 990,34; 0	Stable	1.84	32962.15	4120.27	
CoOK D	Semivariogram	0,81; 0,81; 415,9; 137,04	Stable	2	13145.16	1643.14	
CoOK	Number of Lags	Smoothing Factor					
A	12	1					
A	12	1					
B	12	1					
B	12	1					
C	12	1					
C	12	1					
D	12	1					
D	12	1					
UK	Exploratory Trend Surface Analysis	Order of Polynomial	Kernel Function	Goodness of Fit	Maximum neighbors	Major semiaxis	
	61	0	Exponential	3.23	1.00E+03	5426.6	
	49	0	Gaussian	24.91	1.00E+03	10648.31	
	Minor semiaxis	Anisotropie factor	Bandw idth	Nugget	Parameter	Major Range	
	5426.6	1	4341.28	0.52703	0.78	17205.87	
	10648.31	1	5464.71	0	1.27	9408.45	
	Partial Sill	Lag Size	Number of Lags	Smoothing factor	Major semiaxis	Minor semiaxis	
	0.85	2150.73	12	1	17205.87	17205.87	
	72.35	1176.05	12	1	9408.45	9408.45	
BK	Subset Size	Overlap Factor	Number of Simulation	Transformation	Semivariogram Type	Neighborhood type	
	100	1	100	None	Pow er	Standard Circular	
	100	1	100	None	Pow er	Standard Circular	
	Maximum neighbors	Minimum neighbors	Sector type	Angle	Radius		
	15	10	1	0	35941.81		
	15	10	1	0	41228.19		

Figure S2a: Overview of the model parameters of the interpolation methods used in the study

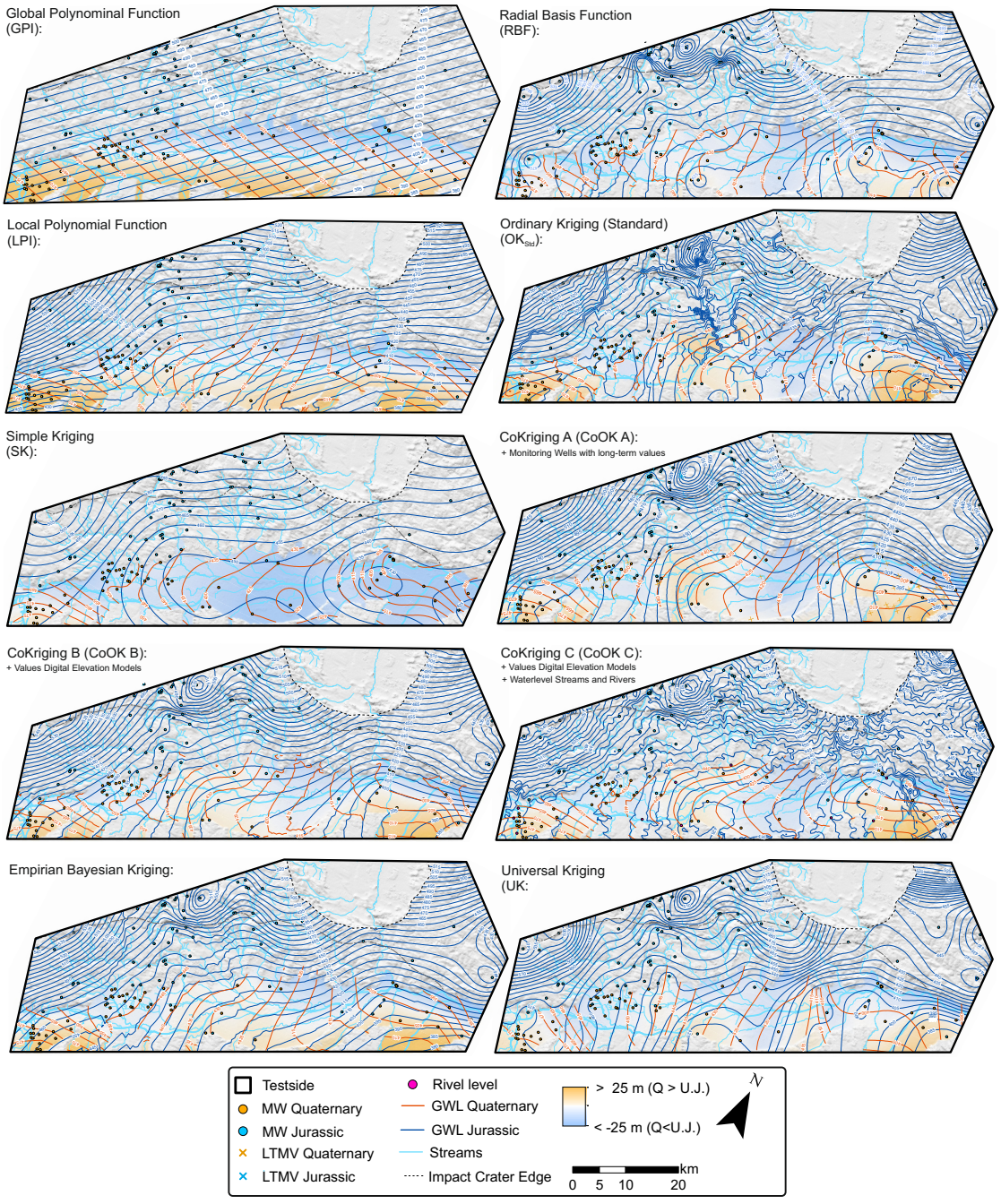


Figure S2b: Estimated GWL contours and differences in potential between Jurassic karst aquifer and Quaternary alluvial aquifer based on the remaining interpolation methods used in this study.

Chapter 3

On the Optimal Spatial Design for Groundwater Level Monitoring Networks

This chapter is based on a study published in the AGU journal *Water Resources Research*. The remainder of this part is an edited reprint of: Ohmer, M., Liesch, T., Goldscheider, N., (2019). On the Optimal Spatial Design for Groundwater Level Monitoring Networks. *Water Resources Research*, 55, 9454-9473, doi: 10.1029/2019WR025728.

Key Points

- A new low-discrepancy method (R_2 method) for an extensible groundwater level monitoring network design is proposed.
- We define a range of groundwater level monitoring network densities with an optimized information/cost ratio.
- We show that global cross-validation error parameters are not suitable for the comparative assessment of different sampling designs

Abstract

Groundwater monitoring networks (GMNs) are essential, as systematic data collected at monitoring wells provide a crucial understanding of the dynamics of hydrogeological systems as well as the basis for many other applications. This study investigates the influence of six groundwater level monitoring network (GLMN) sampling designs (random, grid, spatial coverage, and geostatistical) with varying densities on the accuracy of spatially interpolated groundwater level (GWL). To obtain spatially continuous prediction errors - in contrast to point cross-validation (CV) errors) - we used nine potentiometric groundwater surfaces from three regional MODFLOW groundwater flow models with different resolutions as a priori references. To assess the suitability of frequently used CV error statistics: mean absolute error (MAE), root mean square error (RMSE), root mean square standardized error (RMSSE), average standard error (ASE), and Nash–Sutcliffe model efficiency coefficient (NSE)), we compared them with the absolute prediction error (APE). Additionally, we defined upper and lower thresholds for an appropriate spatial density of monitoring wells. Below the lower threshold, the observation density appears insufficient, and additional monitoring wells significantly improve the results. Additional monitoring wells above the upper threshold lead to minor and inefficient improvements. According to the APE, systematic sampling leads to the best results but is often not suited for GLMN due to its nonprogressive characteristic. Geostatistical and spatial coverage sampling are considerable alternatives, which are in contrast progressive and allow evenly spaced and, in the case of spatial coverage sampling, yet reproducible coverage with accurate results. We found that the global CV error statistics are not suitable for comparing different performances of different sampling designs. However, they allow rough conclusions about the quality of the groundwater level monitoring network.

3.1 Introduction

Groundwater is an important yet spatially extensive, concealed, and inaccessible resource. Therefore, an effective GMN is essential, as systematic data collected at monitoring wells provide a crucial understanding of the dynamics and quality of the hydrogeological system. A GMN is defined by a spatial arrangement of monitoring wells and a temporal sampling frequency (Loaiciga et al., 1992). Economic considerations most strongly influence the number and location of monitoring wells. Therefore, designing an optimal GMN is a task of balancing prediction accuracy with cost minimization (Krivoruchko, 2011). Since a high spatial resolution is usually associated with disproportionate costs, often only domains of high water management importance are adequately

monitored. The design, that is, the selection of the location and number of the monitoring wells, is a vital part of any study involving modeling and prediction based on spatial data. A groundwater model can only be as good as the model input data available. Therefore, poorly distributed monitoring wells can lead to wrong assumptions or a bias of the regional image. Unsuitable interpolation methods can yield drastic overestimates or underestimates of the groundwater level in areas with a low monitoring density, as the parts with a high monitoring density are disproportionately weighted (Ohmer et al., 2017). Groundwater-quality observations are subject to the same problem. However, we focus on the regional GWL as a monitoring parameter in this study.

Varieties of studies dealing with the optimization of GMN have been published in the last 20 years. The literature focuses on optimizing groundwater monitoring network design to observe the groundwater-quality groundwater quality monitoring network, and the number of studies dealing with groundwater level (groundwater level monitoring network) is limited. The majority of approaches for GLMN design optimization are based on geostatistical analysis and, therefore, on minimizing uncertainty in parameter estimation. Several studies apply undifferentiated kriging (Prakash and Singh, 2000; Theodossiou and Latinopoulos, 2006), ordinary kriging (OK; Nunes et al., 2004; Yang et al., 2008), universal kriging (UK; Kambhammettu et al., 2011; Kumar et al., 2005; Olea, 1984). OK and UK (Ahmadi and Sedghamiz, 2007), OK and co-kriging (CoK; Ma et al., 1999), or indicator kriging (IK; Cameron and Hunter, 2002) to interpolate the GWL and use either the mean or maximum kriging variance to determine where additional monitoring wells should be built and/or identify well redundancy. Some recent works use the kriging variance as a part of a multicriteria decision-making analysis (MCA). Chandan and Yashwant (2017) considered multiple parameters in addition to the kriging variance such as GWL fluctuation, land use, hydrology, and recharge lineament density, to optimize an existing GLMN. In Uddameri and Andruss (2014), the kriging variance was linked to a monitoring priority index calculated from a weighted average of several criteria (groundwater-variability, recharge, surface/groundwater interaction, and groundwater fluxes across district boundaries). A similar approach can be found in Zhou (2013). Esquivel et al. (2015) used for their MCA a weighted linear combination that takes groundwater fluctuations, rates of decline, observation density, hydraulic gradients, mountains, and water bodies, for example, into account. Additional studies use information theory (entropy estimation) to evaluate the spatial location and temporal measuring frequency of monitoring wells (Alfonso et al., 2014; Leach et al., 2016; Masoumi and Kerachian, 2008; Mogheir et al., 2009). Further studies apply kriging-based genetic algorithms (Babbar-Sebens and Minsker, 2010; Dhar and Patil, 2012; Kollat and Reed, 2006; Luo et al., 2016; Reed et al., 2007; Yeh et al., 2006), algorithms based on

Kalman filter with space-time covariance matrix as input (Júnez-Ferreira and Herrera, 2013; Wu, 2004; Zhang et al., 2005).

Some authors also take the temporal component of GMNs into account (i.e., frequency of measurements based on groundwater fluctuations; e.g. Ahmadi and Sedghamiz (2007), Cameron and Hunter (2002), Chandan and Yashwant (2017), Kambhammettu et al. (2011), Nunes et al. (2004), and Theodossiou and Latinopoulos (2006). However, the results are always a compromise between the spatial and temporal components. Moreover, the influence of the individual components is difficult to quantify. Since the temporal component (i.e., measuring interval) can be easily varied, whereas the repositioning of sampling points represents a disproportionately greater effort, focusing solely on the spatial arrangement seems justified. Therefore, our study focuses on the spatial component and assumes a steady-state potentiometric groundwater surface to find the optimal spatial arrangement of sampling points regardless of the temporal component.

Most of the studies mentioned here have in common that a large amount of auxiliary data for the respective study areas have been incorporated into the optimizations. Therefore, the results cannot be easily transferred to other sites where these data may not be available. This study aims to find out if there is a universally applicable design approach that can achieve the best possible results without a priori knowledge of the hydrogeological situation. For this reason, we compared six design strategies as general as possible and analyzed their results in nine areas of investigation. We assumed "real" GWL to be known as they are taken from large-scale numerical groundwater models for this simulation experiment. The idea behind this approach is to compare the interpolated surfaces resulting from the respective GLMN design to the "real" surface and thereby compute the "real" error in addition to the CV error.

In detail, the research objectives are to answer the following questions:

- Is there an extensible and transferable GLMN design that allows reliable spatial estimates of GWL with a minimum number of monitoring wells?
- What are the quality differences resulting from the use of various GLMN design approaches?
- At what monitoring well density a reasonable information/cost ratio result?
- Which is the most suitable CV error statistic (MAE, RMSE, RMSSE, ASE, or NSE) to evaluate the quality of interpolated groundwater surfaces?

We applied six design strategies on nine different potentiometric groundwater surfaces to answer these questions, starting with initial ten monitoring wells each. Based on the

GWL of the monitoring wells, a groundwater surface was interpolated and the deviation from the real groundwater surface calculated, along with global CV error statistics. The GMN were then gradually densified up to 500 monitoring wells, the interpolation and error calculation being repeated after each step. The design strategies, which contain random components, were repeated ten times to include the influences caused by them.

3.2 Methods

3.2.1 Interpolation Methods

In a previous study, we examined and compared nine different interpolation methods: inverse distance weighting (IDW), radial basis function (RBF), simple kriging (SK), ordinary kriging (OK), universal kriging (UK), empirical Bayesian kriging (EBK) and co-ordinary kriging (CoOK), as well as local polynomial interpolation (LPI) and global polynomial interpolation (GPI), to find the most suitable method to interpolate a continuous GWL from monitoring well data (Ohmer et al., 2017). The best results were achieved with CoOK based on global CV error statistics. In this type of kriging, additional correlated secondary variables (e.g., digital elevation model (DEM), springs, and wetlands) are used to improve the prediction. Since secondary data are generally not sufficiently available everywhere, we decided not to consider this method here and instead use the second-best method, OK, which is one of the most frequently used geostatistical estimators (e.g., Siska et al., 2005; Wackernagel, 1995).

Geostatistics is based on the work of Krige (1951) and was further developed by Matheron (1963) with his theory of regionalized variables. Kriging is a generic name for a group (e.g., SK, OK, and UK) of generalized least squares regression algorithms (Li and Heap, 2008). Before the prediction, the spatial correlation of the regionalized data is assessed by a semivariogram analysis. The semivariance γ of Z between two points $\mathbf{x}_i\mathbf{x}_o$ separated by distance h is defined as:

$$\gamma = (\mathbf{x}_i\mathbf{x}_o) = \gamma(h) = \frac{1}{2} var[Z(\mathbf{x}_i) - Z(\mathbf{x}_o)] \tag{3.1}$$

An empirical semivariogram is a graphical representation of the semivariance ($\gamma(h)$ vs. h) that represents the spatial autocorrelation of the data points. It quantifies the assumption that nearby data points tend to be more similar than more remote points (First Law of Geography, according to Tobler, 1970). This empirical semivariogram is used as the first estimate of the theoretical semivariogram needed for the spatial interpolation.

Important features of the semivariogram are the nugget, the range, and the sill/partial sill. The nugget effect is a positive value of $\gamma(h)$ at h close to 0. It allows the variogram to assume a nonzero value for two observations with a distance of less than the minimum bin size. It accounts for a sum of measurement error and microscale irregularities. A method that produces an estimate equal to the observed value at the sample points is called exact; all others are called inexact. The sill is the semivariance value at which the semivariogram levels off for stationary data sets (Bohling, 2005). Partial sill results from the difference between sill and nugget. The range is a value of distance at which the sill is reached. Points further away than the range are regarded as spatially independent (Li and Heap, 2008). OK is robust and straightforward and, therefore, probably the most widely used kriging technique (Heuvelink and Pebesma, 2002). Each of the different kriging methods is based on the following basic equation:

$$\hat{z}(\mathbf{x}_0) - \mu(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i [Z(\mathbf{x}_i) - \mu(\mathbf{x}_i)] \tag{3.2}$$

where $\hat{z}$ is the estimated value at a point of interest $\mathbf{x}_0$, n is the total number of observed GWL, and $Z(\mathbf{x}_i)$ is the observed GWL at well $\mathbf{x}_i$. λ_i are the kriging weights derived from a covariance function or semivariogram; μ is, in the case of OK, the Lagrange multiplier that has to be estimated and is considered to be constant over the area to be interpolated (Li and Heap, 2008):

$$\hat{Z}_{OK} = \sum_{i=n}^{n} \lambda_i^{OK}(\mathbf{x}_0) Z(\mathbf{x}_i) \qquad \text{with} \qquad \sum_{i=1}^{n} \lambda_i^{OK}(\mathbf{x}_0) = 1 \tag{3.3}$$

We used an omnidirectional Gaussian semivariogram model, which is flexible, and a good candidate for a default model (Krivoruchko, 2011). Its parabolic behavior at the origin represents very smoothly varying properties (Bohling, 2005). The associated parameters partial sill, range, search neighborhood, and specific search distance were optimized using automated CV diagnostics to minimize the RMSE for each case. A systematically small constant nugget of 0.05 m was used to ensure the stability of the resulting kriging matrices (Coburn et al., 2006; Johnston et al., 2004).

It should be noted that the use of a single variogram model (Gaussian) might not be the optimal way to quantify spatial correlation, especially for nonstationary data. However, it is a necessary simplification owed to the automation process, which still allows comparability of the spatial design methods, while the best possible interpolation result is not the focus of this study.

3.2.2 Data and Data Processing

The spatial and temporal variability of the GWL is generally unknown except for wells, springs, wetlands, and interacting surface waters. Therefore, the accuracy assessment of interpolated/predictive GWL can only take place at these measured locations using CV and error statistics (e.g., MAE, RMSE, NSE, and RMSSE). Therefore, the expected level of fit of these results is primarily dependent on the number and distribution of the monitoring wells.

To quantitatively determine and compare the effects of the different GLMN designs on the accuracy of the predicted GWL, we used nine potentiometric groundwater surfaces extracted from simulation results of three regional MODFLOW groundwater flow models as an a priori reference. The model data are publicly accessible from the USGS (Feinstein et al., 2012; Jones et al., 2017; Parsen et al., 2016). The idea behind this approach was to compare the interpolated surfaces to the "real" surface and thereby compute the "real" error in addition to the CV error. This has been done with completely artificial surfaces to compare different network designs (Aguilar et al., 2005; Heuvelink and Pebesma, 2002; Romero et al., 2005; Wilde, 2009). Still, artificial surfaces may not have the same properties as typical groundwater surfaces in terms of variability, roughness, gradients, and so on. To use surfaces computed by numerical groundwater models, which incorporate the hydraulic properties of the aquifer, are based on physical processes of groundwater flow, and therefore produce not "real" but at least "realistic" surfaces, seems to be the best compromise. The resulting groundwater surfaces each consist of 100 m × 100 pixels with pixel sizes of 100 m × 100 m, 200 m × 200 m, or 500 m × 500 m. The pixel size corresponds approximately to the element size of the groundwater models.

The different resolutions of the surfaces were chosen to assess if and how the resolution affects the results. The resolution can, for example, influence elevation and slope values (Chunmei et al., 2013), as small-scale variabilities below the pixel size are eliminated. However, this does not allow comparing the observation density and resulting errors with each other directly. A detailed overview of essential parameters of the surfaces is given in Tab. S3a, GWL contour maps of the surfaces are shown in Fig. 3.1.

For the simulation experiment, we used the following automated workflow for all surfaces:

1. Ten initial monitoring wells are distributed randomly on every surface from a random number generator (exceptions are the systematic sampling and low discrepancy sampling methods, see section 3.2.3).

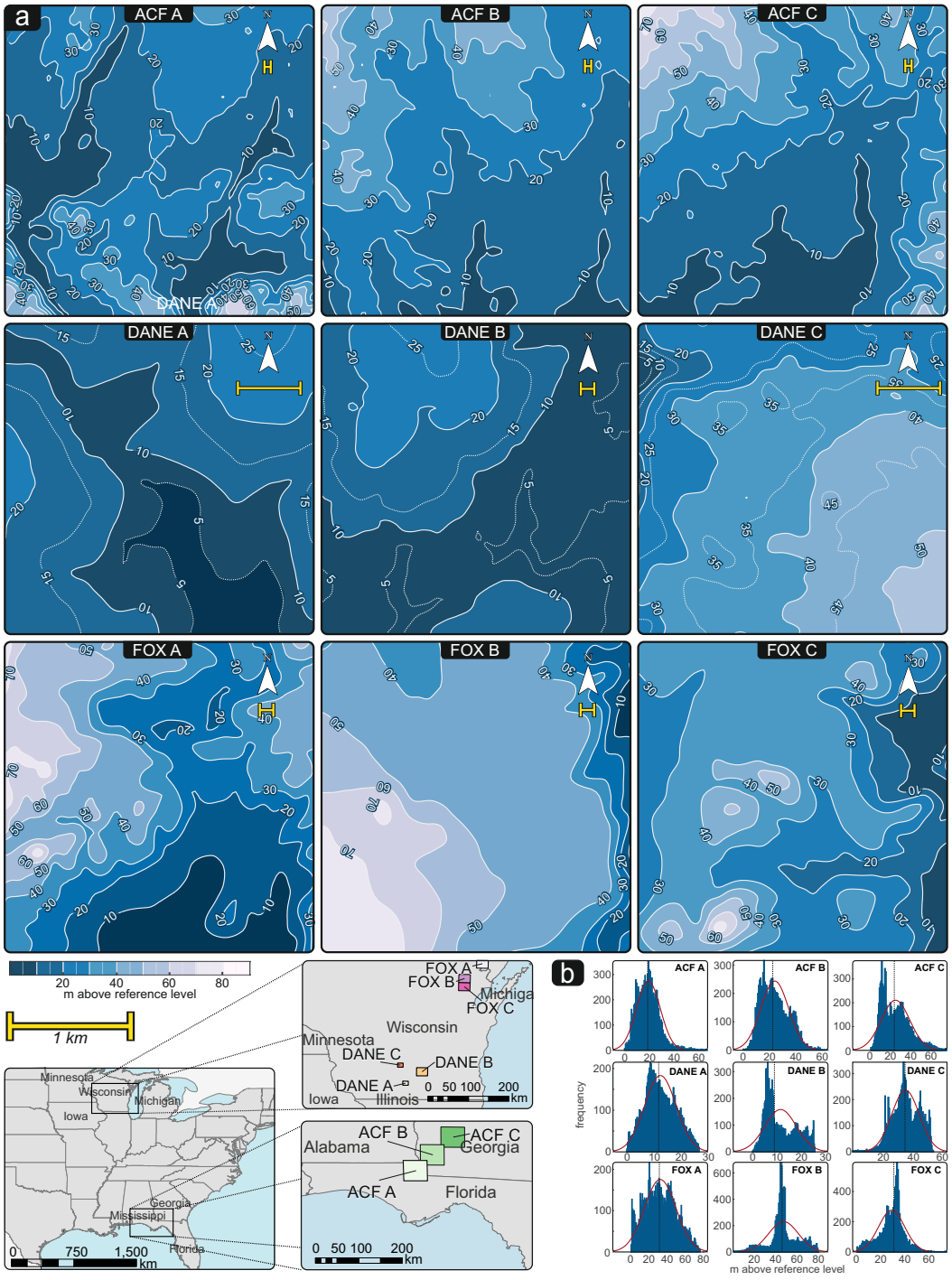


Figure 3.1: GWL contour maps and histograms extracted from three MODFLOW models: lower Apalachicola-Chattahoochee-Flint River Basin, Georgia (Jones et al., 2017), Dane County, Wisconsin (Parsen et al., 2016) und Upper Fox River Basin, Wisconsin (Feinstein et al., 2012), as a priori reference to evaluate the investigated observation network designs qualitatively and quantitatively.

2. Based on the GWL of the available observation points, an empirical variogram is computed, and an optimized (by CV) Gaussian semivariogram model is fitted to the data.
3. Based on the semivariogram, the GWL is interpolated with OK.
4. The prediction error (the difference between the real and the predicted surface) and different CV error statistics are calculated.
5. The location for an additional monitoring well is computed, depending on the used design method(see section 3.2.3),
6. Steps (2) to (4) are repeated until the GMN includes 500 monitoring wells. Steps (1) to (6) are repeated ten times for the methods that include random components, and the results were averaged to consider errors caused by random initialization or random addition of point locations.

3.2.3 Sampling Designs

When planning a new or extending an existing GLMN, one of two fundamentally different strategies must be chosen. One is the design-based sampling approach based on classical sampling theory; the other is the model-based approach based on geostatistics. The main difference between the two is how they deal with the randomness they use to give the inference a stochastic structure (Särndal et al., 1978). The additional monitoring wells (or, in general, samples) in a design-based observation network are selected so that each location within the study area has the same probability of being chosen. Another term for design-based sampling is, therefore, probability sampling. These probabilities provide the foundation for statistical inference from the observations (Gruijter et al., 2006). In a model-based network, a theoretical construction(model) is used to deal with the differential probabilities of the potential observation points. The model is built upon information on prior knowledge and assumptions (Geuna, 2000). It contains the prescription for the statistical inference. A detailed description of sampling theory and the contrast between the two strategies can be found in Särndal et al. (1978), Hansen et al. (1983), Brus and Gruijter (1997) and Brus (2010). Fig. 3.2 shows an overview of the observation network designs compared in this study.

Design-Based Approach: Spatial Statistical Sampling and Probability Sampling

Four standard types of statistical sampling methods are generally used in “classical” statistical surveys (Kish, 1995). These methods are simple random sampling (SRS),

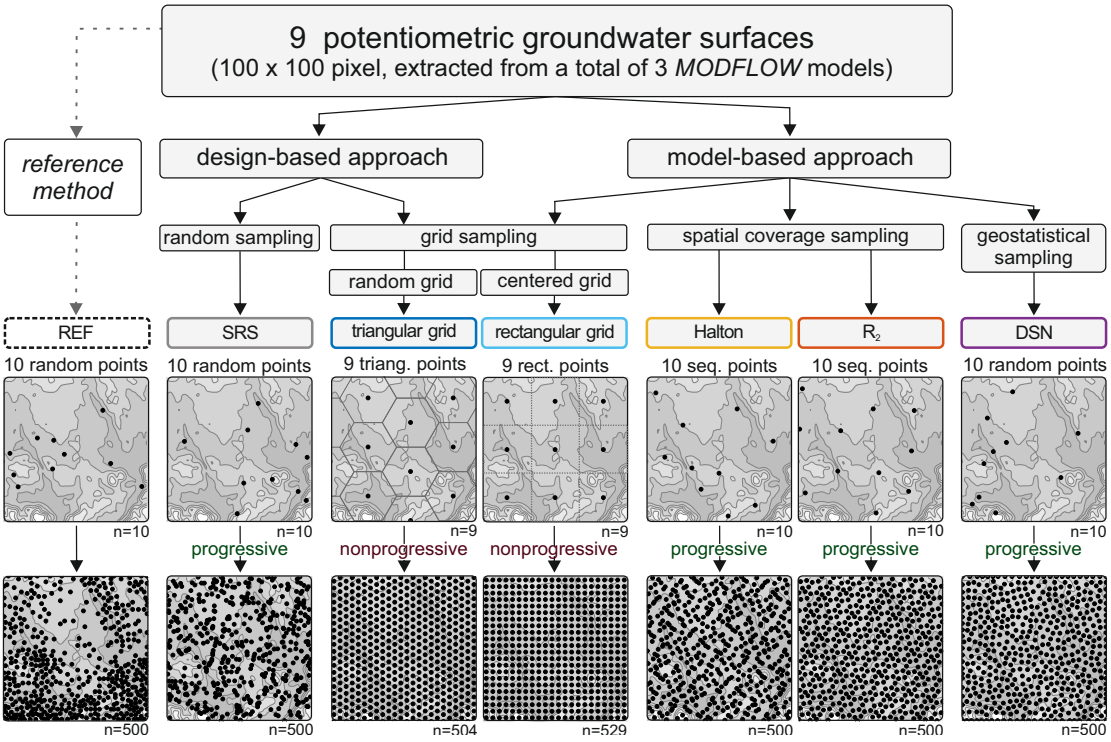


Figure 3.2: Schematic overview of the GMN design approaches compared in this study

random grid sampling (random systematic sampling), stratified sampling, and cluster sampling. For this study, SRS and random grid sampling were chosen. Stratified sampling and cluster sampling are usually used when a heterogeneous distribution of values can be broken down into internally homogeneous parts or when values tend to cluster together (Gilbert, 1987). Neither is the case for GWL.

Simple Random Sampling simple random sampling is the most straightforward and frequently used sampling design approach in survey sampling. It is assumed that n points are located randomly from N potential sites with an "equal probability of selection" (EPS) throughout a domain. The advantages are that it is simple to use and free from bias and prejudice. Disadvantages are poor spatial coverage and the possible occurrence of clustering, redundancy, and regions with no observation points. The additional observation points with SRS were selected in this study using the *ESRI ArcGIS* tool *GenerateRandomPoints*. This tool creates a specified number of random points to a defined extent.

Random Grid Sampling In random grid sampling (or random systematic sampling), samples are taken at regularly spaced intervals over space with the first point m chosen

randomly. The approach is classified as probability sampling since, with the first point selected randomly, each location within the study area initially has the same probability of being selected. Examples of a systematic grid are rectangular, triangular, hexagonal, or radial grids. If the starting point is not chosen randomly, the approach is classified as model-based (see section 2.3.2), as, for example, in regular systematic sampling and centric systematic sampling (Delmelle, 2014). The main advantage of a systematic sampling over SRS is that it can be more conducive to covering more extensive areas through a maximized sampling coverage while clustering and redundancy are prevented. Moreover, it allows to add a degree of system into the process of random selection (Fischer and Nijkamp, 2014). Gridded sampling designs are particularly suitable for large investigation areas, which should be covered with a limited number of sampling points (Gruijter et al., 2006). The disadvantage of this method is that the smallest separation distance is fixed. Since the kriging variance is described as a function of the separation distance, this can lead to unnecessarily large nugget effects for the model semivariogram (Vieux, 2016). Furthermore, the approach is not progressive, so the number of overall observation points must be known beforehand. It is impossible to progressively add more points without breaking the order. Therefore, systematic sampling methods are not suitable for constructing extensible observation networks. We added two methods for comparison, as they are often referred to as the most efficient design for survey sampling (Birch et al., 2007; Olea, 1984).

The systematic random approach used in this study is a triangular grid based on regular hexagonal polygons. The points are placed in the center and at the corners of a hexagon (shared with three other hexagons). The hexagons have a width $w = \sqrt{3} \cdot sidelength$ and a height $h = 2 \cdot sidelength$. Therefore, the lateral distance d_{hz} between adjacent hexagon center points is w while the longitudinal distance is $h \cdot 3/4$. In contrast to other methods, no additional points can be added progressively without breaking the grid symmetry. Instead, the existing points were replaced in the next iteration step by as many points as were necessary to maintain a regular grid with the next-larger number of points.

Model-Based Approach

Centered Grid Sampling Though very similar to random grid sampling, the centered grid sampling is referred to as a model-based approach since, in contrast to random grid sampling, the starting point is not chosen randomly but purposefully so that the area of investigation is well covered, especially when the overall number of points is low. For the rectangular grid used in this study, the investigated area was divided into $n \times n$ square intervals, and the observation points were set in their center (centric systematic

sampling). The shortest distance d between the observation points equals the side length of the square divided by the square root of the sample size $\sqrt{A/n}$ where A is the area of the square. As with all regular grids, this method is not progressive. An existing network cannot be extended under the applicable laws.

Spatial Coverage Sampling Spatial coverage sampling is a technique that optimizes an objective function of the distance between the observation points (Brus, 2010). We have chosen two low-discrepancy sequences as objective functions. In quasi-random or low-discrepancy sampling, the position of sampling points is based on low-discrepancy sequences (also called quasi-random or subrandom sequences). These sequences represent numbers that are better equidistributed than pseudo-random numbers (Dalal et al., 2008). To construct higher-dimension low discrepancy, as in the case of two-dimensional sampling design, several one-dimensional sequences are combined in a component-wise manner, that is, that the $\mathbf{x}$ and $\mathbf{y}$ coordinates of a two-dimensional area are constructed by pairing consecutive numbers of two different low discrepancy series in an $[0, 1] \times [0, 1]$ space and then adjusted to the actual spatial extent of the area to be sampled. Discrepancy refers to the density of points on an area or sampling space in a two-dimensional context. A high discrepancy means that there are large areas of empty space or regions with a disproportionally high point density (as it may be the case in a random distribution). Therefore, SRS can lead to a high discrepancy, while systematic sampling has the lowest discrepancy. Fully deterministic low-discrepancy sequences were developed to optimize Monte Carlo simulations because they fulfill requirements as if they were genuinely random numbers. At the same time, higher accuracy and faster convergence can be achieved with fewer samples than pseudo-random numbers, reducing computational costs. Low-discrepancy sampling methods thus constitute a good compromise of being progressive (like SRS) and having a low discrepancy (like systematic sampling). Therefore, they are frequently used in sampling problems. Furthermore, they allow for a better distribution of sample separation distances than gridded sampling schemes while minimizing sampling bias and clustering. To our knowledge, low-discrepancy sequences have not yet been applied to developing GMN designs before.

A categorization of the different types of low-discrepancy sequences is mainly done by the method of constructing their bias (hyper)parameter. These are either prime numbers (Van der Corput, Halton, and Faure sequence), polynomials (Sobol and Niederreiter sequence), or irrational fractions (Kronecker and R-sequence). In this study, the Halton sequence and the R-sequence were investigated, as they are more suitable for low-discrepancy in two dimensions than other low-discrepancy series (Roberts, 2018).

The Halton sequence (Halton, 1960) is a generalization of the van der Corput sequence (Corput, 1935) to higher dimensions. It uses arbitrary coprime numbers as a base for each dimension. The most frequent selection for two dimensions, due to its apparent simplicity and sensibility, is to select the first primes, which is referred to as the (2,3)-Halton sequence. To generate the sequence for 2, the interval [0,1] is divided in half,then in fourths, eighths, and so forth, which generates $^1/_2$, $^1/_4$, $^3/_4$, $^1/_8$, $^5/_8$, $^3/_8$ and so forth. Equivalently, the Halton sequence for 3 is generated by dividing the [0,1] interval in thirds, ninths, twenty seventh, and so forth, giving $^1/_3$, $^2/_3$, $^1/_9$, $^4/_9$, $^7/_9$, $^2/_9$ and so forth. The coordinates for the sampling points are constructed by placing the $\mathbf{x}$ coordinates according to the 2-Halton sequence and $\mathbf{y}$ coordinates according to the 3-Halton sequence, adjusting the numbers of the sequence to the actual spatial coordinate extent of the area.

The Halton sequence constitutes a good source for a low discrepancy in two dimensions since the selection of small coprime bases ensures a minimal correlation between dimensions (Worley, 2016) and is therefore regularly used in ecological sampling (Brown et al., 2015; Kermorvant et al., 2019).

Recently, a new low discrepancy quasi-random sequence that offers a substantial improvement over current state-of-the-art sequences has been proposed Roberts (2018). The new additive recurrence sequence (R-sequence) is a recurrence method based on irrational numbers (generally called Kronecker sequences), which uses the golden ratio as a basis. For the two-dimensional case (R_2-sequence), it produces more evenly spaced points than any of the other known methods. The generalized version of the golden ratio ϕd is defined as the unique positive root $\mathbf{x}^{d+1} = \mathbf{x} + 1$. That is, for $d = 2$, $\phi_2 = 1.3247\ldots$ This value was conjectured to most likely be the optimal value for a related two-dimensional problem (Hensley and Su, 2001). In two dimensions the $\mathbf{x}$ and $\mathbf{y}$ coordinates of the nth term ($n = 1, 2, 3, \ldots$) are defined as decimal place of:

$$\mathbf{x}_n = \left(0.5 + \frac{1}{\phi_2} \cdot n\right) \qquad \text{and} \qquad \mathbf{y}_n = \left(0.5 + \frac{1}{\phi_2^2} \cdot n\right) \tag{3.4}$$

Geostatistical Sampling

Model-based sampling techniques, specifically geostatistical sampling when the postulated model is a geostatistical model, have been applied in several recent studies to assess and locate additional monitoring wells in a GLMN (see section 3.1).

In geostatistical sampling, a geostatistical model, the model variogram, is used to identify locations for additional sampling points (Gruijter et al., 2006). These are selected based

on the predefined selection criterion, usually by minimizing the mean or the maximum kriging variance (Olea, 1984).

$$\sigma_{OK}^2 = Var(Z) - \sum_{\alpha=1}^{n} \lambda_\alpha C(h_{0\alpha}) \tag{3.5}$$

The kriging variance depends on the covariance model and the data configuration but is independent of the data values. For a given covariance model, two identical sample location distributions would independently yield the same kriging variance regardless of the data (Goovaerts, 1997). Heuvelink and Pebesma (2002) examined the validity of kriging variance by numerical analysis of two-dimensional Gaussian distributed realizations and by mathematical-statistical description. They show that the prediction error variance is independent of the data values. Therefore the kriging variance is still a correct assessment of the local uncertainty even if, in parts of the area, the variations are larger or smaller than elsewhere. However, these findings do not apply to the non-Gaussian case.

We used the maximum prediction standard error (square root of the kriging variance) as the selection criterion for placing additional sampling points. This approach is implemented in the *ArcGIS* Tool *"Densify Sampling Network"* and is therefore referred to as densify sampling network (DSN) (Johnston et al., 2004).

Reference

We introduce a reference method (REF) to evaluate the quality of the tested approaches. After each interpolation, the predicted surface is compared with the "real" surface, and an additional well is set at the point with the largest difference between the two surfaces. The idea behind this approach is that the resulting arrangement is the best possible design for each respective surface. Hence, we assume that it theoretically represents the lower limit for the smallest prediction error achievable with a given number of monitoring wells.

Progressive Versus Nonprogressive Designs

Progressive (sequential) sampling design creates an observation network by optimally adding one or several new points step by step, whereas, in nonprogressive (simultaneous) sampling design, all points have to be added at once. While some methods are either progressive or nonprogressive, others can be used in both ways (though they are usually referred to as progressive only). The latter is the case for SRS and the spatial coverage

sampling strategies (R_2 and Halton), where the results are the same whether one, several, or all points are added at a time. The DSN method constitutes an exception. It can be progressive or nonprogressive, but the results may differ. DSN places additional points at the highest prediction standard error, which depends on son the locations of the existing observations points and the semivariogram model. The semivariogram is recomputed in each densifying process and therefore leads to different results, depending on if a number of points are added one by one (sequential) or all at one time (simultaneously). In practice, the differences in the results are often minor, at least when a certain number of observations is reached since the addition of additional points only leads to minor changes in the modeled semivariogram. The grid sampling methods are nonprogressive, in the sense that only a certain number of points can be added at a time without breaking the grid symmetry (and therefore are usually referred to as nonprogressive only). Whether a progressive or nonprogressive method is preferred may depend from case to case. However, in GLMN design, progressive methods have the advantage of allowing for an extension of the network at a later stage.

3.2.4 Validation Methods

The absolute prediction error is the absolute mean difference between the real GWL and the predicted GWL, divided by the surface area. For the methods with a random component (REF, SRS, triangular random grid, and DSN), a mean value was calculated from 10 runs.

$$APE = \frac{\sum_{i=1}^{n} \left| \hat{Z}_{(\mathbf{x})} - Z_{(\mathbf{x})} \right|}{area} \tag{3.6}$$

The standardized absolute prediction error (SAPE) is the APE divided by the maximum APE of all methods per surface (except REF). The SAPE makes it easier to compare the error propagations of the individual surfaces with each other.

$$SAPE = \frac{APE}{\max APE_{all\,methods}} \tag{3.7}$$

The mean standardized absolute prediction error (MSAPE) is the mean SAPE of the individual design approaches for all nine surfaces. Since the real value for an actual GWL at the time of the prediction is generally unknown, the real prediction error is also unknown. The interpolation can therefore only be validated at the observation points, using CV and error statistics to assess the accuracy of the interpolation. Based on the results of the CV, the following error measures were used to compare the accuracy of

the different interpolation methods, where n is the number of observations, o_i is the observed value, and p_i is the predicted CV value at the position i.

The mean absolute error is the arithmetic mean of the absolute error values. It indicates the magnitude of the error:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|p_i - o_i| \tag{3.8}$$

The root mean square error represents the root of the averaged square error:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2} \tag{3.9}$$

The average standard error is the average of the prediction standard errors:

$$ASE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - (\sum_{i=1}^{n}p_i)/n)^2} \tag{3.10}$$

The root mean square standardized error with p_{si} as the standardized predicted value and o_{si}:

$$RMSSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_{si} - o_{si})^2} \tag{3.11}$$

The Nash–Sutcliffe model efficiency coefficient (Nash and Sutcliffe, 1970)) with μ_o as the arithmetic mean of the measured values is used to quantify how well a model simulation can predict the outcome variable. The NSE ranges from minus infinity to 1 (perfect fit):

$$NSE = 1 - \frac{\sum_{i=1}^{n}[p_i - o_i]^2}{\sum_{i=1}^{n}[o_i - \mu(o)]^2} \tag{3.12}$$

In a detailed review, Li and Heap (2008) have compiled a comprehensive assessment of error statistics. They conclude that MAE and RMSE are similar measures that estimate the average error but do not provide information about the relative size of the average difference or the nature of the difference. In contrast to MAE, RMSE is very sensitive to outliers (Hernandez-Stefanoni and Ponce-Hernandez, 2006; Ikechukwu et al., 2017; Vicente-Serrano et al., 2003; Willmott, 1982). Nonetheless, both are among the best measures of model performance (Willmott, 1982). The following criteria for using error measurements are proposed to assess the performance of spatial interpolation. MAE,

RMSE, and ASE should be as small as possible. ASE and RMSE should be nearly identical, and RMSSE should be close to 1, indicating the estimated prediction uncertainty is consistent. ASE > RMSE or RMSSE > 1 implies an overestimation of the variability of the predicted values; ASE < RMSE or RMSSE < 1 implies an underestimation (Hu et al., 2004).

3.3 Results and Discussion

We divided the presentation of the results into three sections. In section 3.3.1, the spatial distributions of the added monitoring wells from the tested approaches are shown and compared on example (ACF A). Since a priori known potentiometric groundwater surfaces are used, the CV errors and the "real" prediction errors based on the GLMN can be computed. Therefore, in section 3.3.2, a comparison of the tested design approaches based on these real prediction errors (APE) is drawn. Section 3.3.3 finally contains a comparison based on the CV results.

3.3.1 Resulting Spatial Distribution

The GLMN designs resulting from each approach are shown for the example of ACF A in Fig. 3.3. The reference method results in a network design with concentrated observation density and clustering in the more variable southern part of the area. In contrast, there are large regions with a low density even with 500 monitoring wells in the less variable northern part with lower gradients. This clustering is because an additional point is set at the location of the largest deviation between the interpolated and the actual surface (APE). Thus, the first 37 additional monitoring wells are placed exclusively in the south (light blue points). After that, points will also be set further north, although a concentration of additional points will remain in areas of high variability (cyan points).

SRS exhibits point clustering and regions without points. Since the points are placed randomly, this effect can be stronger or weaker from case to case. To take this random component into account, the placement was repeated ten times for SRS. The figure shows one of the ten examples.

To varying degrees, the low discrepancy methods (Halton and R_2 method) and the geospatial method DSN show uniformly distributed arrangements of the observation points. The Halton method shows a very even global point distribution. Locally, however, the method tends to place individual points closer together as the number of monitoring wells increases (cyan and light yellow points). DSN shows a uniform local point

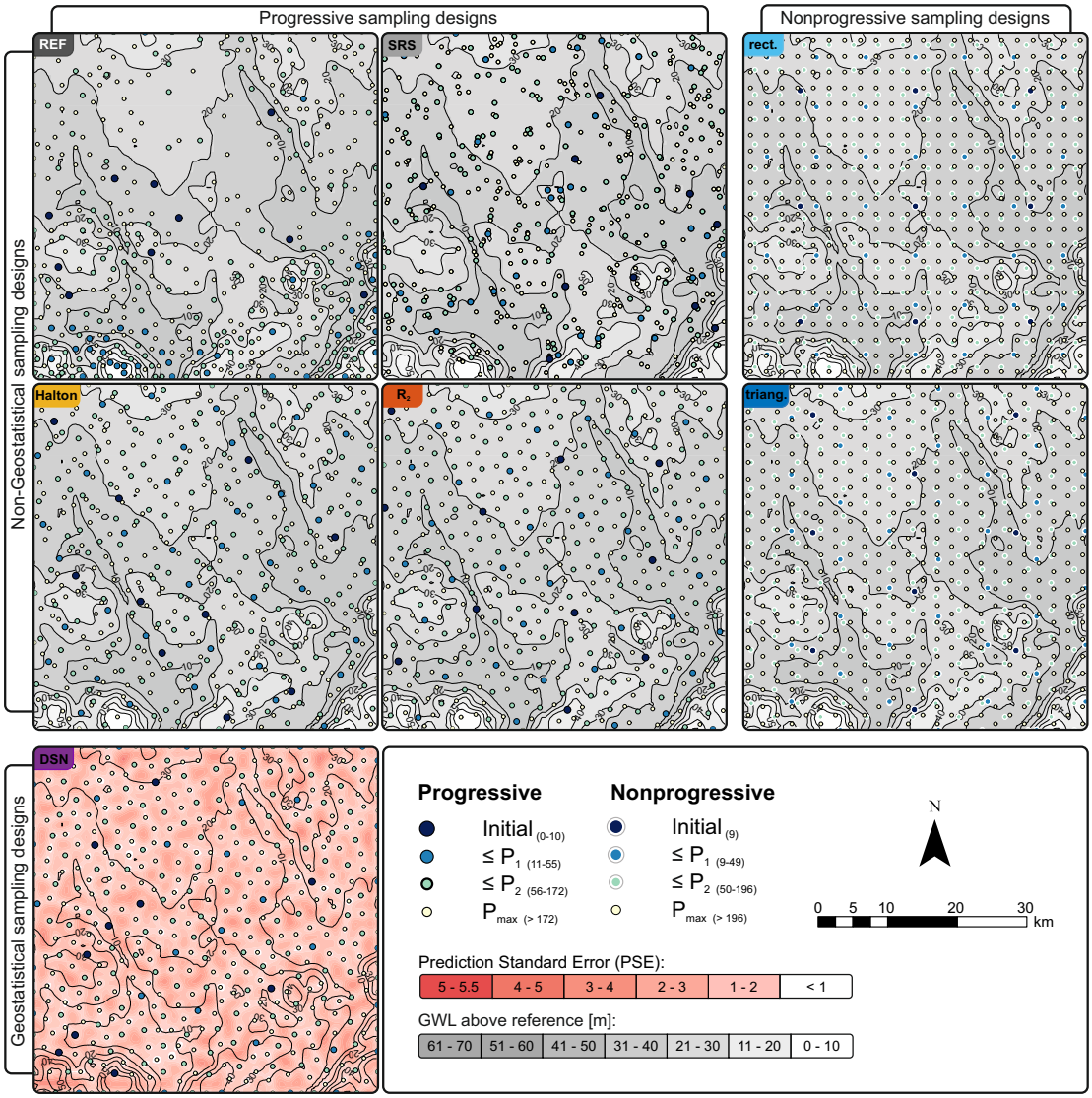


Figure 3.3: Resulting spatial distribution of the monitoring wells for surface ACF A based on the investigated network designs at the start (dark blue), less than 55 ($\leq P_1$, light blue), less than 172 ($\leq P_2$, cyan), and for 500 ($\leq P_{max}$, off white) monitoring wells (the numbers for the nonprogressive methods are adapted according to grid symmetry). Progressive: design that can be extended with n additional monitoring wells. Nonprogressive: design that is not extensible without breaking the symmetry of the grid (points marked by white outlines). The Geostatistical sampling design (DSN) uses the prediction standard error as a selection criterion for placing additional sample points.

distribution with relatively consistent neighboring distances. Since the kriging variance depends on the distance to the closest n-observation points, a high prediction standard error results at the boundary of the research area, and additional points are placed along

that edge. The R_2 method shows both an even global distribution and uniform distances to the neighboring n-observation points across all observation densities.

3.3.2 Absolute Prediction Errors

Fig. 3.4a shows the APE for all tested design approaches for surface ACF A as an example (diagrams for all surfaces are provided as supporting information, Fig. S3c) as a function of the number of monitoring wells. The points P_1, P_1 and P_2 were defined such as the numerical derivatives of the mean SAPE of all methods (excluding REF) are 5×10^{-3} and 1×10^{-3}, respectively. Consequently, a additional monitoring well results in a reduction of the initial mean SAPE by 0.5 % for P_1 and 0.1 % for P_2. Hence, it can be used to define a well density for the respective information/cost ratio.

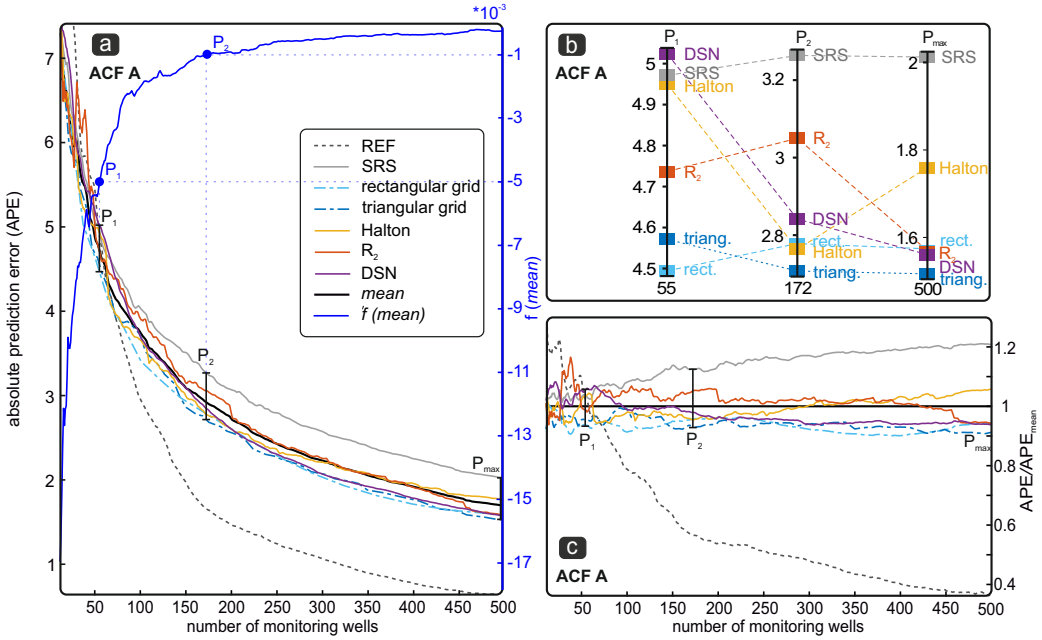


Figure 3.4: (a) APE of the tested methods for ACF A (left y-axes) as a function of the number of monitoring wells and the derivative of the mean SAPE of the methods (excluded REF; right y-axes). (b) Detailed view of P_1, P_2, and P_{max}. P_1 and P_2 are points at which the numerical derivative of the mean SAPE of the methods (excluded REF) has the value 5×10^{-3} and 1×10^{-3}. (c) The ratio of the APE of the individual design approaches to the mean APE of all methods (excluding REF). A value < 1 means that the individual design approach is better than average, > 1 that it is poorer than average.

The idea behind the thresholds $P_1$and P_2 is that below P_1, the observation density seems insufficient, and the placement of additional monitoring wells leads to a substantial improvement of the results. In contrast, above P_2 further monitoring wells show only minor

and thus possibly inefficient improvements. The decision at which information/cost ratio these points are defined can only serve as a rough guideline and must, of course, be adapted to the respective requirements of each site.

The theoretical maximum information content is reached at a density of 1 point per pixel. That is, the value is known for each grid cell. The grid size of the output maps needs to match the sampling density and scale at which the processes of interest occur (Hengl, 2009), since small-scale variabilities can only be displayed up to pixel size, and the coarser the resolution of a GWL, the more small-scale variations disappear. P_1 was achieved for ACF A with 55 observation points and P_2 with 172 observation points. Fig. 3.4b and 3.5b show a detailed view of P_1, P_2, and P_{max} (for the maximum tested number of 500 observation points). Fig. 3.4c and 3.5c show the deviations of the MSAPE of the individual design approaches from the mean MSAPE of all methods (except REF). A value < 1 means that the individual design approach performs better than average, > 1 that it performs worse than average.

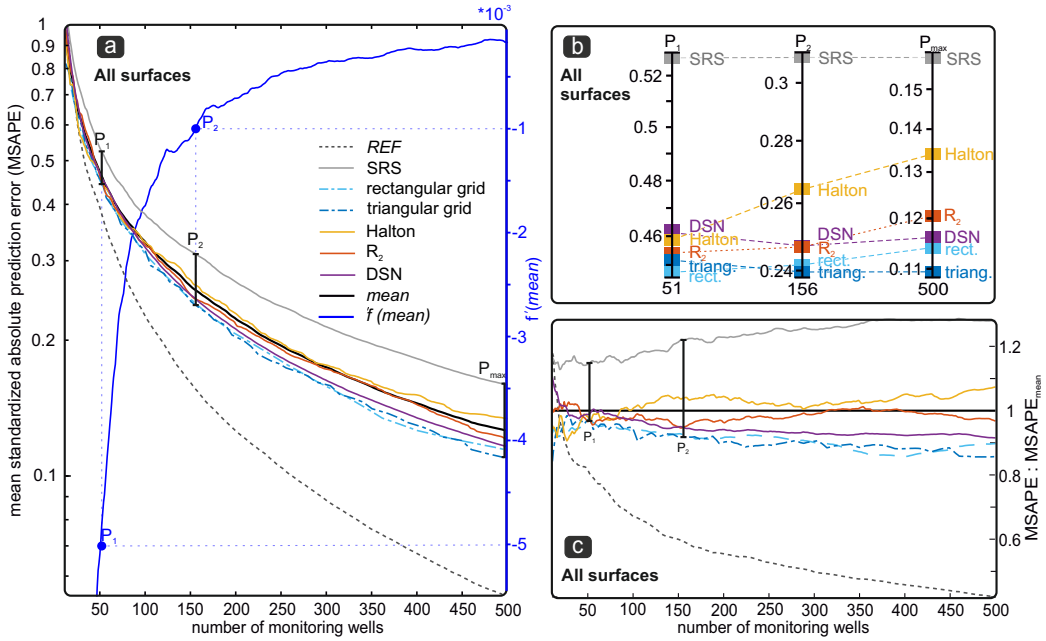


Figure 3.5: (a) MSAPE of the individual methods for all nine surfaces (left y-axes) and the derivative of the averaged MSAPE of the methods (excluded REF; right y-axes). P_1 and P_2 are points at which the numerical derivative of the averaged MSAPE of the methods (excluded REF) has the value 5×10^{-3} and 1×10^{-3}. (b) Detailed view of P_1, P_2, and P_{max}. (c) The ratio of the MSAPE of the individual design approaches to the averaged MSAPE of all methods (excluding REF). A value < 1 means that the individual design approach is better than average, > 1 that it is poorer than average.

3.3.3 Cross-Validation Results

Although assessing the error based on the deviation from the real GWL (APE) seems necessary to determine both the number of monitoring wells required and the best method for their placement, this deviation will not be available in practice. Only the error at the existing observation points can be determined by CV. Thus, we compare the APE with different CV error statistics to examine whether one of them is more appropriate as a proxy for the averaged APE and suitable point densities for assessing an information/cost ratio.

Figure 3.6a shows a comparison of the averaged standardized APE (SAPE) of all design approaches, and the averaged, standardized global CV error estimates (MAE, RMSE, ASE, and NSE), as well as the averaged RMSSE for all surfaces. Along with the averages, the ranges between the minimum and maximum curves are given as shaded areas in the same color. For a better comparison, NSE is shown as 1-NSE. The standardization aims to present the error developments of the different design approaches to be compared with each other. Furthermore, it should be shown whether and to what extent conclusions can be drawn from the shape of the individual CV curves compared to the SAPE curve regarding the assessment of an information/cost ratio.

Figure 3.6b Pearson r between the CV-error estimates and the APE. Except for RMSSE, all methods consistently strongly correlate with APE. RMSE (between 0.950 and 0.995) and MAE (0.965 and 0.997) show the strongest correlation to the APE, making them reliable qualitative error estimators. This is followed by 1-NSE (0.883-0.987). Since the RMSSE should go to 1, the Pearson r of $|1 - \text{RMSSE}|$ instead of RMSSE was calculated. This variable shows the worst correlation with the APE (-0.633 to 0.896) of the tested estimators.

Since both MAE and RMSE have the same units as the estimated quantity, Fig. 3.6c compares them with the APE to examine to what extent their value is suitable for quantifying the actual absolute error. Since RMSE squares the errors before averaging them, it gives a relatively high weight to large errors. Thus, the RMSE is clearly above the APE and the MAE for every surface. The APE corresponds approximately to $1/2$ RMSE and to $2/3$ MAE, which can, according to our data, therefore be used as a rough quantitative estimate for the actual error.

This can also be seen in Fig. 3.6d, which illustrates the development of the ratios of APE and the CV error estimators (APE/MAE,APE/RMSE,APE/ASE, and APE/(1-NSE)). The standard deviation caused by the different design approaches is essentially lower for RMSE (0.09) and MAE (0.08) than for APE (0.22), RMSSE (1.29), and 1-NSE (0.24).

1-NSE shows the most irregular course. The ratio changes from 0.7 at the beginning to 0.03 at 500 observation points.

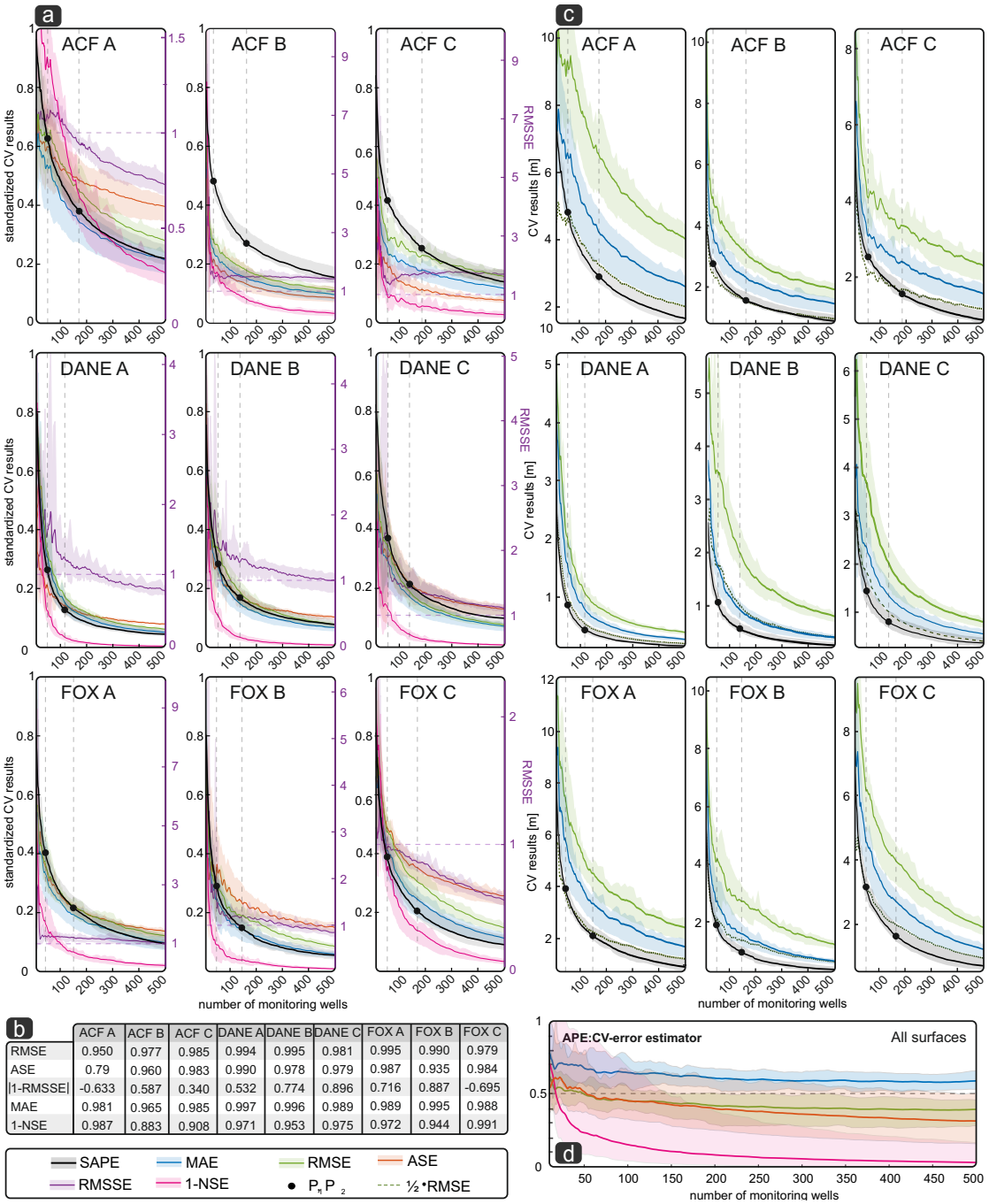
Fig. 3.7 shows a comparison of the APE for all design approaches (Fig. 3.7a) and the individual results of the CV errors (Fig. 3.7b) as a ranking at the points P_1, P_2, and with 500 monitoring wells (P_{max}) for all nine surfaces. The results are ranked according to the error size (RMSSE and NSE ranked by proximity to 1) for better clarity. A table with the actual errors can be found in the supporting information Tab. S3b. The results show that it is not possible to evaluate the different design approaches based on their CV errors. There is a moderate to strong negative Spearman rank correlation with increasing measurement density between the APE-based error ranking and the CV error estimates (except RMSSE). Thus, the SRS design, which led to the largest errors on any surface, is mainly classified as the supposedly best design approach. On average, the triangular grid producing the smallest APE is predominantly classified as the least suitable design. We, therefore, advise not to evaluate different designs based on their CV-error statistics.

3.4 Conclusions

The focus of this study was to compare different design approaches for GLMN to find out if there is an extensible design that allows reliable spatial estimates of GWL with an optimal information/cost ratio. Additionally, we examined what quality differences result from the use of different design approaches and which are the most suitable error statistics to evaluate the quality of the interpolated GWL. For this purpose, we used nine potentiometric groundwater surfaces extracted from three regional MODFLOW groundwater flow models to compare the interpolation results to an a priori reference. The sampling designs examined were random sampling, grid sampling (triangular and rectangular grid), spatial coverage sampling (low-discrepancy methods), and geostatistical sampling (densify sampling network).

The results show that the number of monitoring wells has a more beneficial influence on the interpolation result than their spatial distribution (design), as long as a reasonably even spatial distribution is given. All tested sampling designs led to significantly better results than SRS, but none of these designs proved superior to the others. Which method performs best is primarily dependent on the density of the GLMN and the characteristics of each individual potentiometric groundwater surface.

Interpolated GWL based on systematic sampling approaches (rectangular and triangular grid) showed, on average, the smallest actual APE at all observation densities. Due to



	ACF A	ACF B	ACF C	DANE A	DANE B	DANE C	FOX A	FOX B	FOX C
RMSE	0.950	0.977	0.985	0.994	0.995	0.981	0.995	0.990	0.979
ASE	0.79	0.960	0.983	0.990	0.978	0.979	0.987	0.935	0.984
|1-RMSSE|	-0.633	0.587	0.340	0.532	0.774	0.896	0.716	0.887	-0.695
MAE	0.981	0.965	0.985	0.997	0.996	0.989	0.989	0.995	0.988
1-NSE	0.987	0.883	0.908	0.971	0.953	0.975	0.972	0.944	0.991

Figure 3.6: (a) Comparison of the SAPE, RMSSE, and the standardized averaged MAE, RMSE, and ASE along with the ranges (shaded colors). (b) P_{max} between the SAPE and the respective CV results. (c) Comparison of the APE, the MAE, and RMSE. (d) Development of the ratio of APE and the individual CV results (mean values of all tested methods and areas) for APE:MAE,APE:RMSE,APE:ASE, and APE:(1-NSE).

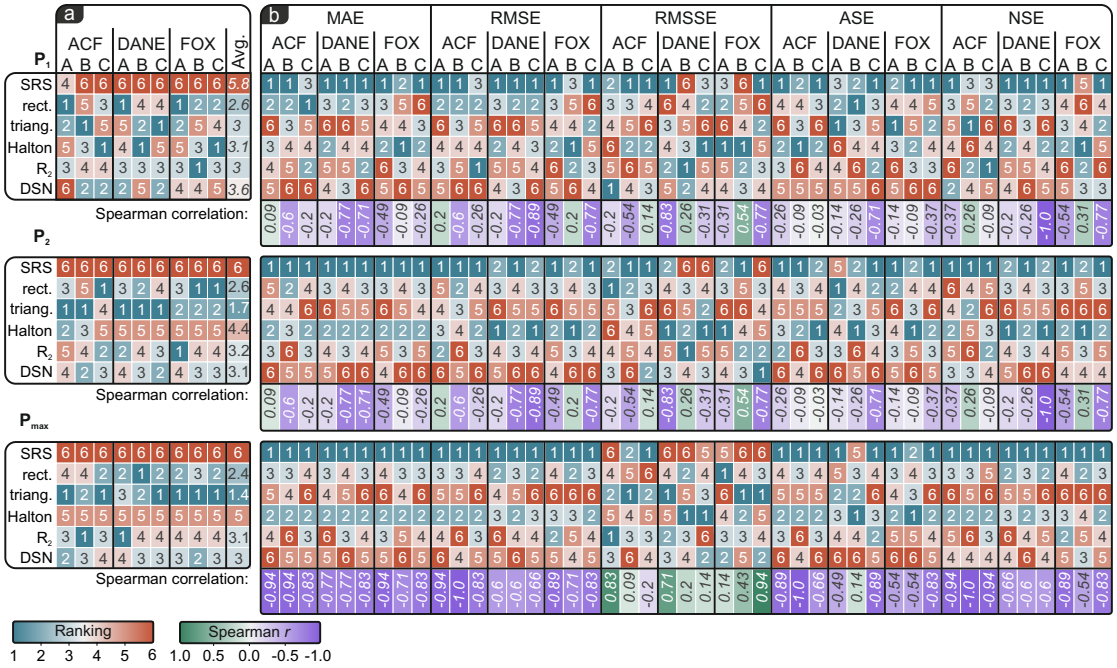


Figure 3.7: Comparison of the (a) ranking of the six investigated methods on all surfaces based on their absolute prediction errors (APE) and (b) the CV error estimates (MAE), (RMSE), (RMSSE) and ASE, NSE, from best (1, blue) to worst (7, red, RMSSE and NSE ranked by proximity to 1), along with the Spearman rank correlation between the APE and the CV error estimates from 1 (green) to -1 (purple).

their nonprogressive nature, they are only suitable for constructing a new GLMN with a defined number of monitoring wells, which will not be extended in the future. On average, the triangular grid design showed better results than the rectangular grid design.

The geostatistical DSN method, in which the location of an additional observation point is selected based on the maximum prediction standard error, resulted in the lowest APE among the tested progressive network designs. Despite its insignificantly higherAPE than the grid designs, DSN has the advantage that the resulting design can be sequentially extended and is, therefore, more appropriate for most GLMNs. However, since DSN selects new observation points based on the maximum prediction standard error, it requires an existing kriging interpolation. Therefore, it can only be used for adding new sampling locations to an existing GMN.

Consistently good results have been achieved with the low-discrepancy methods (Halton and R_2 method), which, to our knowledge, have not yet been used for GLMN designs before. Moreover, their locations are only based on mathematical sequences and can be determined without prior measurements. Furthermore, the placement of the observations points is reproducible and does not change over time (with possibly different

measurements, as is the case for DSN). Among the low-discrepancy methods, the R_2 method delivers better results than the Halton method and should be preferred.

Based on the SAPE, we defined the points P_1 and P_2 where an additional well leads to a reduction of the initial mean SAPE by 0.5 % for P_1 and 0.1 % for P_2. Below P_1, the observation density seems insufficient, and the placement of additional monitoring wells leads to a substantial improvement of the results. In contrast, above P_2 further monitoring wells only lead to minor and thus inefficient improvements. On average, for all surfaces and methods, the observation density for P_1 is 5.1×10^3/pixel and 1.56×10^2/pixel for P_2, respectively, that is, about 0.51 % of all possible sampling options (imaginary pixels of a grid reflecting the assumed variability of the GWL should be sampled, while on average when over 1.56 % of all possible sampling options are sampled, the error reduction becomes considerably less, and the resulting information/cost ratio might become too low. By comparing density per pixel with a real observation density in monitoring wells per km^2, it must be assumed that the resolution of the grid is adapted to the actual variability of the GWL to be expected. Therefore, P_1 and P_2 can be used as rough guideline values for the required number of monitoring wells in the planning of a GLMN as well as the evaluation of an existing one.

From the results of global CV, we conclude that one should avoid comparing different designs based on the global average CV error estimation since there are strong negative correlations between APE and the CV-error estimates, especially at higher observation point densities. Thus, the CV error statistics are not appropriate for evaluating the results of different methods and comparing different design approaches. Which methods perform best can differ significantly from the actual error depending on the surface and the CV statistics used. The basic benefit of CV comes not from using it in a global sense but rather in looking at the spatial distribution of the individual CV errors (correlated residuals, bias, normal distribution, etc.). According to our results, the CV error statistics, especially MAE and RMSE, can be helpful as a rough quantitative estimate for the actual error, though.

Supporting Information

This supporting information provides a comparative overview the important parameters and properties of the GWL model surfaces (S3a), overviews of the investigated methods based on their APE for all nine surfaces (S3b) and a comparison of the actual errors with the CV error estimates (S3c)

Table S3a: Overview of important parameters and properties of the groundwater model surfaces. For easier comparison, all groundwater heights were converted from the original height above sea level to a minimum value of zero.

Name	Low. Apalachicola-Chattahoochee-Flint River Basin, Georgia/Florida, USA			Dane County, Wisconsin, USA			Upper Fox River Basin, Wisconsin, USA		
Author	(Jones et al., 2017)			(Parsen et al., 2016)			(DeSimone et al., 2002)		
Shorthand symbol	ACF A	ACF B	ACF C	DANE A	DANE B	DANE C	FOX A	FOX B	FOX C
Number of pixels	100×100			100×100			100×100		
Pixel size [m]	500			100	200	100	200		
Total Area [km^2]	2,500			100	400	100	400		
Main aquifer system	Unconfined to semiconfined part of the Floridan Aquifer system partially covered with residuum limestone			Unlithified deposits on high permeable fractured sandstones and dolomites (unconfined to semiconfined)			allow, unconfined aquifer-system of glacial/alluvial deposits and Silurian dolomite		
GWS difference$_{max}$ [m]	67.69	64.49	75.33	26.86	25.96	53.79	76.47	80.13	74.86
Mean GWS height μ [m]	19.57	23.77	24.51	12.48	11.67	35.93	32.2	47.12	28.84
Std dev of GWS height, σ [m]	9.66	10.57	14.13	5.86	6.55	9.19	17.39	13.86	10.81
Hydraulic gradientmax$_{max}$ [%]	2.24	1.23	1.31	1.34	0.93	3.26	3.76	2.78	3.64
Hydraulic gradient$_{mean}$ [%]	0.35	0.21	0.25	0.44	0.23	0.57	0.77	0.42	0.62
Hydraulic gradient$_{std\ dev.}$ [%]	0.33	0.14	0.17	0.21	0.14	0.47	0.49	0.45	0.55

Table S3b: Comparison of the (a) Results of the six investigated methods and REF on all surfaces based on their absolute prediction errors (APE) as well as the CV error estimates (b) mean absolute error (MAE), root mean square error (RMSE), root mean square standardized error (RMSSE) average standard error (ASE) and Nash-Sutcliffe model efficiency coefficient (NSE)

a.

	APE									MAE									RMSE								
	ACF			DANE			FOX			ACF			DANE			FOX			ACF			DANE			FOX		
P_1	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
SRS	4.9	2.9	2.9	1.1	1.2	1.5	4.2	2.2	3.6	4.6	2.9	2.9	1.0	1.1	1.5	3.9	2.2	3.4	7.4	4.0	4.0	1.4	1.5	2.3	5.3	4.0	4.9
rect.	4.4	2.8	2.5	0.8	0.9	1.3	3.6	1.7	3.0	6.4	3.1	2.6	1.3	1.4	1.9	5.0	2.9	5.0	9.1	4.3	3.4	1.6	1.8	2.7	6.1	4.6	6.5
triang.	4.5	2.6	2.7	0.8	0.9	1.1	3.7	1.9	3.2	7.8	3.4	3.7	2.0	1.7	2.2	5.2	2.8	4.3	11.1	4.5	5.2	2.6	2.3	3.2	6.6	4.2	5.7
Halton	4.9	2.8	2.4	0.8	0.8	1.3	4.0	1.8	2.8	6.5	3.6	3.0	1.1	1.6	2.0	4.8	1.6	4.3	9.6	4.8	4.4	1.5	2.1	2.7	5.9	2.4	6.5
R_2	4.7	2.8	2.5	0.8	0.9	1.3	3.7	1.7	3.1	6.5	3.7	2.8	1.5	1.7	1.9	6.8	2.4	4.6	9.2	4.8	3.4	2.1	2.1	2.9	8.8	3.9	6.3
DSN	5.0	2.7	2.4	0.8	1.0	1.2	4.0	1.9	3.3	7.4	3.8	3.8	1.4	1.6	2.3	5.5	3.4	4.6	1.6	5.0	5.3	1.9	2.0	3.3	7.2	5.0	6.4
									Pearson correlation:	-0.2	-0.6	-0.1	-0.4	-0.8	0.9	-0.6	0.0	-0.7	-0.1	-0.6	0.0	-0.4	-0.8	-0.9	-0.5	0.1	-0.8
P_2																											
SRS	3.3	1.8	2.0	0.6	0.7	1.0	2.5	1.2	2.1	3.3	1.8	1.9	0.6	0.7	0.9	2.5	1.1	2.0	5.3	2.6	3.0	0.9	1.0	1.5	3.6	2.2	3.2
rect.	2.8	1.5	1.6	0.4	0.5	0.7	2.0	0.8	1.5	4.4	2.2	2.6	0.7	0.9	1.2	3.2	1.7	2.8	6.7	2.9	3.5	1.0	1.2	1.6	4.3	2.9	4.0
triang.	2.7	1.5	1.7	0.4	0.5	0.6	1.9	0.8	1.5	4.4	2.3	2.8	0.9	0.9	1.4	3.8	1.8	2.8	6.6	3.0	3.7	1.2	1.2	2.0	5.0	3.1	4.2
Halton	2.8	1.5	1.8	0.4	0.6	0.8	2.1	1.1	1.7	4.0	2.2	2.2	0.6	0.8	1.0	2.9	1.2	2.4	6.1	3.1	3.1	0.8	1.0	1.4	4.0	2.1	3.6
R_2	3.1	1.5	1.6	0.4	0.5	0.7	1.9	1.0	1.6	4.1	2.5	2.3	0.7	0.9	1.2	3.6	1.5	2.8	6.1	3.3	3.1	1.1	1.1	1.7	4.8	2.7	4.1
DSN	2.8	1.6	1.6	0.4	0.5	0.7	2.0	0.8	1.5	4.8	2.0	2.6	0.8	1.0	1.4	3.5	1.8	3.0	7.7	3.2	3.8	1.1	1.4	2.1	4.6	3.1	4.4
									Pearson correlation:	-0.7	-0.8	-0.6	-0.7	-0.8	-0.9	-0.9	-1.0	-0.9	-0.7	-0.8	-0.5	-0.6	-0.7	-0.7	-0.9	-1.0	-0.9
P_{max}																											
SRS	2.0	1.1	1.1	0.2	0.3	0.5	1.2	0.4	1.0	2.0	1.1	1.2	0.2	0.3	1.2	1.2	0.5	1.0	3.4	1.6	1.8	0.3	0.5	0.7	1.9	1.1	1.6
rect.	1.6	0.8	0.8	0.1	0.2	0.3	0.8	0.3	0.6	2.6	1.5	1.7	0.3	0.4	1.7	1.7	0.6	1.2	3.9	1.9	2.5	0.4	0.6	0.8	2.4	1.1	1.9
triang.	1.5	0.8	0.8	0.1	0.2	0.3	0.7	0.3	0.6	2.9	1.6	1.8	0.3	0.4	2.0	2.0	0.7	1.4	4.3	2.1	2.5	0.4	0.6	0.9	2.7	1.3	2.1
Halton	1.8	0.9	0.9	0.2	0.3	0.4	1.0	0.4	0.8	2.4	1.3	1.4	0.2	0.4	1.5	1.5	0.6	1.1	3.6	1.7	2.0	0.4	0.5	0.8	2.3	1.1	1.8
R_2	1.6	0.8	0.8	0.1	0.3	0.4	0.9	0.4	0.7	2.8	1.6	1.6	0.3	0.4	1.7	1.7	0.7	1.3	4.2	2.1	2.3	0.4	0.6	0.8	2.3	1.3	1.9
DSN	1.6	0.8	0.8	0.1	0.2	0.3	0.8	0.3	0.6	2.9	1.6	1.7	0.3	0.4	1.8	1.8	0.7	1.3	4.4	2.0	2.5	0.4	0.6	0.9	2.5	1.2	2.0
									Pearson correlation:	-1.0	-0.9	-0.9	-0.8	-0.9	-1.0	-1.0	-0.8	-1.0	-0.9	-0.9	-0.9	-0.8	-0.9	-0.9	-0.9	-0.6	-0.9

b.

	|1- RMSSE|									ASE									|1 - NSE								
	ACF			DANE			FOX			ACF			DANE			FOX			ACF			DANE			FOX		
P_1	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
SRS	0.1	0.1	0.0	0.0	0.1	0.2	4.8	0.9	3.9	6.1	2.9	2.2	1.2	1.2	1.1	4.8	2.6	4.9	0.66	0.17	0.09	0.06	0.06	0.07	0.10	0.10	0.24
rect.	0.1	0.5	0.3	3.1	0.4	0.2	5.3	0.4	5.5	8.1	3.0	3.2	1.0	1.2	1.3	5.3	3.1	6.6	0.90	0.18	0.06	0.08	0.07	0.09	0.13	0.11	0.36
triang.	0.1	0.6	1.0	0.6	0.2	0.4	4.8	0.3	4.7	9.8	2.9	3.8	0.7	1.4	1.7	4.8	3.2	6.1	1.01	0.13	0.16	0.20	0.13	0.10	0.12	0.09	0.27
Halton	0.2	0.4	0.1	0.7	0.2	0.2	5.1	0.1	5.5	7.3	2.7	3.0	1.8	1.5	1.5	5.1	2.8	6.5	0.99	0.19	0.11	0.07	0.10	0.09	0.11	0.03	0.37
R_2	0.1	0.8	0.3	0.2	0.2	0.3	6.7	0.1	5.3	7.8	4.8	3.3	1.2	1.8	1.2	6.7	3.0	6.2	1.02	0.16	0.00	0.13	0.10	0.09	0.21	0.08	0.39
DSN	0.1	0.6	0.1	0.7	0.3	0.3	5.8	0.2	5.4	9.7	3.6	3.0	1.5	1.6	1.8	5.8	4.3	6.6	0.79	0.17	0.11	0.08	0.90	0.10	0.16	0.09	0.31
Pearson:	*0.0*	*-0.6*	*0.3*	*-0.5*	*-0.4*	*-0.5*	*-0.4*	*0.8*	*-0.8*	*-0.2*	*0.0*	*-0.5*	*0.0*	*-0.4*	*-0.7*	*-0.4*	*-0.2*	*-0.8*	*-0.5*	*0.6*	*0.2*	*-0.3*	*-0.8*	*-1.0*	*-0.5*	*0.2*	*-0.8*
P_2																											
SRS	0.0	0.1	0.4	0.1	0.2	0.1	3.2	0.0	2.2	5.3	1.8	1.5	0.9	3.2	0.7	1.9	1.9	3.8	0.31	0.01	0.04	0.00	0..2	0.03	0.04	0.03	0.09
rect.	0.0	0.3	0.8	0.4	0.3	0.2	3.6	0.3	3.0	6.8	2.0	1.9	0.8	3.6	0.9	2.2	2.2	4.7	0.49	0.01	0.06	0.00	0.03	0.03	0.06	0.05	0.14
triang.	0.1	0.4	0.9	0.0	0.1	0.3	3.9	0.3	3.2	6.9	2.2	2.0	0.8	3.9	1.2	2.2	2.2	4.9	0.44	0.01	0.07	0.00	0.04	0.04	0.08	0.05	0.16
Halton	0.1	0.4	0.9	0.2	0.2	0.0	3.8	0.2	2.6	6.3	1.8	1.7	0.9	3.8	0.9	1.7	1.7	4.1	0.37	0.01	0.05	0.02	0.02	0.02	0.05	0.02	0.11
R_2	0.0	0.6	0.9	0.0	0.0	0.3	3.9	0.1	3.1	6.3	3.3	1.9	0.8	3.9	0.9	2.3	2.3	4.5	0.45	0.01	0.05	0.00	0.03	0.03	0.07	0.04	0.15
DSN	0.0	0.7	0.7	0.3	0.3	0.2	3.9	0.1	3.4	7.5	2.2	2.2	0.9	3.9	1.9	2.6	2.6	4.8	0.44	0.01	0.06	0.00	0.04	0.04	0.07	0.04	0.15
Pearson	-0.5	-0.6	-0.8	0.0	0.0	-0.7	-0.9	-0.7	-0.9	-0.8	-0.4	-0.7	0.5	-0.9	-0.9	-0.8	-0.8	-0.9	-0.6	-0.3	-0.3	-0.6	-0.7	-0.6	-0.9	-0.9	-0.8
P_{max}																											
SRS	0.3	0.3	0.2	0.6	0.3	0.1	1.9	0.2	0.6	4.7	1.2	1.1	0.5	0.6	0.5	1.9	1.3	2.9	0.13	0.00	0.02	0.00	0.00	0.01	0.01	0.01	0.02
rect.	0.3	0.4	0.9	0.2	0.1	0.0	2.4	0.2	0.9	5.5	1.3	1.3	0.5	0.6	0.6	2.4	1.3	3.4	0.17	0.00	0.03	0.00	0.01	0.01	0.02	0.01	0.03
triang.	0.2	0.3	0.5	0.3	0.1	0.1	2.4	0.1	1.1	5.6	2.4	1.3	0.5	0.6	0.7	2.4	1.3	3.6	0.20	0.03	0.03	0.00	0.01	0.01	0.02	0.01	0.04
Halton	0.3	0.4	0.8	0.2	0.0	0.0	2.2	0.1	0.8	5.0	1.2	1.2	0.5	0.5	0.6	2.2	1.2	3.2	0.14	0.03	0.02	0.00	0.01	0.01	0.02	0.01	0.03
R_2	0.2	0.4	0.6	0.3	0.0	0.0	2.4	0.2	0.9	5.3	2.1	1.3	0.5	0.6	0.6	2.4	1.4	3.4	0.19	0.04	0.03	0.00	0.01	0.01	0.02	0.01	0.03
DSN	0.3	0.5	0.7	0.2	0.0	0.0	2.4	0.2	1.0	5.8	1.5	1.4	0.5	0.6	0.7	2.4	1.5	3.4	0.18	0.30	0.03	0.00	0.01	0.01	0.02	0.01	0.03
Pearson	*0.8*	*-0.3*	*-0.7*	*0.9*	*-0.1*	*-0.1*	*-0.9*	*0.4*	*-0.9*	*-0.9*	*-0.6*	*-0.8*	*-0.8*	*-0.1*	*-0.9*	*-0.9*	*-0.5*	*-1.0*	*-0.9*	*-0.9*	*-0.9*	*-0.9*	*-0.9*	*-0.9*	*-0.9*	*-0.4*	*-0.8*

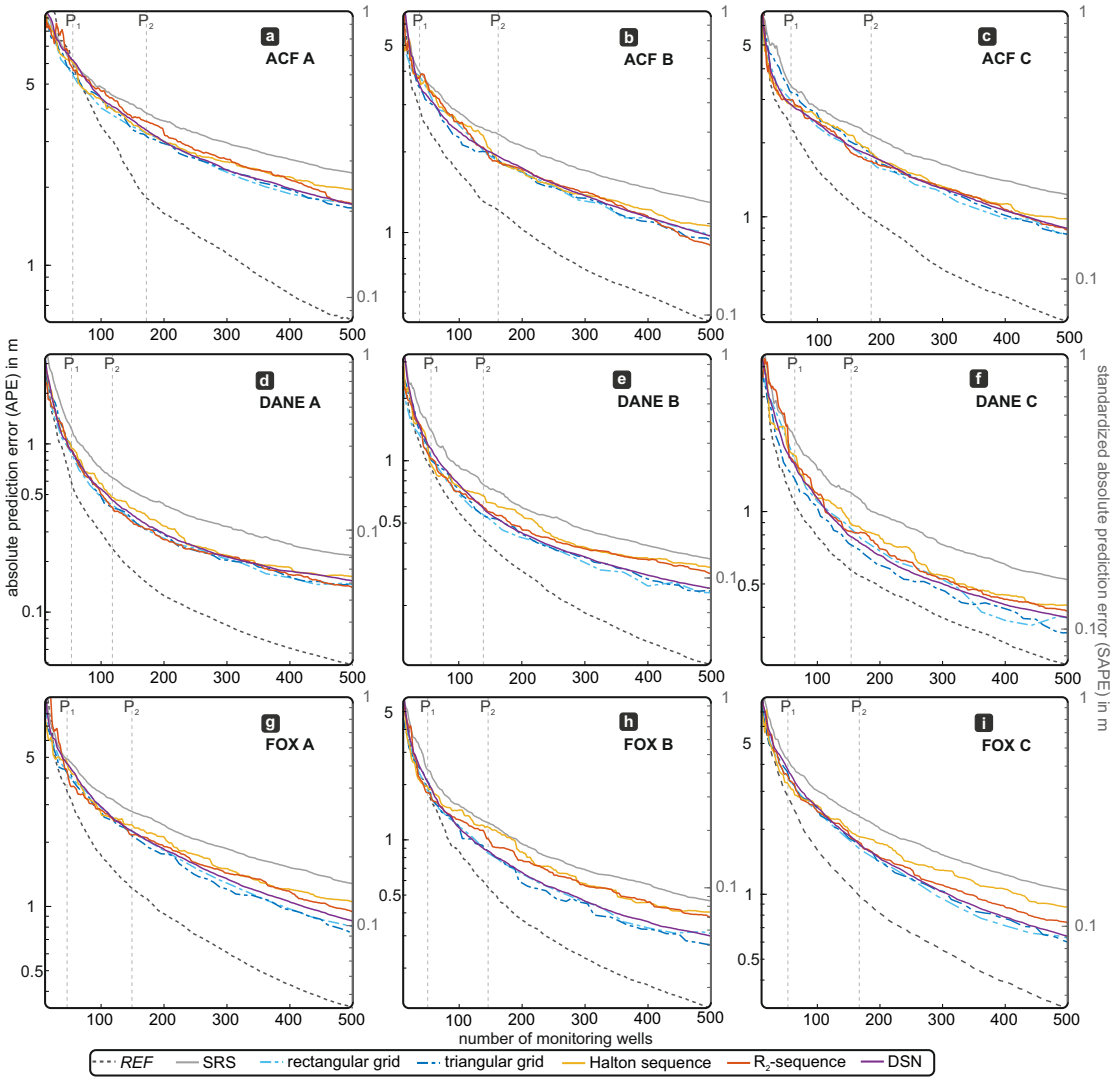


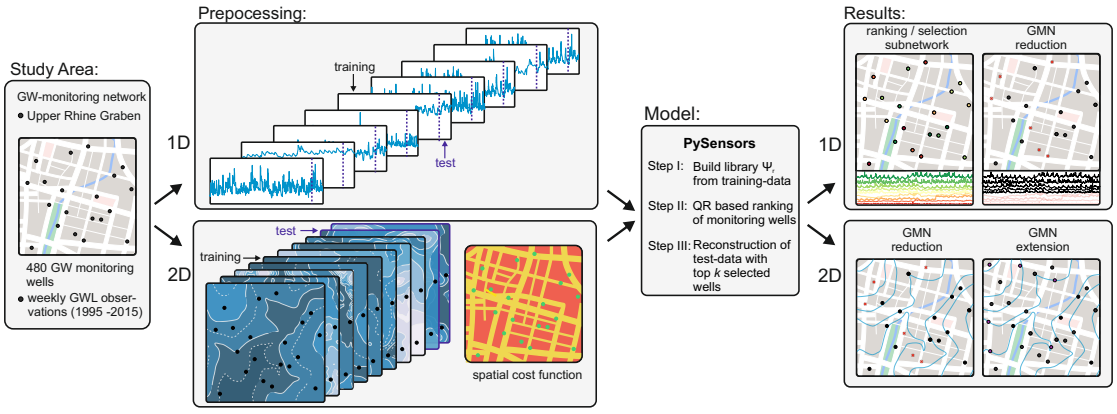
Figure S3c: APE, (semi-log plot) and SAPE of the tested methods for all nine surfaces. P_1 and P_2 are points at which the numerical derivative of the mean APE of the methods (excluded REF) has the value 0.05 and 0.01. Consequently, an additional monitoring point results in a reduction of the initial mean SAPE by 0.5 % for P_1 and 0.1 % for P_2 and hence can be used for the definition of a well density for the respective information/cost ratio.

Chapter 4

Spatiotemporal Optimization of Groundwater Monitoring Networks Using Data-driven Sparse Sensing Methods

This chapter is based on a study published in the Copernicus journal *Hydrology and Earth System Sciences (HESS)*.

The remainder of this part is a reprint of: Ohmer, M., Wunsch, A., Liesch, T., (2022). Spatiotemporal optimization of groundwater monitoring networks using data-driven sparse sensing methods. *Hydrology and Earth System Sciences Discussions*, 26, 4033-4053, doi: 10.5194/hess-26-4033-2022.



Abstract

Groundwater monitoring and specific collection of data on the spatiotemporal dynamics of the aquifer are prerequisites for effective groundwater management and determine nearly all downstream management decisions. An optimally designed groundwater monitoring network (GMN) will provide the maximum information content at the minimum cost (Pareto optimum). In this study, PySensors, a Python package containing scalable, data-driven algorithms for sparse sensor selection and signal reconstruction with dimensionality reduction is applied to an existing GMN in 1D (hydrographs) and 2D (gridded groundwater level contour maps). The algorithm first fits a basis object to the training data and then applies a computationally efficient QR algorithm that ranks existing monitoring wells (for 1D) or suitable sites for additional monitoring (for 2D) in order of importance, based on the state reconstruction of this tailored basis. This procedure enables a network to be reduced or extended along the Pareto front. Moreover, we investigate the effect of basis choice on reconstruction performance by comparing three types typically used for sparse sensor selection (i.e., identity, random projection, and SVD, respectively, PCA). We define a gridded cost function for the extension case that penalizes unsuitable locations. Our results show that the proposed approach performs better than the best randomly selected wells. The optimized reduction makes it possible to adequately reconstruct the removed hydrographs with a highly reduced subset with low loss. With a GMN reduced by 94 %, an average absolute reconstruction accuracy of 0.1 m is achieved, in addition to 0.05 m with a reduction by 69 % and 0.01 m with 18 %.

4.1 Introduction

Groundwater is a vital resource for drinking water supply and industrial, commercial, and agricultural uses. Therefore, effective groundwater management and monitoring practices are critical to ensure the availability and quality of water supplies for future generations. A GMN is defined by a spatial arrangement of groundwater monitoring wells and a temporal sampling frequency (Loaiciga et al., 1992). In most cases, there are economic interests behind groundwater management and thus also behind a monitoring network. As a result, while many monitoring networks meet the basic requirements for groundwater management, they are scientifically insufficient to monitor aquifer dynamics. Considering monitoring costs and monitoring quality (i.e., the information gained by monitoring) as axes in a two-dimensional coordinate system, optimal GMNs lie along a Pareto front on which the maximum information content is achieved for the respective

budget. Moreover, existing GMNs are usually grown historically regarding the locations and number of monitoring wells and are therefore primarily inefficient. This means that the monitoring quality is relatively low for the given costs. Thus, optimization could reduce the operating costs without loss of monitoring quality by optimizing the monitoring network regarding the number of wells and their location (Emmert et al., 2016). Furthermore, directives such as the European Water Framework Directive (WFD; EC, 2000) or the European Nitrates Directive (ND; EC, 1991) demand the integration of regional monitoring networks into national or international networks. Selecting a reasonable subset of these networks capable of capturing the dynamics of the groundwater body is an essential and challenging task.

Usually, GMNs are classified according to their purpose into a groundwater quality monitoring network (GQMN), i.e., mostly multivariate, and groundwater quantity/ groundwater level monitoring network (GLMN), i.e., univariate. This classification does not preclude a GMN from performing both tasks. However, optimization approaches usually address one of the two tasks. To date, there are neither standard regulations for the planning and expansion of existing GMNs nor established methods. Instead, a high degree of subjectivity prevails. In the last few decades, many studies have been published dealing with the optimization of GMNs. The widely varying requirements for optimizing monitoring networks led to various approaches that attempt to meet these requirements differently. The choice of method usually depends on the GMN type (GQMN vs. GLMN), scale (local, regional, and national), uni- or multivariate network, optimization strategy (extension of GMN vs. reduction of redundant wells), consideration of dynamics (spatial vs. spatiotemporal), and, last, purpose of the monitoring network. With the latter, a distinction is usually made between risk-oriented monitoring (mainly concerning groundwater quality in the catchment of waterworks) and surveillance monitoring, e.g., according to the European WFD.

In general, the design of a monitoring network is considered a nonlinear and non-convex optimization problem whose optimal criterion measures the useful information contained in the information matrix of the design (Ushijima et al., 2021). GMN optimization approaches are commonly divided into the following three categories based on the techniques applied: (a) those based on hydrogeological conceptual models and hydrogeological expert knowledge, (b) those based on numerical groundwater flow models (Kim and Lee, 2007; Singh and Datta, 2016; Sreekanth et al., 2017; Thakur, 2017), and (c) those based on data analysis with (geo-)statistical techniques. Many studies have focused on the geostatistical ability of kriging frameworks to determine new monitoring wells based on the reduction of estimation variance as the optimization criterion (Bhat et al., 2015; Li et al., 2011; Nunes et al., 2004; Ohmer et al., 2019; Thakur, 2015; Varouchakis and

Hristopulos, 2013). With the steady increase in computational capacity in recent years, there are a growing number of studies that tackle these optimization problems using traditional datadriven heuristic optimization criteria such as genetic algorithms (GAs; Ayvaz and Elçi, 2018; Dhar and Patil, 2012; Khader and McKee, 2014; Komasi and Goudarzi, 2021; Pourshahabi et al., 2018; Puri et al., 2017; Reed and Kollat, 2013; Yudina et al., 2021), artificial neural networks (ANNs; Alizadeh et al., 2018), particle swarm optimizations (PSOs; De Jesus et al., 2021; Gaur et al., 2013; Guneshwor et al., 2018), support vector machines (SVMs; Asefa et al., 2004; Bashi-Azghadi and Kerachian, 2010) and relevance vector machines (RVMs; Ammar et al., 2008; Khalil et al., 2005), or a combination of these approaches. Further studies use entropy and information-theory-based approaches (Alizadeh and Mahjouri, 2017; Hosseini and Kerachian, 2017; Keum et al., 2017) and Kalman filtering (KF; Júnez-Ferreira et al., 2016; Kollat et al., 2011).

In most of the mentioned studies, the optimization of the GMN amounts to a computationally intensive combinatorial search with innumerable multi-dimensional iteration steps, since many complex physical systems are described by a high-dimensional state [$\boldsymbol{x} \in \mathbb{R}^n$]. Moreover, improved data recording and increasing storage leads to fast and strongly growing system complexities and, therefore, to increased computing time beyond Moore's law (Moore, 1965). However, the dynamics of such complex systems often evolve on a low-dimensional attractor, which can be used to predict and control these systems. Pattern extraction is associated with the search for coordinate transformations that simplify the system dynamic and the computational effort (Brunton and Kutz, 2017). In recent years, powerful new techniques in data science have been developed that are capable of analyzing complex data and extracting essential features and correlations from high-dimensional dynamic systems. Sparse sampling (Baraniuk, 2007; Candes and Wakin, 2008; Candes et al., 2005), sparse reconstruction (Annoni et al., 2018; Castillo and Messina, 2020; Yildirim et al., 2009), and sparse classification (Brunton et al., 2016), enable the recovery of relevant information from remarkably few measurements. Although sparse sampling, such as compressed sensing, is a common and powerful method often used in other fields of science including seismic and medical image processing, fluid dynamics, or remote sensing, to our knowledge, there are only a few studies in the field of hydrogeology that applied sparse sensing for hydrogeological tasks. Hussain and Muhammad (2013) utilized sparse signal extraction methods based on l_1-norm minimization to exploit the spatial sparsity in hydrodynamic models and thereby reduce the number of measurements needed to reconstruct the signal. Lee et al. (2021) used compressed sensing for generating groundwater level (GWL) contour maps based on sparsely sampled or incomplete data from a groundwater model below the Nyquist–Shannon sampling criterion (Shannon, 1949). They found that compressed

sensing performed much better compared to traditional interpolation methods such as kriging. Ushijima et al. (2021) developed an experimental design algorithm to select locations for a network of monitoring wells with maximum information. The combinatorial search was performed with a genetic algorithm (GA) combined with a proper orthogonal decomposition (POD) to reduce the computational cost of using the GA. POD, which is often formulated using the singular value decomposition (SVD), is a dimensionality reduction method that extracts relevant large coherent structures/patterns (lowdimensional features) from high-dimensional data (Pollard et al., 2017).

This study focuses on a data-driven algorithm to optimize a GMN regarding the number and locations of monitoring wells for temporal and spatial GWL reconstruction. The algorithm uses data-driven sparse-sensing techniques and QR-based sensor placement algorithm that ranks sensors, in our case monitoring wells, according to their information content. It is based on work by Silva et al. (2021), Manohar et al. (2018) and Clark et al. (2019), implemented in the *PySensors* package, and has, e.g., been successfully applied in a similar context of sensor placement for sea surface temperature reconstruction, fluid flow data (Clark et al., 2019; Manohar et al., 2018), and wind flow data (Annoni et al., 2018), as well as for classification tasks in, e.g., image recognition or cancer classification by microarrays (Brunton et al., 2016). We have adapted this methodology for the first time for the application to GMN, as we see the following advantages over existing methods: (i) it can simultaneously take spatial and temporal information into account, (ii) it allows the ranking of existing monitoring wells based on their information content, (iii) based on the ranking, an existing network can be reduced, while the values either of the abandoned wells or a spatially continuous GWL can be reconstructed, (iv) it proposes locations for an extension of a network that account for the best possible gain in knowledge, and (v), if necessary, allows the application of a cost function for the extension of the network (Clark et al., 2019), either to prefer more suitable locations (e.g., in terms of infrastructure) or to exclude certain areas (like inaccessible terrains, steep slopes, etc.).

We apply the adapted algorithm to a real-world GLMN to demonstrate its suitability for groundwater monitoring networks in general. The data set used for this purpose consists of weekly GWL monitoring between 1990 and 2015 from 480 monitoring sites in the Upper Rhine Graben (URG)'s upper alluvial aquifer. In particular, we show how the algorithm can be applied to address the following questions regarding the optimization of an existing network:

- What is the ranking of monitoring wells in an existing network in terms of their information content/reconstruction performance, i.e., in which order should the wells be removed if a network reduction is desired?

- How does the reconstruction/interpolation error vary as wells are progressively removed from the monitoring network in the order of the proposed ranking, and how does this compare to the removal of randomly selected wells?
- When the goal is network extension, where should new wells be placed for maximum information gain? How significant is the increase in information, i.e., how much will the spatial reconstruction error be reduced?
- How well does a combined reduction/extension (i.e., replacement) of a certain number of wells perform compared to a straightforward extension?

4.2 Methodology

4.2.1 Mathematical Background

Compressed Sensing

Most multi-dimensional natural signals are compressible (respectively sparsely representable). That mean that when the signals are transformed into a convenient coordinate system (basis), only a limited number of basis modes are active. These basis modes correspond to the large mode amplitudes (Brunton and Kutz, 2017). In data compression, for example, JPEG or MP3 compression, only these values are stored to efficiently reconstruct the input signal with a considerable reduction in data size and little loss of information. A compressible signal $\boldsymbol{x} \in \mathbb{R}^n$ can be written as a sparse vector $s \in \mathbb{R}^n$ on a new orthonormal basis of $\boldsymbol{\Psi} \in \mathbb{R}^{n \times n}$ such that, in the following:

$$\boldsymbol{x} = \boldsymbol{\Psi} s \tag{4.1}$$

Vector $\boldsymbol{s}$ is K-sparse if it is a linear combination of only K basis vectors (exactly K nonzero elements). The theory of compressed sensing uses this principle as it attempts to infer the sparse representation $\boldsymbol{s}$ in a known transformed basis system with a very small, low-dimensional (compressed) subsample.

$$\boldsymbol{y} = \mathbf{C}\boldsymbol{x} = (\mathbf{C}\boldsymbol{\Psi})\boldsymbol{s} = \boldsymbol{\Theta}\boldsymbol{s} \tag{4.2}$$

where the vector $\boldsymbol{y} \in \mathbb{R}^p$ is a set of incoherent observations and $\mathbf{C} \in \mathbb{R}^{p \times n}$ an observation matrix of p linear observations. $\boldsymbol{\Theta}$ is the condition number. The objective in compressed

sensing is to find the l_1-norm of sparsest vector $\hat{\boldsymbol{s}}$ (under a set of conditions) that is consistent with $\boldsymbol{y}$, as follows:

$$\boldsymbol{s} = \arg\min_{\boldsymbol{s'}} \left\|\boldsymbol{s'}\right\|_1 \qquad \text{such that} \qquad \boldsymbol{y} = \boldsymbol{\Theta}\boldsymbol{s'} \tag{4.3}$$

which almost certainly end up with the sparsest possible solution for $\boldsymbol{s}$ (Baraniuk, 2007; Candes and Wakin, 2008; Candes et al., 2005; Donoho, 2006).

Sparse Sensor Placement

While compressed sensing uses random measurements to reconstruct high-dimensional unknown data from a universal basis $\boldsymbol{\Psi} \in \mathbb{R}^{n \times n}$, data-driven sparse sensor placement collects available information about a signal from observed samples to build up a tailored basis $\boldsymbol{\Psi}_r \in \mathbb{R}^{n \times r}$ for the respective signal and thus to identify optimal sensor placements for the reconstruction of this signal with low-losses. Let the full signal be an unknown linear combination of basis coefficients $\boldsymbol{a} \in \mathbb{R}^r$ (vector of mode amplitudes of $\boldsymbol{x}$ in basis Ψ):

$$\boldsymbol{x} = \sum_{k=1}^{r} \psi_k \boldsymbol{a}_k = \boldsymbol{\Psi}_r \boldsymbol{a} \tag{4.4}$$

The central challenge is to design an incoherent (i.e. rows of $\mathbf{C}$ not correlated with columns ψ of $\boldsymbol{\Psi}_r$) measurement matrix $\mathbf{C}$ that allows to identify the optimal p observations $\boldsymbol{y}_i$ to accurately reproduce the signal $\boldsymbol{x}$, as follows:

$$\boldsymbol{y} = \mathbf{C}\boldsymbol{x} = (\mathbf{C}\boldsymbol{\Psi}_r)\boldsymbol{a} = \boldsymbol{\Theta}\boldsymbol{a} \tag{4.5}$$

For n sensor observations and a given p sensor budget, the sampling matrix $\mathbf{C}$ must be structured as follows:

$$\mathbf{C} = [\mathbf{e}_{\gamma 1} \quad \mathbf{e}_{\gamma 2} \quad \dots \quad \mathbf{e}_{\gamma p}]^T \tag{4.6}$$

Here $\mathbf{e}_j \in \mathbb{R}^n$ are the canonical basis vectors with unit entry at index j and zeros elsewhere. Thus each row of $\mathbf{C}$ only observes from a single spatial location, corresponding to the sensor location. The observations are made up of p elements selected from $\boldsymbol{x}$,

$$\boldsymbol{y} = \mathbf{C}\boldsymbol{x} = [\boldsymbol{x}_{\gamma 1} \quad \boldsymbol{x}_{\gamma 2} \quad \dots \quad \boldsymbol{x}_{\gamma p}]^T \tag{4.7}$$

with $\gamma \in \mathbb{N}^p$ as an index set of the sensor locations that designates the index with cardinality $|\gamma| = p$ and, additionally, the number of sensors $n \geq r$ of $\boldsymbol{\Psi}$ for a well-defined linear inverse problem (Manohar et al., 2018). The unknown $\boldsymbol{x}$ can thus be reconstructed by approximating $\boldsymbol{a}$ with the Moore-Penrose pseudoinverse of (4.5) to the following:

$$\mathbf{C}^\star = \underset{\mathbf{C} \in \mathbb{R}^{p \times n}}{\arg\min} |\boldsymbol{x} - \boldsymbol{\Psi}(\mathbf{C}\boldsymbol{\Psi})^\dagger \boldsymbol{y}|_2^2 \tag{4.8}$$

Where † denotes the Moore-Penrose pseudoinverse, it is assumed that optimal sensor selection $\mathbf{C}^\star$ is mostly a sparse subset selection operator, and the nonzero entries in the rows represent the monitoring wells.

Taylored Basis $\boldsymbol{\Psi}_r$

As described above, in data-driven sparse sensing, the universal basis $\boldsymbol{\Psi}$ is replaced by a tailored basis $\boldsymbol{\Psi}_r$, which is built from the training data $\mathbf{X}^{tr}$, e.g., by using dimensionality reduction techniques. In this study, we are using the following three basis types which are typically used for sparse sensor selection:

- **Identity basis**: Centered raw data is used directly without dimensionality reduction. $\boldsymbol{\Psi}_r = \mathbf{X}^{tr}$. Since no low-rank approximation of the data is performed, no information is lost. However, this comes at the cost of a longer computation time (Silva et al., 2021).

- **Random Projection basis**: Dimensionality is reduced by projecting the input data onto a randomly generated matrix $\boldsymbol{\Psi}^r = \mathbf{G}\mathbf{X}^{tr}$ where the entries $\mathbf{G} \in \mathbb{R}^{2p \times m}$ are drawn from a Gaussian density function with mean zero and variance $1/n$ (Dasgupta, 2000; Li et al., 2006).

- **SVD/principle component analysis (PCA)**: Linear dimensionality reduction is performed using a truncated SVD. SVD is a numerically robust and efficient method for extracting dominant patterns from low-dimension (Golub and Kahan,

1965; Halko et al., 2011). For a matrix $\mathbf{X} \in \mathbb{C}^{n \times n}$, the SVD is given by the following:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \mathbf{\Psi}\mathbf{\Sigma}\mathbf{V}^T \approx \mathbf{\Psi}_r\mathbf{\Sigma}_r\mathbf{V}_r^T \qquad \text{where} \qquad \mathbf{U} \in \mathbb{R}^{n \times r}, \mathbf{\Sigma} \in \mathbb{R}^{r \times r}, \mathbf{V} \in \mathbb{R}^{m \times r}, \tag{4.9}$$

The columns of $\mathbf{\Psi}$ are the left singular vector of $\mathbf{X}$. They are often termed as spatial correlations, principal components, features or POD modes of the data set. Where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices and $\mathbf{\Sigma}$ is a diagonal matrix.

QR Pivoting for sparse sensors

While the previous steps serve to best fit the basis to the training data, the next steps aim to determine the resulting optimal sensor locations that minimize the reconstruction error. This optimization problem is solved using an approximate greedy solution using reduced QR factorization with column pivoting (Brunton and Kutz, 2017; Halko et al., 2011). The QR factorization decomposes a matrix $\boldsymbol{a} \in \mathbb{R}^{m \times n}$ into a unitary matrix $\mathbf{Q}$ and an upper-triangular matrix $\mathbf{R}$ and a column permutation matrix $\mathbf{C}$ (4.6), such that $\mathbf{A}\mathbf{C}^T = \mathbf{Q}\mathbf{R}$. The diagonal inputs of $\mathbf{R}$ are determined by selection of the pivot columns with maximum l_2-norm within all modes in the library. Subsequently, the orthogonal projection of the pivot column is then subtracted from all other columns, and the process is iteratively repeated over all columns. Thus, QR factorization with column pivoting yields r column indices (which correspond to sensor locations) that best sample the r basis modes (columns) $\mathbf{\Psi}_r^T$.

$$\mathbf{\Psi_r}^T\mathbf{C}^T = \mathbf{Q}\mathbf{R} \tag{4.10}$$

Since the pivot columns represent the sensors, the QR factorization results in a hierarchical list of all n pivots, where the first p pivots are optimized for the reconstruction of $\mathbf{\Psi}_r$. This means that, in the GMN optimization based on hydrograph data, all monitoring wells are ranked based on their information content. When using spatial input data, e.g., from interpolation or model results, all gridded input data cells are ranked based on their information content. Thus, it allows recommendations for the placement of additional monitoring wells at locations with supposedly high information content. The used QR decomposition approach includes a cost constraint function (Clark et al., 2019). This constraint allows different costs to be considered when selecting sensor placement, such as favoring or excluding certain areas.

4.2.2 Application Cases

In principle, there are the following two possible application cases of the algorithm regarding groundwater monitoring data: (i) the application to the observed data at the wells (i.e., hydrographs) only and (ii) the application of spatially continuous gridded information that has been regionalized based on the well data (e.g., by interpolation). While the optimization based on hydrographs only serves to rank the individual wells of the network according to their information content and thus to identify and eliminate redundant wells, the spatially continuous input data also allow a GMN extension at optimal locations and the best possible reduction of the spatial prediction error.

4.2.3 Error metrics

The calculation of the reconstruction error for a given set of measurements is done using the root mean square error (RMSE) for the scoring function. We further used the following metrics widely used for calibration and evaluation of hydrological models: mean absolute error (MAE), Nash–Sutcliffe model efficiency coefficient (NSE), Kling-Gupta efficiency (KGE), squared Pearson's correlation coefficient (R^2), and relative Bias (rBias). In the following equations, o stands for observed values, r for the reconstructed values, cov is the covariance, σ is the standard deviation, μ is the arithmetic mean, n stands for the number of measurements.

The RMSE is one of the most commonly used error index statistics. In general, the lower the RMSE, the better the model performance. It is useful for comparing different models performances for a given time series. However, only the relative root mean square error (rRMSE) is meaningful in comparing the model performance between different time series.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[o_i - r_i]^2} \quad \text{and} \quad rRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{o_i - r_i}{o_{max} - o_{min}}\right]^2} \tag{4.11}$$

Analogous to the RMSE, the smaller the MAE, the better the performance, as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|o_i - r_i| \tag{4.12}$$

The NSE (Nash and Sutcliffe, 1970) is a widely used goodness of fit measure of hydrologic models, as it normalizes model performance into an interpretable scale (Knoben et al., 2019). The NSE ranges between -∞ and 1, where 1 indicates a perfect correspondence

between observations and reconstructions, while a NSE = 0 indicates that the model has the same explanatory power as $\mu(o)$.

$$NSE = 1 - \frac{\sum_{i=1}^{n}[o_i - r_i]^2}{\sum_{i=1}^{n}[o_i - \mu(o)]^2} \tag{4.13}$$

The KGE (Gupta et al., 2009) was proposed as an alternative to the NSE because it addresses several shortcomings of the NSE (Knoben et al., 2019). Like the NSE, a KGE = 1 indicates a perfect model correspondence. However, explicit statements on benchmark performance have varied so far.

$$KGE = 1 - \sqrt{[r-1]^2 + [\alpha - 1]^2 + [\beta - 1]^2} \tag{4.14}$$

with

$$r = \frac{cov(o,r)}{\sigma(o)\sigma(r)}, \quad \alpha = \frac{\sigma(r)}{\sigma(o)}, \beta = \frac{\mu(r)}{\mu(o)} \tag{4.15}$$

where r is the linear correlation between o and r, α is a measure of the variability error, and β is a bias term.

We use the squared Pearson (r, eq. 4.15) correlation coefficient as a general coefficient R^2. It describes the degree of collinearity between measured and reconstructed data. R^2 ranges from 0 to 1, with higher values indicating lower error variance. In general, values above 0.5 are considered acceptable.

$$R^2 = \left[\frac{cov(o,r)}{\sigma(o)\sigma(r)}\right]^2 \tag{4.16}$$

The relative bias is a measure for a systematic over- or underestimation of a model. The optimal rBias is 0. Positive values indicate model underestimation of bias; negative values indicate model overestimation of bias (Gupta et al., 1999).

$$rBias = \frac{1}{n}\sum_{i=1}^{n}[\frac{\mathrm{o_i} - \mathrm{r_i}}{o_{max} - o_{min}}] \tag{4.17}$$

Statements about model performance in section 4.3 section are based on Moriasi et al. (2007) guidelines for model evaluation.

4.2.4 Data and Study Area

Hydrogeological Framework

The Upper Rhine Graben (URG), also known as Rhine Rift Valley, is a 300 km long and, on average, 50 km wide structural trough. It was formed in the Oligocene in response to the alpine orogenesis and subsequently filled with fluvial to lacustrine sediments of the Late Miocene, Pliocene, and Quaternary (Przyrowski and Schäfer, 2015). The Pliocene and Quaternary alluvial gravels and sands represent the largest groundwater reservoir in central Europe (LUBW, 2006). Based on their permeability and the appearance of finegrained horizons, the Pliocene and Quaternary gravels are subdivided into three (locally also more) aquifers, partly separated by fine-grained sediments (Wirsing and Luz, 2007). The study region is in the Baden-Württemberg part of the URG (Fig. 4.1). The Rhine forms the western boundary, Kaiserstuhl volcano complex is the southern boundary. To the East, the URG is bounded by a rift flank uplift composed of a system of troughs and highs, which follow the ENE structural grain of the Variscan fold belt (Derer, 2003). Along the study area these are, from South to North, the Black Forest high, Kraichgau-basin, and Odenwald-Spessart-high. Groundwater recharge occurs predominantly through lateral inflow and infiltration of streams from the Black Forest valleys in the east, the Freiburg basin in the southwest, and the infiltration of the Rhine and other surface waters.

Data and Preprocessing

The data set used in this study consists of weekly GWL measurements from 480 wells in the uppermost aquifer within the Quaternary sand/gravel deposits of the URG, covering the period from 1990 to 2015 (i.e., 1304 time steps). Data values that deviate by more than $\pm 3\ \sigma$ from the moving average (with a window size of 11 values) are considered outliers and were removed from further processing. Data gaps were subsequently filled based on information from highly correlated neighboring hydrographs using the clustering results of Wunsch et al. (2021). Where this did not yield plausible results or was not possible due to similarly missing data in neighboring hydrographs, an alternative PCHIP (piecewise cubic hermite interpolating polynomial) interpolation was performed. To avoid possible bias, the measurement data are globally and locally centered. The data set is split into two subsets: the first 80 % (January 1990-December 2009, 1043 time steps) is used to train the algorithm and the last 20 % for test/validation (January 2009-December 2014, 261 time steps). The well data are publicly available from the Baden-Württemberg State Office for Environment web service (LUBW, 2021b).

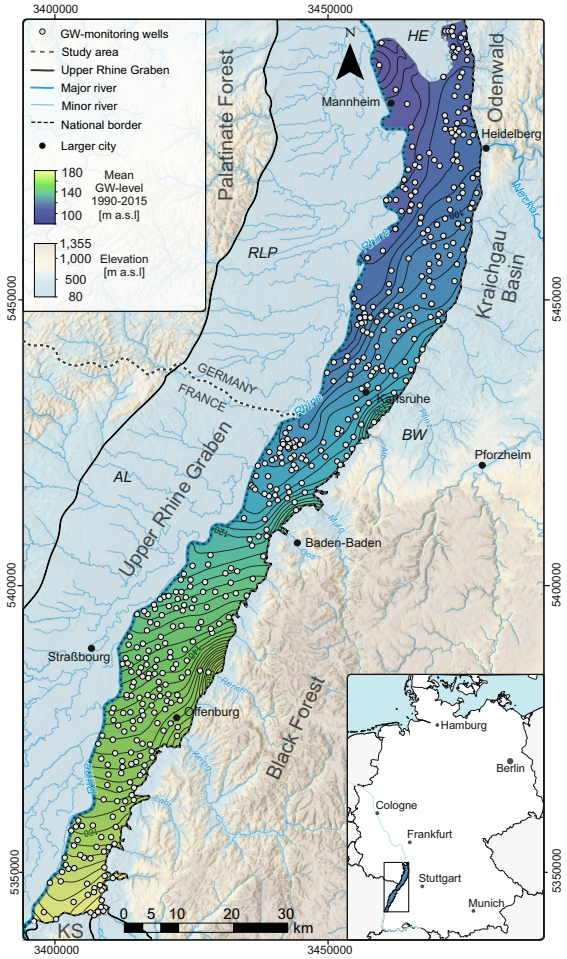


Figure 4.1: Study area within the URG and the 480 monitoring wells used for optimization. RLP: Rhineland-Palatinate; He: Hesse, BW: Baden-Württemberg, Al: Alsace (France); KS: Kaiserstuhl volcanic complex.

Groundwater level contour maps

We used the hydrograph data to generate 1304 weekly GWL contour maps with a grid size of 50×50 m using ordinary kriging. In addition to finding a sparse set of monitoring wells for the optimal temporal reconstruction of other hydrographs, the objective is to identify monitoring wells that allow optimal spatiotemporal reconstruction of GWLs from a subset of the wells. Moreover, the spatially continuous information of the gridded contour maps is used to suggest additional locations for an extended network. We used an isotropic Gaussian semivariogram model for interpolation, which is flexible and, therefore, a good candidate for a standard model (Krivoruchko, 2011). The associated parameters partial sill (42.7 m), range (17 853 m), lag size (1485 m), and nugget (0.05 m) were optimized using automatic cross-validation (CV) diagnostics to achieve the

lowest mean square error. It should be noted that the use of a single variogram model (Gaussian) may not be the optimal way to quantify spatial correlation, especially for nonstationary data. However, this is a necessary simplification due to the automation process that still produces comparable interpolation results, while the best possible interpolation result is not the focus of this study. Just as with the hydrographs, the first 80 % (January 1990–December 2009; 1043 time steps) of the contour maps were used to train the algorithm and the last 20 % for test/validation (January 2009–December 2014;261 time steps).

4.3 Results and Discussions

The following section is structured as follows. First, the grid search results regarding the three types of basis used, the number of basis modes, and a varying number of sensors are presented and discussed. Since we are applying the presented sensor placement approach to a groundwater monitoring well optimization, we use the term "well" as a synonym for sensors in the following. This is followed by the results of the GMN optimization based solely on the hydrograph data set. Finally, the results of the GMN optimization with the interpolated GWL as inputs are shown.

4.3.1 Grid Search Results

Fig. 4.2a shows the RMSE between the estimated and actual GWLs for the validation set as a function of the number of wells that the network is reduced to, the type of tailored basis (identity basis, random projection, and SVD basis), and the number of basis modes used. Since the number of wells and basis modes interacts, it is necessary to determine the appropriate number of basis modes that will result in the lowest reconstruction error for a given number of wells. If the number of wells is close to the number of basis modes, the reconstruction error increases significantly for all three basis mode types used. While the number of basis modes for identity and random projection is theoretically open-ended, the dimensionality for SVD, and thus the number of basis modes, must be less than the number of wells. Although there are SVD methods (e.g., randomized SVD; Halko et al., 2011) that allow oversampling with an additional number of random vectors, our results show that the accuracy of SVD generally decreases as the number of basis modes increases, which is consistent with the findings of Silva et al. (2021). Therefore, we decided against an SVD method with oversampling and opted for the truncated SVD implemented in PySensors (Silva et al., 2021), thus using a maximum of 480 basis modes. All 1043 time steps (in the training set) were used as

the maximum number of basis modes for identity and random projection. According to previous studies, the number of basis modes should be at least equal to the number of wells $p + 10$ (Silva et al., 2021). Clark et al. (2019) used $2p$ basis modes, which in our case equals a maximum of 960 (for all 480 wells) and, thus, is covered by the maximum of 1043 basis modes in the grid search.

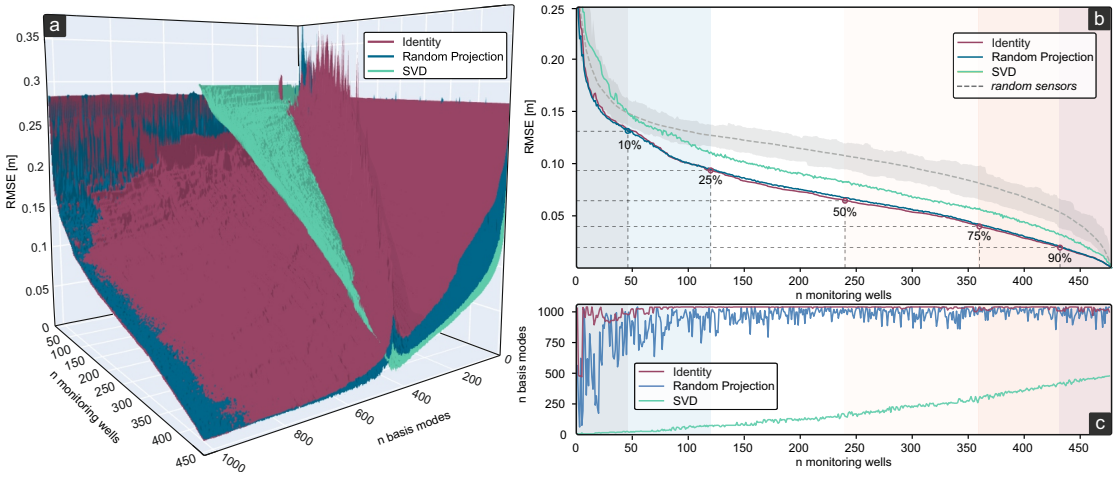


Figure 4.2: Grid search CV results. (a) Reconstruction error (RMSE) vs. the number of monitoring wells and the number of basis modes for the three basis projections described in basis: identity ($\mathbf{\Psi}_r = \boldsymbol{x}^{tr}$), random projection ($\mathbf{\Psi}_r = \mathbf{G}\mathbf{X}^{tr}$), SVD ($\mathbf{\Psi}_r = \mathbf{U}^{tr}$). (b) Minimum RMSE achieved with a given number of QR-selected monitoring wells (Pareto front) and the same number of randomly selected wells (median of 100 runs, grey shading represents the total range) for comparison. (c) The corresponding number of basis modes, that leads to the lowest RMSE in b. Accordingly, b and c represent a section through a at the respective minimum on the axis n monitoring wells.

The results show that, with only a few basis modes, SVD has the highest accuracy (Fig. 4.2a). As the number of basis modes exceeds the number of remaining wells, the identity and random projection basis perform better. In general, random projection and identity basis perform similarly. With fewer than 950 basis modes, slightly better results are obtained with random projection; above 950, the identity basis performs marginally better. Fig. 4.2b shows the lowest RMSE per number of wells achieved in the grid search, and Fig. 4.2c shows the corresponding number of basis modes. The gray dashed line in Fig. 4.2b shows the median of reconstruction errors from 100 iterations, with a random well selection as a benchmark. Except for SVD basis with fewer than 50 wells, all three basis types perform considerably better compared to reconstruction with the randomly placed wells and independently of the number of wells. Our findings are consistent with those of Clark et al. (2019) and Manohar et al. (2018), and Silva et al. (2021), where SVD consistently underperforms compared to random projection and identity basis. However, the latter two show almost identical results with an optimized number of basis modes

(Fig. 4.2b), with random projection performing marginally better with a small number of wells and identity basis performing slightly better with a larger number of wells. For consistency reasons, and based on the grid search results, we decided to compute the optimization steps shown in the following uniformly with identity basis and a fixed number of 1043 basis modes. This combination, on average, shows the best results for any chosen number of remaining wells (Fig. 4.2c).

As examples, all results of the following section are presented using five reduction stages: 10 %, 25 %, 50 %, 75 %, and 90 %. Thus, in the 25 % stage, the 25 % wells with the lowest information content are removed, and their hydrographs are predicted using the remaining optimal 75 % of the monitoring wells. The reduction stages are also shown in the grid search results (Fig. fig:2b and c). The color scheme of the reduction stages is kept for all remaining figures.

4.3.2 Ranking of Wells and Network Reduction with Hydrograph Data

Figure 4.3 shows the result of the ranking of the monitoring wells computed with an identity basis and using 1043 basis modes. The ranks are assigned from 1 (essential well with high information content) to 480 (most redundant well); thus, lower numbers mean a higher ranking concerning their information content and importance to reconstructing potentially removed redundant wells.

The reduction stage at which the respective wells are removed is indicated in parentheses. The color scheme ranges from dark red for redundant monitoring wells that can be eliminated with a minor loss in prediction accuracy to dark blue for important monitoring wells that contain essential information about the system and are needed for the accurate reconstruction of signals at other monitoring wells. In addition, Fig. 4.4 shows the centered hydrographs of the most important 10 % and the most redundant 10 % of all hydrographs. Most of the important wells (blue; removal at a reduction of > 90 %) show a pronounced flashiness (i.e., high frequency and rapidity of shortterm changes) and strong irregular patterns during the recording period. These dynamics indicate a strong interaction with surface waters or boundary inflows, for example, from side valleys of the rift flanks, which can also be seen in Fig. 4.3 from the location of the wells. Additionally, the most important wells include those with a distinct trend, which can be best seen for the two highest-ranked wells at the bottom, showing an upward trend over the considered period.

In contrast, the redundant wells show low flashiness and also include wells with high seasonality, though most of the signals seem to be dominated by interannual variations.

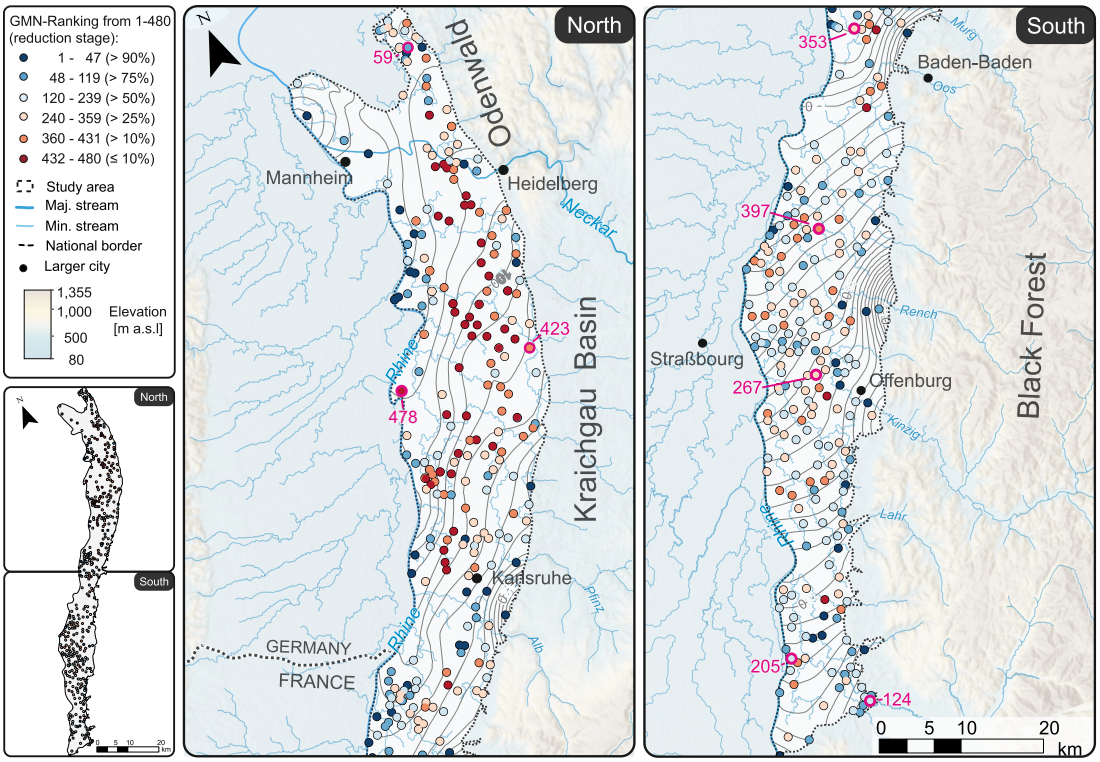


Figure 4.3: QR based ranking of monitoring wells based on hydrograph information content, from 1 (high information content, blue) to 480 (redundant, red). The hydrographs of the highlighted wells (pink) are shown in Fig. 4.6.

Most of these wells are located in the northern part of the study area within the URG. Since the eastern boundary in this area is the Kraichgau basin, the landscape profile is less pronounced than in the Black Forest hill range in the south and the Odenwald in the north. Therefore, less recharge occurs through stream infiltration, which is often the reason for more pronounced short-term variations or flashiness. The hydrographs of all wells can be found in the Appendix. Overall, the ranking shows that the most important wells include the ones with a noticeable unusual behavior, i.e., patterns that are not present in many of the other wells (like flashiness, trends, and jumps) and thus are hard to reconstruct. Overall, the more redundant wells either show a higher seasonality or tend to show low variability. Both patterns are common to a larger number of the wells and can be reconstructed more easily.

While the ranking itself already contains essential information and could be used, for example, to equip higherranked monitoring wells with higher-quality sensors or measure them with a higher time frequency, we use the ranks here to reduce the original network well by well, with most of the results shown only for the five abovementioned reduction stages, i.e., GMN reduction by 10 %, 25 %, 50 %, 75 %, and 90 %.

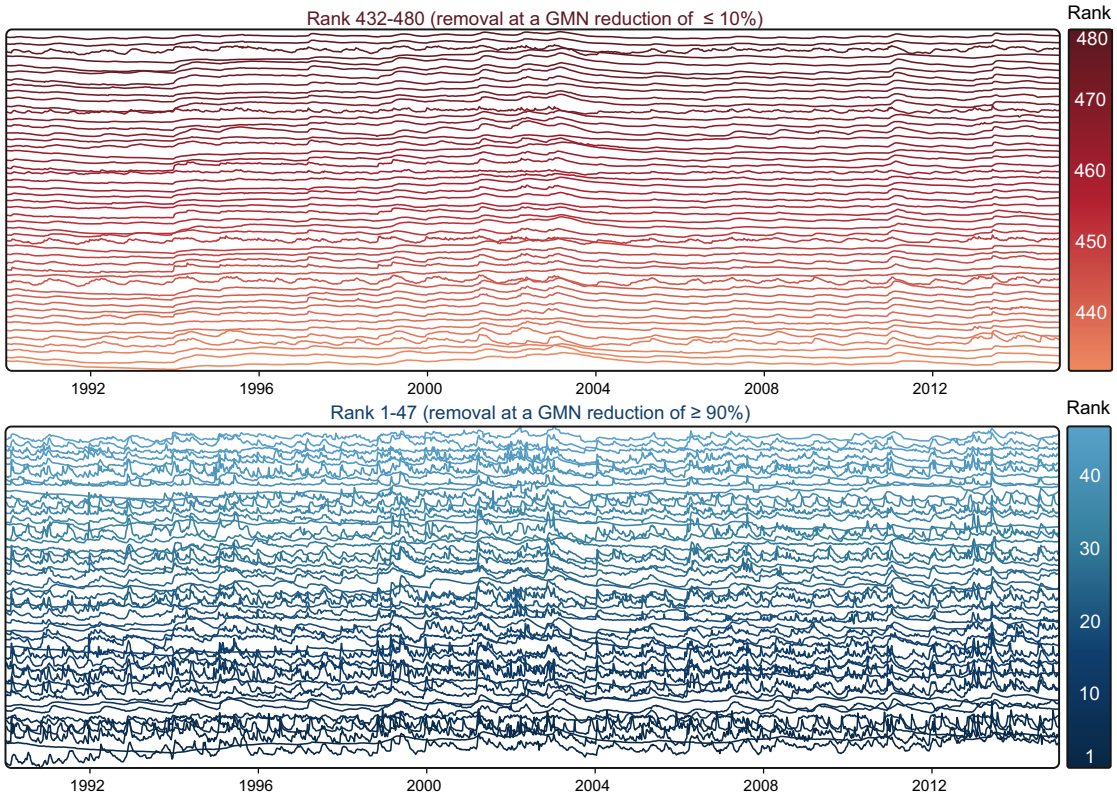


Figure 4.4: Stacked z-transformed hydrographs of the 10 % "most important" monitoring wells (blue, bottom) and the 10 % "most redundant" monitoring wells (red, top). Coloring and stacked order reflecting ranking order.

The upper part of Fig. 4.5 shows the development of the prediction accuracy of the GMN reduction for the error measures NSE, KGE, and R^2 (left), as well as rBias, RMSE, and MAE (right), for the validation data set (mean and ranges of the reconstructed validation period of all predicted/removed wells). Even with only a few optimally selected wells, the predictive power is considerably higher than the mean value of the time series (NSE = 0). An average performance for the validation period of all predicted removed wells rated as satisfactory (NSE > 0.5) is already achieved with only nine remaining wells (corresponds to a reduction of 98.1 %), those rated as good (NSE > 0.65) with 21 remaining wells (95.6 %), and those rated as very good (NSE > 0.75) with 54 remaining wells (88.7 %). With more than 191 wells (60.2 %), the NSE rises above 0.9. KGE and R^2 behave in much the same way as NSE. A KGE of 0.75 is achieved with nine wells (98.1 %) and 0.9 with 144 wells (70.0 %). R^2 of 0.75 is achieved with 22 wells (95.4 %) and of 0.9 with 155 wells (67.7 %), respectively. A MAE of 0.1 m is achieved with 31 monitoring wells (93.5 %) remaining, 0.05 m with 147 wells (69.4 %), and 0.01 m with 394 wells (17.9 %). From a reduction of more than about 75 %, the removal of each subsequent well leads to a disproportionate decrease in accuracy, with a very steep drop from about 95 % on.

When the reduction is less than 75 %, the gradient shows a nearly linear course, meaning a linear (but small) performance increase with more monitoring wells. The rBias also approaches zero at a reduction below 75 %. Thus, we conclude that about 25 % of the wells could be seen as a kind of absolute minimum that is required to adequately describe the system dynamics in the considered study area, despite the average NSE of the reconstruction still being rated as good for only the optimally selected 10 % of the wells.

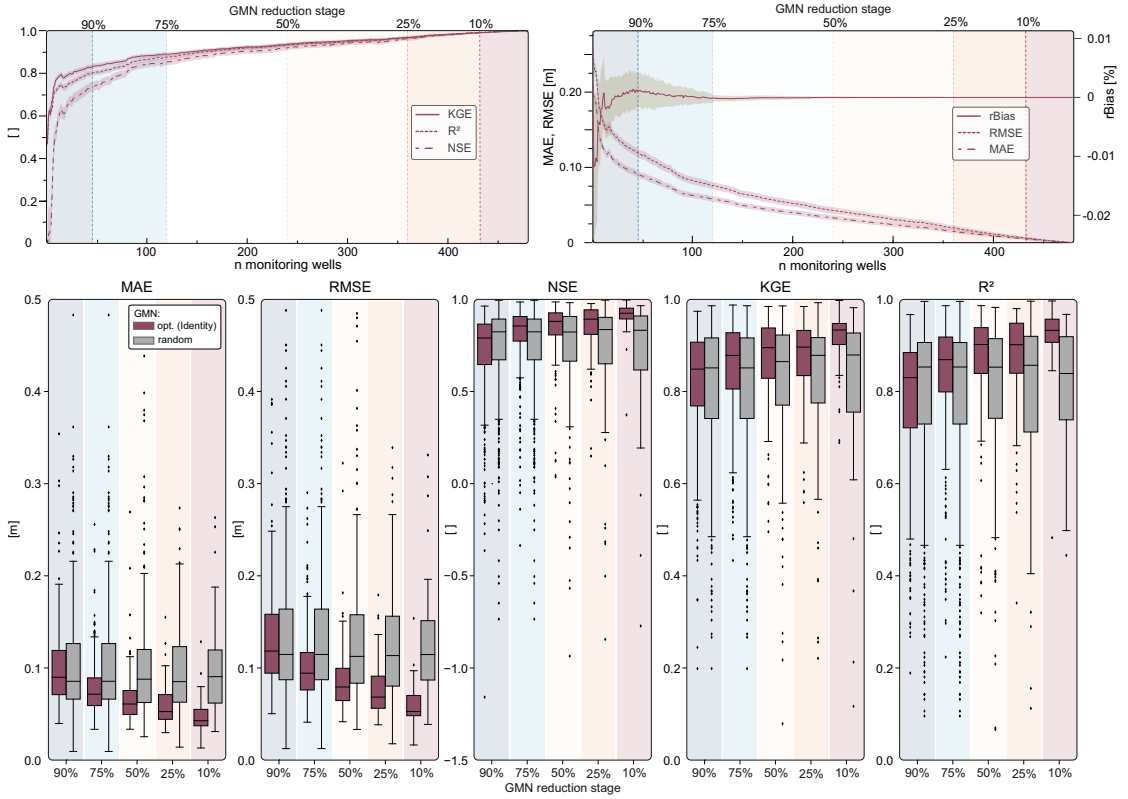


Figure 4.5: Top: Reconstruction error metrics (KGE, R^2, NSE, MAE, RMSE, rBias) as a function of the number of QR-ranked remaining monitoring wells for the identity basis and 1043 basis modes (lines are mean values, shading represents total range of recontstruction errors over all removed wells). Bottom: Same reconstruction error metrics at the reductions steps 10 %, 25 %, 50 %, 75 %, and 90 %, as boxplots over all removed wells, and compared to the same number of randomly removed monitoring wells.

The lower part of Fig. 4.5 displays the performance of the QR-optimized wells compared to an equal number of randomly selected remaining wells for a reduction by 10 % (48 wells) to 90 % (432 wells) of the GMN. Removing wells based on the ranking results leads to a lower prediction loss. Only at a reduction of more than 90 % are the errors in the same range as for the randomly removed wells. Therefore, the information content of these 10 % remaining wells is probably not sufficient to reflect the overall

dynamics. Again, from about 75 % downwards, the performance differences become more pronounced, with a considerably higher average, 25 % quantile, and minimum NSE, KGE, and R^2 values (lower MAE and RMSE, respectively). Below a reduction of 25 %, the 75 % quantile and minimum NSE, KGE, and R^2 values are also clearly higher (lower for MAE and RMSE, respectively). This clearly shows that the advantages of the data-driven optimization method come into play, especially for moderate to smaller reductions of a GMN.

Figure 4.6 shows the temporal reconstruction accuracy at the considered reduction stages from 10 % to 90 % for eight selected wells (see also Fig. 4.3). These wells were chosen to reflect the dynamics spectrum and represent the full ranking range. The reconstruction is always based on the higherranked remaining wells (but keeping the chosen reduction stages). Consequently, well 154-304-1, the highest-ranked well shown with rank 59 (bottom), which could theoretically be reconstructed with a maximum of 58 remaining wells, is reconstructed with 10 % of the wells (48). Similarly, well 132-257-4, the lowest-ranked well with rank 478 (top), which could theoretically be reconstructed with a maximum of 477 wells, is reconstructed with 90 %, 75 %, 50 %, 25 %, and 10 % remaining wells (432, 360, 240, 120, and 48, respectively) for a comparison.

The results show that the individual dynamics of the hydrographs can already be adequately reconstructed with a 10 % subset of the monitoring network. As expected, as the number of wells increases, the accuracy improves on average, hydrographs are reproduced more consistently, and short-term peaks are reproduced more accurately. Though these seem to be only, comparatively, slight improvements, considering the overall dynamics for some wells and time steps, the absolute errors can be up to several tens of centimeters, albeit achieved by many additional wells. Whether this justifies the increased operating costs of the monitoring network depends on the task at hand. The reconstruction results for the other wells can be found in the Appendix.

4.3.3 Network Reduction and Extension Based on Gridded GWL Maps

This application case is based on spatially continuous gridded weekly GWL contour maps from 1990 to 2015. Analogous to the hydrograph data set, the first 80 % of this period was used for model training and the last 20 % for evaluation. According to the ranking, we investigate how well the GWL can be reconstructed with two reduction stages in which 10 % and 20 % of the GMN are removed. We have selected these reduction stages since, in the analysis of the reduction stages with hydrographs, the advantages of the datadriven optimization method were more pronounced for moderate to smaller reductions of a GMN. Moreover, reducing an existing network by more than

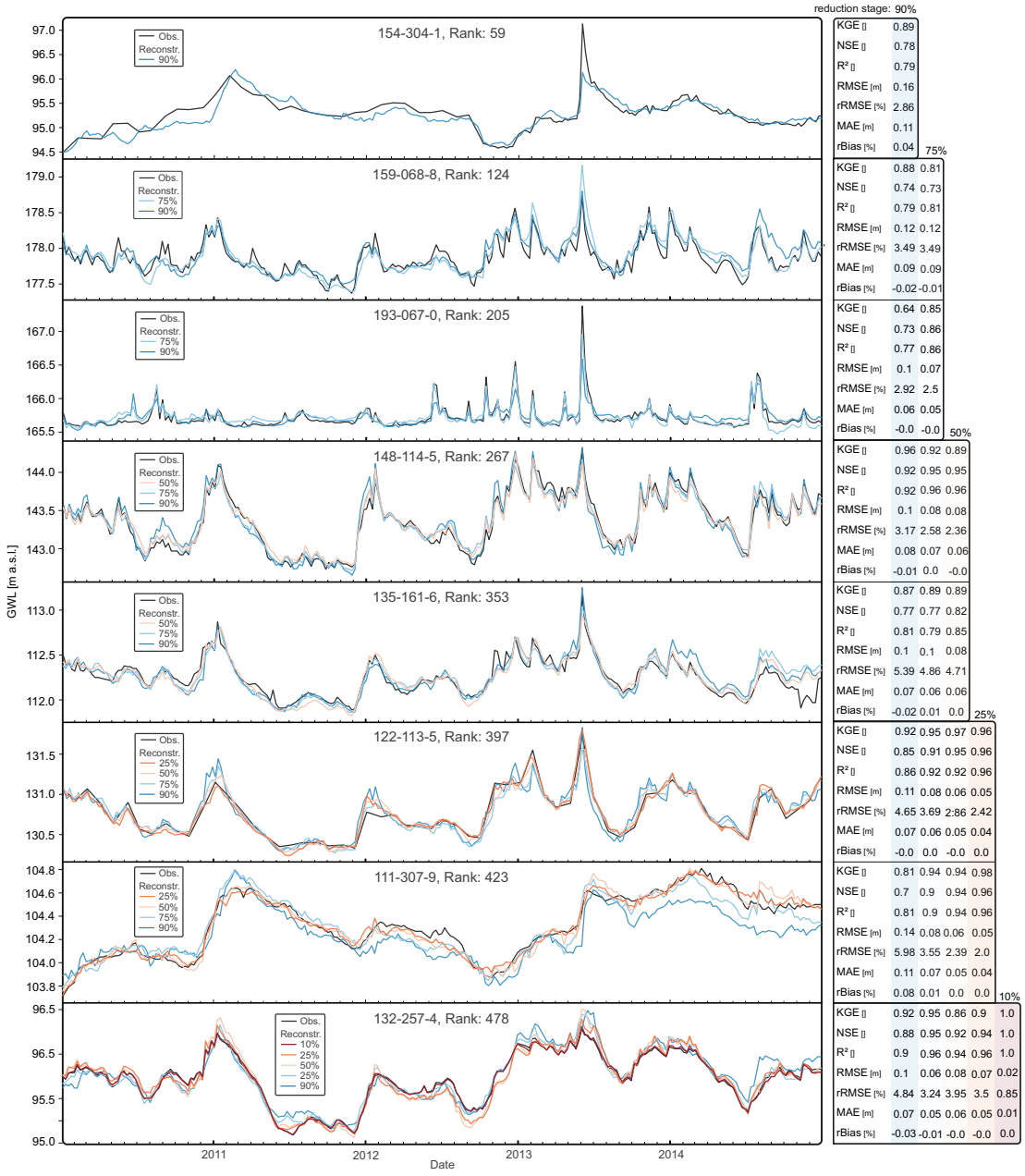


Figure 4.6: Hydrograph reconstruction at the five reduction stages 10 % - 90 % exemplary shown for eight monitoring wells, along with the respective error measures.

20 % seems unrealistic in practice. Furthermore, we extend the existing network by 10 % and 20 % wells and analyze where new wells are placed to supposedly improve the GMN. To account for more or less suitable locations (e.g., with regard to infrastructure), we apply costs with a non-uniform spatial step function on a 50×50 grid (corresponding to the used GWL grid) into the QR factorization. The cost function grid is assigned zero (no additional costs) at existing monitoring well locations to ensure that existing wells

remain since, technically, the extension (by e.g., 10 %) is realized such that 110 % of new wells are placed on the gridded GWL maps. At potentially well-suited additional sites, which are defined within 50 m of roads and paths, outside surface waters, and where the slope is less than 20 %, we assigned a cost value of 21. For all other areas that are considered as not suitable, the cost weighting is set to 22. Alternatively, a gradual cost function can be used, where the weighting increases with distance to the infrastructure or similar. It should be noted that the weighting depends on the system, basis, and cost function and must be adjusted for the particular case (Clark et al., 2019). We assigned the mentioned weighting factors iteratively until it resulted in the desired behavior. With the weightings chosen in this way, it was possible to achieve a result such that the first 480 wells are placed at existing monitoring wells, and all subsequent wells are placed at suitable locations, while the algorithm avoids the other locations. Finally, we combine a reduction/extension scenario, where the original number of wells is kept, but the 10 % and 20 % most redundant wells are removed and replaced afterward. Technically, this is done in a two-step procedure, consisting of the reduction step followed by the above-described extension, where the cost function is adapted for the existing wells.

The maps on the left side of Fig. 4.7$a_{1\text{-}3}$ show the spatial distribution of the monitoring wells as a result of a 10 % (yellow dots) and 20 % (red dots) reduction (a_1), and a 10 % (light blue dots) and 20 % (dark blue dots) extension (a_2) of the GMN, as well as a combined reduction/extension of the GMN (a_3). For the latter, redundant 10 % and 20 % monitoring wells were eliminated and replaced with wells at optimal locations. We should note that the ranking based on the spatially interpolated data is different from the ranking based on the hydrographs alone (see Appendix S4a).

This variation can be explained by the ranking reflecting the information content regarding the reconstruction with the lowest possible error. While, in the case of hydrographs, the goal is to reconstruct the hydrographs of the removed wells, here the goal is to reconstruct the interpolated surface (which constitutes a best guess of spatially continuous GWL based on the available data).

While the first 10 % of reduced wells are evenly distributed across the study area, the subsequent removal step (i.e., additional 10 %; thus, 20 % removed wells in total) eliminates well clusters in the central and northern regions. This seems conclusive because clusters of nearby wells tend to show similar dynamics and thus do not add much information to an interpolation, according to Tobler's law. Optimal locations for additional wells are identified primarily along the western and eastern margins, i.e., along the Rhine and downstream of the alluvial valley aquifers of the adjacent Black Forest. These are areas with expected higher groundwater dynamics (e.g., high seasonal magnitudes and high flashiness) and, on the other hand, due to the elongated geometry of the URG,

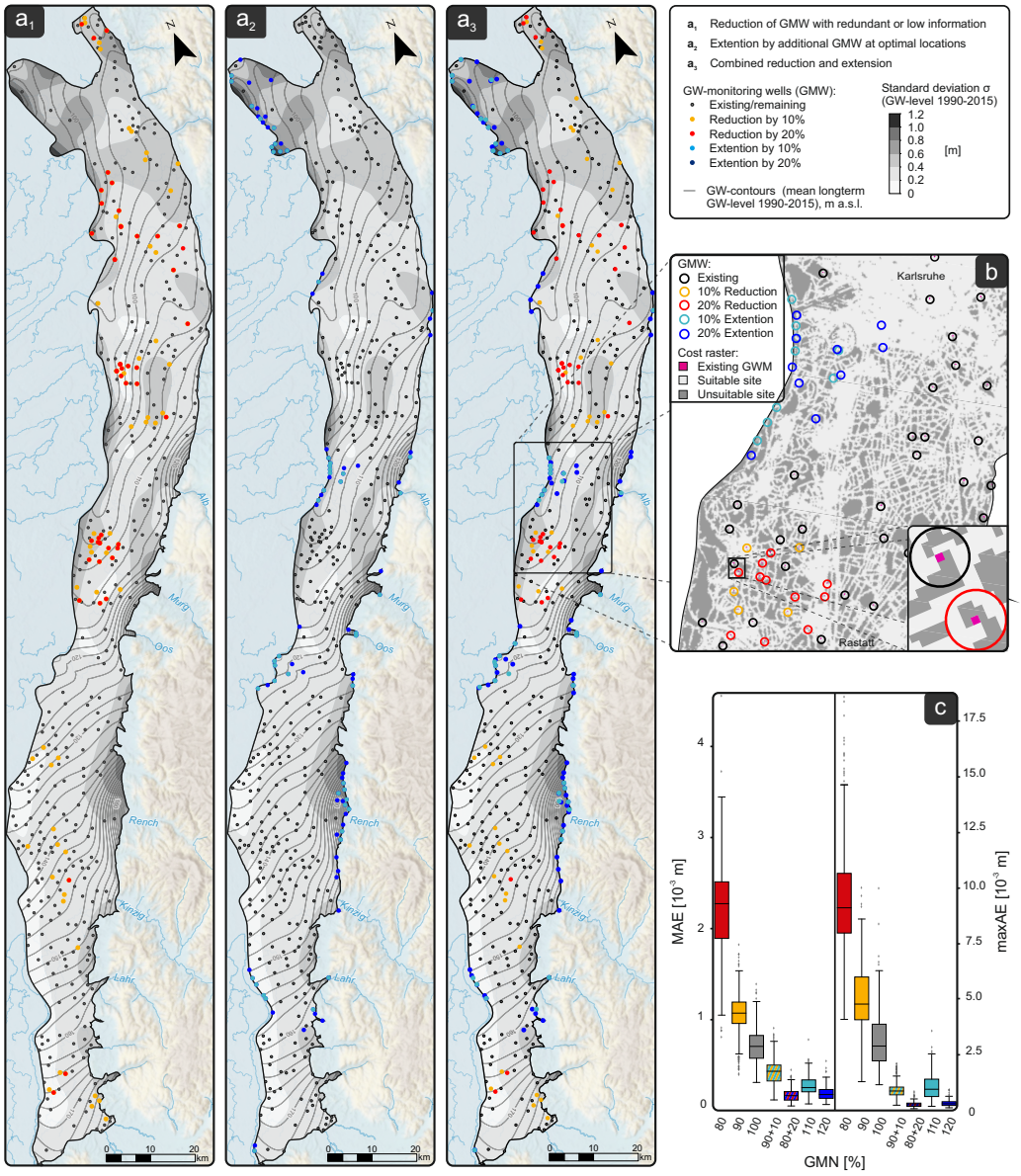


Figure 4.7: (a): Location of removed monitoring wells in a 10 % (yellow) and 20 % (red) QR-based monitoring well reduction (left map, a_1), in a 10 % (light blue) and 20 % extension (center map, a_2), and a combined reduction/extension in which 10 % and 20 %, respectively, of the monitoring wells were removed and replaced with wells at optimal locations (right map, a_3). (b): Cost function grid used for the GMN extension. (c): Boxplots show the mean and max. absolute error of the reconstruction of the 216 GWL contour maps of the test set obtained with the mentioned GMN reduction/extension.

areas with increased interpolation uncertainty (transition from interpolation to extrapolation). Optimal well locations are primarily, but not exclusively, located in areas of

increased variability (standard deviation of the interpolated GWL; see Fig. 4.7a_2 and a_3).

The box plots in Fig. 4.7c show the mean (left) and maximum (right) absolute error of the reconstructed 261 GWL contour maps of the evaluation set for all abovementioned scenarios. It has to be noted that the MAE is now taken as the mean over the spatial axis, i.e., for each of the reconstructed 261 GWL contour maps separately (whereas, with the hydrographs, the MAE was taken as the mean over the time axis for each reconstructed well). This was done because, in the application case, the focus is on the error of a spatial reconstruction of GWL contours and not explicitly on time series. Correspondingly, the maximum absolute error (maxAE) is the maximum over the spatial axis for each of the reconstructed 261 GWL contour maps. We therefore also refer to them as mean and maximum spatial reconstruction errors. Thus, the boxes in Fig. 4.7c show the variability in the mean absolute error and maximum absolute error over the 261 time steps.

On average, the model can reconstruct the GWL contour maps with very high accuracy, with mean absolute errors far below 1 cm. This seems very low compared to the reconstruction of the hydrographs. However, this is due to the fact that the reconstruction of a large number of many similar values (i.e. raster pixels) is much easier for the model, for the following two reasons: (i) there are many more training patterns for each type of dynamics than with the hydrographs alone, and (ii) the overall dynamics are reduced by the interpolation itself, which smooths the data spatially and temporally. Taking these limitations due to spatially interpolated data into account, it seems more reasonable to focus on the maximum absolute error, which allows the identification of areas with higher errors, where the model (with the existing data) cannot produce reliable reconstructions and additional wells would bring the most information.

When comparing the maxAE (Fig. 4.7c; right) for all scenarios, we see that a reduction in the network increases the spatial reconstruction error by a factor of about 2 for 10 % reduction and about 3 to 4 for 20 %. For comparison, the gray box (100 %) shows the reconstruction errors with an unchanged GMN (this error results from the fact that the model is trained with the first 80 % of all time steps, but the reconstruction is performed for the unknown 20 % of the evaluation data set). An extension of the network by 10 % can considerably reduce the spatial reconstruction error to about less than two-thirds, while an extension by 20 % reduces it further to 1/10 of the initial value. Most interestingly, the reconstruction errors for the combined reduction/extension scenarios with 90 %/10 % and 80 %/20 %, respectively (thus an unchanged number of 480 wells in total), are slightly below the straightforward GMN extension with 110 % (528 wells) and 120 % (576 wells). To a lesser degree, this also applies to the mean absolute errors,

at least for the 80 %/20 % scenario, which performs slightly better than an extension by 20 %, and considerably better than a 10 % extension. In practice, that means that, with a combined reduction/extension, for example, sensors/data loggers that become available can be used elsewhere at better locations. This reduces the installation costs of the additional wells and the operating costs of the GMN and, moreover, performs about the same as or even better than a pure extension.

4.4 Conclusions

This study investigated data-driven sparse sensing approaches based on the work of Silva et al. (2021), Clark et al. (2019), and Manohar et al. (2018) and adapted them to optimize an existing GLMN. The algorithm fits a tailored basis to the training data, subsequently used in a QR decomposition to rank the monitoring wells by importance based on reconstruction performance. This approach allows us to remove groundwater monitoring wells with low information content if needed, equip monitoring wells with higher rank with higher quality sensors, or measure with a higher time frequency. When using spatially continuous input data (by interpolation or numerical simulation), the ranking is performed according to the same scheme for all locations. This rank can be used as a decision-making aid to search for locations for additional monitoring wells. We incorporated a cost function to eliminate inaccessible locations from the site selection process. Adjusting the cost constraint allows a specific adaptation to the individual problem definition.

Our results show that identifying redundant, low-ranking monitoring sites would allow a drastic reduction of the monitoring network, with a minor loss of information, compared to a random reduction (which corresponds to a reduction based on other criteria, as is often the case in practice). In the case of a desired network extension, the reconstruction quality can benefit from the additional removal of unsuitable wells.

As in related previous studies (Clark et al., 2019; Manohar et al., 2018), using the identity data basis (raw data without dimensionality reduction) and the total number of available base modes yielded a lower reconstruction error for a given number of wells compared to other basis mode types and numbers. This is because no information is lost when constructing a low-ranking approximation to the data. However, for larger data sets than the one used in this study, an optimization without previous dimensionality reduction can lead to impractically long computation times. Just as in the work of Clark et al. (2019), and Manohar et al. (2018), a randomized projection of the data in our study performed, on average, only slightly worse than the raw data and may be worthwhile

for large data sets or multiple computational runs due to lower computational costs. Even though the widely used SVD basis gave the worst results for our data set, the reconstruction errors are still lower than for a random network optimization.

In addition to GLMNs, this approach can also optimize groundwater quality or multivariate monitoring networks. As with all data-driven methods, the quality of the results depends strongly on the availability of the input data (spatial and primarily temporal). Since this approach relies on detecting patterns in data and placing monitoring locations based on those patterns, it benefits from large data sets. Therefore, we see the main application of this technique in optimizing monitoring networks of regional-scale groundwater systems, where a comprehensive overview of the variability and quantity of groundwater bodies and the assessment of long-term changes in natural conditions is the monitoring objective.

Overall, we could demonstrate that modern data-driven methods of sparse sensing are well suited for the application to GMNs, as long as there is a good historic data basis. The applied method can be used for an optimization regarding the number of wells and their location, for a network reduction and extension, or for both combined. Using hydrographs (1D) as input data, the applied approach allows an information-based assessment of an operated monitoring network. The outcomes can be used to identify representative key wells for selecting expressive subnetworks, equip the critical wells with improved data loggers, or release installed sensors/loggers at redundant wells for more suitable locations. The spatial dependency structures and the sphere of influence of wells can be considered in optimizing with two-dimensional input data, both for reduction and for an extension of monitoring networks tailored to the dynamics of the aquifer. Although optimized reduction can generally lead to greater cost efficiency, it should always be done judiciously and in combination with expert knowledge of the system.

Code and Data Availability *Pysensor* Software is available at: `https://github.com/dynamicslab/pysensors`.The well data are publicly available at the web service of the Baden-Württemberg State Office for Environment (LUBW, 2021b). Our Python code files are available on Ohmer (2022)

Author Contribution M.O.: Conceptualization, Methodology, Software, Formal analysis, Validation, Investigation, Visualization, Writing – original draft preparation. T.L. Conceptualization, Methodology, Writing – review and editing, Supervision. A.W.: Data preprocessing and curation, Writing – review and editing

Acknowledgements This study is a contribution to the project: Nitrate Monitoring 4.0 - Intelligent Systems for Sustainable Reduction of Nitrate in Groundwater (NiMo 4.0), funded by the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU) on the basis of a resolution of the German Bundestag.

Supporting Information

This supporting information provides a comparison of the QR-based ranking with 1D hydrograph data and with 2D interpolated GWL contour maps (S4a) and stacked z-transformed hydrographs of monitoring wells in order of model ranking (S4b-S4c).

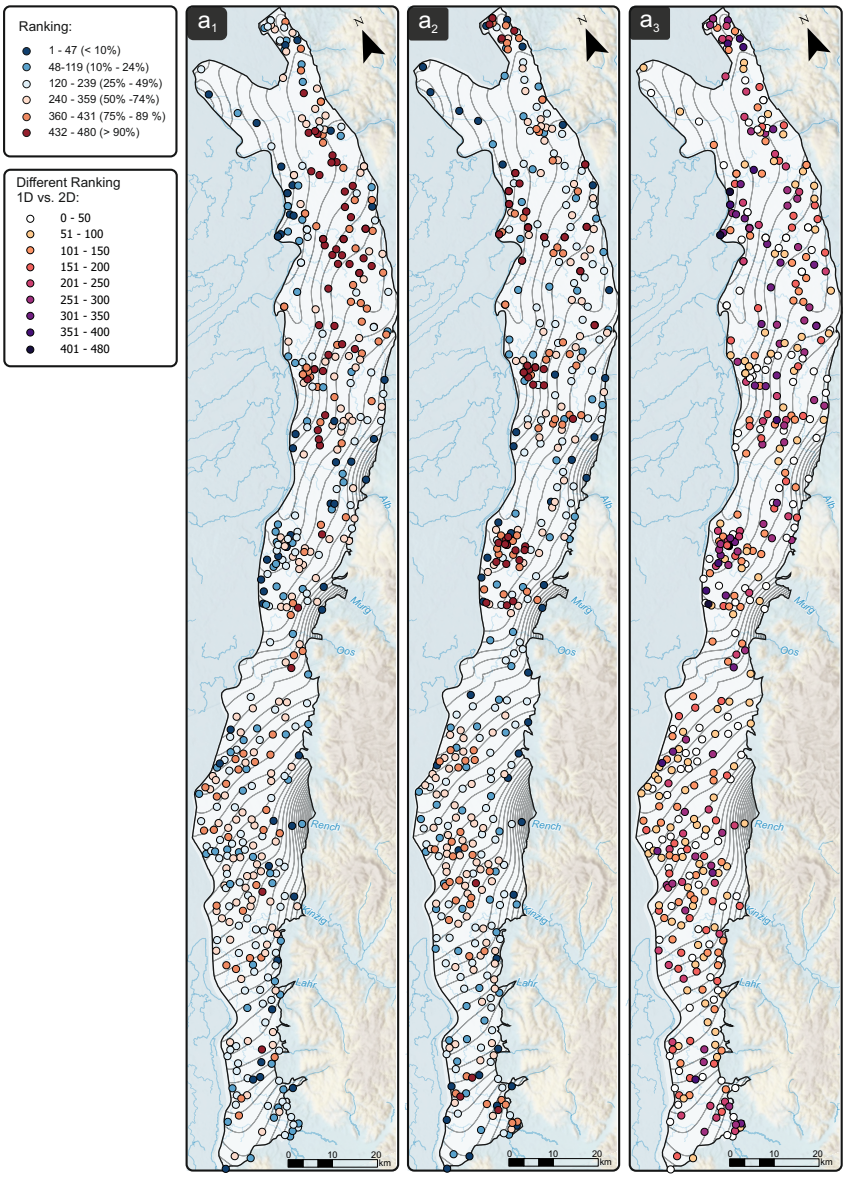


Figure S4a: Comparison of QR-based ranking with 1D hydrograph data (left, a_1), with 2D interpolated GWL contour maps (middle, a_2), and differences of rankings between 1D and 2D (right, a_3).

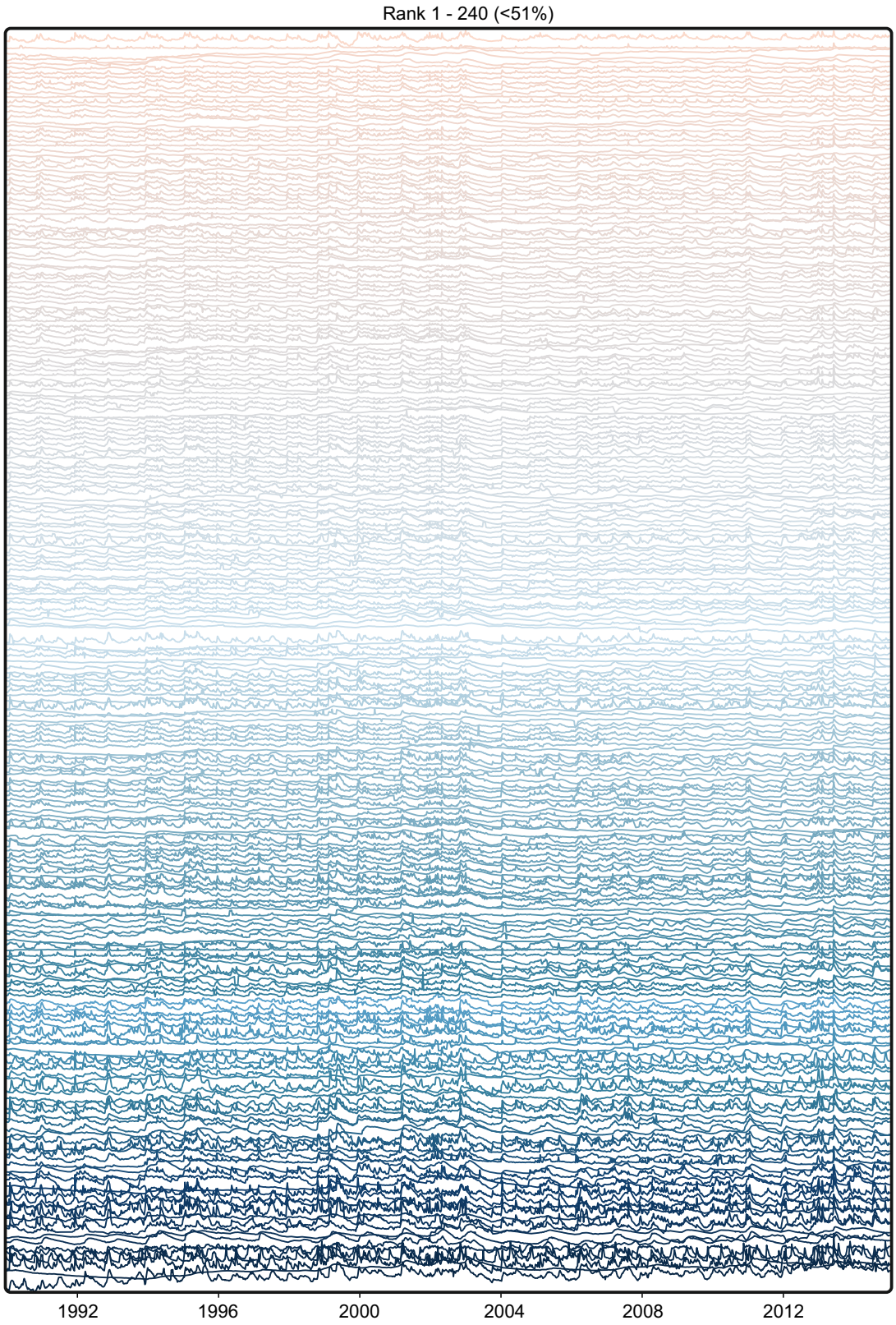


Figure S4b: Stacked z-transformed hydrographs of monitoring wells. Rank 1 -240

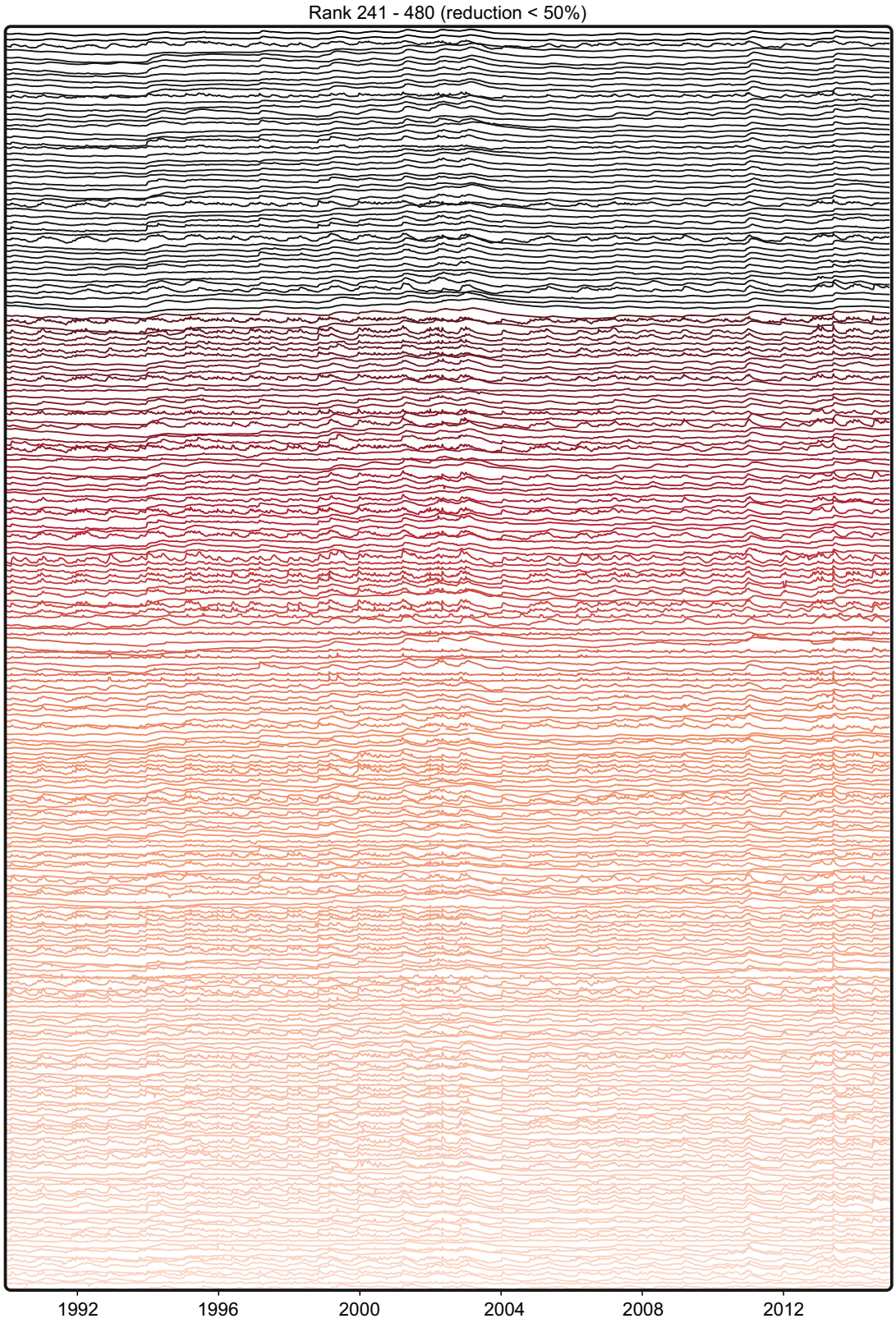


Figure S4c: Stacked z-transformed hydrographs of monitoring wells. Rank 241 -480

Chapter 5

Synthesis and Outlook

5.1 Summary and Synthesis

This cumulative work pursued two central objectives. The first objective focused on exploring strategies to improve the design of large-scale groundwater level monitoring networks in terms of their ability to cost-effectively capture the spatial and temporal dynamics of the hydrogeologic system and provide spatially continuous estimates of the quantitative status of groundwater. The second objective aimed to investigate and evaluate existing interpolation techniques for spatial prediction of this quantitative state from existing monitoring networks.

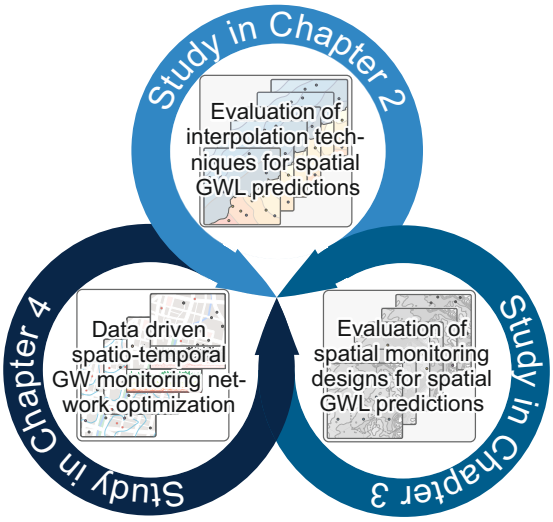


Figure 5.1: Overview of the studies carried out within the scope of this thesis at a glance and their objectives.

For the latter objective, Chapter 2 provides a comprehensive evaluation of the most established deterministic and geostatistical interpolation techniques for groundwater con-

tour mapping conducted in a study area with a complex interaction between karst and an alluvial aquifer. The results led to the following key findings. Co-ordinary kriging (CoOK), the multivariate variant of the ordinary kriging (OK), with the incorporation of additional correlated variables (e.g., topography, springs, river levels), provided the most plausible results and the lowest CV errors for both aquifers. Especially in hilly terrain where observation wells are primarily located in valley regions, digital elevation data led to significantly improved prediction accuracy since the GWL is often a subdued replica of the topography. In contrast, other techniques overestimated the GWL in the valley and underestimated the GWL below ridges due to the lack of terrain information. Although cross-validation is often the only way to assess interpolation quality, its results should never be based on only a single error metric. Thus, the CV error metrics differ in their ability to reflect the variations in interpolation results with additional validation methods. For example, a $R^2 >$ of 0.9 indicates good interpolation accuracy for almost all scenarios. At the same time, calculated aquifer exchange rates based on worst-case results vary up to a factor of 10 and even lead to hydraulic gradient reversals. This illustrates that the choice of an inappropriate interpolation technique can lead to severe misconceptions that propagate and increase in the course of further analyses. In summary, the following conclusions can be drawn from the study. An inadequate GMN cannot be replaced by an efficient interpolation method. When observational data are too sparse, the underlying assumptions about variation between samples can vary widely, and results can differ significantly depending on the interpolation method chosen. At the same time, the predictive accuracy of the various techniques converges as the density of observations increases. The spacing between monitoring wells must be consistent with groundwater variability. Otherwise, observations may be too sparsely distributed to identify autocorrelation. In this case, correlated secondary variables incorporated in co-kriging can benefit greatly.

The study presented in Chapter 3 is a simulation experiment to investigate the prerequisites of a capable GMN for spatial prediction of GWLs. For this purpose, state-of-the-art and recently developed network design approaches with varying monitoring densities are compared on groundwater surfaces computed by numerical models with variable spatial dynamics, hydraulic properties, and scaling. The numerical surfaces provide a realistic "a priori" reference that allows evaluating the spatial prediction error in addition to the CV error and thus assessing the suitability of the commonly used cross-validation error statistics (MAE, RMSE, RMSSE, ASE, and NSE). The study's findings revealed that adequate spatial coverage of the network of observation wells beyond the regions of water management importance is the most critical factor for an enhanced spatial understanding of the groundwater system and, thus, has a significant positive effect on the

estimation of groundwater quantity and consequently on all derived parameters such as flow direction, hydraulic gradients, flow velocities, and flow rates. The results show that the number of monitoring wells is significantly more critical for prediction accuracy than their optimal spatial distribution, mainly when the scale level of monitoring density is larger than that of spatial groundwater variability. All designs investigated showed significantly lower prediction errors than the randomly placed wells (SRS) design, which is closest to existing, historically grown GMNs. However, the quality improvement of the design approaches compared to SRS primarily depends on the monitoring density and spatial groundwater variability and increases with increasing monitoring density. The grid designs enabled slightly lower prediction errors than the other designs. However, this is not commensurate with the disadvantages they bring. They are not extensible and may not be feasible to implement in reality without violating the grid structure. In addition, wide spatial coverage can lead to aliasing effects if the spatial dynamics are higher than the grid spacing or unevenly distributed. In this study, the densify sampling network (DSN) method, which uses the maximum prediction standard error as a selection criterion for placing additional wells, provided the best prediction accuracy of the progressive designs across all model surfaces and monitoring densities. However, this approach is more appropriate for incremental extension with a single monitoring well at a time and subsequent re-calculation of the prediction standard error. Otherwise, the resulting network will perform worse. The spatial coverage sampling methods with low discrepancy sequences are appropriate for designing entirely new groundwater monitoring networks from the ground up. In particular, the novel 2-dimensional additive recurrence sequences (R_2), a recursion method based on irrational numbers, leads to nearly comparable results to DSN. The method was applied for the first time in this study to monitoring design optimization. Since site selection is based solely on quasi-random mathematical sequences, these methods have the advantages that they (i) can be applied without prior knowledge through observations to establish an effective and uniformly distributed monitoring network, and they (ii) provide reproducible design results regardless of the number of additional wells (single or multiple). The study demonstrates further that global CV statistics are not suitable for assessing the results and comparing the different GMNs design approaches. CV exhibits a significant pessimistic bias in the designs studied depending on the degree of randomness in site selection. This bias leads to a negative relationship between the results based on the CV and the actual error.

The study presented in Chapter 4 investigates the prerequisites of an efficient GMN for temporal and spatial prediction of GWLs on a well-developed GLMN with long-term hydrograph data (Upper Rhine Graben, Germany). In this study, a data-driven greedy QR algorithm is applied for the first time to optimize groundwater monitoring networks. The

approach uses unsupervised dimensionality reduction techniques to identify existing coherent structures to find wells with maximum variance in the training data and thus rank wells according to their information content to reconstruct the overall signal. Depending on the dimension of the input data, either monitored well data or regionalized variables, such as groundwater levels, the approach can be used as a tool for identifying monitoring wells with redundant information content, designation of representative sub-networks, and planning sites for additional monitoring wells providing maximum information content. The results show that identifying a drastically reduced subset of key monitoring wells capable of capturing the overall dynamics of the groundwater system is sufficient to enable low-loss predictions of the temporal dynamics of the residual monitoring wells with redundant or low information content. This frees up sensors in redundant wells to be used in previously unmonitored or newly drilled wells. This approach could be a valuable tool in planning and combining multiple local GMNs into regional, national, or international GMNs by identifying meaningful and representative wells from these subnetworks that provide nearly all temporal dynamic information content.

For a conclusive summary of the findings of this thesis, the following section provides brief answers to the research questions posed in Chapter 1.5:

Study Chapter 2:

RQ 1: Which interpolation technique provides the best results for the studied alluvial and karst aquifer?

- Co-ordinary kriging with the additional use of secondary parameters (e.g., topography, springs, river levels) produced the lowest cross-validation errors and the most plausible results for both aquifers studied. Empirical Bayesian kriging (EBK) also delivered good results for both aquifers.
- If the spatial autocorrelation is not very low, the kriging methods are generally more effective than deterministic methods.

RQ 2: How do the different methods deal with the:
i: spatially inhomogeneous distribution/patterns of the existing groundwater monitoring network?

- In contrast to deterministic methods, kriging considers the spatial correlation of the measurements, e.g., well clusters are weighted more diminutive, and the bias reduces.

ii: different surface types change from a pronounced hilly topography to a flat riverine landscape?

- In hilly terrain, methods that can exceed the minimum and maximum measured values showed more plausible results (kriging). Especially in hilly terrain, incorporating digital elevation data (co-kriging) provided the most benefits.

- Also, in a riverine landscape, correlated data such as terrain information, wetland, or river levels increase the prediction accuracy.

ii: different hydraulic pressure conditions within the aquifer which fluctuate from unconfined to artesian?

- Ordinary kriging and co-ordinary kriging results best met hydrogeologic expectations in this region.

RQ 3: What are the possible influences of the chosen methods on further computations, namely the calculation of the estimated vertical groundwater exchange between different aquifer systems?

- They can be drastic! The standard deviation of the calculated exchanges based on the methods is abaout 750 million m^3/year, the range between about 270 million m^3/year and -2,770 million m^3/year.

RQ 4: Which are the most suitable error statistics to compare the performance of the methods?

- In principle, an error metric should never be considered isolated, as they all have advantages and disadvantages (representation of bias, outliers, etc.). It is recommended to use at least MAE and RMSE and a metric that provides information about bias (e.g., ME, Bias). MAE is intuitive and measures the average magnitude of the error. Comparison with RMSE shows the variation in errors. The greater the RMSE compared to the MAE, the greater the variation.

RQ 5: How can the results be validated with additional eco-hydrogeological data?

- Geographical locations of karst springs, wetlands, and surface waters with groundwater interaction are suitable to verify the results locally. The calculated groundwater depth indicates accuracy, and the comparison of locations with springs or receiving waters with the calculated streamlines gives plausibility indications.

Study Chapter 3:

RQ 6: Is there an extensible and transferable GLMN design that allows reliable spatial estimates of GWL with a minimum number of monitoring wells?

- The best prediction accuracies with extensible designs were obtained using the geostatistical DSN method (maximum prediction standard error) and the low-discrepancy R_2 method. The latter offers the advantage of being independent of the actual monitoring data and hence reproducible. It is also more versatile, as it allows for any number of additional wells at once.

RQ 7: What are the quality differences resulting from the use of various GLMN design approaches?

- They depend on the monitoring density and spatial groundwater variability. The quality differences increase with increasing monitoring density. The best design (hexagonal grid) prediction error was 31.2%, DSN 24.7%, and R_2 23.8% below the simple random sampling design for the maximum number of monitoring wells investigated.

RQ 8: At what observation well density does a reasonable information/cost ratio result?

- This depends on the scaling and groundwater variability. A reasonable ratio was achieved on average when about 0.51% (with a range of 0.37-0.64%) of all possible sampling options (imaginary pixels of a grid reflecting assumed groundwater variability) were monitored.

RQ 9: Which is the most suitable CV error statistic (MAE, RMSE, RMSSE, ASE, or NSE) to evaluate the quality of interpolated groundwater surfaces?

- According to our results, the CV error statistics, especially MAE and RMSE, can be helpful as a rough quantitative estimate for the actual error. From the results, we conclude that one should avoid comparing different design approaches based on the error estimate of the global average CV.

Study Chapter 4:

RQ 10: What is the ranking of monitoring wells in an existing network in terms of their information content/reconstruction performance, i.e., in which order should the wells be removed if a network reduction is desired?

- The algorithm employed ranks monitoring wells based on the reconstruction error. If the other wells can accurately predict the temporal variability at this point, the respective well has no unique information content. In our case, these were mainly wells with a pronounced seasonal cycle and without short-term fluctuations and trends.

RQ 11: How does a reconstruction/interpolation error develop when a given number of monitoring wells are reduced? How does the error of reducing wells according to information content compare to a random reduction?

- The results show that even with a drastic reduction, the hydrographs of the removed wells can be reliably reconstructed (average absolute reconstruction accuracy of 0.1 m is achieved with a subset of 6% of the wells, 0.05 m with 31%, and 0.01m with 82% of the wells). Compared to randomly selected wells, removal based on ranking results in an average 39% reduction in reconstruction error.

RQ 12: When the goal is network extension, where should new wells be placed for maximum information gain? How significant is the increase in information, i.e., how significant will the spatial reconstruction error be reduced?

- The algorithm places additional monitoring wells mainly in areas with low monitoring density and higher expected groundwater dynamics (e.g., high seasonal fluctuations, high flash floods). A network extension by 10% reduced the spatial reconstruction error to less than two-thirds, while an extension by 20% further reduced it to one-tenth. However, since the data basis is based on interpolated values, this statement has certain limitations.

RQ 13: How well does a combined reduction/extension (i.e., replacement) of a certain number of wells perform compared to a straightforward extension?

- Surprisingly, the reconstruction error of the combined reduction/expansion scenarios (by 10% in each case), i.e., an unchanged total number of wells, is slightly lower than that of the straightforward expansion by 10%.

5.2 Perspective and Outlook

Developing an optimal groundwater monitoring network is a chicken-or-egg problem: Its scale and spatial distribution must capture the temporal and spatial variability of groundwater (including the complex nature of geologic, hydrologic, and other environmental or human influences) for monitoring programs, often with multiple competing objectives. In addition, the relationship between monitoring costs and system knowledge should be Pareto optimal so that there is no spatial and temporal redundancy. Since each groundwater system is individual in its spatial and temporal dynamics, there is no universal planning answer regarding well placement and spacing. In general, higher dynamics require higher monitoring density and frequency. However, all this leads to a paradox since most of our knowledge about the dynamics is based on data we have gath-

ered from existing GMNs. An inadequate monitoring network can lead to this dynamic being underestimated. Therefore, sufficient observation wells are needed to determine which are necessary and which are not. Nevertheless, new wells are often associated with high initial investments, as they usually require extensive drilling. For this reason, wells are not traditionally constructed without specific justification.

A pure hydrogeological approach best describes how new wells are usually planned in practice. This approach is based on qualitative and quantitative hydrogeological information (e.g., pumping tests, modeling of flow and transport processes) and expert knowledge and concepts. Expert knowledge is the most significant element in the planning of optimal GMNs. However, the isolated application of this approach often seems to lead to the emergence of monitoring networks that are mainly concentrated in small areas with water management significance and often characterized by an expectation bias. For instance, a nitrate monitoring network where wells are located exclusively in agricultural regions is suitable to demonstrate agricultural impact. However, the general state of the aquifer is represented in a biased manner. Statistical approaches (including simulation-, variance-, and probability-based approaches), such as those examined in the chapters 2 and 3, have proven extremely useful in addressing these issues because they account for conceptual uncertainties regarding the underlying hydrogeology and are free of bias.

Besides the task of spatial optimization, there is also the challenge of temporal and spatio-temporal GMN optimization. As a result of increasingly affordable data loggers, there is a growing number of long-term groundwater hydrographs. In-situ groundwater quality sensors are also becoming more affordable and are expected to increase in use. More comprehensive recording of temporal variations in nitrate and other pollutants will provide data in the future that will allow more profound insights into possible sources, pathways, transport, and degradation processes. Currently, groundwater quality (GWQ) data are not yet sufficient to apply data-driven models such as those described in Chapter 4. Unfortunately, however, the trend that can be observed in Germany is that the temporal resolution of monitoring is decreasing, despite more cost-effective data loggers, and many long-term hydrographs are no longer measured continuously. Moreover, it appears that decisions to reduce monitoring are based on random factors rather than optimization strategies (e.g., broken data loggers are not replaced). Though optimization can generally lead to more cost efficiency, it should be done wisely. It should be kept in mind that only long-term hydrographs provide a better understanding of long-term changes in groundwater recharge and storage, climatic variations, regional impacts of groundwater development, changes in groundwater flow directions, and allow for statistical analysis of water level trends. Therefore, the optimization results should

not justify a drastic downsizing of the monitoring networks but rather increase the focus on the especially relevant wells or the selection of meaningful monitoring sites from local monitoring networks to create a representative regional, national or international monitoring network. Especially concerning possible influences of climate change on groundwater bodies and the associated changes in boundary conditions, a recurring assessment of the quality of the monitoring network and sufficiently reliable data are necessary.

For this purpose, it may be worthwhile to look outside the box to other fields to gain insights. The goal of optimal placement of in-situ measurements or sensors to monitor or control spatiotemporally dynamic systems under cost constraints arises in virtually all scientific and engineering fields under terms such as optimal experimental design and variable selection. Although the sensors and parameters differ, they all provide local observational data from which global properties shall be inferred and are selected from a more extensive set of possible locations. The overall objectives of each optimization approach is dictated by data availability. In contrast to hydrogeology, where data availability is limited, many water-related sciences (e.g., monitoring precipitation, runoff, or flow dynamics) and soil mapping extensively use data-driven methods (e.g., machine learning and artificial intelligence) to sampling design optimization and parameters regionalization. For instance, Wadoux et al. (2019) investigated optimal sampling designs for soil mapping with random forest (RF) algorithms. Sagan et al. (2020) applied deep learning (DL) techniques to make inferences from satellite data to measurements from in situ water quality sensors used for inland water quality monitoring. Such a methodology could be employed, for instance, in regions with extensive artificial agricultural irrigation from groundwater and undersized monitoring to conclude storage changes from satellite data (e.g., NDVI, normalized difference vegetation index). The data-driven QR algorithm investigated in Chapter 4 proved to be highly effective for monitoring network optimization when sufficient data were available. A recent study by Williams et al. (2022) effectively combines this approach with artificial neural networks. Optimal/essential sensors are identified by QR decomposition, and reconstruction is subsequently performed using nonlinear shallow decoder networks (SDN). The reconstruction accuracy could thus be further increased in this way. Assuming that in the future, groundwater quality parameters are collected by low-cost sensors with a frequency similar to GWL, this methodology could be used to determine an optimal GMN for multiple monitoring targets through multivariate analysis. The data-based identification from Chapter 4 based on two-dimensional input data (e.g., contour maps, model data) has also proved very promising. It can identify essential monitoring wells in an existing network and potential locations for new wells. Combined with numerical groundwater management

models, synergies could emerge, and monitoring networks can be iteratively optimized. In addition to the data-based identification of suitable sites, data constraints from model calibration indicate at which locations a well can provide extensive information gain.

For the spatial estimation of GWL and GWQ parameters from point measurements, machine learning methods have been considered inferior to geostatistical interpolation methods such as kriging. Reasons for this, besides data limitations, are that they do not take into account the spatial dependency structures in the data (Hengl et al., 2022). A growing number of studies utilize machine learning that incorporates spatial dependence structures with various geographic features (e.g., elevation, geographical proximity to sampling locations, receiving water, and watersheds) as covariates (e.g., Behrens et al., 2018; Hengl et al., 2018; Knoll et al., 2019). This leads to an increasing improvement in predictive performance with results already similar to those obtained with kriging. Therefore, it is likely that future methods will outperform kriging through optimized techniques and suitable covariates. Limitations such as stationarity assumptions or manual variogram fitting (including transformation or anisotropy parameters) would be overcome, and automated predictions facilitated.

Effective groundwater monitoring networks are essential for making groundwater "visible". In most cases, the knowledge we gain from them is the basis for many subsequent calculations and conceptions. This work has shown that false knowledge due to inappropriate monitoring methods can propagate and be drastically amplified. Effective monitoring networks are still rare, but modern cost-effective methods exist, including technology for real-time data recording. The best way to overcome the above-mentioned chicken-or-egg problem to optimal GMN is to combine site-specific expert hydrogeological knowledge with approved geostatistical and new data-driven machine learning techniques, which are developing rapidly. Machine learning techniques, in particular, benefit enormously from sufficient, adequate data, and therefore omissions in collection and survey should be made up now. Consequently, more collaboration between research, practice, and policymakers along with easier access to collected data will likely be needed in order to implement the knowledge gained into practice. Thus, according to J. Cherry (2020), effective, modern groundwater monitoring with data freely available data is one of the five essential actions governments can take to protect groundwater from the impacts of a global climate crisis.

Acknowledgements

I could not have done this work without the inspiration and support of several notable people. Here I would like to take the opportunity to express my gratitude to them.

First of all, I would like to thank my supervisor Nico Goldscheider for making this work possible, especially since it was done mainly independently of funded projects. I appreciate that I was given the freedom to work freely on this exciting topic and to realize my ideas, but at the same time, I could always count on your wise advice.

My deepest gratitude goes to Tanja Liesch. Your consistent support, encouragement, motivation, trust, and great ideas have made up a considerable part of this work. This work would not have been possible without your patient guidance.

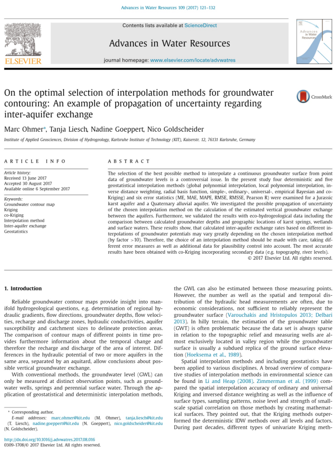
I am very grateful to Prof. Ty Ferré for spontaneously being my co-referee and to my PhD committee for their willingness and time.

My appreciation goes to Petra Linder and Christine Mackert for their great help in all organizational matters.

Thanks, Andreas Wunsch, for the great climbing trips and for holding the rope tight when necessary. Thank you, Nikolai Fahrmeier, for the always enjoyable and entertaining working atmosphere in our office. I would like to thank all other colleagues of the Departments of Hydrogeology and Engineering Geology who have accompanied me during these years for the great times and conversations, be it in the hallway, at the Thursday coffee break, lunch, BBQ, or Ph.D. party. Namely, these are Simon Frank, Julian Xanke, Anna Enders, Markus Merk, Zhao Chen, Dominik Richter and many more.

Thank you, Lisa, for the primarily funny, sometimes profound, often silly, but always great adventures we have had and will continue to have together.

Declaration of Authorship



Advances in Water Resources 109 (2017) 121–132

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

On the optimal selection of interpolation methods for groundwater contouring: An example of propagation of uncertainty regarding inter-aquifer exchange

Marc Ohmer*, Tanja Liesch, Nadine Goeppert, Nico Goldscheider

Institute of Applied Geosciences, Division of Hydrogeology, Karlsruhe Institute of Technology (KIT), Kaiserstr. 12, 76131 Karlsruhe, Germany

ARTICLE INFO

Article history:
Received 13 June 2017
Accepted 30 August 2017
Available online 6 September 2017

Keywords:
Groundwater contour map
Kriging
co-Kriging
Interpolation method
Inter-aquifer exchange
Geostatistics

ABSTRACT

The selection of the best possible method to interpolate a continuous groundwater surface from point data of groundwater levels is a controversial issue. In the present study four deterministic and five geostatistical interpolation methods (global polynomial interpolation, local polynomial interpolation, inverse distance weighting, radial basis function, simple-, ordinary-, universal-, empirical Bayesian and co-Kriging) and six error statistics (ME, MAE, MAPE, RMSE, RMSSE, Pearson R) were examined for a Jurassic karst aquifer and a Quaternary alluvial aquifer. We investigated the possible propagation of uncertainty of the chosen interpolation method on the calculation of the estimated vertical groundwater exchange between the aquifers. Furthermore, we validated the results with eco-hydrogeological data including the comparison between calculated groundwater depths and geographic locations of karst springs, wetlands and surface waters. These results show, that calculated inter-aquifer exchange rates based on different interpolations of groundwater potentials may vary greatly depending on the chosen interpolation method (by factor >10). Therefore, the choice of an interpolation method should be made with care, taking different error measures as well as additional data for plausibility control into account. The most accurate results have been obtained with co-Kriging incorporating secondary data (e.g. topography, river levels).

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Reliable groundwater contour maps provide insight into manifold hydrogeological questions, e.g. determination of regional hydraulic gradients, flow directions, groundwater depths, flow velocities, recharge and discharge zones, hydraulic conductivities, aquifer susceptibility and catchment sizes to delineate protection areas. The comparison of contour maps of different points in time provides furthermore information about the temporal change and therefore the recharge and discharge of the area of interest. Differences in the hydraulic potential of two or more aquifers in the same area, separated by an aquitard, allow conclusions about possible vertical groundwater exchange.

With conventional methods, the groundwater level (GWL) can only be measured at distinct observation points, such as groundwater wells, springs and perennial surface water. Through the application of geostatistical and deterministic interpolation methods, the GWL can also be estimated between those measuring points. However, the number as well as the spatial and temporal distribution of the hydraulic head measurements are often, due to economic considerations, not sufficient to reliably represent the groundwater surface (Varouchakis and Hristopulos 2013; Delbari 2013). In hilly terrain, the estimation of the groundwater table (GWT) is often problematic because the data set is always sparse in relation to the topographic relief and measuring wells are almost exclusively located in valley region while the groundwater surface is usually a subdued replica of the ground surface elevation (Hoeksema et al., 1989).

Spatial interpolation methods and including geostatistics have been applied to various disciplines. A broad overview of comparative studies of interpolation methods in environmental science can be found in Li and Heap (2008). Zimmerman et al. (1999) compared the spatial interpolation accuracy of ordinary and universal Kriging and inversed distance weighting as well as the influence of surface types, sampling patterns, noise level and strength of small-scale spatial correlation on those methods by creating mathematical surfaces. They pointed out, that the Kriging methods outperformed the deterministic IDW methods over all levels and factors. During past decades, different types of univariate Kriging meth-

* Corresponding author.
E-mail addresses: marc.ohmer@kit.edu (M. Ohmer), tanja.liesch@kit.edu (T. Liesch), nadine.goeppert@kit.edu (N. Goeppert), nico.goldscheider@kit.edu (N. Goldscheider).

http://dx.doi.org/10.1016/j.advwatres.2017.08.016
0309-1708/© 2017 Elsevier Ltd. All rights reserved.

Paper I (Chapter 2)

Citation: Ohmer, M., Liesch, T., Goeppert, N., Goldscheider, N. (2017) On the optimal selection of interpolation methods for groundwater contouring: An example of propagation of uncertainty regarding inter-aquifer exchange. Advances in Water Resources. Vol. 109. 121-132. 10.1016/j.advwatres.2017.08.016

Author Contribution: M. Ohmer (MO) & T. Liesch (TL) conceptualized the study, all authors contributed to the methodology. MO performed the investigation and validation of the results and the visualization of the results. MO wrote the original paper draft, N. Goldscheider (NG), N. Goeppert, & TL contributed to draft review and editing. TL & NG supervised the work.

Water Resources Research

RESEARCH ARTICLE

On the Optimal Spatial Design for Groundwater Level Monitoring Networks

M. Ohmer[1], T. Liesch[1], and N. Goldscheider[1]

[1]Institute of Applied Geosciences, Division of Hydrogeology, Karlsruhe Institute of Technology, Karlsruhe, Germany

Abstract Effective groundwater monitoring networks are important, as systematic data collected at observation wells provide a crucial understanding of the dynamics of hydrogeological systems as well as the basis for many other applications. This study investigates the influence of six groundwater level monitoring network (GLMN) sampling designs (random, grid, spatial coverage, and geostatistical) with varying densities on the accuracy of spatially interpolated groundwater surfaces. To obtain spatially continuous prediction errors (in contrast to point cross-validation errors), we used nine potentiometric groundwater surfaces from three regional MODFLOW groundwater flow models with different resolutions as a priori references. To assess the suitability of frequently-used cross-validation error statistics (MAE, RMSE, RMSSE, ASE, and NSE), we compared them with the actual prediction errors (APE). Additionally, we defined upper and lower thresholds for an appropriate spatial density of monitoring wells. Below the lower threshold, the observation density appears insufficient, and additional wells lead to a significant improvement of the results. Above the upper threshold, additional wells lead to only minor and inefficient improvements. According to the APE, systematic sampling lead to the best results but is often not suited for GLMN due to its nonprogressive characteristic. Geostatistical and spatial coverage sampling are considerable alternatives, which are in contrast progressive and allow evenly spaced and, in the case of spatial coverage sampling, yet reproducible coverage with accurate results. We found that the global cross-validation error statistics are not suitable to compare the performance of different sampling designs, although they allow rough conclusions about the quality of the GLMN.

1. Introduction

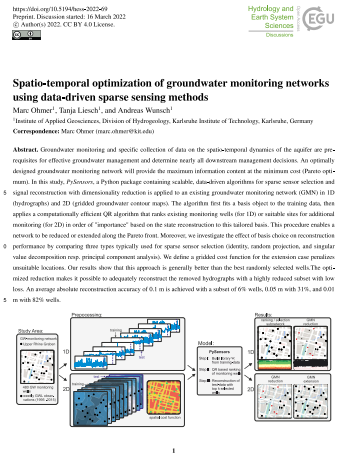
Groundwater is an important, yet spatially extensive, concealed, and inaccessible resource. Therefore, an effective groundwater monitoring network (GMN) is important, as systematic data collected at observation wells provide a crucial understanding of the dynamics and quality of the hydrogeological system. A GMN is defined by a spatial arrangement of monitoring sites and a temporal sampling frequency (Loaiciga et al., 1992). Economic considerations most strongly influence the number and location of monitoring wells. Designing an optimal GMN is, therefore, a task of balancing prediction accuracy with cost minimization (Kollat et al., 2011). Since a high spatial resolution is usually associated with disproportionate costs, often only domains of high water management importance are adequately monitored. The design, that is, the selection of the location and number of the monitoring wells, is a vital part of any study involving modeling and prediction based on spatial data. A groundwater model can only be as good as the model input data available. Therefore, poorly distributed monitoring wells can lead to wrong assumptions or to a bias of the regional image. Unsuitable interpolation methods can yield drastic overestimates or underestimates of the groundwater level in areas with a low monitoring density, as the parts with a high monitoring density are disproportionately weighted (Ohmer et al., 2017). GW-quality observations are subject to the same problem. In this study, however, we focus on the regional groundwater level as a monitoring parameter.

Varieties of studies dealing with the optimization of GMN have been published in the last 20 years. The literature focuses on the optimization of GMN design to observe the groundwater quality (GQMN), and the number of studies dealing with groundwater level (GLMN) is limited. The majority of approaches for GLMN design optimization are based on geostatistical analysis and therefore on minimizing uncertainty in parameter estimation. Several studies apply (multi)factorial kriging (Prakash & Singh, 2000; Theodossiou & Latinopoulos, 2006), ordinary kriging (OK; Nunes et al., 2004; Yang et al., 2008), universal kriging (UK; Kanevski et al., 2011; Kumar et al., 2007; Olea, 1984), OK and UK (Ahmadi & Sedghamiz, 2007), OK and co-kriging (CG; Mo et al., 1990), or indicator kriging (IK; Cameron & Hunter,

Paper II (Chapter 3)

Citation: Ohmer, M., Liesch, T., Goldscheider, N. (2019). On the Optimal Spatial Design for Groundwater Level Monitoring Networks. Water Resources Research, 55, 9454–9473. 10.1029/2019WR025728

Author Contribution: M. Ohmer (MO) & T. Liesch (TL) conceptualized the study, all authors contributed to the methodology. MO & TL developed the software code, performed formal analysis, investigation and validation. MO performed formal analysis, and validation of the results, visualized the results and wrote the original paper draft with contributions from TL. TL & N. Goldscheider both supervised the work.



https://doi.org/10.5194/hess-2022-69
Preprint. Discussion started: 16 March 2022

Hydrology and Earth System Sciences Discussions

Spatio-temporal optimization of groundwater monitoring networks using data-driven sparse sensing methods

Marc Ohmer[1], Tanja Liesch[1], and Andreas Wunsch[1]
[1]Institute of Applied Geosciences, Division of Hydrogeology, Karlsruhe Institute of Technology, Karlsruhe, Germany
Correspondence: Marc Ohmer (marc.ohmer@kit.edu)

Abstract. Groundwater monitoring and specific collection of data on the spatio-temporal dynamics of the aquifer are prerequisites for effective groundwater management and determine nearly all downstream management decisions. An optimally designed groundwater monitoring network will provide the maximum information content at the minimum cost (Pareto optimum). In this study, *PySensors*, a Python package containing scalable, data-driven algorithms for sparse sensor selection and signal reconstruction with dimensionality reduction is applied to an existing groundwater monitoring network (GMN) in 1D (hydrographs) and 2D (gridded groundwater contour maps). The algorithm first fits a basis object to the training data, then applies a computationally efficient QR algorithm that ranks existing monitoring wells (for 1D) or suitable sites for additional monitoring (for 2D) in order of "importance" based on the state reconstruction to this tailored basis. This procedure enables a network to be reduced or extended along the Pareto front. Moreover, we investigate the effect of basis choice on reconstruction performance by comparing three types typically used for sparse sensor selection (identity, random projection, and singular value decomposition resp. principal component analysis). We define a gridded cost function for the extension case penalizes unsuitable locations. Our results show that this approach is generally better than the best randomly selected wells.The optimized reduction makes it possible to adequately reconstruct the removed hydrographs with a highly reduced subset with low loss. An average absolute reconstruction accuracy of 0.1 m is achieved with a subset of 6% wells, 0.05 m with 31%, and 0.01 m with 82% wells.

Paper III (Chapter 4)

Citation: Ohmer, M., Wunsch A., Liesch, T. (2022). Spatio-temporal optimization of groundwater monitoring networks using data-driven sparse sensing methods. Hydrology and Earth System Sciences Discussions (HESS).1–29. 10.5194/hess-2022-69

Author Contribution: M. Ohmer (MO) & T. Liesch (TL) conceptualized the study and developed the methology and developed the software code. MO performed formal analysis, and validation of the results, visualized the results. A. Wunsch (AW) performed the data preprocessing and curation. MO wrote the original paper draft with contributions from TL and AW. TL supervised the work.

Other published works

- Liesch, T., **Ohmer, M**. (2016). Comparison of GRACE data and groundwater levels for the assessment of groundwater depletion in Jordan. *Hydrogeol Journal* 24, 1547–1563. https://doi.org/10.1007/s10040-016-1416-9.
- Frank, S, Goeppert, N, **Ohmer, M.**, Goldscheider, N. (2019). Sulfate variations as a natural tracer for conduit-matrix interaction in a complex karst aquifer. *Hydrological Processes*. ; 33: 1292–1303 https://doi.org/10.1002/hyp.13400.
- **Ohmer, M**., Klester, A., Kissinger, A., Mirbach, S., Class, H., Schneider, M., Lindenlaub, M., Bauer, M., Liesch, T., Menberg, K., Blum, P. (2022). Berechnung von Temperaturfahnen im Grundwasser mit analytischen und numerischen Modellen. *Grundwasser - Zeitschrift der Fachsektion Hydrogeologie*, 113–129, https://doi.org/10.1007/s00767-022-00509-2.
- **Ohmer, M**, Liesch, T., Goldscheider N., (2022). Extended hydrologic impacts of karst discharge zone confinement - a modeling study. *Hydrological Processes*, preprint, 1-16, https://doi.org/110.22541/au.165334713.38201184/v1 .
- Xanke, J., **Ohmer, M**, Liesch, T. (2022). Konzepte zur Wiederverwendung von aufbereitetem Abwasser im Unteren Jordantal. *WasserWirtschaft*, Ausgabe 7-8/2022, 61-67.

References

Aguilar, F. J., F. Agüera, M. A. Aguilar, and F. Carvajal (2005). "Effects of Terrain Morphology, Sampling Density, and Interpolation Methods on Grid DEM Accuracy". In: *Photogrammetric Engineering & Remote Sensing* 71.7, pp. 805–816. ISSN: 00991112. DOI: 10.14358/PERS.71.7.805.

Ahmadi, S. H. and A. Sedghamiz (2007). "Geostatistical Analysis of Spatial and Temporal Variations of Groundwater Level". In: *Environmental Monitoring and Assessment* 129.1-3, pp. 277–294. ISSN: 0167-6369, 1573-2959. DOI: 10.1007/s10661-006-9361-z.

Alfonso, L., E. Ridolfi, S. Gaytan-Aguilar, F. Napolitano, and F. Russo (2014). "Ensemble Entropy for Monitoring Network Design". In: *Entropy* 16.3, pp. 1365–1375. ISSN: 1099-4300. DOI: 10.3390/e16031365.

Alizadeh, Z., J. Yazdi, and A. Moridi (2018). "Development of an Entropy Method for Groundwater Quality Monitoring Network Design". In: *Environmental Processes* 5.4, pp. 769–788. ISSN: 2198-7491, 2198-7505. DOI: 10.1007/s40710-018-0335-2.

Alizadeh, Z. and N. Mahjouri (2017). "A spatiotemporal Bayesian maximum entropy-based methodology for dealing with sparse data in revising groundwater quality monitoring networks: the Tehran region experience". In: *Environmental Earth Sciences* 76.12, p. 436. ISSN: 1866-6280, 1866-6299. DOI: 10.1007/s12665-017-6767-6.

Ammar, K., A. Khalil, M. McKee, and J. Kaluarachchi (2008). "Bayesian deduction for redundancy detection in groundwater quality monitoring networks". In: *Water Resources Research* 44.8, p. 15. ISSN: 00431397. DOI: 10.1029/2006WR005616.

Anderson, M., W. Woessner, and R. J. Hunt (2015). *Applied groundwater modeling: simulation of flow and advective transport*. Second edition. London ; San Diego, CA: Academic Press. ISBN: 978-0-12-058103-0.

Annoni, J., T. Taylor, C. Bay, K. Johnson, L. Pao, P. Fleming, and K. Dykes (2018). "Sparse-Sensor Placement for Wind Farm Control". In: *Journal of Physics: Conference Series* 1037, p. 11. ISSN: 1742-6588, 1742-6596. DOI: 10.1088/1742-6596/1037/3/032019.

Anonymous (1881). *Bericht der meteorologischen Commission des naturforschenden Vereines in Bruenn ueber die Ergebnisse der meteorologischen Beobachtungen im*

Jahre 1881 (Report of the meteorology Commision of the Society of nature research in Brno on the results of meteorology monitoring in 1881). Bruenn: W. Burkart – Verlag des Vereines.

Arslan, H. (2014). "Spatial and temporal distribution of areas with drainage problems as estimated by different interpolation techniques: Spatial and temporal distribution of areas". In: *Water and Environment Journal* 28.2, pp. 203–211. ISSN: 17476585. DOI: 10.1111/wej.12026.

Asefa, T., M. W. Kemblowski, G. Urroz, M. McKee, and A. Khalil (2004). "Support vectors-based groundwater head observation networks design". In: *Water Resources Research* 40.11, p. 14. ISSN: 00431397. DOI: 10.1029/2004WR003304.

Ayvaz, M. T. and A. Elçi (2018). "Identification of the optimum groundwater quality monitoring network using a genetic algorithm based optimization approach". In: *Journal of Hydrology* 563, pp. 1078–1091. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2018.06.006.

Babbar-Sebens, M. and B. Minsker (2010). "A Case-Based Micro Interactive Genetic Algorithm (CBMIGA) for interactive learning and search: Methodology and application to groundwater monitoring design". In: *Environmental Modelling & Software* 25.10, pp. 1176–1187. ISSN: 13648152. DOI: 10.1016/j.envsoft.2010.03.027.

BAFU (2019). *Zustand und Entwicklung Grundwasser Schweiz*. Nr. 1901 Stand 2016. Bern: Bundesamt für Umwelt BAFU, p. 138.

Baraniuk, R. (2007). "A Lecture on Compressive Sensing v". In: *IEEE signal processing magazine* 24.4, pp. 118–121.

Bashi-Azghadi, S. N. and R. Kerachian (2010). "Locating monitoring wells in groundwater systems using embedded optimization and simulation models". In: *Science of The Total Environment* 408.10, pp. 2189–2198. ISSN: 00489697. DOI: 10.1016/j.scitotenv.2010.02.004.

Behrens, T., K. Schmidt, R. A. Viscarra Rossel, P. Gries, T. Scholten, and R. A. MacMillan (2018). "Spatial modelling with Euclidean distance fields and machine learning: Spatial modelling with Euclidean distance fields". In: *European Journal of Soil Science* 69.5, pp. 757–770. ISSN: 13510754. DOI: 10.1111/ejss.12687.

Bhat, S., L. H. Motz, C. Pathak, and L. Kuebler (2015). "Geostatistics-based groundwater-level monitoring network design and its application to the Upper Floridan aquifer, USA". In: *Environmental Monitoring and Assessment* 187.1, p. 15. ISSN: 0167-6369, 1573-2959. DOI: 10.1007/s10661-014-4183-x.

Bierer, S. (1987). *75 Jahre Landeswasserversorgung: 1912 -1987*. Zweckverband Landeswasserversorgung. Stuttgart: Ernst Klett Verlag.

Birch, C. P., S. P. Oom, and J. A. Beecham (2007). "Rectangular and hexagonal grids used for observation, experiment and simulation in ecology". In: *Ecological Modelling* 206.3-4, pp. 347–359. ISSN: 03043800. DOI: 10.1016/j.ecolmodel.2007.03.041.

Blum, A., A. Winchel, and A. Laurent (2013). "Groundwater monitoring over the decades". In: *the Briefs* 6, p. 16.

BMU (2016). *Nitratbericht 2016 - Gemeinsamer Bericht der Bundesministerien für Umwelt, Naturschutz, Bau und Reaktorsicherheit sowie für Ernährung und Landwirtschaft*. Tech. rep. Bonn: Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit, p. 143.

Bohling, G. (2005). "Introduction to geostatistics and variogram analysis". In: C&PE 940, p. 20.

Brown, J. A., P Robertson, and T. Mc Donald (2015). "Spatially Balanced Sampling: Application to Environmental Surveys". In: *Procedia Environmental Sciences* 27, pp. 6–9. ISSN: 1878-0296. DOI: 10.1016/j.proenv.2015.07.108.

Brunton, B. W., S. L. Brunton, J. L. Proctor, and J. N. Kutz (2016). "Sparse Sensor Placement Optimization for Classification". In: *SIAM Journal on Applied Mathematics* 76.5, pp. 2099–2122. ISSN: 0036-1399, 1095-712X. DOI: 10.1137/15M1036713.

Brunton, S. L. and J. N. Kutz (2017). *Data-driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press. ISBN: 978-1-108-38069-0. DOI: 10.1017/9781108380690.

Brus, D. J. (2010). "Design-based and model-based sampling strategies for soil monitoring". In: *19th World Congress of Soil Science, Soil Solutions for a Changing World*. Brisbane, Australia. DOI: 10.1029/2007WR006123.

Brus, D. and J. de Gruijter (1997). "Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion)". In: *Geoderma* 80.1-2, pp. 1–44. ISSN: 00167061. DOI: 10.1016/S0016-7061(97)00072-4.

Cameron, K. and P. Hunter (2002). "Using spatial models and kriging techniques to optimize long-term ground-water monitoring networks: a case study". In: *Environmetrics* 13.5-6, pp. 629–656. DOI: 10.1002/env.582.

Candes, E. and M. Wakin (2008). "An Introduction To Compressive Sampling". In: *IEEE Signal Processing Magazine* 25.2, pp. 21–30. ISSN: 1053-5888. DOI: 10.1109/MSP.2007.914731.

Candes, E., J. Romberg, and T. Tao (2005). "Stable Signal Recovery from Incomplete and Inaccurate Measurements". In: *Pure Appl. Math*, pp. 1207–1223. DOI: 10.1002/cpa.20124.

Castillo, A. and A. R. Messina (2020). "Data-driven sensor placement for state reconstruction via POD analysis". In: *IET Generation, Transmission & Distribution* 14.4, pp. 656–664. ISSN: 1751-8695, 1751-8695. DOI: 10.1049/iet-gtd.2019.0199.

Chandan, K. S. and B. K. Yashwant (2017). "Optimization of groundwater level monitoring network using GIS-based geostatistical method and multi-parameter analysis: A case study in Wainganga Sub-basin, India". In: *Chinese Geographical Science* 27.2, pp. 201–215. ISSN: 1002-0063, 1993-064X. DOI: 10.1007/s11769-017-0859-9.

Cherry, J. A. (2020). "Interview with Dr. John Cherry: 5 Things Goverments Can Do To Protect Groundwater". In: *SIWI Stockholm Waterfront*. Stockholm International Water Institute (SIWI) No. 2, pp. 4–9. ISSN: 1102 7053.

Chilton, J. (2009). "Groundwater Monitoring Approaches at International Level". In: *Groundwater monitoring*. Ed. by P. Quevauviller. Water quality measurements series. Chichester, U.K: J. Wiley, pp. 25–31. ISBN: 978-0-470-77809-8.

Chinchuluun, A., P. Pardalos, A. Migdalas, L. Pitsoulis, and P. Grant, eds. (2008). *Pareto Optimality, Game Theory And Equilibria*. Vol. 17. Springer Optimization and Its Applications. New York, NY: Springer New York. ISBN: 978-0-387-77246-2 978-0-387-77247-9. DOI: 10.1007/978-0-387-77247-9.

Chung, J. and J. D. Rogers (2012). "Interpolations of Groundwater Table Elevation in Dissected Uplands". In: *Groundwater* 50.4, pp. 598–607. ISSN: 0017467X. DOI: 10.1111/j.1745-6584.2011.00889.x.

Chunmei, W., Y. Qinke, L. Hongyan, G. Weiling, D. L. Jupp, and L. Rui (2013). "Influence of resolution on elevation and slope at watershed scale in Loess Plateau". In: *SpringerPlus* 2.S1, S13. ISSN: 2193-1801. DOI: 10.1186/2193-1801-2-S1-S13.

Clark, E., T. Askham, S. L. Brunton, and J. Nathan Kutz (2019). "Greedy Sensor Placement With Cost Constraints". In: *IEEE Sensors Journal* 19.7, pp. 2642–2656. ISSN: 1530-437X, 1558-1748, 2379-9153. DOI: 10.1109/JSEN.2018.2887044.

Clutterbuck, J., J. Dickinson, L. Playfair, S. Homersham, F. Braithwaite, J. Horn, and F. Ward (1850). "On the periodical alternations, andprogressive permanent depression, of the Chalk waterlevel under London". In: *Proceedings of the Institution of Civil Engineers* 9, pp. 151–155.

Coburn, T. C., J. M. Yarus, and R. L. Chambers (2006). *Stochastic Modeling and Geostatistics: Principles, Methods, and Case Studies, Volume II*. American Association of Petroleum Geologists. ISBN: 978-1-62981-036-2. DOI: 10.1306/CA51063.

Conybeare, W. and W. Phillips (1822). *Outlines of the geology of England and Wales, with an introductory compendium of the general principles of that science, and comparative views of the structure of foreign countries. Part 1*. London: William Phillips.

Cooper, H. M., C. Zhang, and D. Selch (2015). "Incorporating uncertainty of groundwater modeling in sea-level rise assessment: a case study in South Florida". In: *Climatic*

Change 129.1-2, pp. 281–294. ISSN: 0165-0009, 1573-1480. DOI: 10.1007/s10584-015-1334-1.

Corput, J. G. v. d. (1935). "Verteilungsfunktionen. I". German. In: *Proceedings. Akadamie van Wetenschappen Amsterdam* 38. Publisher: North-Holland, Amsterdam, pp. 813–821. ISSN: 0370-0348.

Dalal, I. L., J. Harwayne-Gidansky, and D. Stefan (2008). "On the fast generation of long-period pseudorandom number sequences". In: *2008 IEEE Long Island Systems, Applications and Technology Conference*. Farmingdale, NY: IEEE, pp. 1–9. ISBN: 978-1-4244-1731-5. DOI: 10.1109/LISAT.2008.4638953.

Darcy, H. (1856). *Les fontaines publiques de la ville de Dijon*. Qual des Augustins 49. Paris: V. Dalmont.

Dasgupta, S. (2000). "Experiments with Random Projection". In: *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pp. 143–151. DOI: 10.48550/arXiv.1301.3849.

Dassargues, A., O. Batelaan, and A. Anceau (2021). "The First Potentiometric Map". In: *Groundwater* 59.5, pp. 772–779. ISSN: 0017-467X, 1745-6584. DOI: 10.1111/gwat.13123.

Davenport, R. J., M. Satchell, and L. M. W. Shaw-Taylor (2019). "Cholera as a 'sanitary test' of British cities, 1831–1866". In: *The History of the Family* 24.2, pp. 404–438. ISSN: 1081-602X, 1873-5398. DOI: 10.1080/1081602X.2018.1525755.

De Jesus, K. L. M., D. B. Senoro, J. C. Dela Cruz, and E. B. Chan (2021). "A Hybrid Neural Network–Particle Swarm Optimization Informed Spatial Interpolation Technique for Groundwater Quality Mapping in a Small Island Province of the Philippines". In: *Toxics* 9.11, p. 273. ISSN: 2305-6304. DOI: 10.3390/toxics9110273.

Delbari, M. (2014). "Accounting for exhaustive secondary data into the mapping of water table elevation". In: *Arabian Journal of Geosciences* 7.10, pp. 4221–4233. ISSN: 1866-7511, 1866-7538. DOI: 10.1007/s12517-013-0986-2.

Delmelle, E. M. (2014). "Spatial sampling". In: *In M. M. Fischer & P. Nijkamp (Eds.), Handbook of Regional Science*. Berlin, Heidelberg: Springer, pp. 1385–1399. ISBN: 978-3-642-23431-6.

Derer, C. E. (2003). "Tectono-sedimentary evolution of the northern Upper Rhine Graben (Germany), with special regard to the early syn-rift stage". PhD thesis. Bonn: Rheinischen Friedrich-Wilhelms-Universität Bonn.

DeSimone, L. A., D. A. Walter, J. R. Eggleston, and M. T. Nimiroski (2002). *Simulation of ground-water flow and evaluation of water-management alternatives in the upper Charles River basin, eastern Massachusetts*. Report 2002-4234. DOI: 10.3133/wri024234.

Dhar, A. and R. S. Patil (2012). "Multiobjective Design of Groundwater Monitoring Network Under Epistemic Uncertainty". In: *Water Resources Management* 26.7, pp. 1809–1825. ISSN: 0920-4741, 1573-1650. DOI: 10.1007/s11269-012-9988-1.

Donoho, D. (2006). "Compressed sensing". In: *IEEE Transactions on Information Theory* 52.4, pp. 1289–1306. ISSN: 0018-9448. DOI: 10.1109/TIT.2006.871582.

Dumont, G. (1856). *Rapport fait par le représentant du Peuple Daunou*. Liège: Labalte et Durand.

Dupuit, J. (1863). *Etudes theoriques et pratiques sur le mouvement des eaux dans les canaux decouverts et a travers les terrains permeables*. Paris: Dunod.

EC (1991). *Council Directive 91/676/EC of 12 December 1991 concerning the protection of waters against pollution caused by nitrates from agricultural sources*. European Commission.

EC (2000). *Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy*. European Commission.

— (2006). *Directive 2006/118/EC of the European Parliament and of the Council of 12 December 2006 on the protection of groundwater against pollution and deterioration*. European Commission.

— (2007). *Guidance on groundwater monitoring No 15*. European Commission. Luxembourg: OPOCE. ISBN: 978-92-79-04558-5.

Emmert, M., N. Zigelli, F. Haakh, F. Bode, and W. Nowak (2016). "Risikiobasiertes Grundwassermonitoring für Wasserschutzgebiete". In: *energie/wasser-praxis*, pp. 68–73.

Esquivel, J. M., G. P. Morales, and M. V. Esteller (2015). "Groundwater Monitoring Network Design Using GIS and Multicriteria Analysis". In: *Water Resources Management* 29.9, pp. 3175–3194. ISSN: 0920-4741, 1573-1650. DOI: 10.1007/s11269-015-0989-8.

Feinstein, D., M. Fienen, J. Kennedy, C. Buchwald, and M. Greenwood (2012). *Development and application of a groundwater/surface-water flow model using MODFLOW-NWT for the Upper Fox River Basin, southeastern Wisconsin*. Report 2012-5108. Reston, VA, p. 138. DOI: 10.3133/sir20125108.

Fischer, M. M. and P. Nijkamp, eds. (2014). *Handbook of Regional Science*. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-642-23429-3 978-3-642-23430-9. DOI: 10.1007/978-3-642-23430-9.

Flinspach, D., F. Haakh, A. Locher, Mäck, B. Röhrle, and R. Scheck (1997). *Das württembergische Donauried - Seine Bedeutung für Wasserversorgung, Landwirtschaft und Naturschutz*. Stuttgart: Zweckverband Landeswasserversorgung (self published).

Frank, S., N. Goeppert, M. Ohmer, and N. Goldscheider (2019). "Sulfate variations as a natural tracer for conduit-matrix interaction in a complex karst aquifer". en. In: *Hydrological Processes* 33.9, pp. 1292–1303. ISSN: 0885-6087, 1099-1085. DOI: 10.1002/hyp.13400.

Gaur, S., S. Ch, D. Graillot, B. R. Chahar, and D. N. Kumar (2013). "Application of Artificial Neural Networks and Particle Swarm Optimization for the Management of Groundwater Resources". In: *Water Resources Management* 27.3, pp. 927–941. ISSN: 0920-4741, 1573-1650. DOI: 10.1007/s11269-012-0226-7.

Geuna, S. (2000). "Appreciating the difference between design-based and model-based sampling strategies in quantitative morphology of the nervous system". In: *Journal of Comparative Neurology* 427.3, pp. 333–339. DOI: 10.1002/1096-9861.

Gilbert, R. O. (1987). *Statistical methods for environmental pollution monitoring*. New York: Van Nostrand Reinhold Co. ISBN: 978-0-442-23050-0.

Golub, G and W Kahan (1965). "Calculating the Singular Value and Pseudo-Inverse of a Matrix". In: *J. SIAM Numer. Anal.* Ser. B, Vol. 2, No.2, pp. 205–224. DOI: 10.1137/0702016.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press. ISBN: 978-0-19-511538-3.

Graaf, I. E. M. de, T. Gleeson, L. P. H. (Rens) van Beek, E. H. Sutanudjaja, and M. F. P. Bierkens (2019). "Environmental flow limits to global groundwater pumping". In: *Nature* 574.7776, pp. 90–94. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-019-1594-4.

Gribov, A. and K. Krivoruchko (2004). "Geostatistical Mapping with Continuous Moving Neighborhood". In: *Mathematical Geology* 36.2, pp. 267–281. ISSN: 0882-8121. DOI: 10.1023/B:MATG.0000020473.63408.17.

Gruijter, J. J. de, M. Bierkens, D. J. Brus, and M. Knotters (2006). *Sampling for Natural Resource Monitoring*. 1st ed. Berlin Heidelberg: Springer-Verlag. ISBN: 978-3-540-33161-2.

Guekie, S., T. Aubin, A. Marache, R. Lastennet, and D. Breysse (2016). "Geostatistical investigations for suitable mapping of the water table: the Bordeaux case (France)". In: *Hydrogeology Journal* 24.1, pp. 231–248. DOI: 10.1007/s10040-015-1316-4.

Guneshwor, L., T. I. Eldho, and A. Vinod Kumar (2018). "Identification of Groundwater Contamination Sources Using Meshfree RPCM Simulation and Particle Swarm Optimization". In: *Water Resources Management* 32.4, pp. 1517–1538. ISSN: 0920-4741, 1573-1650. DOI: 10.1007/s11269-017-1885-1.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009). "Decomposition of the mean squared error and NSE performance criteria: Implications for improving

hydrological modelling". In: *Journal of Hydrology* 377.1-2, pp. 80–91. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2009.08.003.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1999). "Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration". In: *Journal of Hydrologic Engineering* 4.2, pp. 135–143. ISSN: 1084-0699, 1943-5584. DOI: 10.1061/(ASCE)1084-0699(1999)4:2(135).

Halko, N., P.-G. Martinsson, and J. A. Tropp (2011). "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions". In: *SIAM Rev., Survey and Review section* Vol. 53, num. 2, pp. 217–288. DOI: 10.1137/090771806.

Halton, J. H. (1960). "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals". In: *Numerische Mathematik* 2.1, pp. 84–90. ISSN: 0029-599X, 0945-3245. DOI: 10.1007/BF01386213.

Hansen, M. H., W. G. Madow, and B. J. Tepping (1983). "An Evaluation of Model-Dependent and Probability-Sampling Inferences in Sample Surveys". In: *Journal of the American Statistical Association* 78.384. Publisher: Taylor & Francis, pp. 776–793. ISSN: 0162-1459. DOI: 10.1080/01621459.1983.10477018.

Hengl, T., M. Nussbaum, M. Wright, G. Heuvelink, and B. Gräler (2018). "Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables". In: *PeerJ* 6, p. 49. ISSN: 2167-8359. DOI: 10.7717/peerj.5518.

Hengl, T., L. Parente, and I. Wheeler (2022). "Spatial sampling and resampling for Machine Learning". In: DOI: 10.5281/ZENODO.5886677.

Hengl, T. (2009). *A Practical Guide to Geostatistical Mapping*. Vol. 2nd Edition. Amsterdam: JRC Scientific and Technical Reports. Amsterdam. ISBN: 978-90-90-24981-0.

Hensley, D. and F. E. Su (2001). "Random walks with badly approximable numbers". In: *DIMACS volume "Unusual Applications of Number Theory"*, p. 7. DOI: 10.48550/arXiv.math/0102206.

Hernandez-Stefanoni, J. L. and R. Ponce-Hernandez (2006). "Mapping the Spatial Variability of Plant Diversity in a Tropical Forest: Comparison of Spatial Interpolation Methods". In: *Environmental Monitoring and Assessment* 117.1-3, pp. 307–334. ISSN: 0167-6369, 1573-2959. DOI: 10.1007/s10661-006-0885-z.

Heuvelink, G. B. M. and E. J. Pebesma (2002). *Is the ordinary kriging variance a proper measure of interpolation error?* Melbourne: RMIT University: Editors G. Hunter and K. Lowell.

Hoeksema, R. J., R. B. Clapp, A. L. Thomas, A. E. Hunley, N. D. Farrow, and K. C. Dearstone (1989). "Cokriging model for estimation of water table elevation". In: *Water Resources Research* 25.3, pp. 429–438. ISSN: 00431397. DOI: 10.1029/WR025i003p00429.

Holmes, S. (1985). *History of Water Resources activities of the U.S. Geological Survey (water fact sheet)*. Report 85-646. DOI: 10.3133/ofr85646.

Hosseini, M. and R. Kerachian (2017). "A Bayesian maximum entropy-based methodology for optimal spatiotemporal design of groundwater monitoring networks". In: *Environmental Monitoring and Assessment* 189.9, 433 (24). ISSN: 0167-6369, 1573-2959. DOI: 10.1007/s10661-017-6129-6.

Hu, K. L., B. G. Li, Y. Z. Lu, and F. R. Zhang (2004). "Comparison of various spatial interpolation methods for non-stationary regional soil mercury content". In: *Huanjing Kexue / Chinese Journal of Environmental Science* 25.3, pp. 132–137. ISSN: ISSN 0250-3301.

Hua, Z., M. Dehai, and W. Cheng (2009). "Optimization of the Spatial Interpolation for Groundwater Depth in Shule River Basin". In: *2009 International Conference on Environmental Science and Information Application Technology*. Wuhan, China: IEEE, pp. 415–418. ISBN: 978-0-7695-3682-8. DOI: 10.1109/ESIAT.2009.478.

Hussain, Z. and A. Muhammad (2013). "Sample size reduction in groundwater surveys via sparse data assimilation". In: *2013 IEEE 10th International Conference on Networking, Sensing and Control (ICNSC 2013)*. Evry: IEEE, pp. 176–182. ISBN: 978-1-4673-5200-0 978-1-4673-5198-0 978-1-4673-5199-7. DOI: 10.1109/ICNSC.2013.6548732.

IGRAC (2020). *Groundwater monitoring programmes: A global overview of quantitative groundwater monitoring networks*. Tech. rep. Delft: International Groundwater Resources Assessment Centre (IGRAC), p. 144.

Ikechukwu, M. N., E. Ebinne, U. Idorenyin, and N. I. Raphael (2017). "Accuracy Assessment and Comparative Analysis of IDW, Spline and Kriging in Spatial Interpolation of Landform (Topography): An Experimental Study". In: *Journal of Geographic Information System* 09.03, pp. 354–371. ISSN: 2151-1950, 2151-1969. DOI: 10.4236/jgis.2017.93022.

Jasechko, S., D. Perrone, K. Befus, M. Bayani Cardenas, G. Ferguson, T. Gleeson, E. Luijendijk, J. McDonnell, R. Taylor, Y. Wada, and J. Kirchner (2017). "Global aquifers dominated by fossil groundwaters but wells vulnerable to modern contamination". In: *Nature Geoscience* 10.6, pp. 425–429. ISSN: 1752-0894, 1752-0908. DOI: 10.1038/ngeo2943.

Johnston, K., J. M. V. Hoef, K. Krivoruchko, and N. Lucas (2004). *Using ArcGIS™ Geostatistical Analyst*. Redlands, California: Esri Press.

Jones, L. E., J. A. Painter, J. H. LaFontaine, N. Sepúlveda, and D. F. Sifuentes (2017). *Groundwater-flow budget for the lower Apalachicola-Chattahoochee-Flint River Basin in southwestern Georgia and parts of Florida and Alabama, 2008–12*. Report 2017-5141. Reston, VA, p. 88. DOI: 10.3133/sir20175141.

Jousma, G. and F. Roelofsen (2004). *World-wide inventory on groundwater monitoring*. Tech. rep. GP 2004-1. Utrecht: International Groundwater Resources Assessment Centre.

Júnez-Ferreira, H. E. and G. S. Herrera (2013). "A geostatistical methodology for the optimal design of space–time hydraulic head monitoring networks and its application to the Valle de Querétaro aquifer". In: *Environmental Monitoring and Assessment* 185.4, pp. 3527–3549. ISSN: 0167-6369, 1573-2959. DOI: 10.1007/s10661-012-2808-5.

Júnez-Ferreira, H. E., G. S. Herrera, L. González-Hita, A. Cardona, and J. Mora-Rodríguez (2016). "Optimal design of monitoring networks for multiple groundwater quality parameters using a Kalman filter: application to the Irapuato-Valle aquifer". In: *Environmental Monitoring and Assessment* 188.1, 39 (22). ISSN: 0167-6369, 1573-2959. DOI: 10.1007/s10661-015-5036-y.

Kambhammettu, B. V. N. P., P. Allena, and J. P. King (2011). "Application and evaluation of universal kriging for optimal contouring of groundwater levels". In: *Journal of Earth System Science* 120.3, pp. 413–422. ISSN: 0253-4126, 0973-774X. DOI: 10.1007/s12040-011-0075-4.

Kermorvant, C., N. Caill-Milly, B. Noëlle, and F. D'Amico (2019). "Optimizing cost-efficiency of long term monitoring programs by using spatially balanced sampling designs: The case of manila clams in Arcachon bay". In: *Ecological Informatics* 49, pp. 32–39. ISSN: 1574-9541. DOI: 10.1016/j.ecoinf.2018.11.005.

Keum, J., K. Kornelsen, J. Leach, and P. Coulibaly (2017). "Entropy Applications to Water Monitoring Network Design: A Review". In: *Entropy* 19.11, 613 (21). ISSN: 1099-4300. DOI: 10.3390/e19110613.

Khader, A. I. and M. McKee (2014). "Use of a relevance vector machine for groundwater quality monitoring network design under uncertainty". In: *Environmental Modelling & Software* 57, pp. 115–126. ISSN: 13648152. DOI: 10.1016/j.envsoft.2014.02.015.

Khalil, A., M. N. Almasri, M. McKee, and J. J. Kaluarachchi (2005). "Applicability of statistical learning algorithms in groundwater quality modeling". In: *Water Resources Research* 41.5. ISSN: 00431397. DOI: 10.1029/2004WR003608.

Kim, K.-H. and K.-K. Lee (2007). "Optimization of groundwater-monitoring networks for identification of the distribution of a contaminant plume". In: *Stochastic Environmental Research and Risk Assessment* 21.6, pp. 785–794. ISSN: 1436-3240, 1436-3259. DOI: 10.1007/s00477-006-0094-x.

Kinzelbach, W. and H. Kunstmann (1998). "Nachhaltige Grundwassernutzung: Kriterien und Strategien". In: *Spektrum der Wissenschaft* 4, p. 94.

Kish, L. (1995). *Survey sampling*. New York: Wiley-Interscience. ISBN: 978-0-471-10949-5.

Knoben, W. J. M., J. E. Freer, and R. A. Woods (2019). "Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores". In: *Hydrol. Earth Syst. Sci.* 23, pp. 4323–4331. DOI: 10.5194/hess-2019-327.

Knoll, L., L. Breuer, and M. Bach (2019). "Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning". In: *Science of The Total Environment* 668, pp. 1317–1327. ISSN: 00489697.

Kollat, J. and P. Reed (2006). "Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design". In: *Advances in Water Resources* 29.6, pp. 792–807. ISSN: 03091708. DOI: 10.1016/j.advwatres.2005.07.010.

Kollat, J., P. Reed, and R. Maxwell (2011). "Many-objective groundwater monitoring network design using bias-aware ensemble Kalman filtering, evolutionary optimization, and visual analytics". In: *Water Resour Res* 47.W02529, p. 18. DOI: 10.1029/2010WR009194.

Komasi, M. and H. Goudarzi (2021). "Multi-objective optimization of groundwater monitoring network using a probability Pareto genetic algorithm and entropy method (case study: Silakhor plain)". In: *Journal of Hydroinformatics* 23.1, pp. 136–150. ISSN: 1464-7141, 1465-1734. DOI: 10.2166/hydro.2020.061.

Koreimann, C, J Grath, G Winkler, W Nagy, and W. R. Vogel (1996). *Groundwater Monitoring in Europe*. Tech. rep. Copenhagen: European Environment Agency, p. 143.

Krešic, N. (2007). *Hydrogeology and groundwater modeling*. 2. ed. Boca Raton: CRC Press. ISBN: 978-0-8493-3348-4.

Krige, D. G. (1951). "A statistical approach to some basic mine valuation problems on the Witwatersrand". In: *Journal of the Chemical, Metallurgical and Mining Society of South Africa* 52.6, pp. 119–139.

Krivoruchko, K. (2011). *Spatial Statistical Data Analysis for GIS Users*. Redlands, California: Esri Press.

Kumar, S., S. K. Sondhi, and V. Phogat (2005). "Network design for groundwater level monitoring in Upper Bari Doab Canal tract, Punjab, India". In: *Irrigation and Drainage* 54.4, pp. 431–442. ISSN: 1531-0353, 1531-0361. DOI: 10.1002/ird.194.

Leach, J. M., P. Coulibaly, and Y. Guo (2016). "Entropy based groundwater monitoring network design considering spatial distribution of annual recharge". In: *Advances in Water Resources* 96, pp. 108–119. ISSN: 03091708. DOI: 10.1016/j.advwatres.2016.07.006.

Lee, T.-W., J. Y. Lee, J. E. Park, H. Bellerova, and M. Raudensky (2021). "Reconstructive Mapping from Sparsely-Sampled Groundwater Data Using Compressive Sensing". In: *Journal of Geographic Information System* 13.03, pp. 287–301. ISSN: 2151-1950, 2151-1969. DOI: 10.4236/jgis.2021.133016.

Li, J. and A. D. Heap (2008). *A Review of Spatial Interpolation Methods for Environmental Scientists*. Tech. rep. ISBN: 978 1 921498 30 5. Canberra: Geoscience Australia, p. 137.

— (2011). "A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors". In: *Ecological Informatics* 6.3-4, pp. 228–241. ISSN: 1574-9541. DOI: 10.1016/j.ecoinf.2010.12.003.

Li, J., A. Bárdossy, L. Guenni, and M. Liu (2011). "A Copula based observation network design approach". In: *Environmental Modelling & Software* 26.11, pp. 1349–1357. ISSN: 13648152. DOI: 10.1016/j.envsoft.2011.05.001.

Li, P., T. J. Hastie, and K. W. Church (2006). "Very sparse random projections". In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06*. Philadelphia, PA, USA: ACM Press, p. 287. ISBN: 978-1-59593-339-3. DOI: 10.1145/1150402.1150436.

Liesch, T. and M. Ohmer (2016). "Comparison of GRACE data and groundwater levels for the assessment of groundwater depletion in Jordan". In: *Hydrogeology Journal* 24.6, pp. 1547–1563. ISSN: 1431-2174, 1435-0157. DOI: 10.1007/s10040-016-1416-9.

Lindinger, H. and A. Scheidleder (2004). *(WEU01) Nitrate in groundwater. Indicator Fact Sheet*. European Nitrates Directive (91/676/EEC). Kopenhagen: European Environment Agency (EEA), p. 8.

Liu, J. and C. Zheng (2016). "Towards Integrated Groundwater Management in China". In: *Integrated Groundwater Management*. Ed. by A. Jakeman, O. Barreteau, R. J. Hunt, J. Rinaudo, and A. Ross. Cham: Springer International Publishing, pp. 455–475. ISBN: 978-3-319-23575-2 978-3-319-23576-9. DOI: 10.1007/978-3-319-23576-9_18.

Loaiciga, H. A., R. J. Charbeneau, L. G. Everett, G. E. Fogg, B. F. Hobbs, and S. Rouhani (1992). "Review of Ground-Water Quality Monitoring Network Design". In: *Journal of Hydraulic Engineering* 118.1, pp. 11–37. ISSN: 0733-9429, 1943-7900. DOI: 10.1061/(ASCE)0733-9429(1992)118:1(11).

LUBW (2006). *Hydrogeologischer Bau und hydraulische Eigenschaften - INTERREG III A-Projekt MoNit „Modellierung der Grundwasserbelastung durch Nitrat im Oberrheingraben"*. Tech. rep. Karlsruhe: LUBW Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg.

LUBW (2021a). "Grundwasserüberwachungsprogramm. Ergebnisse 2018 und 2019". In: Reihe Grundwasserschutz Bd. 61, 2020, p. 76. ISSN: 1437-0131.

— (2021b). *Umwelt-Daten und -Karten Online (UDO)*. URL: https://udo.lubw.baden-wuerttemberg.de/public/ (visited on 12/30/2021).

Lucas, J. (1876). "Artesian system of the Thames basin". In: *Journal of the Society of Arts* 25. London, United Kingdom, Royal Society of Arts., p. 8. ISSN: 0035-9114.

Luo, Q., J. Wu, Y. Yang, J. Qian, and J. Wu (2016). "Multi-objective optimization of long-term groundwater monitoring network design using a probabilistic Pareto genetic algorithm under uncertainty". In: *Journal of Hydrology* 534, pp. 352–363. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2016.01.009.

Ma, T.-S., M. Sophocleous, and Y.-S. Yu (1999). "Geostatistical Applications in Ground-Water Modeling in South-Central Kansas". In: *Journal of Hydrologic Engineering* 4.1, pp. 57–64. ISSN: 1084-0699, 1943-5584. DOI: 10.1061/(ASCE)1084-0699(1999)4:1(57).

Manohar, K., B. Brunton, K. Kutz, and S. Brunton (2018). "Data-Driven Sparse Sensor Placement for Reconstruction: Demonstrating the Benefits of Exploiting Known Patterns". In: *IEEE Control Systems* 38.3, pp. 63–86. ISSN: 1066-033X, 1941-000X. DOI: 10.1109/MCS.2018.2810460.

Masoumi, F. and R. Kerachian (2008). "Assessment of the groundwater salinity monitoring network of the Tehran region: application of the discrete entropy theory". In: *Water Science and Technology* 58.4, pp. 765–771. ISSN: 0273-1223. DOI: 10.2166/wst.2008.674.

Mather, J. and G. S. o. London (2004). *200 Years of British Hydrogeology*. Geological Society special publication 200 years of British hydrogeology. Geological Society. ISBN: 978-1-86239-155-0.

Matheron, G. (1963). "Principles of geostatistics". In: *Economic Geology* 58.8, pp. 1246–1266. ISSN: 0361-0128. DOI: 10.2113/gsecongeo.58.8.1246.

Mitas, L. and H. Mitasova (2011). "Spatial interpolation". In: *Geographic information systems & science*. Ed. by P. Longley. 3rd ed. OCLC: ocn502414068. Hoboken, NJ: Wiley, pp. 481–492. ISBN: 978-0-470-72145-2 978-0-470-72144-5.

Mogheir, Y., J. L. M. P. de Lima, and V. P. Singh (2009). "Entropy and Multi-Objective Based Approach for Groundwater Quality Monitoring Network Assessment and Redesign". In: *Water Resources Management* 23.8, pp. 1603–1620. ISSN: 0920-4741, 1573-1650. DOI: 10.1007/s11269-008-9343-8.

Moore, G. E. (1965). "Cramming more components onto integrated circuits". In: 38.8, p. 4.

Moriasi, D., J. G. Arnold, M. W. V. Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith (2007). "Model Evaluation Guidelines for Systematic Quantification of Accuracy in

Watershed Simulations". In: *Transactions of the ASABE* 50.3, pp. 885–900. ISSN: 2151-0040. DOI: 10.13031/2013.23153.

Muzikar, R. (2013). "Hydrogeology in the Czech Republic". In: *History of hydrogeology*. First issued in paperback. International contributions to hydrogeology 28. Boca Raton London New York: CRC Press, pp. 47–58. ISBN: 978-0-367-57662-2 978-0-415-63062-7.

Möhler, F., S. Dinse, and A. Hermsdorf (2014). "Grundwassergleichenplan für Brandenburg – Interpolation mittels Kriging mit externer Drift". In: *Grundwasser* 19.3, pp. 189–199. ISSN: 1430-483X, 1432-1165. DOI: 10.1007/s00767-014-0255-7.

Nash, J. and J. Sutcliffe (1970). "River flow forecasting through conceptual models part I — A discussion of principles". In: *Journal of Hydrology* 10.3, pp. 282–290. ISSN: 00221694. DOI: 10.1016/0022-1694(70)90255-6.

Nunes, L. M., M. C. Cunha, and L. Ribeiro (2004). "Groundwater Monitoring Network Optimization with Redundancy Reduction". In: *Journal of Water Resources Planning and Management* 130.1, pp. 33–43. ISSN: 0733-9496, 1943-5452. DOI: 10.1061/(ASCE)0733-9496(2004)130:1(33).

O'Callaghan, J. F. and D. M. Mark (1983). "The Extraction of Drainage Networks from Digital Elevation Data". In: *Computer Vision, Graphics, and Image Processing* 28.3, pp. 323–344. ISSN: 0734-189X. DOI: 10.1016/S0734-189X(84)80011-0.

Ohmer, M., T. Liesch, and N. Goldscheider (2019). "On the Optimal Spatial Design for Groundwater Level Monitoring Networks". In: *Water Resources Research* 55.11, pp. 9454–9473. ISSN: 0043-1397, 1944-7973. DOI: 10.1029/2019WR025728.

— (2022a). "Extended hydrologic impacts of karst discharge zone confinement - a modeling study". In: Hydrological Processes.submitted, p. 14. DOI: 10.22541/au.165334713.38201184/v1.

Ohmer, M. (2022). *0.1.0 marcohmer/GMNO: Initial Release*. Zenodo. URL: https://doi.org/10.5281/zenodo.6075863.

Ohmer, M., A. Klester, A. Kissinger, S. Mirbach, H. Class, M. Schneider, M. Lindenlaub, M. Bauer, T. Liesch, K. Menberg, and P. Blum (2022b). "Berechnung von Temperaturfahnen im Grundwasser mit analytischen und numerischen Modellen : Calculation of temperature plumes in groundwater with analytical and numerical models". In: *Grundwasser*. Publisher: Springer. ISSN: 1430-483X, 1432-1165. DOI: 10.1007/s00767-022-00509-2.

Ohmer, M., T. Liesch, N. Goeppert, and N. Goldscheider (2017). "On the optimal selection of interpolation methods for groundwater contouring: An example of propagation of uncertainty regarding inter-aquifer exchange". In: *Advances in Water Resources* 109, pp. 121–132. ISSN: 03091708. DOI: 10.1016/j.advwatres.2017.08.016.

Olea, R. A. (1984). "Sampling design optimization for spatial functions". In: *Mathematical Geology* 16, pp. 369–392. DOI: 10.1007/BF01029887.

Oliver, M. A., R. Webster, and S. P. McGRATH (1996). "Disjunctive Kriging for Environmental Management". In: *Environmetrics* 7.3, pp. 333–357. ISSN: 11804009, 1099095X. DOI: 10.1002/(SICI)1099-095X(199605)7:3<333::AID-ENV209>3.0.CO;2-V.

Parsen, M. J., K. R. Bradbury, R. J. Hunt, and D. T. Feinstein (2016). *The 2016 groundwater flow model for Dane County, Wisconsin*. Bulletin / Wisconsin Geological and Natural History Survey 110. Madison, Wisconsin: Wisconsin Geological and Natural History Survey. ISBN: 978-0-88169-992-0.

Pollard, A., L. Castillo, L. Danaila, and M. Glauser, eds. (2017). *Whither Turbulence and Big Data in the 21st Century*. Cham: Springer International Publishing. ISBN: 978-3-319-41215-3 978-3-319-41217-7. DOI: 10.1007/978-3-319-41217-7.

Pourshahabi, S., N. Talebbeydokhti, G. Rakhshandehroo, and M. R. Nikoo (2018). "Spatio-Temporal Multi-Criteria Optimization of Reservoir Water Quality Monitoring Network Using Value of Information and Transinformation Entropy". In: *Water Resources Management* 32.10, pp. 3489–3504. ISSN: 0920-4741, 1573-1650. DOI: 10.1007/s11269-018-2003-8.

Prakash, M. and V. Singh (2000). "Network design for groundwater monitoring - a case study". In: *Environmental Geology* 39.6, pp. 628–632. ISSN: 0943-0105. DOI: 10.1007/s002540050474.

Price, M. (2004). "Dr John Snow and an early investigation of groundwater contamination". In: *Geological Society, London, Special Publications* 225.1, pp. 31–49. ISSN: 0305-8719, 2041-4927. DOI: 10.1144/GSL.SP.2004.225.01.04.

Przyrowski, R. and A. Schäfer (2015). "Quaternary fluvial basin of northern Upper Rhine Graben". In: *Zeitschrift der Deutschen Gesellschaft für Geowissenschaften* 166.1, pp. 71–98. ISSN: 1860-1804. DOI: 10.1127/1860-1804/2014/0080.

Puri, D., K. Borel, C. Vance, and R. Karthikeyan (2017). "Optimization of a Water Quality Monitoring Network Using a Spatially Referenced Water Quality Model and a Genetic Algorithm". In: *Water* 9.9, p. 704. ISSN: 2073-4441. DOI: 10.3390/w9090704.

Reed, P., J. B. Kollat, and V. Devireddy (2007). "Using interactive archives in evolutionary multiobjective optimization: A case study for long-term groundwater monitoring design". In: *Environmental Modelling & Software* 22.5, pp. 683–692. ISSN: 13648152. DOI: 10.1016/j.envsoft.2005.12.021.

Reed, P., B. Minsker, and A. J. Valocchi (2000). "Cost-effective long-term groundwater monitoring design using a genetic algorithm and global mass interpolation". In: *Water Resources Research* 36.12, pp. 3731–3741. DOI: 10.1029/2000WR900232.

Reed, P. M. and J. B. Kollat (2013). "Visual analytics clarify the scalability and effectiveness of massively parallel many-objective optimization: A groundwater monitoring design example". In: *Advances in Water Resources* 56, pp. 1–13. ISSN: 03091708. DOI: 10.1016/j.advwatres.2013.01.011.

Richter, O. (1943). *Johann Gregor Mendel wie er wirklich war*. Brünn: mähr. Landesbehörde, der Landeshauptstadt Brünn und der Deutschen Akademie der Wissenschaften in Prag.

Rivoirard, J. (1990). *Introduction to disjunctive kriging and nonlinear geostatistics*. Fontainebleau: Centre de Géostatistique.

Roberts, M. (2018). *The Unreasonable Effectiveness of Quasirandom Sequences. A new low discrepancy quasirandom sequence that offers many substantial improvements over other popular sequences such as the Sobol and Halton sequences*. URL: http://extremelearning.com.au/unreasonable-effectiveness-of-quasirandom-sequences (visited on 04/07/2022).

Romero, V., R. Slepoy, L. Swiler, A. Giunta, and T. Krishnamurthy (2005). "Error Estimation Approaches for Progressive Response Surfaces". In: *46th AIAA/ASME /ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*. Austin, Texas: American Institute of Aeronautics and Astronautics. ISBN: 978-1-62410-065-9. DOI: 10.2514/6.2005-1822.

Sadat Noori, S. M., K. Ebrahimi, A.-M. Liaghat, and A.-H. Hoorfar (2013). "Comparison of different geostatistical methods to estimate groundwater level at different climatic periods: Comparison of different geostatistical methods". In: *Water and Environment Journal* 27.1, pp. 10–19. ISSN: 17476585. DOI: 10.1111/j.1747-6593.2012.00321.x.

Sagan, V., K. T. Peterson, M. Maimaitijiang, P. Sidike, J. Sloan, B. A. Greeling, S. Maalouf, and C. Adams (2020). "Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing". In: *Earth-Science Reviews* 205, p. 103187. ISSN: 0012-8252. DOI: 10.1016/j.earscirev.2020.103187.

Shannon, C. (1949). "Communication in the Presence of Noise". In: *Proceedings of the IRE* 37.1, pp. 10–21. ISSN: 0096-8390. DOI: 10.1109/JRPROC.1949.232969.

Siebert, S., J. Burke, J. M. Faures, K. Frenken, J. Hoogeveen, P. Döll, and F. T. Portmann (2010). "Groundwater use for irrigation – a global inventory". In: *Hydrology and Earth System Sciences* 14.10, pp. 1863–1880. ISSN: 1607-7938. DOI: 10.5194/hess-14-1863-2010.

Silva, B. M. de, K. Manohar, E. Clark, B. W. Brunton, S. L. Brunton, and J. N. Kutz (2021). "PySensors: A Python Package for Sparse Sensor Placement". In: *Journal of Open Source Software* 6(58).2828, p. 4. DOI: 10.21105/joss.02828.

Singh, D. and B. Datta (2016). "Linked Optimization Model for Groundwater Monitoring Network Design". In: *Urban Hydrology, Watershed Management and Socio-Economic Aspects*. Vol. 73. Series Title: Water Science and Technology Library. Cham: Springer International Publishing, pp. 107–125. ISBN: 978-3-319-40194-2 978-3-319-40195-9. DOI: 10.1007/978-3-319-40195-9_9.

Siska, P. P., P. Goovaerts, I.-K. Hung, and V. M. Bryant (2005). "Predicting ordinary kriging errors caused by surface roughness and dissectivity". In: *Earth Surface Processes and Landforms* 30.5, pp. 601–612. ISSN: 0197-9337, 1096-9837. DOI: 10.1002/esp.1164.

Snow, J. (1854). *On the mode of communication of cholera*. John Churchill.

Sreekanth, J., H. Lau, and D. E. Pagendam (2017). "Design of optimal groundwater monitoring well network using stochastic modeling and reduced-rank spatial prediction: Optimal Monitoring Network Design". In: *Water Resources Research* 53.8, pp. 6821–6840. ISSN: 00431397. DOI: 10.1002/2017WR020385.

Sun, Y., S. Kang, F. Li, and L. Zhang (2009). "Comparison of interpolation methods for depth to groundwater and its temporal and spatial variations in the Minqin oasis of northwest China". In: *Environmental Modelling & Software* 24.10, pp. 1163–1170. ISSN: 13648152. DOI: 10.1016/j.envsoft.2009.03.009.

Särndal, C.-E., I. Thomsen, J. M. Hoem, D. V. Lindley, O. Barndorff-Nielsen, and T. Dalenius (1978). "Design-Based and Model-Based Inference in Survey Sampling". In: *Scandinavian Journal of Statistics* 5.1, pp. 27–52. ISSN: 03036898, 14679469.

Tapoglou, E., G. P. Karatzas, I. C. Trichakis, and E. A. Varouchakis (2014). "A spatio-temporal hybrid neural network-Kriging model for groundwater level simulation". In: *Journal of Hydrology* 519, pp. 3193–3203. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2014.10.040.

Thakur, J. (2015). "Optimizing Groundwater Monitoring Networks Using Integrated Statistical and Geostatistical Approaches". In: *Hydrology* 2.3, pp. 148–175. ISSN: 2306-5338. DOI: 10.3390/hydrology2030148.

Thakur, J. K. (2017). "Hydrogeological modeling for improving groundwater monitoring network and strategies". In: *Applied Water Science* 7.6, pp. 3223–3240. ISSN: 2190-5487, 2190-5495. DOI: 10.1007/s13201-016-0469-1.

Theodossiou, N. and P. Latinopoulos (2006). "Evaluation and optimisation of groundwater observation networks using the Kriging methodology". In: *Environmental Modelling & Software* 21.7, pp. 991–1000. ISSN: 13648152. DOI: 10.1016/j.envsoft.2005.05.001.

Thiem, A. (1870). "Über die Ergiebigkeit artesischer Bohrlöcher, Schachtbrunnen und Filtergalerien". In: *Journ. f. Gasbeleuchtung und Wasserversorgung* 14, pp. 450–467.

Tobler, W. R. (1970). "A Computer Movie Simulating Urban Growth in the Detroit Region". In: *Economic Geography* 46, p. 234. ISSN: 00130095. DOI: 10.2307/143141.

Uddameri, V. and T. Andruss (2014). "A GIS-based multi-criteria decision-making approach for establishing a regional-scale groundwater monitoring". In: *Environmental Earth Sciences* 71.6, pp. 2617–2628. ISSN: 1866-6280, 1866-6299. DOI: 10.1007/s12665-013-2899-5.

Udluft, P. (2000). *Das Grundwasser im schwäbischen Donautal*. München: Bayerischer Industrieverband Steine und Erden E. V. ISBN: 3-9807169-0-2.

USGS (2014). *Shuttle Radar Topography Mission (SRTM) 1 Arc-Second. N48E009V3 - N48E010V3*. URL: https://earthexplorer.usgs.gov/.

Ushijima, T. T., W. W. G. Yeh, and W. K. Wong (2021). "Constructing robust and efficient experimental designs in groundwater modeling using a Galerkin method, proper orthogonal decomposition, and metaheuristic algorithms". In: *PLOS ONE* 16.8. Ed. by S. Mirjalili, p. 27. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0254620.

Varouchakis, E. A. and D. Hristopulos (2013). "Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins". In: *Environmental Monitoring and Assessment* 185.1, pp. 1–19. ISSN: 0167-6369, 1573-2959. DOI: 10.1007/s10661-012-2527-y.

Vicente-Serrano, S., M. Saz-Sánchez, and J. Cuadrat (2003). "Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature". In: *Climate Research* 24, pp. 161–180. ISSN: 0936-577X, 1616-1572. DOI: 10.3354/cr024161.

Vieux, B. E. (2016). *Distributed Hydrologic Modeling Using GIS*. Vol. 74. Water Science and Technology Library. Dordrecht: Springer Netherlands. ISBN: 978-94-024-0928-4 978-94-024-0930-7. DOI: 10.1007/978-94-024-0930-7.

Villinger, E. (1977). *Über Potentialverteilung und Strömungssysteme im Karstwasser der Schwäbischen Alb (Oberer Jura, SW-Deutschland)*. Stuttgart, Germany: Schweizerbart Science Publishers. ISBN: 978-3-510-96238-9.

Wackernagel, H. (1995). "Ordinary Kriging." In: *Multivariate Geostatistics*. Springer Berlin Heidelberg. ISBN: 978-3-662-03100-1. DOI: 10.1007/978-3-662-03098-1_11.

Wadoux, A. M. J.-C., D. J. Brus, and G. B. M. Heuvelink (2019). "Sampling design optimization for soil mapping with random forest". In: *Geoderma* 355, p. 113913. ISSN: 0016-7061. DOI: 10.1016/j.geoderma.2019.113913.

Wang, S., G. H. Huang, Q. G. Lin, Z. Li, H. Zhang, and Y. R. Fan (2014). "Comparison of interpolation methods for estimating spatial distribution of precipitation in Ontario, Canada". In: *International Journal of Climatology* 34.14, pp. 3745–3751. ISSN: 08998418. DOI: 10.1002/joc.3941.

Wilde, B. (2009). "Minimizing Error Variance in Estimates by Optimum Placement of Samples – A Comparison of Optimization Techniques". In: *Paper 406, CCG Annual Report 11*, p. 10.

Williams, J., O. Zahn, and J. N. Kutz (2022). "Data-driven sensor placement with shallow decoder networks". en. In: *arXiv*. arXiv: 2202.05330. DOI: 10.48550/arXiv.2202.05330.

Willmott, C. J. (1982). "Some Comments on the Evaluation of Model Performance". In: *Bulletin of the American Meteorological Society* 63.11, pp. 1309–1313. ISSN: 0003-0007, 1520-0477. DOI: 10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2.

Winkler, H. A. G. (1972). *Das Grundwasser im Nördlinger Ries unter Berücksichtigung der hydrologischen und hydrochemischen Beziehungen zum Speichergestein.* München: Technische Universität München, Fakultät f. Allg. Wissenschaften.

Wirsing, G. and A. Luz (2007). *Hydrogeologischer Bau und Aquifereigenschaften der Lockergesteine im Oberrheingraben (Baden-Württemberg)*. 19. Freiburg: LGRB - Infortmationen. ISBN: 1619-5329.

World Water Assessment Programme (2009). *The United Nations World Water Development Report 3: Water in a Changing World.* Paris: UNESCO, and London: Earthscan. ISBN: 978-92-3-104095-5.

Worley, B. (2016). "Subrandom methods for multidimensional nonuniform sampling". In: *Journal of Magnetic Resonance* 269, pp. 128–137. DOI: 10.1016/j.jmr.2016.06.007.

Wu, Y. (2004). "Optimal design of a groundwater monitoring network in Daqing, China". In: *Environmental Geology* 45.4, pp. 527–535. ISSN: 0943-0105, 1432-0495. DOI: 10.1007/s00254-003-0907-x.

Wunsch, A., T. Liesch, and S. Broda (2021). "Feature-based Groundwater Hydrograph Clustering Using Unsupervised Self-Organizing Map-Ensembles". In: *Water Resources Management*. ISSN: 0920-4741, 1573-1650. DOI: 10.1007/s11269-021-03006-y.

Xiao, Y., X. Gu, S. Yin, J. Shao, Y. Cui, Q. Zhang, and Y. Niu (2016). "Geostatistical interpolation model selection based on ArcGIS and spatio-temporal variability analysis of groundwater level in piedmont plains, northwest China". In: *SpringerPlus* 5.1, p. 425. ISSN: 2193-1801. DOI: 10.1186/s40064-016-2073-0.

Xie, Y., T.-b. Chen, M. Lei, J. Yang, Q.-j. Guo, B. Song, and X.-y. Zhou (2011). "Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: Accuracy and uncertainty analysis". In: *Chemosphere* 82.3, pp. 468–476. ISSN: 00456535. DOI: 10.1016/j.chemosphere.2010.09.053.

Yang, F.-g., S.-y. Cao, X.-n. Liu, and K.-j. Yang (2008). "Design of Groundwater Level Monitoring Network with Ordinary Kriging". In: *Journal of Hydrodynamics* 20.3, pp. 339–346. ISSN: 1001-6058, 1878-0342. DOI: 10.1016/S1001-6058(08)60066-9.

Yao, L., Z. Huo, S. Feng, X. Mao, S. Kang, J. Chen, J. Xu, and T. S. Steenhuis (2014). "Evaluation of spatial interpolation methods for groundwater level in an arid inland oasis, northwest China". In: *Environmental Earth Sciences* 71.4, pp. 1911–1924. ISSN: 1866-6280, 1866-6299. DOI: 10.1007/s12665-013-2595-5.

Yeh, M.-S., Y.-P. Lin, and L.-C. Chang (2006). "Designing an optimal multivariate geostatistical groundwater quality monitoring network using factorial kriging and genetic algorithms". In: *Environmental Geology* 50.1, pp. 101–121. ISSN: 0943-0105, 1432-0495. DOI: 10.1007/s00254-006-0190-8.

Yildirim, B., C. Chryssostomidis, and G. Karniadakis (2009). "Efficient sensor placement for ocean measurements using low-dimensional concepts". In: *Ocean Modelling* 27.3-4, pp. 160–173. ISSN: 14635003. DOI: 10.1016/j.ocemod.2009.01.001.

Yudina, E., A. Petrovskaia, D. Shadrin, P. Tregubova, E. Chernova, M. Pukalchik, and I. Oseledets (2021). "Optimization of Water Quality Monitoring Networks Using Metaheuristic Approaches: Moscow Region Use Case". In: *Water* 13.7, p. 888. ISSN: 2073-4441. DOI: 10.3390/w13070888.

Zaisheng, H. and C. Mengxiong (2013). "History of hydrogeology in China". In: *History of hydrogeology*. International contributions to hydrogeology 28. Boca Raton London New York: CRC Press. ISBN: 978-0-367-57662-2 978-0-415-63062-7.

Zhang, Y., G. F. Pinder, and G. S. Herrera (2005). "Least cost design of groundwater quality monitoring networks". In: *Water Resources Research* 41.8. ISSN: 00431397. DOI: 10.1029/2005WR003936.

Zhou, Y. (2013). "Upgrading a regional groundwater level monitoring network for Beijing Plain, China". In: *Geoscience Frontiers* 4.1, pp. 127–138. ISSN: 1674-9871. DOI: 10.1016/j.gsf.2012.03.008.

Zimmerman, D., C. Pavlik, A. Ruggles, and M. P. Armstrong (1999). "An Experimental Comparison of Ordinary and Universal Kriging and Inverse Distance Weighting". In: *Mathematical Geology* 31, p. 16. ISSN: 0882-8121. DOI: 10.1023/A:1007586507433.