The Challenge of Interdisciplinarity at the Intersection of Groundwater Management and Visualization Research

Laura Lotteraner [ID], *Faculty of Computer Science and Environmental Geosciences, University of Vienna, 1090, Vienna, Austria*

Thilo Hofmann [ID], *Environmental Geosciences, University of Vienna, 1090, Vienna, Austria*

Torsten Möller [ID], *Faculty of Computer Science and Research Network Data Science, University of Vienna, 1090, Vienna, Austria*

This design study presents an analysis and abstraction of temporal and spatial data, and workflows in the domain of hydrogeology and the design and development of an interactive visualization prototype. Developed in close collaboration with a group of hydrogeological researchers, the interface supports them in data exploration, selection of data for their numerical model calibration, and communication of findings to their industry partners. We highlight both pitfalls and learnings of the iterative design and validation process and explore the role of rapid prototyping. Some of the main lessons were that the ability to see their own data changed the engagement of skeptical users dramatically and that interactive rapid prototyping tools are thus powerful to unlock the advantage of visual analysis for novice users. Further, we observed that the process itself helped the domain scientists understand the potential and challenges of their data more than the final interface prototype.

Groundwater is one of the primary sources of drinking water, covering around 45% of the demand globally and 70% in the European Union. As climate change endangers water resources worldwide, the importance of groundwater as a resource is increasing further.[1]

Monitoring and protecting groundwater quality and quantity is crucial in ensuring safe drinking water.[1] Various tools and approaches have been established for this purpose. In recent decades, for example, geographic information systems (GIS) have been widely used in combination with statistical methods to help monitor water quality from a spatial perspective.

However, they are limited in dealing with, analyzing and visualizing time series data.[2] This temporal variability in hydrochemical water quality data is paramount for adequately managing resources and promptly reacting to unforeseen events. Both academic researchers and water suppliers have developed different solutions according to their specific needs, but the community lacks a standard tool for visualizing temporal and spatial data.

In this design study, we explore the temporal and spatial aspects of data used in hydrogeological research, the needs of academic researchers for visualizing these data, and the tasks the data are used for.

We then present the design and development process of an interactive visualization prototype for water management that provides insight into chemical and hydrological data from a spatial and temporal perspective through a combination of dashboards. Our focus is on the difficulties and pitfalls that we encountered during the design process, building on

and extending the work of Sedlmair et al.[3] Our experiences highlight both the importance and lack of visualization literacy of domain experts and motivate alternative approaches to the conventional design study methodology.

Having been implemented using Tableau, the interface also demonstrates the usefulness of traditional visualization approaches. It serves as a proof-of-concept for using existing domain-agnostic visualization software to support groundwater research and management instead of developing custom tools.

DOMAIN BACKGROUND

Hydrogeology is an interdisciplinary field studying the behavior of water in the ground. Various approaches, including hydrogeological and numerical models, have been developed to describe and solve different problems to sustain healthy groundwater systems and the best possible groundwater quality to ensure drinking water supply.[4]

Numerical models are based on the physical behavior of water in the specific geology of the modeling area and are calibrated using, for example, hydraulic heads (groundwater levels). Flow models describe the movement of water in the subsurface, and transport models the movement of substances.[5] While these models are used extensively to make predictions and to identify best resource management options based on (limited) field observations[6] some problems remain open. Since all groundwater models are underdetermined, i.e., there are orders of magnitude fewer field observations available than parameters to be fitted, traditional calibration of models based on hydraulic heads cannot uniquely constrain the parameter field and yields more than one possible solution.[5] Additional independent data significantly reduce the uncertainty of such models and help make better decisions. For example, including and calibrating chemical data in addition to groundwater levels may drastically improve a model.[5]

Hydrogeologists, both academic researchers and practitioners, have particular skill sets. Often, they are specialized to use a specific type of model and the software that comes with it. The group we worked with were experts in numerical modeling software, such as FEFLOW,[7] and relied on MS Excel for all additional data analysis and visualizations. Aside from using Python to automatize data input and output for their FEFLOW models, they had little experience in writing code.

The community lacks a standard tool, known and used across institutions, that helps researchers efficiently and effectively select parameters to improve their models and calibration based on their available data. Water utilities measure hundreds of different parameters, often at high time resolution over many years, so scientists are confronted with selecting suitable parameters from tens of thousands of data points for better system understanding and model calibration. Their current selections are guided by prior knowledge and experiences and are thus limited. Furthermore, numerical models only cover some aspects of hydrogeological research. When it comes to finding and describing correlations and relationships between different (chemical and hydrological) parameters, a tool for efficient data exploration is crucial. Not only does it enable researchers to compare many different parameters quickly, it also opens new research directions as the large amounts of data collected by water utilities become accessible.

RELATED WORK

Discussing all tools and approaches developed for and used in hydrogeology would fill a book by itself and is therefore out of scope for this article. Instead, we will focus on software relevant to the problems at hand.

For spatial information, hydrogeologists have been working with GIS software for many years. GIS tools are used to visualize groundwater and surface water, hydrological and chemical data.[2] Many scientists have built their applications based on existing GIS software such as, among others, QGIS[8] and ArcGIS.[9] Often these applications are integrated with computational models in the backend, e.g., numerical simulations,[10] statistical methods,[11] and machine learning models. GIS-based applications do not only visualize and display data; they also include data storage, manipulation, and export functionalities.[12]

Temporal data (i.e., time series) are often visualized directly in the software used for numerical modeling, such as, among many others, FEFLOW[7] or simply in MS Excel. Even though visual analysis of time series plots plays an important role[13] in hydrogeological research, computational time series analysis approaches, combined with custom visualizations, are gaining relevance.[14]

While spatial and temporal data individually contain valuable information, often a combination of both is of high interest. Visualization of spatiotemporal data has been studied from different perspectives, and various approaches have been introduced.[15,16]

To visualize spatiotemporal hydrogeological data many powerful custom approaches, tools, and applications have been developed that offer a wide range of

functionalities (e.g., see Benedict et al.[17] Castrellón et al.[18] Guzman et al.[19] and Mazher[20]). However, none of them covered the needs of the domain scientists we were working with, and, more importantly, none of them were known nor accessible to them. This motivated our interest in using widely known, domain-agnostic visualization software, such as Tableau in hydrogeology, a research direction that has only recently emerged.[21] All custom tools share the combination of maps with time series plots of points selected on the map. Often, maps are continuously colored to display hydrogeological values[19,20] while others use the map only to localize the measuring points.[18]

PROCESS

For data and requirements analysis, feedback and validation, we talked to three researchers from our case study who have been closely collaborating with a water utility for a year and their supervisor overseeing the cooperation for several years (who also became a co-author of this article).

Over six months, 15 interviews and feedback sessions in different combinations of the four participants took place (see Figure 1). We talked to the researchers both in groups and in one-on-one meetings.

The first set of one-on-one requirement interviews was based on a fixed list of questions found in the supplementary material. During these interviews, the researchers were not able to articulate their needs beyond the statement that they had large amounts of data and lacked the tools to analyze them efficiently.

To help the researchers formulate requirements, we decided to take an iterative approach and worked with prototypes that could already load in "real" data from the very beginning. Tableau, as a rapid prototyping framework, allowed us to iterate on prototypes

FIGURE 1. The interface was developed in an iterative process over six months, starting with individual requirements interviews, continuing with prototyping and open discussion rounds, and ending with individual validation interviews.

and incorporate feedback quickly, yet enabled us to show the actual data of interest to the potential users early on. That created an agency for the users and got them more involved. We also switched to open discussions during the design phase as it became apparent that more ideas were generated this way.

The first prototypes were based on the needs and ideas that were vaguely stated in the first round of requirement interviews and our own experiences trying to get an overview of the data. These prototypes had the goal of providing a basis for more productive discussions and interviews. Screenshots and descriptions of the prototypes can be found in the supplementary material. The prototypes were presented to the research group supervisor to narrow down requirements. The final prototype was built based on the requirements established in this phase and presented to the entire team of researchers in individual sessions and improved in an iterative process. New ideas and suggestions emerged in each discussion, supporting the original observation that the researchers could not pinpoint what they needed. Development was also complicated by contradicting requests, such as "we need a map on every dashboard" versus "maps take a lot of space, only include one on the overview dashboard," which were often voiced by the same person in different conversations. Throughout many interviews and discussions, we filtered out relevant requirements instead of temporary preferences.

Besides prototype discussions, weekly group discussions about the researchers' current work and attending their project meetings helped refine requirements. After implicitly validating the interface during the design phase, we concluded with another questionnaire to get feedback on all components and aspects of the interface from all four participants. The questionnaire can be found in the supplementary material.

The prototyping process of the interface significantly overlapped with the case study, as it was the process itself more than the final interface that helped the domain scientists gain insight into their data.
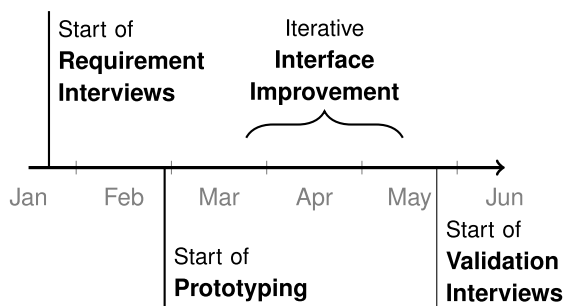
DATA AND TASK ANALYSIS

In this section, we discuss data sources and structures, workflows and tasks for this design study. All specific examples we are referring to are from the water utility from our case study but from the perspective of the academic researchers collaborating with the practitioners.

Data Abstraction
At first, the data seem relatively simple. They primarily come as time series of measured values, each linked
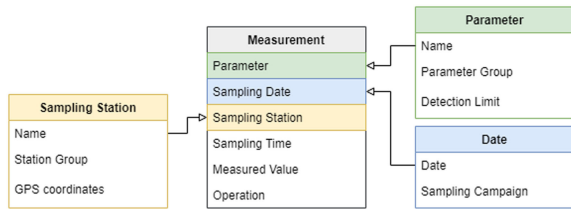
FIGURE 2. Each data point consisting of parameter, sampling date and time, sampling station, and the measured value is complemented with additional information.

to a spatial location (sampling station) and a chemical compound (parameter).

Data are complemented with metadata; see Figure 2. Parameters are grouped into parameter groups according to their properties. Some chemical and microbial parameters have a detection limit, i.e., the lowest possible concentration (value). Concentrations below that limit cannot be determined and are recorded as some fixed value below the limit. Sampling stations are grouped into station groups according to their location, i.e., rivers, groundwater, or water utility infrastructure. Some dates are part of a sampling campaign. Dates are also linked to the Boolean operation parameter indicating whether or not the water utility is currently delivering water to the city.

Data Challenges

Researchers face the challenge of different data sources and, therefore, a variety of data structures. Parameters are measured either by the water utility or the academic researchers themselves, in different locations and in different temporal granularity.

For hydrogeological time series, water utilities combine frequent field sampling with on- and offline data loggers and probes in the infrastructure. Especially, rivers supplying the area, and groundwater observation wells collect data in different frequencies, ranging from once per minute to once per month. In the groundwater wells, they continuously measure and store hydraulic heads (groundwater levels) and water temperature. In the water collection infrastructure, they measure water levels, extraction at the water utility, and electrical conductivity. In the rivers, they measure discharge (water flow), temperature, and water level. Data from the infrastructure data loggers are automatically transferred to the water utility's database and used for operation management. In all other locations, data are stored locally and have to be extracted manually in the field, which is usually only done upon request.

Chemical and microbial data are collected at distinct sampling points in the water collection infrastructure, groundwater observation wells, and rivers. Sampling frequencies range from daily to once per year or are triggered by events, such as floods. Measurements are stored in several databases and are provided to the researchers as exported MS Excel sheets. Although chemical data are not used in daily operations, many chemical parameters have to be tracked in order to comply with legal regulations regarding drinking water quality.[22]

As some research questions require a finer grid of sampling stations for a specific point in time, researchers measure chemical parameters in additional groundwater wells and rivers in their own sampling campaigns.

The hydrological and chemical parameters that are actively measured by the water utility and the researchers are complemented by data freely available from public authorities (e.g., precipitation, river levels, and discharge) or other sources (e.g., geocoordinates from Google Maps). Other data that have been discussed but not yet included are hydrological and chemical measurements referring to a wastewater treatment plant upstream of the study area. Although this example is specific to our case study, a similar situation can be assumed for other water utilities in densely populated areas. Discharge from water treatment plants can heavily influence the concentration of chemicals in groundwater. Wastewater data can therefore help interpret measurements in groundwater and rivers.

Different data sources and formats are not the only challenges to overcome. Since researchers use different data for different applications, they often have their own, not standardized storage systems, ranging from MS Access databases to MS Excel sheets. Details on the data preparation needed for our case study can be found in the supplemental information.

Existing Workflows

Our collaborators and future users are hydrogeologists, not data engineers. They are experienced with hydrogeological numerical models, using their modeling software FEFLOW,[7] and have a high understanding of hydrological data. Although they occasionally use Python for basic scripting, their data analysis is done in MS Excel or directly in the MS Access database.

Based on the discussions inspired by the early prototypes, we identified five workflows that can benefit from efficient visualizations.

W1 (Project Onboarding): The collaboration with the water utility in our case study has been going on for several years. However, there is natural

fluctuations in the research group (students graduate, new students join). The three researchers we interviewed for this study have been working with this specific water utility for less than two years. Currently, there is no other way to get familiar with the project and the data than reading old reports and manually searching the database and various files. New team members need to spend a significant amount of time getting familiar with the available data.

W2 (Numerical Modeling): To calibrate their numerical models, the researchers use data from the hydraulic heads saved as MS Excel files. The calibration can be improved by combining hydraulic with chemical parameters.[5] The selection of parameters is often based on researchers' experiences from prior projects and might not take into account all of the available data.

The models are calibrated with data from selected time periods. To increase the validity of the models for describing the groundwater system in its baseline state, it is important to select data from representative years for calibration, i.e., avoid events such as floods. The current process of comparing data from different years manually is very time-consuming and error-prone.

W3 (Parameter Comparison): Besides numerical models, the researchers are interested in comparing different parameters and understanding the relationships between them. To do so, currently, the following steps are carried out.

1) Researchers select a small number of parameters they are interested in based on prior knowledge.
2) All parameters of interest are copied into one MS Excel sheet. If they come from different sources, the format is manually adjusted.
3) Plots and calculations are done directly in the Excel sheet, which is then saved.

Since most of the work must be done manually, the researchers only compare parameters if they expect a correlation. New correlations are rarely discovered. Large amounts of data significantly slow down Excel sheets, so the researchers use new files every time they want to compare parameters. This leads to large numbers of files that need to be managed. In particular, researchers often want to compare measurements and whether or not the water utility is "in operation," i.e., supplying the city with water. Instead of directly including this information in all of their plots (e.g., by

color coding them), they create individual plots for each operation timeline.

W4 (Station Comparison): When the researchers want to compare one parameter across different locations visually, they copy data into a GIS, such as QGIS or ArcGIS. This must be done manually for each parameter of interest.

W5 (Customer Presentation): Besides their research, the scientists also use data visualizations to communicate with their customers, i.e., water utility practitioners. As described in the previous workflows, creating visualizations requires many manual steps and cannot easily be done interactively. All visualizations to be shared with customers are therefore prepared before meetings.

INTERFACE DESIGN
Both requirements and design decisions presented in this section are the result of the prototyping process.

Design Requirements
While normally, requirements are collected before prototypes (as we also tried to do), it became clear during the feedback phase of the early prototypes that the initial requirements did not fully match the actual needs. Based on feedback for the early prototypes and the workflows previously discussed, three main design requirements were established.

R1 (Data Availability): Users need to visualize data availability by parameter, sampling station, and date. It is easy to look at the availability of several parameters and sampling stations at once.

R2 (Data Analysis): The interface enables users to understand their data across time and space by providing functionalities for the following tasks.

(a) A visual overview of the locations of sampling stations.
(b) Compare parameter values at all sampling stations on specific dates (sampling campaigns).
(c) Compare the temporal development of different (any desired number of) parameters at different (any desired number of) sampling stations.
(d) Compare the temporal development of any desired number of parameters at any desired number of sampling stations across years.
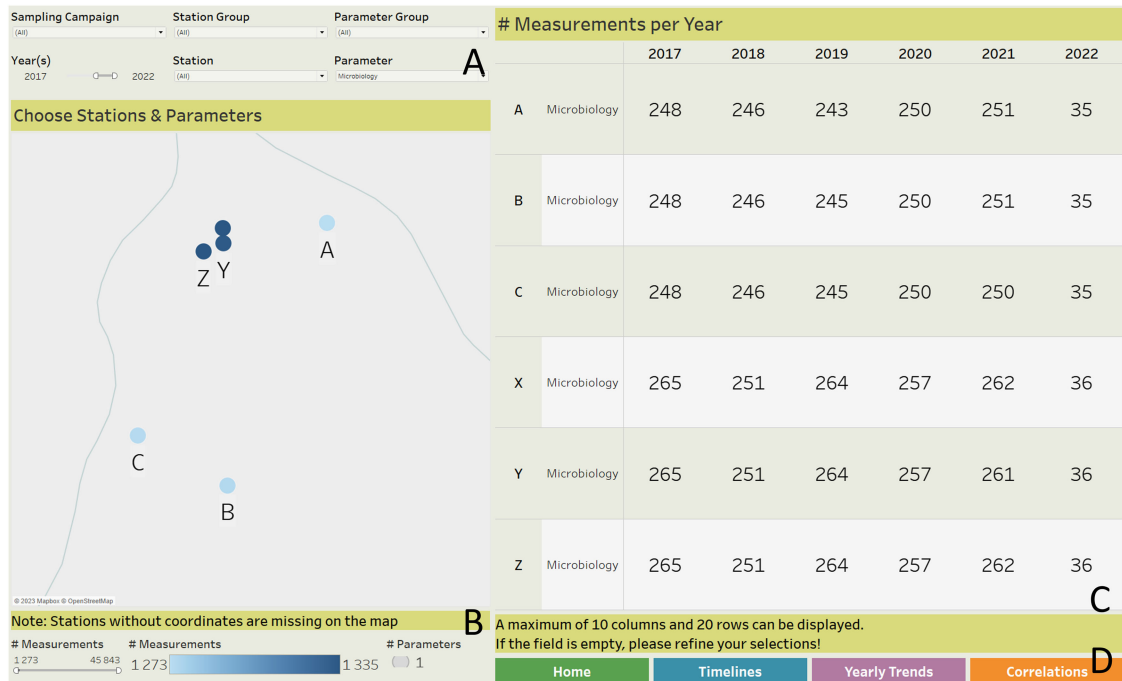(e) Relate parameter measurements to the water utility's operation.

FIGURE 3. The *overview* dashboard consists of four sections. (A) Filters: Users can filter by sampling campaign and then time, station group and then station, parameter group, and then parameter. (B) Map: All stations in the selection are displayed on the map, with color indicating the number of measurements in total and size indicating the number of different parameters. (C) Data availability: The number of measurements is displayed per station, parameter, and year. (D) Navigation: Color-coded buttons navigate to other dashboards.

(f) See if parameter values are below the parameter's detection limit and distinguish between missing values and measured values below zero.

(g) Visually identify trends and correlations between different parameters at different sampling stations.

(h) Group sampling stations and parameters according to configuration files provided by the users. Sampling campaigns can group dates according to configuration files provided by the users.

R3 (Usability): Since the users' skills lie in hydrogeology and not data engineering, designing an interface that is easy to use without much experience in programming is key. The interface lets users easily switch between different visualizations and views. Data can easily be updated, and plots can easily be saved.

Design Decisions

All design decisions presented refer to the final prototype that, while it was the final product of our development process, is far from a polished dashboard and still has room for improvement. Users are able to dynamically choose any number of sampling stations and parameters they are interested in. In contrast to the common practice of combining everything into one single view, we made the conscious choice to allow users to switch between four separate dashboards as follows.

› *Overview* (see Figure 3).
› *Timelines* (see Figure 4).
› *Yearly Trends* (see Figure 5).
› *Correlations* (see Figure 6).

The different dashboards serve different purposes, and for many problems, researchers only access one of them, so we prioritized increasing space for individual visualizations. An overview of how the individual dashboards satisfy the design requirements can be found in Table 1.

All four dashboards have a similar structure and share the following components.

Maps: The *overview*, *timelines*, and *yearly trends* dashboards include OpenStreetMap satellite image

TABLE 1. Four different dashboards each satisfy some of the requirements linked to data availability (R1) and analysis (R2). All of them are easy to use (R3).

		Overview	Timelines	Yearly Trends	Correlations
R1: Availability		✓			
	(a)	✓	✓	✓	
	(b)		✓	✓	
	(c)		✓	✓	
	(d)			✓	
R2: Analysis	(e)		✓		
	(f)		✓	✓	
	(g)				✓
	(h)	✓	✓	✓	✓
R3: Usability		✓	✓	✓	✓

maps (hidden in Figures 3–5 for data privacy reasons) displaying the sampling stations. Different statistical characteristics are useful to be displayed on the map for different parameter types, so it proved difficult to choose one. We agreed on:

› the number of different parameters and the total number of measurements in the *overview* map;
› the average values and the number of measurements for a selected parameter in the *timelines* map; and
› the maximum yearly means and their variance for a selected parameter in the *yearly trends* map.

We decided against an option for users to select a statistical characteristic to avoid further "cluttering" the interface. In all maps, it is possible for users to choose their preferred color scheme.

Filtering: Users can filter their data by time (specific sampling campaigns or years), space (station group and individual sampling station), and parameter (group and individual parameter).

Filters are synchronized among each other, i.e., selecting a sampling campaign, station group, and parameter group reduces the number of stations and parameters available. Filters are also synchronized between the *overview*, *timelines*, and *yearly trends* dashboards. To keep the filter options simple, time can only be filtered by years and individual sampling campaigns.

Data Graphs: Each dashboard displays a different aspect of the data as follows.

› The *overview* dashboard (see Figure 3) presents the number of measurements per sampling station, parameter, and year. The visual representations of these numbers used in prototypes P1a and P1b were removed following user feedback and recent findings that tables are a widely accepted form of data display.[23] Combined with their knowledge about a parameter's sampling frequency, the absolute number of measurements by year, presented in plain text, is the most relevant information.
› The *timelines* dashboard (see Figure 4) displays time series for the entirety of the selected time span, with detection limits marked by a solid orange line where available. To increase readability with respect to missing values, we opted for a dot instead of a line graph. Upon users' request, two different views are provided as follows.
  1) The standard view shows every selected station/parameter combination in a separate plot. Colors indicate whether or not the water utility is in operation.
  2) The overlay view plots data for all selected stations into one graph per parameter. Colors indicate stations, and operation info is omitted.
› The *yearly trends* dashboard (see Figure 5) shows the time series for all selected years in the same January–December line graph, with colors indicating the year. Yearly trends are displayed in separate plots for each station-parameter combination. Years are indicated by color; operation information is omitted. Users can choose between dots and lines, as dots are better to distinguish the actual values, and lines are preferred to see trends.
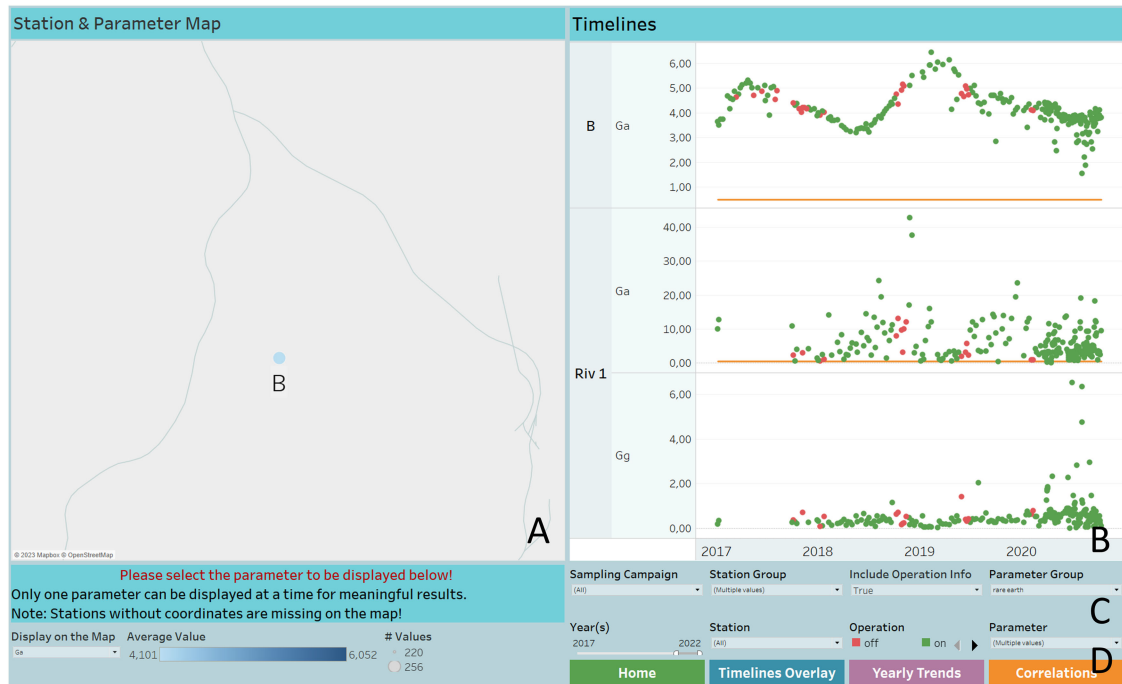› The *correlations* dashboard displays scatterplots of all selected combinations in a scatterplot

FIGURE 4. The *timelines* dashboard consists of four sections. (A) Map: Users select one single parameter to be displayed. All stations in the selection are displayed on the map, with color indicating the average value and size indicating the number of measured values. (B) Timelines: The standard view shows every selected station/parameter combination in a separate plot. Colors indicate whether or not the water utility is in operation. The parameter's detection limit is displayed in orange. (C) Filters: Users can filter by sampling campaign and then time, station group and then station, parameter group, and then parameter and choose to display operation information. (D) Navigation: Color-coded buttons navigate to other dashboards.

matrix with linear trend lines (R2g). There is an option to exclude values below the detection limit from scatterplots, as they distort correlations.

Our users argued that units of measurement were common knowledge that did not need to be specifically stated, so we did not include units in the plots.

CASE STUDY

This case study describes the use of the interface and insights from the development process by a group of researchers collaborating with a water utility. In the context of a long-standing cooperation between the university and a water utility, the individual researchers on the team have been working on the current project for a year. Their main goal is to understand how water quality relates to both operation of the water utility and environmental factors, such as water levels.

The main focus of the research group is numerical groundwater modeling. Two chemicals *Ga* and *Gg*

were chosen as chemical parameters for the model calibration based on their chemical properties and data availability. Other chemical parameters were not yet further investigated in this context. Besides numerical models to describe groundwater movements, the research group is interested in the chemical properties of the water measured in different locations over time. Their main issue is the amount of heterogeneous data from different sources. With some of the data saved in MS Excel sheets (hydrological time series) and some in an MS Access database (chemical data), they struggle with having an overview of data availability and quality.

Working on and using the visualization interface revealed major issues with data quality. Using the *overview* dashboard (see Figure 3), the group found discrepancies between the number of measurements they expected and the actual number for several parameters, leading to a significant amount of time spent cleaning and updating the database.

When comparing different stations and parameters (workflows W3 and W4) in the *timelines*
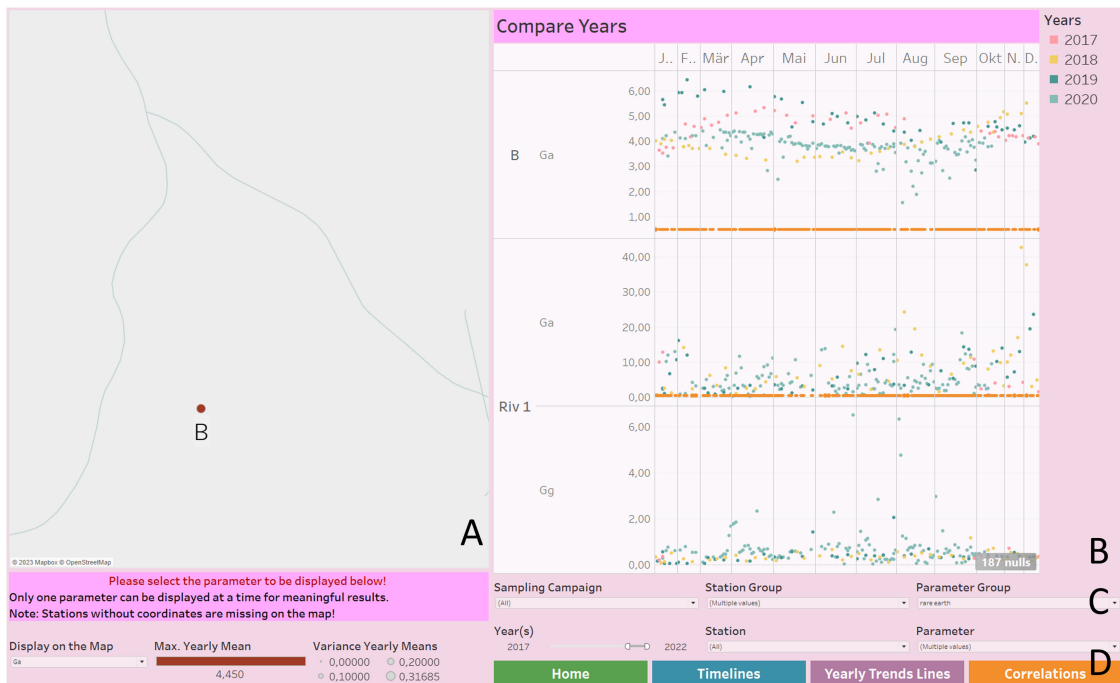
FIGURE 5. The *yearly trends* dashboard consists of four sections. (A) Map: Users select one single parameter to be displayed. All stations in the selection are displayed on the map, with color indicating the maximum yearly mean and size indicating the variance of yearly means. (B) Timelines: For each selected station/parameter combination, the temporal development in all the selected years is displayed in one chart, color indicating the year. The parameter's detection limit is displayed in orange. (C) Filters: Users can filter by sampling campaign and then time, station group and then station, parameter group, and then parameter. (D) Navigation: Color-coded buttons navigate to other dashboards.

dashboard (see Figure 4), the group found inconsistencies in both parameter and sampling station names which led to mismatches and wrong interpretation of time series. In particular, measurements for the parameters HCO_3 and ANC, which describe the same hydrochemical property, were saved as independent time series, covering different time periods. These inconsistencies likely happened due to different data sources and the extent of manual work involved in collecting and preparing data.

The interface was mainly used to support the group in their ongoing research. One researcher used the interface, in particular, the *timelines* (see Figure 4) and *yearly trends* (see Figure 5) dashboards, to select a representative time period of chemical *Ga* measurements to use for their model. In particular, they selected sampling campaigns that had measurements of the chemical in a large number of monitoring wells while the groundwater levels were stable. By filtering, they restricted the data displayed to chemical *Ga*, and selected sampling stations relevant to their model and the years they were interested in.

Two different researchers explicitly mentioned temperature amplitudes as one aspect they gained a better understanding of with the interface's help. One of them compared the amplitudes of temperature changes in different groundwater wells in the *timelines* dashboard in order to understand better the infiltration of river water into the groundwater (workflow W4). Without the interface, they had struggled with comparing temperature curves from many different wells at once. The other found a potential relationship between the amplitude of temperature changes and the hardness of the water in their study area. This question will be pursued in a future project (workflow W3).

Visualizing chemical measurements on a map (workflow W4) also helped the researchers to define different "types of water" in their area of interest. In the past, this was determined based on the electric conductivity values, i.e., the hardness of the water. Using the interface to visualize a variety of parameters quickly allowed the researchers to define "types of water" based on the measured concentrations of different chemical
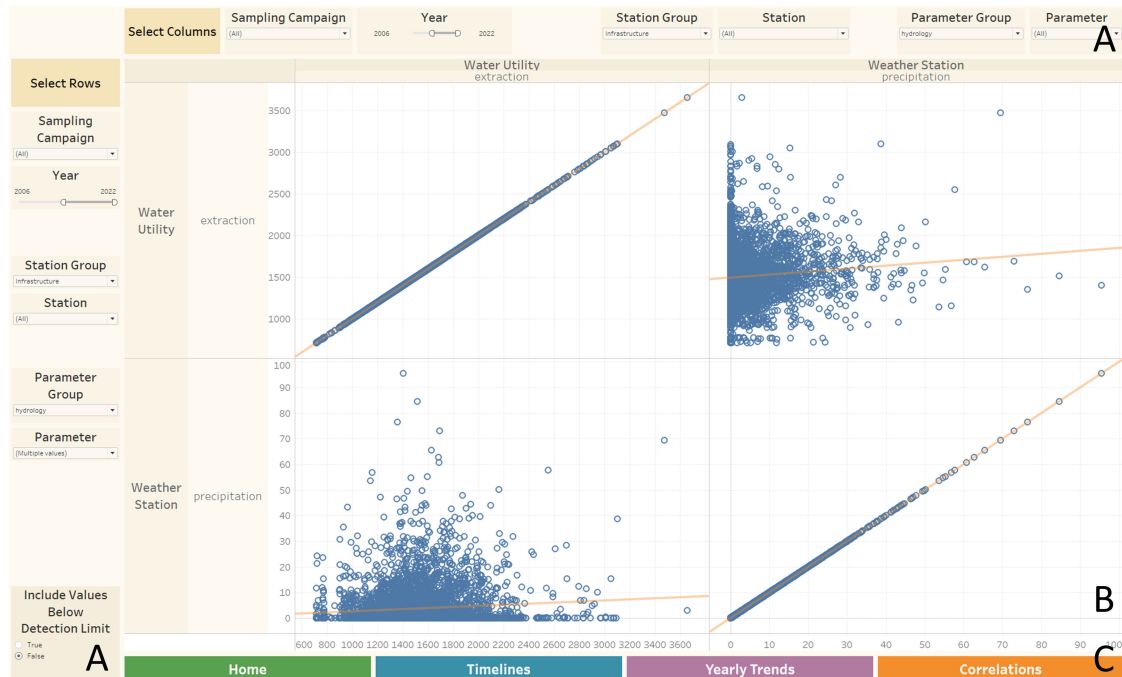
FIGURE 6. The *correlation* dashboard consists of four sections. (A) Filters: Users filter columns and rows independently by sampling campaign and then years, station group and then station, parameter group and then parameter. Users choose whether or not to include values below the detection limit in the scatterplots. (B) Scatterplot matrix: Scatterplots of all selected combinations are displayed in a scatterplot matrix with linear trend lines. (C) Navigation: Color-coded buttons navigate to other dashboards.

compounds. In general, the interface, in particularly the *timelines* (see Figure 4) and the *correlations* (see Figure 6) dashboard, was used to generate hypotheses about potential indicator parameters for certain events, such as floods and resulting changes in water types.

The group sees potential applications of such an interface not only in their research but also in onboarding new team members (workflow W1) and in customer presentations (workflow W5).

LESSONS LEARNED

In this section, we discuss the insights we gained during the design and development process. We focus on the insights concerning the process and only briefly mention insights about the interface itself.

In the following paragraphs, we link some of our insights about the design process to the stages of a design study and the pitfalls (PF) identified by Sedl-mair et al.[3] and expand them. While we did not find any freely available software solving exactly the problems our domain scientists were facing, a lot of visualization tools already exist in hydrogeology. A deeper literature review at the beginning of our study would

have been helpful in shaping the scope of our study (PF-2).

Our experiences emphasized the importance of the winnowing stage of a design study (PF-3) as, in hindsight, a design study was not the best approach to the problems at hand for several reasons.

Whereas the group leader was enthusiastic about the idea of a design study, the other researchers were unsure about the procedure and possible outcomes. They, thus, lacked the time and enthusiasm to engage in the project as much as needed for a successful outcome (PF-5). In particular, at the beginning of the study, they gave feedback when asked and used the interface for a few specific tasks but showed limited interest in exploring the interface. When encountering bugs during prototype testing, they stopped their work and waited for the next scheduled interviews to report them, delaying the feedback. While it would be easy to say that in hindsight, this lack of enthusiasm for the design study should have been the end of the cooperation, it could potentially have been overcome with appropriate communication. We believe that it was at least partly rooted in their limited knowledge about

the power of visualization beyond the tools and graphs they usually work with. Unless specializing in that direction, hydrogeologists, in particular numerical modelers, are not used to thinking about visualizations. They simply apply them as a tool. After the first round of individual requirements interviews proved inconclusive, taking a stronger lead and, for example, conducting a structured workshop with a detailed introduction to the potential of a design study and a collective brainstorming session could have been useful. We propose the addition of a new pitfall (PF-33) linked to misjudging the working environment and not flexibly adapting the study to that environment.

We realized too late that some of the challenges in motivation and engagement were also rooted in misunderstandings due to individual and cultural differences (PF-11) in communication. The group is culturally diverse, which added to the communication issues common in an interdisciplinary setting.[24]

Ultimately most of their problems did not require visualization research but data and visualization engineering (PF-8). One of our main insights is that the process of interviews, developing prototypes, and finally seeing and interacting with their data changed their minds and attitudes. With every iteration, they became more engaged, gaining an understanding of the potential of a drag-and-drop tool and the ability to see a lot of their data side-by-side, visually. This highlights how part of their need was help with data wrangling and analysis, not an elaborate visualization interface.

The final prototype was not used as much as anticipated. This was partly because many of the most relevant questions had been answered during the development process already, but mainly because the research group we were working with disintegrated only months after the project was completed, leaving us without real users (PF-25). While this (researchers leaving for industry, collaborations with industry partners ending, and shifts in research interests) can always happen in an academic setting, we recommend raising this issue before the start of a design study project. Including the water utility as stakeholders with a long-term interest could have been beneficial to the study, so we propose the addition of another pitfall (PF-34) linked to focusing on a limited audience early in the process and thus missing the actual needs of the wider community. Including requirements from a larger group of future users requires more effort on the designer's side but will ultimately increase the chances of the tool being used in the entire community.

Some other difficulties we encountered during the core phase of our design study (discover – design – implement – deploy) were not linked to the group of researchers we worked with but to the way we set up our study.

During the requirements analysis, we relied strongly on the results of the first round of individual interviews, thus missing a lot of information that only became apparent during the design phase (PF-16). Two major ideas worth a design study, namely the interpolation and continuous visualization of parameter values across the entire area of interest, and graphs incorporating the delay of concentrations between different sampling points, came up only when the final prototype was almost finished and it was already becoming clear that we would not further pursue the study. By, instead, conducting a larger number of consecutive individual interviews with the same people over the course of several weeks or months, we expect discussions at different stages of their work, helping understand different aspects. We also recommend increasing the frequency of interviews after the first round of prototyping, when researchers can already actively work with the prototypes. Starting out with functional prototypes that could already work with "real" data to increase engagement in discussions and refine requirements proved very useful for us, questioning the usual separation of low- and high-fidelity prototypes.

In addition, most of the conversations at the beginning of the design phase were already centered around potential visualizations instead of focusing on the problems at hand (PF-17). One reason for this could be that the process was not seen as a collaborative development but as delegating implementation work.

After presenting the first three different prototypes, we quickly committed to the final prototype. In hindsight, better results could have been obtained with a more diverse set of prototypes (PF-20). Even though Tableau ultimately was an appropriate framework for our quick-prototyping process, some more (and maybe better) ideas could have been generated by exploring other software options (PF-21).

Looking back, many of these difficulties could have been avoided had we concentrated on the design study process and adapted it as needed. Besides the workshop mentioned previously, we believe that better results could also have been obtained with a combination of more frequent interviews, a more rigorous feedback timeline, more prototype options, and the exploration of other visualization techniques.

The most interesting insight into visualization in hydrogeology was the fact that there seem to be few design standards in the field. This corresponds to the observation that researchers contradicted each other and also themselves between interviews regarding the prominence of maps and filtering options and also their color and plot preferences. In consequence, we allowed for customizations in plots, line style, and colors.

For example, two researchers preferred diverging color schemes to visualize measured values on the map. The values are not in a diverging range, and there is no universal ranking of values that would justify a "green = good" to "red = bad" color scheme. However, affording this choice lowered the threshold for using the tool by these skeptical users.

CONCLUSION

While this study started out as a traditional design study, in many ways it was not.

› *We had to spend time convincing our future users:* It was apparent early on in the project that, whereas the group leader was a clear champion for our approach, the other potential users did not share his vision. They seemed to be expecting a data engineer to help them with coding (as that skill was not strong among these researchers). They had limited expectations, if any, of the power of visual analysis and visualization research.

› *No clear separation between requirements analysis and prototyping:* The diverging expectations made a proper requirements analysis very difficult, resulting in a mixed phase of prototyping and requirements analysis.

› *Functional early prototypes:* Only through the development of early prototypes that were functional in the sense of being able to interact with the actual data in question we were able to understand the needs of the researchers better and at the same time communicate the potential a visual analysis platform could offer.

› *No original interface based on the final prototype:* Due to the fact that the main insights were already generated through the development process itself, the group of domain scientists disintegrated shortly after the prototype was finished, and our finding that Tableau was sufficient for most of the domain scientists' problems, we never developed an interface that was not based on Tableau.

Still, outside the framework of a design study, our study was very informative and raised several interesting questions to be answered in future research.

The iterative process would have been much harder without a rapid-prototyping tool, such as Tableau. At every step, even during our initial prototyping phase, we were able to keep the researchers engaged by showing them their own data. While we have successfully used low- and high-fidelity prototyping for many other projects, in this particular case, we benefited greatly from using "real" data even during the initial prototyping phase. This experience was one of this study's most valuable insights, and we see potential in analyzing the drawbacks of the conventional design study methodology and possible alternative approaches.

Our study also demonstrated that many problems can be solved using off-the-shelf, domain-agnostic software, such as Tableau, making the development of custom tools unnecessary. While some specific problems will always require custom solutions, the potential and drawbacks of off-the-shelf solutions and what is necessary for them to be used by the hydrogeological community should be further investigated. Two possible drawbacks to be addressed are scalability and performance, as the community often deals with large datasets. We also see an interesting challenge in assessing visual analysis literacy in the community and exploring how it can be improved.

The different needs of hydrogeological researchers that we observed in the small group we worked with posed a challenge to the development of a single interface that accommodates just a few researchers of the same research group. A future study could explore how these findings can be generalized to the wider field of hydrogeology, focusing on the different needs and the underlying common ground.

While the final prototype itself was welcomed by the domain researchers we worked with, it was mainly the process of discussing requirements, preparing data, and testing prototypes that helped them better understand their data. This highlights that early and thorough cooperation of data scientists with domain experts as well as improved training of domain experts in data wrangling and analysis would advance finding solutions more than the introduction of new tools. Instead of developing an interface, as we tried to do in this study, we suggest a study on what hydrogeologists need to work efficiently with their data. The result of such a study could be educational material to improve their (visual) data literacy, a collection of workflows and guidelines, and a concrete tool for data analysis.

REFERENCES

1. J. Margat and J. Van der Gun, *Groundwater Around the World: A. Geographic Synopsis*. Boca Raton, FL, USA: CTC Press, 2013.

2. D. Machiwal, V. Cloutier, C. Güler, and N. Kazakis, "A review of GIS-integrated statistical techniques for groundwater quality evaluation and protection," *Environ. Earth Sci.*, vol. 77, no. 19, 2018, Art. no. 681.

3. M. Sedlmair, M. Meyer, and T. Munzner, "Design study methodology: Reflections from the trenches and the stacks," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 12, pp. 2431–2440, Dec. 2012.

4. K. Hiscock and V. Bense, *Hydrogeology: Principles and Practice*. Oxford, U.K.: Blackwell Sci. Ltd, 2005.

5. M. P. Anderson, W. W. Woessner, and R. J. Hunt, *Applied Groundwater Modelling: Simulation of Flow and Advective Transport*, 2nd ed. New York, NY, USA: Academic Press/Elsevier, 2015.

6. A. Singh, "Groundwater resources management through the applications of simulation modeling: A review," *Sci. Total Environ.*, vol. 499, pp. 414–423, 2014, doi: 10.1016/j.scitotenv.2014.05.048.

7. C. P. Kumar, "An overview of commonly used groundwater modelling software," *Int. J. Adv. Res. Sci., Eng. Technol.*, vol. 6, no. 1, pp. 7854–7865, 2019.

8. R. Criollo et al., "AkvaGIS: An open source tool for water quantity and quality management," *Comput. Geosci.*, vol. 127, pp. 123–132, 2019, doi: 10.1016/j. cageo.2018.10.012.

9. M. Alcaraz, "A loosely coupled GIS and hydrogeological modeling framework," *Environ. Earth Sci.*, vol. 76, pp. 1–13, 2017.

10. L. Foglia et al., "FREEWAT, A free and open source, GIS-integrated, hydrological modeling platform," *Groundwater*, vol. 56, no. 4, pp. 521–523, 2018.

11. A. A. Konaté, S. F. Konaté, M. Zaheer, A. Geophrey, and F. Oularé, "Geostatistical analysis of hydrogeological data of the Baraka gold zone, prefecture of Siguiri, Republic of Guinea," *Environ. Earth Sci.*, vol. 81, 2022, Art. no. 232, doi: 10.1007/ s12665-022-10340-w.

12. A. Ashraf and Z. Ahmad, "Integration of groundwater flow modeling and GIS," in *Water Resources Management and Modeling*. Rijeka, Croatia: InTech, 2012, pp. 239–262.

13. R. Barthel, E. Haaf, M. Nygren, and M. Giese, "Systematic visual analysis of groundwater hydrographs: Potential benefits and challenges," *Hydrogeol. J.*, vol. 30, no. 2, pp. 359–378, 2022, doi: 10.1007/s10040-021-02433-w.

14. W. Zaadnoordijk, S. Bus, A. Lourens, and W. Berendrecht, "Automated time series modeling for piezometers in the national database of the Netherlands," *Groundwater*, vol. 57, no. 6, pp. 834–843, 2019.

15. W. Aigner, S. Miksch, H. Schumann, and C. Tominski, *Visualization of Time-Oriented Data*, 2nd ed. Berlin, Germany: Springer, 2023. [Online]. Available: https:// timeviz.net

16. N. Andrienko, G. Andrienko, and P. Gatalsky, "Exploratory spatio-temporal visualization: An analytical review," *J. Vis. Lang. Comput.*, vol. 14, no. 6, pp. 503–541, 2003.

17. S. T. Benedict et al., "Data visualization, time-series analysis, and mass-balance modeling of hydrologic and water-quality data for the McTier Creek watershed," U.S. Geological Survey, Baltimore, MD, USA, U.S. Geological Survey, Open-File Report 2011–1209, 2012.

18. M. G. Castrellón, R. A. Pauloo, I. Popescu, and J. Fábrega, "ONASP: A web application for groundwater data visualization in Panama," in *Proc. IOP Conf. Ser.: Earth Environ. Sci.*, 2023, vol. 1136, Art. no. 012028.

19. J. A. Guzman, D. N. Moriasi, M. L. Chu, P. J. Starks, J. L. Steiner, and P. H. Gowda, "A tool for mapping and spatio-temporal analysis of hydrological data," *Environ. Model. Softw.*, vol. 48, pp. 163–170, 2013, doi: 10.1016/j.envsoft.2013.06.014.

20. A. Mazher, "Visualization framework for high-dimensional spatio-temporal hydrological gridded datasets using machine-learning techniques," *Water*, vol. 12, no. 2, 2020, Art. no. 590.

21. K. Nešetřil and J. Šembera, "Business intelligence and geographic information system for hydrogeology," in *Proc. 12th Int. Symp. Environ. Softw. Syst.*, 2017, pp. 163–170.

22. European Commission, "Directive (EU) 2020/2184 of the European parliament and of the council of 16 December 2020 on the quality of water intended for human consumption," *Official J. Eur. Union*, vol. L 435/ 1-61, 2020.

23. L. Bartram, M. Correll, and M. Tory, "Untidy data: The unreasonable effectiveness of tables," *IEEE Trans. Vis. Comput. Graphics*, vol. 28, no. 1, pp. 686–696, Jan. 2022.

24. R. Freeth and G. Caniglia, "Learning to collaborate while collaborating: Advancing interdisciplinary sustainability research," *Sustain. Sci.*, vol. 15, no. 1, pp. 247–261, 2020.

LAURA LOTTERANER is a Ph.D. student at the University of Vienna, 1090, Vienna, Austria. Her research interests include applications of data visualization and data analysis in the environmental sciences. Lotteraner received her M.Sc. degree in technical mathematics from Vienna University of Technology, Vienna, Austria. She is the corresponding author of this article. Contact her at laura.lotteraner@univie.ac.at.

THILO HOFMANN leads the Environmental Geosciences team at the University of Vienna, 1090, Vienna, Austria, focusing on emerging pollutant behavior in water, such as microplastics, nanoparticles, PFAS, and anthropogenic tracers. Hofmann received his Ph.D. degree in aquatic geochemistry from Bremen University, Bremen, Germany. Contact him at thilo.hofmann@univie.ac.at.

TORSTEN MÖLLER is the head of the Visualization and Data Analysis Research Group at the Faculty of Computer Science, University of Vienna, 1090, Vienna, Austria. His research interests include visualization, computer graphics, and data science. Möller received his Ph.D. degree in computer and information science from the Ohio State University, Columbus, OH, USA. He is a senior member of the IEEE Computer Society. Contact him at torsten.moeller@univie.ac.at.