

A database application framework toward data-driven vertical connectivity analysis of rivers

Beatriz Negreiros^{a,*}, Sebastian Schwindt^a, Federica Scolari^a, Ricardo Barros^b, Alcides Aybar Galdos^c, Markus Noack^c, Stefan Haun^a, Silke Wieprecht^a

^a Institute for Modelling Hydraulic and Environmental Systems, University of Stuttgart, Stuttgart, 70569, Baden-Württemberg, Germany

^b Institute for Water and River Basin Management, Karlsruhe Institute of Technology, Karlsruhe, D-76131, Baden-Württemberg, Germany

^c Institute of Applied Research, Karlsruhe University of Applied Sciences, Karlsruhe, 76133, Baden-Württemberg, Germany

ARTICLE INFO

Dataset link: <https://osf.io/kt53d/>, <https://riveranalyst.github.io/app>

Keywords:

Restoration science
Riverbed clogging
Data management
Python programming
Web app
Hydroinformatics

ABSTRACT

The description of complex river environments requires interdisciplinary approaches to collect and manage manifold data types and sources. Deriving comprehensive knowledge from complex data sources is challenging and necessitates not only knowledge of environmental science but also statistics and Software engineering. This study introduces a relational database framed in an application called River Analyst for creating and managing river data with open-source standards (Python3 and Django). We conceptualize data models of river environments, which describe sediment characteristics and hydraulics related to hyporheic exchange. River Analyst enabled us to derive novel insights for restoring rivers affected by so-called riverbed clogging, notably, fine sediment infiltration in the hyporheic zone. The database analysis reveals that clogging is not a dominant control process when the fraction of fine sediment exceeds 50%–55%. In conclusion, the new Software holds promise for data-informed advancements in augmenting knowledge to restore ecologically functional hydro-environments.

Software and data availability

• Program size: 26.8 MB

- Name of the Software: River Analyst
- Developers: Beatriz Negreiros, Federica Scolari, Ricardo Barros, and Sebastian Schwindt
- Contact information: river.analyst.software@gmail.com
- Hardware required: personal computer or server
- Year first available: 2023
- Program Language: Python 3 (35.7%), Javascript (35.4%), CSS (21.4%), HTML (7.5%)
- Cost: free
- Software availability: <https://riveranalyst.github.io/app> | (Negreiros et al., 2023b)

1. Introduction

The analysis of data from natural water resources is experiencing major improvements with new opportunities arising from information management techniques (Wojda et al., 2010; Yi et al., 2018). Additionally, advances in techniques for measuring ground truth and remote sensing are generating vast amounts of data (Hernandez et al., 2012). Computer models can use these data to simulate hydro-environmental processes at almost any spatiotemporal scale, resulting in even greater amounts of data. Such floods of information underscore a new era in

* Corresponding author.

Linkedin: [beatriz-negreiros](#) (B. Negreiros), [sebastian-schwindt](#) (S. Schwindt), [federica-scolari-81b6a214b](#) (F. Scolari), [ricardo-barros-918750163](#) (R. Barros), [markus-noack-649a641bb/](#) (M. Noack), [stefan-haun-671b3283/](#) (S. Haun).

E-mail addresses: beatriz.negreiros@iws.uni-stuttgart.de (B. Negreiros), sebastian.schwindt@iws.uni-stuttgart.de (S. Schwindt), federica.scolari@iws.uni-stuttgart.de (F. Scolari), ricardovobarros@live.com (R. Barros), alcides.aybar_galdos@h-ka.de (A.A. Galdos), markus.noack@h-ka.de (M. Noack), stefan.haun@iws.uni-stuttgart.de (S. Haun), wieprecht@iws.uni-stuttgart.de (S. Wieprecht).

URLs: <https://beatriznegreiros.com/> (B. Negreiros), <https://sebastian-schwindt.org/> (S. Schwindt), <https://www.iws.uni-stuttgart.de/en/institute/team/Scolari/> (F. Scolari), <https://ricardovobarros.github.io/> (R. Barros), <https://www.h-ka.de/die-hochschule-karlsruhe/organisation-personen/personen-a-z/person/alcides-aybar-galdos> (A.A. Galdos), <https://www.h-ka.de/die-hochschule-karlsruhe/organisation-personen/personen-a-z/person/markus-noack> (M. Noack), <https://www.iws.uni-stuttgart.de/institut/team/Haun-00001/> (S. Haun), <https://www.iws.uni-stuttgart.de/en/institute/team/Wieprecht-00003/> (S. Wieprecht).

<https://doi.org/10.1016/j.envsoft.2023.105916>

Received 30 July 2023; Received in revised form 27 November 2023; Accepted 1 December 2023

Available online 5 December 2023

1364-8152/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

which technological development and increased familiarity with hydroinformatics are crucial for analyzing and interpreting data (Abbott, 1991; Benson et al., 2010; Vitolo et al., 2015; Yi et al., 2018). Still, data management and sharing are often complicated because of lacking common designs, standards, and methods (Wojda et al., 2010; Plana et al., 2019). More efforts are thus necessary to improve data management in hydro-environmental research and engineering, both regionally and globally (Rieger et al., 2004; Carrera-Hernández and Gaskin, 2008; Horsburgh et al., 2016). For example, new tools are required to manage and analyze rapidly growing river data originating from extensive field surveys conducted by various entities. In addition, to improve scientific data exchange, dataset management should be low-maintenance, scalable (to include additional datasets), and transparent (Hilty et al., 2006; Carrera-Hernández and Gaskin, 2008).

Relational databases enable systematic analyses of large and complex environmental datasets, providing valuable overviews of system parameters (Pokorný, 2006; Mooij et al., 2014). Pioneering research on water resources has already shown the relevance of relational databases for understanding spatial patterns of modeled and observed data (Swain et al., 2015). For instance, water resources web applications were developed to handle interactions with databases, such as to modify, store, visualize, and analyze data (Swain et al., 2015). However, previous studies have primarily developed applications to handle hydrological data (e.g., Goodrich et al., 2011; Brooking and Hunter, 2013; Delipetrev et al., 2014), where little attention has been paid to the state of an environment defined by complex hydraulic, sedimentological, and biotic parameters at multiple spatiotemporal scales. To this end, a relational database approach is introduced here to overcome challenges related to storing, organizing, and retrieving large amounts of hydro-environmental information. To provide an easily accessible interface, we have used the Django web application framework (WAF) in Python (Python Software Foundation, 2023) to generate a web application (hereafter referred to as “app”) that can be used online or locally to interact with any relational database. The development goals were to provide free and open-source software that reduces common hurdles of developing hydro-environmental database apps, such as finding suitable software packages and database architectures (Swain et al., 2016).

The creation of a database framework was also motivated by the challenge that the theory of using dimensionless parameters for inter-site comparison of measurement data (Yalin, 1971; Barenblatt, 1987, 1996) is often limited in complex hydro-environments. For example, to describe and restore vertical connectivity of river networks, multiple studies focused on isolated considerations of specific parameters of the so-called hyporheic zone (e.g., Huston and Fox, 2015; Dubuis and De Cesare, 2023). The hyporheic zone represents the subsurface region that exchanges water and matter with surface water (Orghidan, 1959). Fish require a functional hyporheic zone with sufficient pore space, both for spawning and foraging for macrozoobenthos that thrive in the shelter of coarse sediment (Boulton et al., 1998; Kondolf et al., 2006; Tonina and Buffington, 2009). The porous matrix of the hyporheic zone may become filled by fine sediments infiltrating into the riverbed in response to hydro-sedimentological regime changes, resulting in riverbed clogging, which can be exacerbated by anthropogenic interventions, such as the construction of dams and land use changes leading to high non-natural fine sediment inputs. Thus, clogging often causes a reduction in riverbed permeability (Jin et al., 2019) and interstitial dissolved oxygen (IDO) in the hyporheic zone (Greig et al., 2007). In a comprehensive review of (dimensionless) parameters to describe clogging of the hyporheic zone, Dubuis and De Cesare (2023) examined flume experiments and field data with multiple approaches to combine them through dimensional analysis. While Dubuis and De Cesare (2023) derive some meaning from a dimensionless clogging depth, they also find contradictory trends when using dimensionless numbers. For example, substrates with different sediment compositions result in conflicting magnitudes of the clogging depth

relative to the mean grain size diameter. The dimensionless clogging depth reaches several multiples of the mean grain size for substrates consisting of fine gravel, but it only reaches a few times the mean grain size diameter for substrates consisting of coarse gravel and cobble (Schälchli, 1993; Dubuis and De Cesare, 2023). Other studies used a multiparameter approach to quantify clogging (MultiPAC) and vertical hyporheic connectivity to derive characteristic grain sizes, porosity, depth-explicit IDO, and depth-explicit hydraulic conductivity (Seitz, 2020; Negreiros et al., 2023a). Still, to the author’s best knowledge, no study could derive coherent dimensionless parameters to describe clogging. In this study, we hypothesized that sedimentological riverbed clogging only occurs under environmental conditions dominated by particularly coarse sediment. This hypothesis was inspired by previous studies suggesting that sedimentary clogging can only occur in mountainous hydro-environments where gravel and cobble constitute the dominant mass of the riverbed (Einstein, 1968; Schälchli, 1992; Huston and Fox, 2015; Wharton et al., 2017). With a new database framework and its built-in routines to perform principal component analysis (PCA), we investigated this hypothesis based on extensive data from 19 field sites to overcome the incapacity of typical dimensional analysis (e.g., Barenblatt, 1987) to explain vertical connectivity disruptions in the form of clogging.

2. Database and web app

2.1. Relational databases

Relational databases are designed to store, organize, and retrieve large amounts of interlinked information, which is one of the major challenges in processing rapidly growing data (Holt et al., 2015). A relational database contains data in tabular form, with each row representing a record and each column representing an attribute. Tables, also referred to as data models, are linked to each other through three relational rule types, the so-called *foreign keys*: one-to-one, one-to-many, and many-to-many, indicating if one or many rows of a table is/are linked with one or many rows of another table. Each row in a data model has by default a *primary key* that is non-null and unique in the data table, while *foreign keys* establish relationships between two data models (Melton, 1996). A relational database is managed through a database management system (DBMS) that uses structured query language (SQL) by default (Severance, 2014). Specifically, SQL is a database programming language that runs commands to create, retrieve, update, and delete (CRUD) data from a database (Groff et al., 2002).

2.2. The Django web application framework (WAF)

A web application (app) builds on request–response cycles between a browser (user side, frontend) and a server (dynamic backend) where the app is hosted with a hypertext transfer protocol (HTTP). Users send requests to the server, which are processed by a controller and returned to the user in the form of a response (i.e., a rendered page). HTTP responses are presented to the user through hypertext markup language (HTML) pages. Django is a powerful framework in Python for database and web development based on object-relational mapping (ORM). A Python framework is a collection of modules, libraries, and tools that provides a predefined structure and a set of conventions for developing software applications. Thus, a web application framework (WAF) constitutes a standardized way to build, deploy, and develop web apps linked to databases. The Django WAF provides capabilities for web development, database connectivity, and user authentication to build an interface between a database and the user side (Dauzon et al., 2016). Similarly to other Python libraries (e.g., NumPy; Harris et al., 2020), Django is only wrapped in Python code and not natively written in it, keeping the backend computationally efficient. In addition, Django outperforms other WAFs (e.g., Ruby on Rails) regarding the ease of user

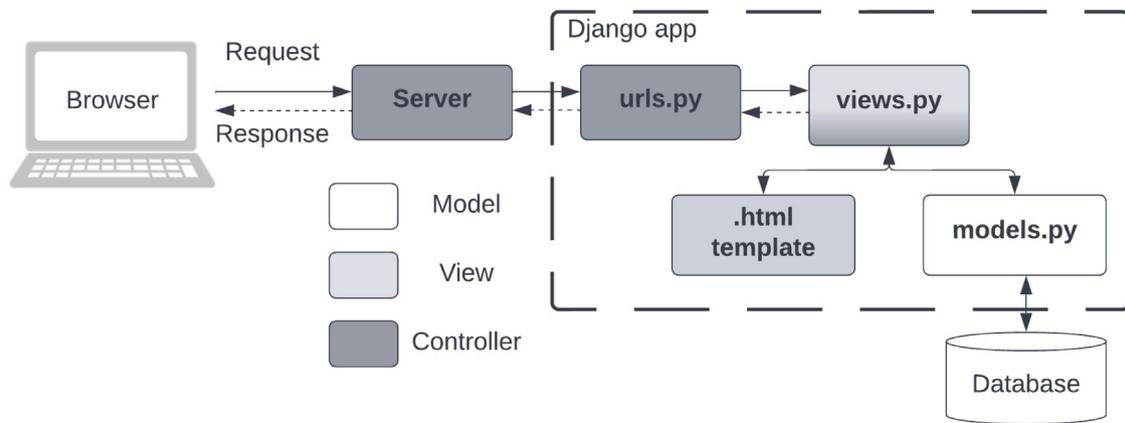


Fig. 1. Django application architecture (in line with Bioco and Rocha, 2019).

interface development, maintainability, community support, and even marketability (Plekhanova, 2009; Dauzon et al., 2016). The Django WAF uses a three-layered architecture called model-view-controller (MVC) to build web apps (Severance, 2009):

- (i) Model: an object that communicates with the database.
- (ii) View: a component that renders responses into visual elements constituting the look of the app through HTML and style files (e.g., cascading style sheets, CSS).
- (iii) Controller: code that controls how request routing and operations occur in the app.

The MVC architecture enables running a request–response cycle (Fig. 1) starting with the browser (user) sending an HTTP request to the server, which can be a *GET* or *POST* request. A *GET* request only looks up data through the app. A *POST* request makes changes to the database within the app. Any request contains a URL to display a page that is directed to the *urls.py* script of a Django app. The *urls.py* script calls a *view* function corresponding to the URL invoked by a *views.py* script. The *views.py* script contains both a *controller* agent for database access and a *view* agent configuring visual elements from HTML templates. *views.py* also provides routines to open a connection with the database through the *models.py* script, which is closed at the end of a request. Thus, the Django WAF has methods that wrap SQL code to communicate with database entities, which are termed Django models. Several DBMS can be used with Django, such as SQLite, MySQL, and Oracle, among others. We used SQLite, as it is suitable for small databases <200 TB, which is typical for data from rivers.

3. Software description & capacities

3.1. Database architecture

The design of a database depends on the nature of the data and organizational needs (Holt et al., 2015). Designing a database involves the use of data models to represent selected information of a system (Webb et al., 2015). To represent complex and datatype-variable hydro-environmental data, we used data models describing hydraulic, morphological, and biochemical characteristics. To enhance data consistency and redundancy, data normalization was performed by decomposing data into smaller, manageable tables and subsequently establishing links between them through inter-table relationships (i.e., foreign keys). Fig. 2 shows the relationships among the Django data models (i.e., database entities). The data model MeasPosition (abbreviation of “measurement position”) characterizes a snapshot of hydro-environmental data at an x–y location with a timestamp. Thus, every measurement position can contain data on surface and/or subsurface sediments, hydraulics, free-flowing water quality, ecology, and depth-explicit riverbed characteristics. Examples of depth-explicit riverbed

characteristics are hydraulic conductivity k_f and IDO (Seitz, 2020; Negreiros et al., 2023a). The complete list of attributes with descriptions and data types can be seen in the *models.py* script documented in the user manual (River Analyst, 2023).

3.2. Structure and usage

River Analyst uses Python3 and HTML templates styled with CSS and Javascript for backend and frontend development, respectively (Fig. 3). The software can be hosted either locally on a personal computer or publicly on the internet. Local hosting makes the app run on a local computer (port 8888 by default), which opens automatically in the default browser. Online hosting requires a proprietary server that allows users to query and interact with the database through a webpage. While version control platforms, such as GitHub and GitLab are good solutions for centralized code version management, they are unsuitable for large files or when a dynamic backend is required. Suitable dynamic backend solutions for hosting database files (*.sqlite3) are, for example, the Amazon web services relational database service (AWS RDS), Google cloud SQL, or a self-hosted server. We provide an example for connecting River Analyst with a database hosted in AWS RDS in the user manual, and more suitable hosting platforms for deploying web apps are listed in the Supplementary Material.

The River Analyst app contains four interactive tabs representing modules for querying, changing, managing, and analyzing data from the database, notably the *query*, *upload*, *admin*, and *analysis* modules, respectively. An additional *home* module provides users with generic database information and field methods. In addition, a suite of algorithms for deriving sediment characteristics of the riverbed is implemented and described below. The *query* module filters data models and exports the tables to comma-separated value (CSV) or other spreadsheet formats (e.g., .xlsx). The query also features georeferenced point visualization of filtered (queried) measurement positions using the scatter mapbox function of the Plotly Python library (Plotly Technologies Inc., 2022b). In addition, interactive plots of the available data for a selected measurement position can be displayed through the *query* tab, such as cumulative grain size distribution curves or depth profiles of k_f and IDO (Figure 1 in the Supplementary Material). The *upload* module provides an option to read CSV-data with template-specific formatting to append it to Django models, thus creating new rows in the database tables. The *admin* module is the built-in Django administration page for database maintenance (cf. Plekhanova, 2009), such as CRUD (create, retrieve, update, delete) operations corresponding to queries and uploads with administrator rights. A history page listing the changes made to an object shows the timestamp and username of the admin who ran the CRUD operation. Also, user rights, groups, and authentication are managed through the *admin* module. The *analysis* module contains a built-in principal component analysis (PCA) tool.

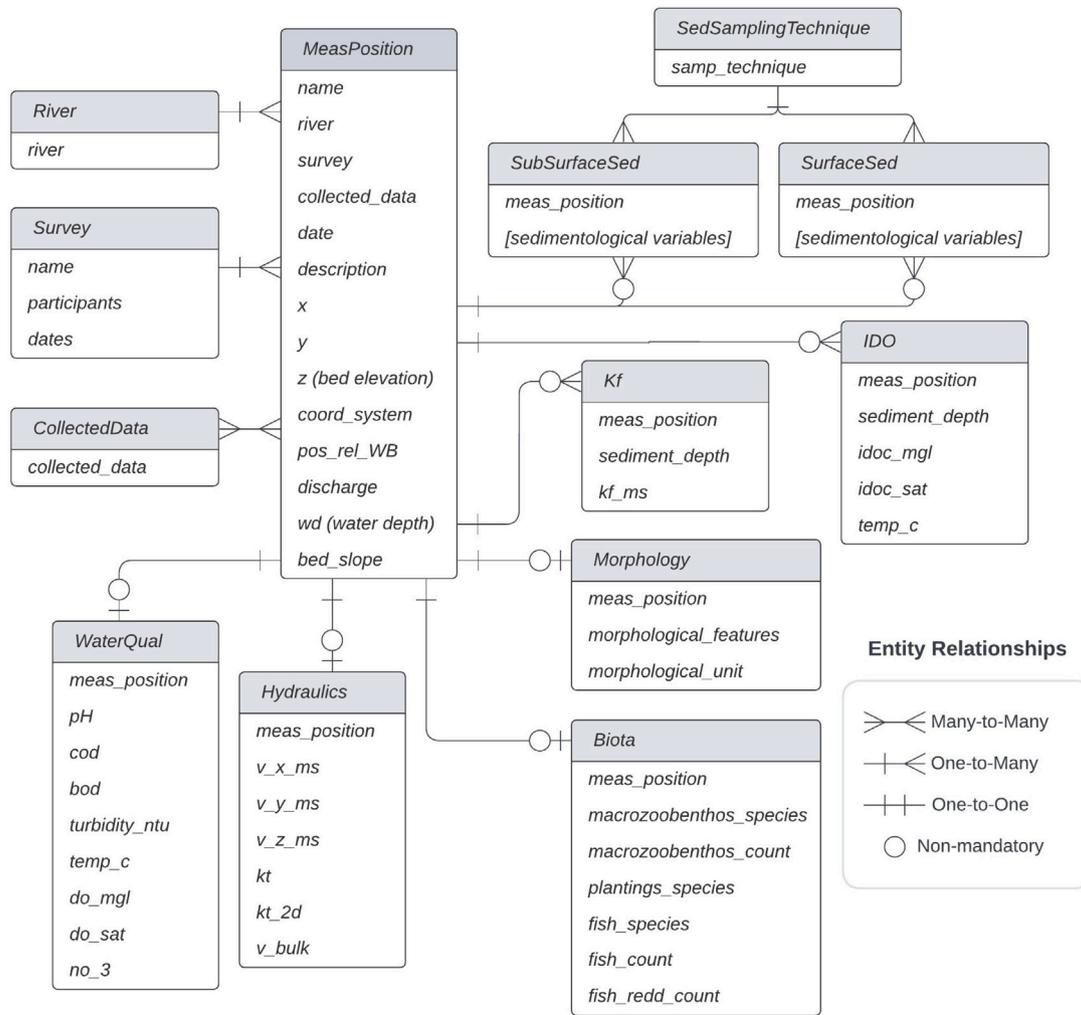


Fig. 2. River Analyst database entity relationship diagram (ERD).

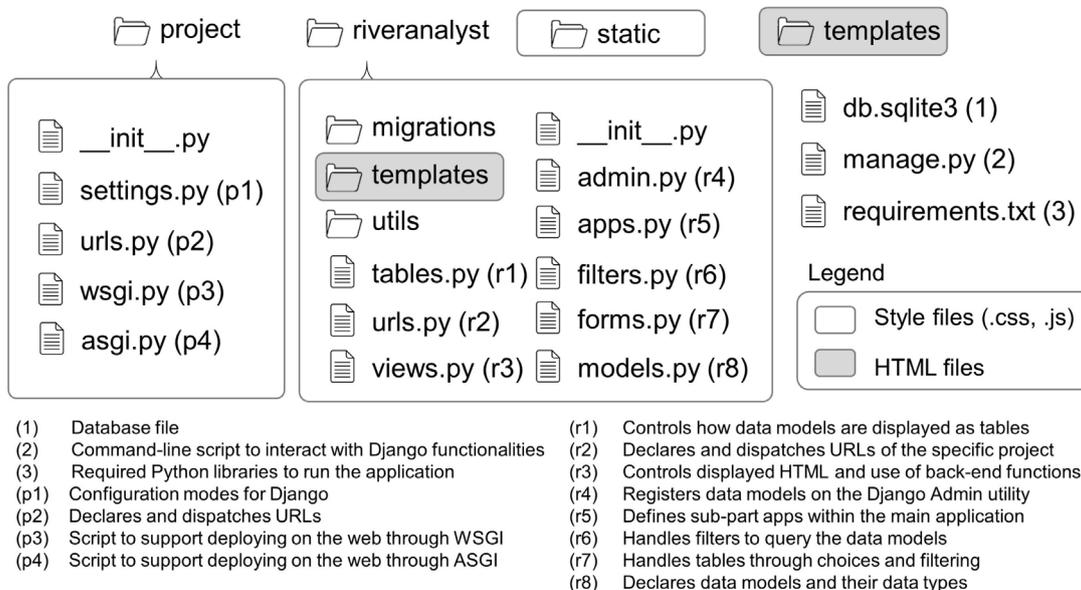


Fig. 3. Software structure with its main constituents and definitions.

Table 1
Output parameters and methods used to compute characteristics of sediment grain size distributions.

Parameter	Equation	Sources
Characteristic grain sizes	$d_{\%c}^a$	Linear interpolation
Fine sediment fractions	FSF < 2 mm; FSF < 0.5 mm	Cumulative percentages
Mean grain size	$d_m = \frac{1}{N} \sum_i d_{\%i}^b$	
Standard deviation of grain sizes	$\sigma_d = \sqrt{\sum_i (d_{\%i} - d_m)^2 / N^b}$	
Skewness; Kurtosis	$\frac{\sum_i (d_{\%i} - d_m)^3}{(N-1)\sigma_d^3}; N \cdot \frac{\sum_i (d_{\%i} - d_m)^4}{\sum_i (d_{\%i} - d_m)^2}$	
Geometric mean grain size	$d_g = \sqrt{d_{84} \cdot d_{16}}$	Kondolf and Wolman (1993) and Yang (1996)
Geometric standard deviation	$\sigma_\phi = \sqrt{\sum_k f_k (\phi_k - \sum_k f_k \phi_k)^2 c}$	Frings et al. (2011) and Krumbain (1934)
Sorting coefficient	$S_0 = \sqrt{d_{84} / d_{16}}$	Bunte and Abt (2001)
Fredle index for spawning gravel	FI = S_0 / d_g	Lotspeich and Everest (1981)
Uniformity coefficient	$C_u = d_{60} / d_{10}$	DIN 18196 (DIN, 2006)
Curvature coefficient	$C_c = d_{30}^2 / (d_{60} \cdot d_{10})$	DIN 18196 (DIN, 2006)
Porosity η	Multiple equations from Carling and Reader (1982), Wu and Wang (2006), Wooster et al. (2008) and Frings et al. (2011), see Supplementary Material (Section 3)	
Estimated hydraulic conductivity	$k_{f,est}$ from Kozeny–Carman equation, see Kozeny (1927), Carman (1956, 1937) and Carrier III (2003), see Supplementary Material (Section 4)	

^a For $\%c \in \{10, 16, 25, 30, 50, 60, 75, 84, 90\}$.

^b $d_{\%i}$ denotes the interpolated grain size diameter for the i th cumulative size class, $i \in \{1, 2, 3, \dots, 399, 400\}$, $N = 400$, and the size classes $\%i$ were $\{0, 0.25\%, 0.50\%, \dots, 99.75\%, 100\%\}$.

^c f_k denotes the fraction content of sediment in the k th size class; ϕ_k is the characteristic sediment diameter for size class k expressed on the ϕ scale.

3.3. Grain size analysis

The sedimentary attributes of a river environment are key to fluvial sediment transport (Einstein, 1950), morphological conditions (Eaton et al., 2004), and habitat suitability of rivers (Bovee, 1986; Kondolf and Wolman, 1993; Stalnaker et al., 1995; Noack et al., 2013). Statistics describing sediment characteristics compose the *Surface* and *SubSurfaceSed* data models in our database (Fig. 2). Such characteristics include, for instance, the grain size $d_{\%}$ (diameter of which a % value of a sample is finer), porosity (estimated), and sorting coefficient S_0 . To obtain these substrate parameters to populate (i.e., upload data to) the database, we developed customized functions to analyze sediment samples using the Dash Python library (Plotly Technologies Inc., 2022a) and embedded them in an additional *sediment* module. A CSV template in the code repository provides guidance for preparing the input files (see also Figure 2 in the Supplementary Material), along with a video tutorial featuring instructions on inserting parsing information for a drag & drop data input (River Analyst, 2023). The *sediment* module computes linearly interpolated grain size characteristics, such as cumulative percentages (0 to 100%) with a step width of 0.25%, which also enables deriving characteristic grain sizes (d_{10} , d_{16} , d_{25} , d_{30} , d_{50} , d_{60} , d_{75} , d_{84} , or d_{90}). Table 1 shows a complete list of the sediment-related outputs and implemented calculation methods.

3.4. Principal component analysis (PCA)

Principal component analysis (PCA) is a variance-based method to identify the directions of the highest variance in a dataset and reveal data patterns by reducing dimensionality. In River Analyst, PCA was implemented through the scikit-learn Python library (Pedregosa et al., 2011). In contrast to traditional physics-based dimensional analysis (Barenblatt, 1987) of SI units of measured parameters, PCA uses the statistical variance in data structures of measured parameters for transforming them into a new set of parameters, the principal components (PCs). Therefore, the PCs are linear combinations of measured parameters (hereafter called features, in line with machine learning terminology). The first PC is defined as the linear combination of features and explains the largest amount of variance in the dataset. Each subsequent PC is orthogonal to its preceding PC and explains the next highest amount of variance. Thus, the total amount of data

variance is increasingly explained with a higher number of PCs, but a too high number of PCs leads to unwanted overfitting. Overfitting with too many PCs leads to adding more complexity (i.e., PCs) that may not reflect the most meaningful patterns or variations in the data, but instead capture noise or minor fluctuations.

Every PC is expressed as a function of *loadings* indicating the degree to which each feature contributes to the general variance explained by the PC. Thus, the magnitude of the loading represents a correlation strength between features and PCs, where the sign indicates whether the feature has a positive or negative contribution to the PC (Bro and Smilde, 2014). The optimum number of PCs can be determined through a heuristic method called “elbow” cutoff, which consists of plotting the amount of variance explained by every PC against the number of PCs. The resulting curve of variance explained versus the number of PCs has a knick-point corresponding to an “elbow” suggesting an optimum number of PCs. Adding more PCs beyond this point will not significantly improve the variance explained but tend to result in overfitted PCs.

Although normality is not a requirement for PCA, data standardization (also referred to as normalization or feature scaling) plays a crucial role in handling features with varying magnitudes and units. Unequal absolute magnitudes can, for instance, lead to PCs that are dominated by features with high absolute variances, regardless of their actual contribution to the data structure. Therefore, standardization ensures that the data structures have comparable features with a zero mean and comparable minima and maxima through scaling with the feature's standard deviation. The equation to standardize each feature X to its standardized form X_{norm} is as follows (Spiegel, 1990):

$$X_{norm} = (X - \mu) / \sigma \quad (1)$$

where μ is the arithmetic mean and σ is the standard deviation of a feature.

Although PCA is based on statistics, it can potentially reveal valuable insights into physical processes. For instance, a PC strongly loaded with nutrient concentrations and dissolved oxygen may be associated with ecosystem metabolism rates and possibly eutrophication in the water body. Analogously, a PC heavily loaded with features such as turbidity and suspended sediment concentration may point to strong effects of sediment erosion processes in the watershed.

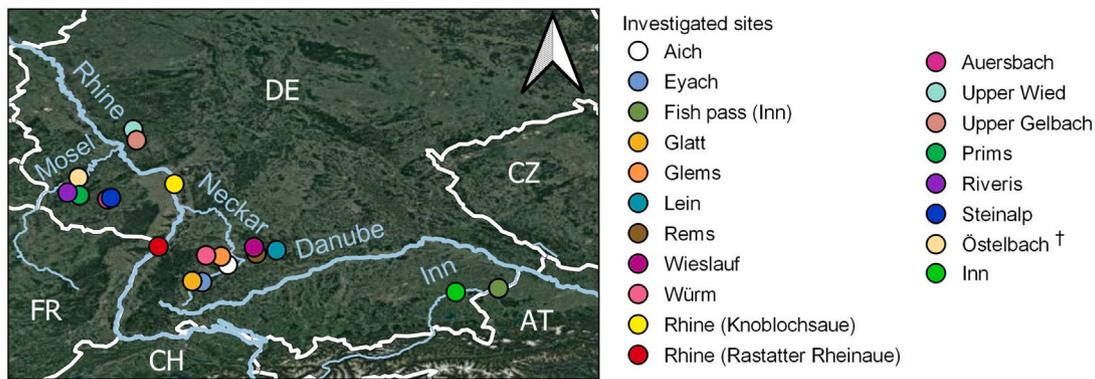


Fig. 4. Locations of investigated rivers and sites between 2018 and 2022. Map data: Landsat/Copernicus through Google Satellite. ©Google. † Also referred to as “Rommelsbach” in the past.

4. Test approach

4.1. Study data

This study builds on multiple, extensive field surveys of rivers in Germany, conducted between 2018 and 2022 (cf. Supplementary Material, and published data in Seitz, 2020; Negreiros et al., 2023a). Specifically, the study sites were at the Rhine River (average discharges of 642.9 to 680.2 m³/s), small rivers of the Rhine Basin, notably the Upper Wied (2.63 m³/s), tributaries of the Mosel River: Riveris (unknown discharge), Östelbach (unknown discharge), and Prims (0.45 m³/s), tributaries of the Glan River: Auersbach (6.1 m³/s) and Steinalp (1.01 m³/s), one tributary of the Lahn River: Upper Gelbach (2.4 m³/s), small tributaries of the Neckar River, notably, the Aich (1.3 m³/s), Eyach (3.2 m³/s), Glatt (4.9 m³/s), Glems (1.1 m³/s), Lein (3.6 m³/s), Rems (5.5 m³/s), Wieslauf (0.9 m³/s), Würm (4.7 m³/s), the Inn River (101 m³/s), and a near-natural morphodynamic fish pass at the Inn River (2 m³/s) near the city of Simbach am Inn (details of the sites in Fig. 4, and the Supplementary Material Section 5). The surveys included measurements of the surface sediment composition, hydrodynamic acoustic Doppler velocimetry (ADV), and MultiPAC (Negreiros et al., 2023a). MultiPAC yielded grain size measurements from frozen sediment cores of the subsurface, and depth-explicit riverbed profiles of IDO and so-called *slurping rates* (belonging to the K_f data model). Slurping rates represent flow rates of interstitial water when connected to atmospheric pressure. Sediment samples (frozen cores and surface samples) were dried and sieved in a lab to obtain grain size distributions that can be expressed with the *Surface-* and *SubSurfaceSed* data models (cf. Fig. 2). A total of 221 *Hydraulic*, 2085 *IDO* (197 depth profiles), 1979 slurping rates (from 195 depth profiles), 246 *SubSurfaceSed*, and 77 *SurfaceSed* objects were created, where each object represented a row of the data model (table). Note that multiple objects (rows) belonged to a measurement position (*MeasPosition* object; see Fig. 2).

4.2. Algorithmic implementation for riverbed data analysis

To investigate vertical connectivity in the riverbeds, we applied PCA with features potentially describing vertical connectivity (i.e., hyporheic exchange) and its disruption through clogging. Specifically, the queried features of interest for sedimentological riverbed clogging (Dubuis and De Cesare, 2023; Negreiros et al., 2023a) were the characteristic grain size d_{84} , IDO, slurping rates (as a replacement for the hydraulic conductivity k_f), the ϕ -scale geometric standard deviation of grain sizes σ_ϕ , and the amount of fine sediment expressed as the fine sediment fraction smaller than 2 mm and 0.5 mm (FSF < 2 mm and FSF < 0.5 mm, respectively). Regarding surface hydraulics, only measurements of water depth h were available at every site

for consideration in the PCA. A too-low amount of measurements of free-surface hydrodynamics, such as flow velocity or turbulent kinetic energy (Nikora and Goring, 1998; Kundu and Cohen, 2008), did not allow for their consideration in the PCA, which is why these measurements are only provided in the Supplementary Material (Section 6). The biochemical parameter IDO was factored in because of its known relevance in describing riverbed clogging in any form (Negreiros et al., 2023a). Furthermore, we used a raw MultiPAC feature, the so-called slurping rate as a proxy for riverbed permeability for water. In addition, a back-calculation of riverbed porosity (η) was not reliably possible due to biases of available empirical equations based on grain sizes only. Instead, σ_ϕ served as a universal indicator of available pore space, similar to porosity. High σ_ϕ points to wide grain size distributions, where fine grains fill the voids between coarse grains, and therefore decrease pore space. The σ_ϕ feature has a known physical dependence on FSF < 2 mm (Frings et al., 2011), which represents redundancy in the data structure. However, such redundancy is uncomplicated for PCA because it effectively mitigates spurious correlation (Aggarwal et al., 2015). Specifically, the orthogonal transformation performed in PCA produces uncorrelated PCs that capture the maximum variance in the data. The maximum-variance criterion prioritizes components that explain the most significant information and discards less relevant (i.e., redundant) information, thus reducing the influence of redundant parameters (Aggarwal et al., 2015). Still, the involvement of redundant parameters should be possibly small to avoid computing inefficiency and hampering the interpretability of PCs.

A differentiation between fine and coarse sediment-dominated riverbeds was of particular interest for testing the hypothesis of whether sedimentological riverbed clogging only occurs under environmental conditions dominated by particularly coarse sediment. Thus, additional PCAs were performed to classify sites accordingly. The classification was implemented as a function of a sediment size threshold value of FSF < 2 mm of 50% (hereafter referred to as ψ_{FSF}), where a FSF < 2 mm larger or smaller than ψ_{FSF} corresponded to fine sediment-dominated or coarse sediment-dominated riverbeds, respectively. Also, ψ_{FSF} was varied between FSF < 2 mm ranging from 35% to 65% (in 5% steps) to assess a potential amount of fine sediment that leads to different feature loadings of PCs. The variation not only served for testing the hypothesis but also as a sensitivity analysis of ψ_{FSF} .

Data was queried with River Analyst and post-processed in a Pandas dataframe (McKinney, 2010) using measurement positions and vertical riverbed depth as merging keys (not to be confused with foreign keys). Before the PCA, the data was standardized according to Eq. (1). The depth-explicit features IDO and slurping rate were depth-averaged to enable comparisons with bulk sediment parameters that were not measured over depth, such as grain size characteristics. To ensure consistency in the averaging process, only rows with both IDO and slurping rate measurements were considered.

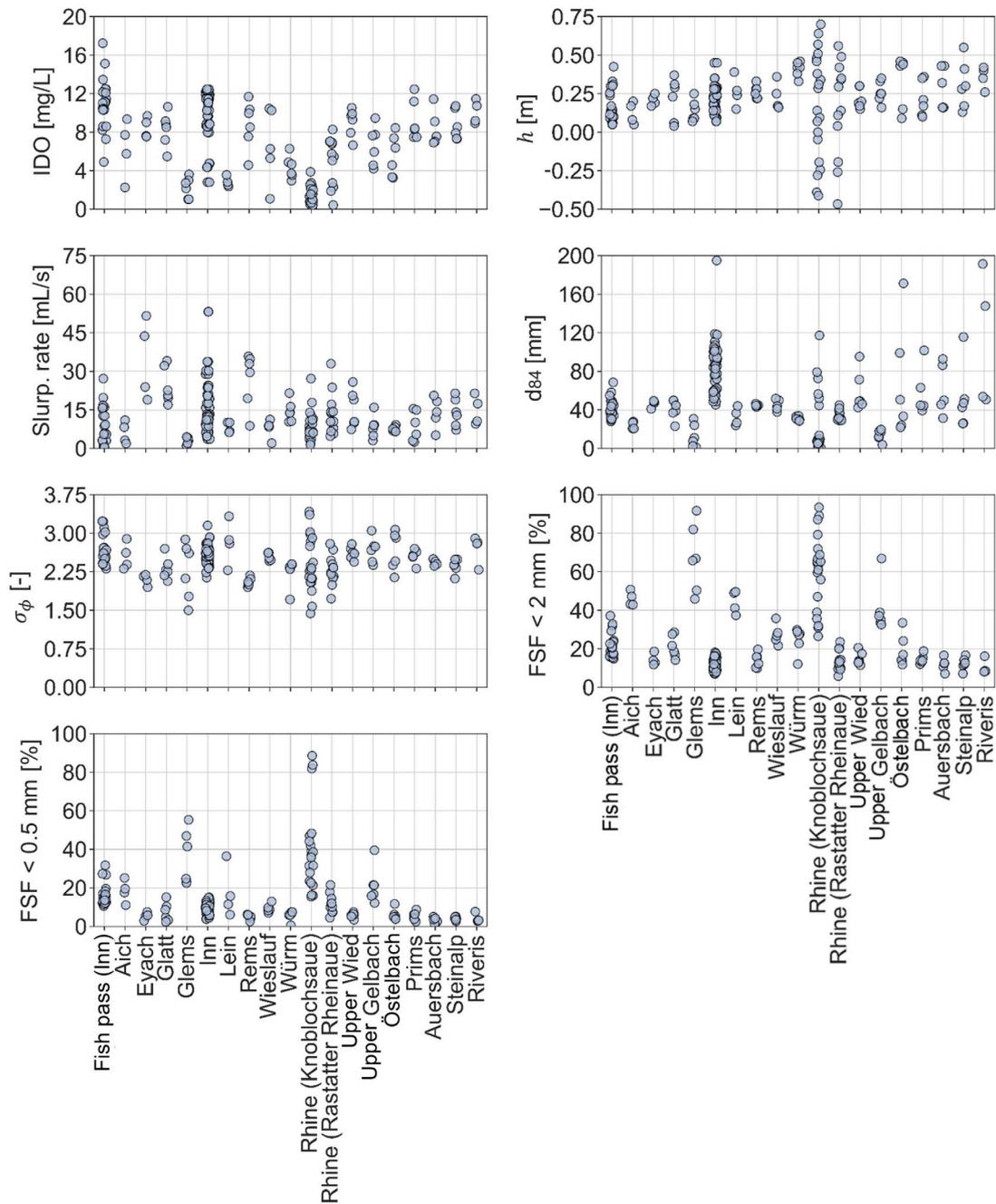


Fig. 5. Strip plots of the available features measured at the sites.

5. Results

5.1. Feature characteristics

The River Analyst database was populated with a total of 465 measurement positions. Querying the database for PCA yielded an array of feature values from 186 measurement positions (Fig. 5). The pre-standardization feature values of IDO, slurping rate, geometric standard deviation σ_ϕ of grain sizes, fine sediment fractions (FSF < 2 mm and FSF < 0.5 mm), characteristic grain size d_{84} , and water depth h exhibited data structures that may be attributed to the absolute discharge (Rhine), and absolute sediment size (e.g., FSF and d_{84} at the Inn and Riveris versus Rhine (Knoblochsau) and Glerns). IDO ranged from 0.2 to 17.2 mg/L, indicating the presence of both hypoxic (below 2 mg/L; Hawley et al., 2006) and supersaturated conditions (above

9 mg/L; Roegner et al., 2011) in the interstitial water. The water depths ranged from -0.5 to 0.7 m, where negative values represented a water table below the terrain surface (i.e., measurements where the sediments at the surface of the riverbed were not saturated, see Figure 5 in the Supplementary Material). The slurping rate ranged from 0.3 to 53.1 mL/s, the d_{84} from 0.5 to 194.9 mm, the geometric standard deviation of grain sizes σ_ϕ from 1.4 to 3.4 , the FSF < 2 mm from 5.8 to 93.5% , and the FSF < 0.5 mm from 0.5 to 88.6% . All minima and maxima, means and standard deviations are provided in the Supplementary Material, Table 2 and 3, respectively. The sampling sites encompassed predominantly sand (0.063 to 2 mm), and gravel (2 to 64 mm) to cobble (64 to 256 mm) substrates (Wentworth, 1922). The absolute maxima regarding IDO occurred at the morphodynamic fish pass, h (0.7 m) at the Rhine (Knoblochsau), slurping rate and d_{84} at the Inn River, and σ_ϕ , FSF < 2 mm, and FSF < 0.5 mm at the Rhine

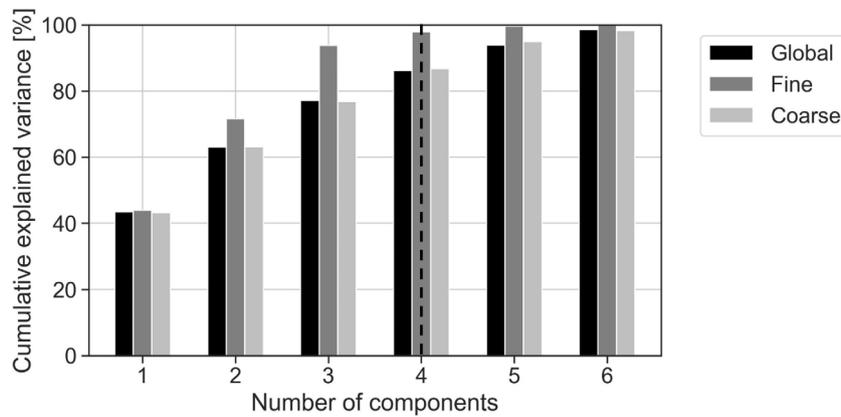


Fig. 6. Cumulative explained variance resulting from PCAs explaining the global dataset, fine sediment-dominated subset, coarse sediment-dominated subset. The selected cut-off point according to the “elbow” method is indicated with the dashed line.

(Knoblochsaue). Absolute minima occurred with respect to IDO at the Rhine (Knoblochsaue), h at the Rhine (Rastatter Rheinaue), slurping rate at the fish pass, d_{84} and σ_ϕ at the Rhine (Knoblochsaue), FSF < 2 mm at the Inn River, and FSF < 0.5 mm at the Würm River sites. Histograms of the features are provided in the Supplementary Material Section 6.

5.2. Principal component analysis (PCA)

5.2.1. Global dataset

The PCA of the standardized dataset provided insights into the variance explained by the loadings of the PCs. The “elbow” method yielded an optimum number of four PCs, explaining 86.3% of the total variance of the dataset (Fig. 6). The PC loadings for each feature are correlation strengths between features and PCs (Fig. 7). PC 1 (global), accounting for 43.4% of the data variance, exhibited strong positive loadings with fine sediment features (FSF < 2 mm and FSF < 0.5 mm loadings >0.8, cf. Fig. 7) and pronounced negative loadings (−0.75) with the characteristic grain size d_{84} and IDO. PC 2 (global), explaining 19.6% of the variance, showed a heavy positive loading (0.88) with the geometric standard deviation of grain sizes (σ_ϕ) and a negative loading (−0.66) with the slurping rate, indicating a decrease in permeability associated with wide grain size distributions and small pore space. PC 3 (global) accounted for 14.1% of the dataset variance and displayed a significant negative loading (−0.93) with water depth h while having only small loadings (<0.22) with other features. PC 4 (global), explaining 9.1% of the variance, exhibited primary loadings (between 0.37 and 0.44) with the slurping rate, FSF < 0.5 mm, and d_{84} . Notably, PC 4 (global) represented coarse substrate with high levels of very fine sediment (FSF < 0.5 mm) and high slurping rates, deviating from the expected physical causality that high fine sediment fractions reduce riverbed permeability.

Unlike PC 1 (global), PC 4 (global) featured positively correlated loadings of very fine sediment (FSF < 0.5 mm), slurping rates, and characteristic grain size d_{84} with positive signs. The inverse relationships between PC 1 (global) and PC 4 (global) are depicted in Fig. 8, which illustrates the transformation of the dataset into PC coordinates of every site-specific position. Thus, the scatter plots in Fig. 8 reveal site-specific trends among the PCs, providing insights into the importance of each PC at every site. A comparison of PC 1 (global), PC 2 (global), and PC 3 (global) across the sites identifies point clouds of positions where the Inn and Rhine (Knoblochsaue) sites constituted prominent spikes in the coordinate directions, suggesting insignificant statistical correlations between these sites. Conversely, a comparison between PC 1 (global) and PC 4 (global) shows clear positive trends observed at the sandy sites at the Glems and the Rhine (Knoblochsaue), and evident negative trends at all other sites, particularly the generally clogged Inn site.

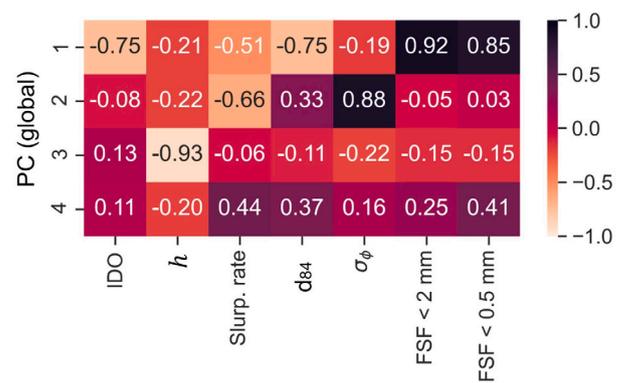


Fig. 7. Feature loadings resulting from the principal components analysis (PCA) of the global dataset (see features in Section 4).

5.2.2. Fine and coarse sediment subsets

The strong contrasting trends between PC 1 and PC 4 among sites primarily characterized by fine sediments and sites with considerable coarse sediment sizes indicated that these two sedimentological environments should be considered in separate PCAs. Based on a sediment size threshold of $\psi_{FSF} = 50\%$ (FSF < 2 mm), the resulting total variances explained by four PCs for fine sediment and coarse sediment environments were 91.3 and 85.2% (Fig. 6), respectively. PC 1 (fine) and PC 1 (coarse) explained 38.0% and 40.1% of the variance in each subset, respectively, followed by PC 2 (fine) and PC 2 (coarse) explaining 24.1% and 19.6%, PC 3 (fine) and PC 3 (coarse) explaining 15.8% and 13.9%, and finally, PC 4 (fine) and PC 4 (coarse) explaining 13.4% and 11.5% of the variance in the subsets.

The differentiated PCAs yielded similar primary components (PC 1s) with regard to fine sediment and d_{84} loadings, but very different loadings regarding all other features (Fig. 9). Specifically, in the fine sediment subset, PC 1 (fine) had a heavy negative σ_ϕ loading (−0.85 in Fig. 9a) making PC 1 (fine) embody particularly fine-grained positions with narrow grain-size distributions. In contrast, in the coarse sediment subset, PC 1 (coarse) had positive σ_ϕ (0.54 in Fig. 9b) and heavily negative slurping rate (−0.68) and IDO (−0.55) loadings, making PC 1 (coarse) represent coarse sediment positions with high fine sediment share and low permeability. PC 2 (fine) was mainly loaded with the water depth h (0.82) and slurping rate (0.85). In contrast, PC 2 (coarse) was heavily loaded with the d_{84} (0.76) and σ_ϕ (0.79). PC 3 (fine) was positively loaded with primarily the FSF < 0.5 mm (0.65) followed by d_{84} and σ_ϕ (both 0.40) and FSF < 2 mm (0.29), and negatively loaded with IDO (−0.46) and the slurping rate (−0.27). In the coarse sediment-dominated subset, PC 3 (coarse) was almost exclusively and

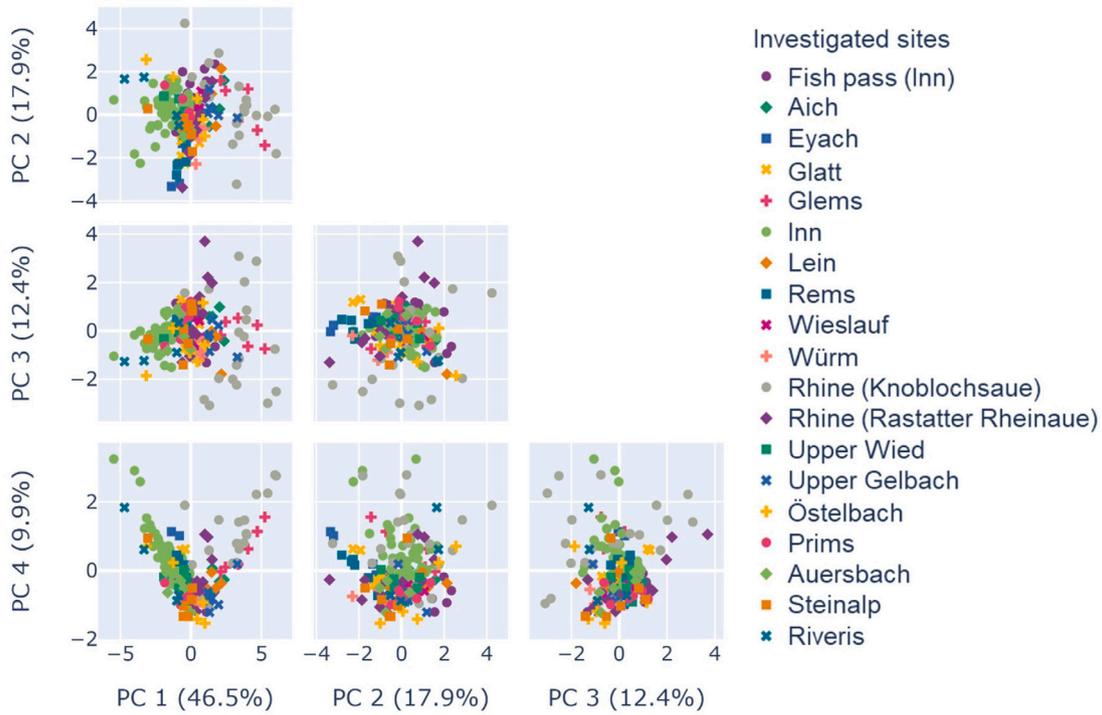


Fig. 8. Pairwise scatter plots of global principal component (PC) coordinates resulting from the principal component analysis (PCA) of the positions, sorted by the measurement sites.

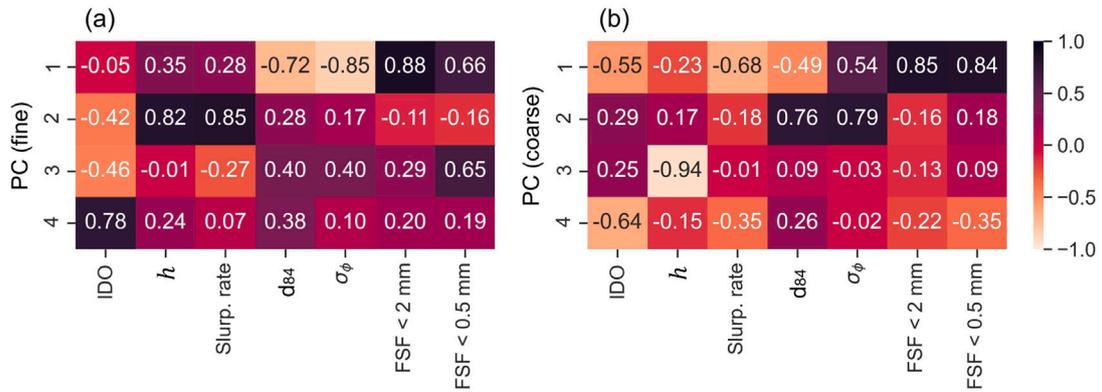


Fig. 9. Feature loadings resulting from applying PCA to the (a) fine sediment-dominated and (b) coarse sediment-dominated subsets, defined by $\psi_{FSF} = 50\%$ (FSF < 2 mm).

negatively loaded with the water depth h (-0.94), and secondarily with IDO (0.25). PC 4 (fine) was predominantly loaded with positive IDO (0.78) and d_{84} (0.38), with weaker positive loadings (≤ 0.24) of all other features. Also, PC 4 (coarse) was primarily but negatively loaded with IDO (-0.64), and less importantly, with negative slurping rates and FSF < 0.5 mm (both -0.35). Only d_{84} had a positive loading on PC 4 (coarse).

5.3. Variation of the sediment size threshold ψ_{FSF}

The characterization of fine and coarse sediment-dominated environments was a function of FSF < 2 mm variations in the form of the sediment size threshold ψ_{FSF} . To investigate the hypothesis that clogging only occurred in environmental conditions dominated by coarse sediment, the PC loadings of particular features can be expected to positively correlate. Notably, for describing a riverbed with functional vertical connectivity (i.e., a non-clogged riverbed state), the IDO, water depth h , and slurping rate loadings were expected to positively correlate (i.e., equal loading sign) with each other. The water depth h and slurping rate loadings should correlate in the absence of clogging

because increases in the hydraulic head (higher water depths) yield higher slurping (outflowing) rates of interstitial water (Seitz, 2020). For describing a clogged riverbed, the FSF < 0.5 mm, FSF < 2 mm, and σ_ϕ loadings were expected to positively correlate with each other. Furthermore, the IDO and slurping rate loadings were expected to exhibit opposing signs to the FSF < 2 mm, FSF < 0.5 mm, and σ_ϕ loadings to characterize a clogged riverbed.

In a fine sediment-dominated environment defined by FSF < 2 mm $\geq \psi_{FSF}$, the PC 1 (fine) loadings of the vertical connectivity features water depth h and slurping rates (light blue lines in Fig. 10) displayed correlational strength with the same sign for $\psi_{FSF} \geq 40\%$. Only the IDO loadings (dark blue line in Fig. 10), generally had different signs. The loadings of PC 1 (fine) with the feature set indicating clogging (red lines in Fig. 10) never showed a positive correlation between fine sediment fractions and σ_ϕ . The loadings of PC 2 (fine) showed strong agreement between the vertical connectivity features h and slurping rate. The clogging feature set did not clearly exhibit positive correlation strength of PC 2 (fine) loadings for any ψ_{FSF} . The vertical connectivity and clogging feature set loadings of PC 3 (fine) displayed generally

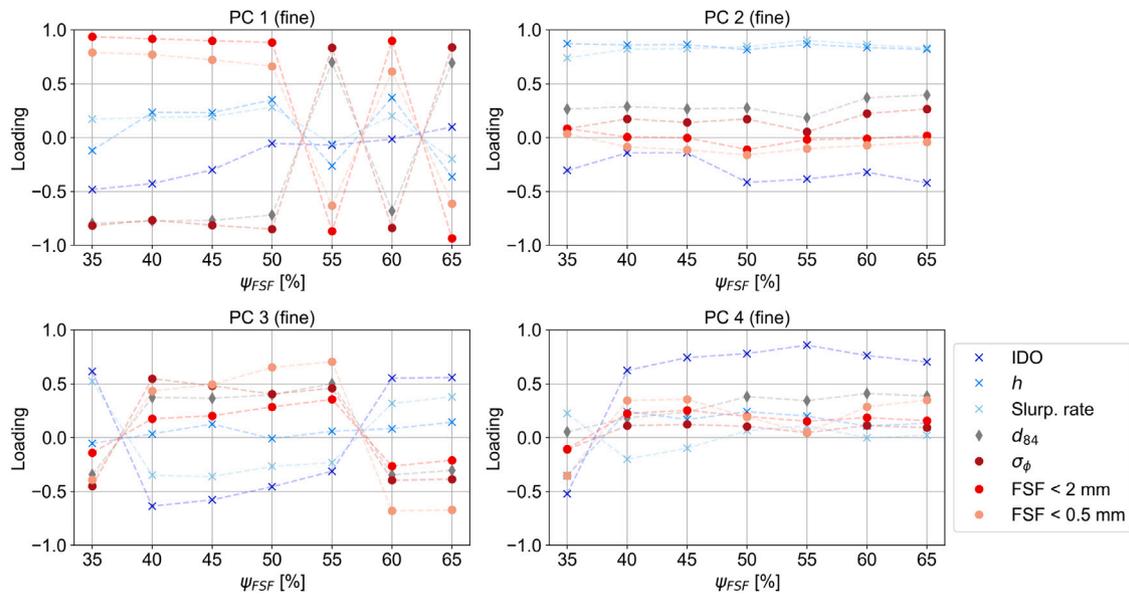


Fig. 10. Feature loadings resulting from PCA applied to fine sediment-dominated subsets defined by ψ_{FSF} . Positions with fine sediment dominance corresponded to FSF < 2 mm larger than ψ_{FSF} .

correlating loadings. For $\psi_{FSF} \geq 50\%$, PC 4 (fine) had overall positively correlating loadings with the clogging and vertical connectivity features, where both groups strictly had the same signs.

When an environment dominated by coarse sediments was defined by $FSF < 2 \text{ mm} < \psi_{FSF}$, the PC 1 (coarse) loadings of both clogging and vertical connectivity-indicating feature sets (red and blue lines, respectively, in Fig. 11) generally showed positive correlation strength with the same sign. Yet, the strength of the loadings correlation of the clogging-indicating feature set weakened considerably for $\psi_{FSF} \geq 55\%$. PC 2 (coarse) was mostly loaded with opposing signs within the clogging and vertical connectivity feature groups. The loadings of PC 3 (coarse) were significantly different at $\psi_{FSF} = 35\%$, with little change for $\psi_{FSF} \geq 40\%$, where the vertical connectivity-indicating features h and IDO exhibited clearly opposing loadings. The clogging-indicating feature loadings of PC 3 (coarse) were weakly correlated regarding FSF < 2 mm and σ_ϕ . PC 4 (coarse) was loaded with weakly correlating clogging features for $\psi_{FSF} \geq 50\%$, and negative-signed trends of vertical connectivity features for $\psi_{FSF} \leq 60\%$. Only for $\psi_{FSF} = 65\%$, all feature loadings of PC 4 (coarse) switched to a positive sign.

6. Discussion

6.1. Insights from the database: drivers of riverbed clogging

Sedimentary riverbed clogging manifests in reduced pore space and consequently low porosity (Frings et al., 2011; Huston and Fox, 2015) corresponding to high σ_ϕ and reduced permeability of the riverbed (Schälchli, 1992) with low correlation between the water depth h and slurping rates. Therefore, a fully clogged riverbed was characterized by low IDO, low slurping rates, high σ_ϕ , and high fine sediment fractions (i.e., FSF < 0.5 mm and FSF < 2 mm). Thus, a system that is susceptible to clogging should be described by PCs loaded with feature trends corresponding to clogging conditions. The higher the importance of such PC, the higher is also the relevance of riverbed clogging in the considered system.

The PCA of the global dataset produced two primary PCs that exhibited positively associated loadings with clogging features. Specifically, the primary important PC 1 (global) was heavily positively loaded with fine sediment features and negatively loaded with the IDO and slurping rate (Fig. 7). The only exception was a negative loading with σ_ϕ , which should, however, be positively correlated with FSF < 0.5 mm and FSF

< 2 mm to express clogging (Einstein, 1968; Schälchli, 1992; Huston and Fox, 2015; Negreiros et al., 2023a). Such high and positive σ_ϕ loading was observed in PC 2 (global) (σ_ϕ loading = 0.88), which was negatively loaded with the slurping rate (−0.66) and slightly with IDO (−0.08). The less important PC 3 (global) and PC 4 (global) had little relevance to indicate clogging.

PCAs of riverbeds dominated by coarse sediments also yielded a primary PC 1 (coarse) with similar, but compared with PC 1 (global), more pronounced loadings that are descriptive of clogging (Fig. 9). Specifically, PC 1 (coarse) was not only heavily positively loaded with FSF < 0.5 mm and FSF < 2 mm (0.84 and 0.85, respectively), but also with σ_ϕ (0.54), while the feature set pointing to functional vertical connectivity contributed purely with negative loadings (−0.23 to −0.68). However, PC 2, 3, and 4 (coarse) were not descriptive for clogging conditions because neither σ_ϕ and the fine sediment features, nor the vertical connectivity features (IDO, h , and slurping rate), had the same signs, respectively. Thus, the important PC 1 (coarse and global) and PC 2 (global) can be considered strong indicators of sedimentary riverbed clogging in environments where coarse (i.e., at least fine gravel) and finer sediment fractions are present.

For the fine sediment-dominated subset, PC 1 (fine) was partially loaded with features pointing to clogging (heavily negative σ_ϕ against positive FSF < 0.5 mm and FSF < 2 mm loadings), but the vertical connectivity features of water depth h and slurping rate had the same sign and similar magnitudes. This observation was even more pronounced regarding PC 2 (fine), which was heavily loaded with both vertical connectivity features (h and slurping rate). In particular, in a permeable riverbed with functional vertical hyporheic connectivity, the hydrostatic pressure exerted by the water depth h increases the interstitial water fluxes vertically. Thus, the heavily positive h and slurping rate loadings of PC 2 (fine) suggest that higher water depths were associated with higher slurping rates in fine sediment-dominated environments. Thus, both PC 1 (fine) and PC 2 (fine) suggest that fine sediment deposition did not influence fine sediment environments toward reducing riverbed permeability. Still, the IDO loadings were not in line with the observed permeability-rich patterns of h and the slurping rate loadings of PC 1 (fine) and PC 2 (fine). This variability in IDO may be detached from fine sediment controls because of upwelling and downwelling effects in the riverbed, which, however, were not quantified in the dataset. Only PC 3 (fine) can be considered a descriptor of clogging with the feature set of FSF < 0.5 mm, FSF < 2 mm, and

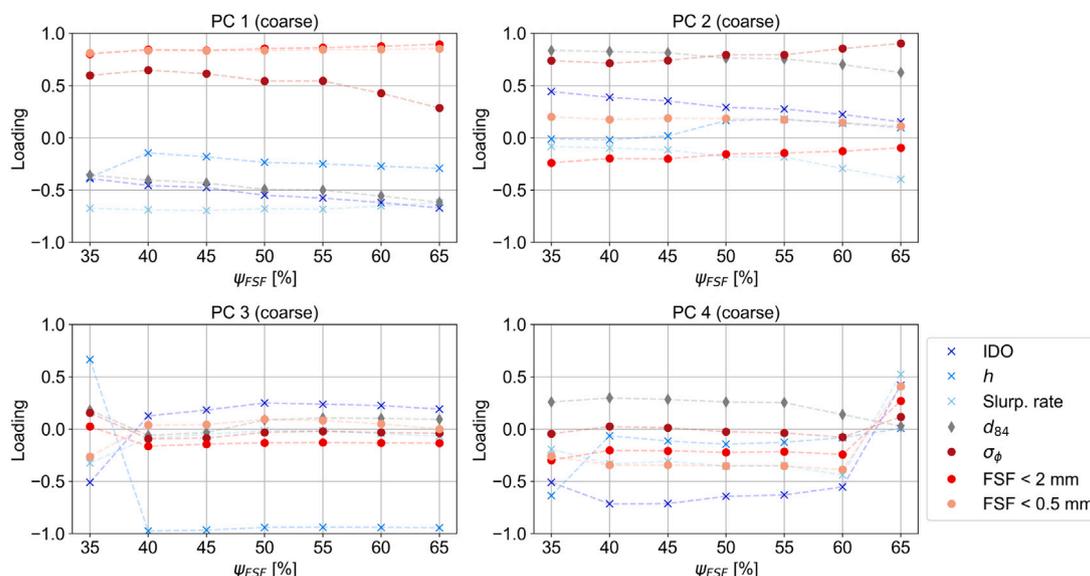


Fig. 11. Feature loadings resulting from PCA applied to coarse sediment-dominated subsets defined by ψ_{FSF} . Positions with coarse sediment dominance corresponded to FSF < 2 mm smaller than ψ_{FSF} .

σ_ϕ being inversely related to the slurping rate and IDO (Fig. 9a). Thus, in fine sediment environments, PC 1 (fine) and PC 2 (fine) suggested positive correlations describing high substrate permeability, and only the less important PC 3 (fine) could be interpreted as an indicator of clogging with weak correlations of feature loadings.

PC 1 (coarse) was the most important component driving riverbed variability in coarse sediment environments and could be considered a clogging descriptor. For fine sediment environments, a clogging descriptor could only be observed in one of the least important PCs, notably PC 3 (fine). These interpretations stemmed from $\psi_{FSF} = 50\%$ and thus were related to the definition of ψ_{FSF} . For this reason, PC loadings were examined as a function of $\psi_{FSF} \in [35\%, 65\%]$, meaning that substrate processes in an environment were considered to be dominated by fine sediment when its FSF < 2 mm was greater than ψ_{FSF} . Vice versa, processes in an environment were considered to be dominated by coarse sediment when its FSF < 2 mm was smaller than ψ_{FSF} . The variations also express the sensitivity of the PCA loadings to ψ_{FSF} . The trends in the loadings of PCs 1–4 (fine) and PCs 1–4 (coarse) generally confirmed the observations of environments dominated by fine (Fig. 10) and coarse (Fig. 11) sediments, though with refinements regarding the observations. For instance, the strong magnitude of clogging-indicating loadings of PC 1 (coarse) weakened above $\psi_{FSF} \geq 55\%$, which points to less relevance of clogging mechanisms when FSF < 2 mm exceeds 55%. PC 2 (coarse) was not associated with clogging because the clogging and vertical connectivity-indicating loadings presented contradictory within-set signs. Notably, σ_ϕ loadings of PC 2 (coarse) were negative, whereas the fine sediment feature loadings were positive. Similar to PC 2 (coarse), PC 3 (coarse) and PC 4 (coarse) could not be related to clogging (Fig. 11). In addition, the relevance for PC 1 (fine) and PC 2 (fine) being indicators for the absence of clogging was only provided for $\psi_{FSF} \geq 50\%$ (Fig. 10). For $\psi_{FSF} \geq 50\%$ in PC 1 (fine), IDO loadings became negligible and h and slurping rate loadings have the same sign, which implies that interstitial oxygen was not a function of fine sediment increases nor of the permeability.

For clogging to be designated as a relevant phenomenon in the considered system, clogging patterns should be indicated by the dominant PCs explaining most of the variance of the data. Those are primarily the first and second PCs (Fig. 6). Thus, the two most important components, PC 1 (global and coarse), can be considered indicators for clogging, in particular, when the FSF < 2 mm was smaller than 55%. In contrast, the next-important PC 2 (fine) represented an indicator for the absence

of clogging when the FSF < 2 mm was larger than 40%, in particular, because of non-clogged riverbeds reflecting the hydrostatic pressure (expressed by the water depth h) in the slurping rate. Evidence can be seen in the nearly identical h and slurping rate loadings of PC 2 (fine) for FSF < 2 mm larger than 40% in Fig. 10. Thus, the most important statistical descriptors of coarse and fine sediment environments can be interpreted as pointers to riverbed clogging and the absence of clogging characteristics, respectively. Accordingly, environments with mixtures of coarse sediment with grain diameters greater than 2 mm (i.e., gravel and coarser) and fine sediment (sand and finer) can be expected to be affected by clogging. Therefore, it is expected that the presence of clogging characteristics becomes less dominant when FSF < 2 mm exceeds approximately 50%–55%. In conclusion, this application of River Analyst provided evidence that supports the hypothesis that sedimentological riverbed clogging only occurs under environmental conditions dominated by particularly coarse sediment with FSF < 2 mm smaller than 50%–55%.

6.2. Consequences for river restoration

The sand-dominated sites at the Glems and Rhine (Knoblochsaue) River (average FSF < 2 mm larger than 60.1%; Table 3 in the Supplementary Material) had on average low slurping rates (3.1–8.9 mL/s; Table 3 in the Supplementary Material) related to small pore space, but their PC loadings still suggested functional vertical connectivity (e.g., PC 1 and 2 (fine) in Fig. 10). These site characteristics are typical for low-energy environments with fine sediment and wide grain size distributions of lowland riverbeds (Boggs, 2009; Mooneyham and Strom, 2018). In contrast, mountain rivers naturally manifest only a few low-energy patches (e.g., slackwater units, cf. Wyrick and Pasternack, 2014), where fine sediment deposition can clog their generally coarse sediment matrix. Mountain river sites (according to definitions in Wohl, 2000) with low sedimentological clogging defined by high IDO, high slurping rates, and low fine sediment fractions were found in this study at multiple positions of the fish pass (Inn), Eyach, and Riveris. In contrast, the Inn River site embraced multiple high-clogging positions, originating from segmentation by dams and other centuries-old river training structures (Kunz et al., 2021), which artificially impose a low-energy environment. Also, changes in the catchment can lead to increased suspended sediment loads (Gurnell and Bertoldi, 2022), and

therefore, contribute to turning gravel-dominated into fine sediment-dominated rivers. Such influences generate unnatural vertical discontinuity in the form of clogging that cannot be easily restored into a high-energy, dynamic environment because the missing hydraulic energy being converted into low-emission hydropower (Schleiss, 2017), and morphodynamics coming to a quasi-standstill due to lacking coarse sediment supply (Kondolf et al., 2006).

The insights from this study suggest that high content of sandy material (FSF < 2 mm of more than 50%–55%), for example, recruited from the alluvial plains of a river, characterize riverbeds with functional vertical connectivity. To this end, river widening by removing bank protections or terraforming could not only serve to immediately increase local high-quality hydraulic habitat patches (Schwindt et al., 2019) but also improve vertical connectivity by new sediment recruitment. Yet, high fine sediment fractions may deteriorate spawning gravel habitat quality (e.g., for salmonids), even when the fine sediment is recruited from the natural riverbanks (Kondolf and Wolman, 1993). Ultimately, river restoration to a pristine historic environmental state is hardly feasible in light of legacies, which is why sustainable river restoration gives particular importance to ecological functionality and morphodynamic connectivity (Wohl, 2019). However, the scientific baseline for the restoration of ecologically functional river environments through improving their connectivity still needs ample enrichment to achieve restoration goals, as, for example, defined in the European Water Framework Directive (WFD, 2000/60/EC, cf. Dale and Beyeler, 2001; Wuijts et al., 2023). The River Analyst algorithms have the potential to bolster insights from restoration science to accomplish targeted ecological standards with data-driven evidence.

6.3. Database application and traditional sites comparisons

State-of-the-art techniques for comparing measurements between sites and scaled experiments are based on dimensional analysis, which reduces physical quantities (here: features) to dimensionless numbers (Barenblatt, 1987, 1996). However, dimensional analysis is only meaningful for describing a (static) state of a specific environment, and it cannot inherently differentiate between environments dominated by different physical controls. For instance, any differentiation between fine and coarse sediment-dominated environments was not possible, which also explains why previous studies could not derive a dimensionless description of riverbed clogging (Schälchli, 1993; Cui et al., 2008; Dubuis and De Cesare, 2023). To this end, bulk dimensionless numbers, such as a fine sediment Rouse number (Dubuis and De Cesare, 2023), or a permeability Reynolds number (Fries and Trowbridge, 2003; Huston and Fox, 2015; Voermans et al., 2018) cannot recognize the environment and associated dominant physical processes to which they are applied. For processes in environments dominated by even finer, cohesive material, other physical quantities are relevant system controls, such as capillary forces (Baumgartner and Liebscher, 1996), but a dimensional analysis cannot inherently control for absolute sediment sizes.

In contrast to dimensional analysis, the PCAs performed with the River Analyst algorithms and database efficiently showed that weak trends in data structures may be linked to the comparison of incomparable environments, which explains the incapacity of dimensional analysis to describe clogging. Specifically, the database WAF facilitates the examination of measurements in environments dominated by different control conditions. The centralized database management system with geospatial context enabled a workflow for identifying potential patterns in the structure of measurement data, which can then be verified with PCA. Although PCA can be applied on top of traditional dimensional analysis, it would not change the statistical structure of the data. Still, efficient data analysis and visualization including PCA or similar unsupervised learning techniques, require expert knowledge of environmental engineering, and, at the same time, proficiency in

programming (e.g., Python and WAFs) and (SQL) database management. In this light, the here-introduced River Analyst represents a novel tool for environmental engineers and other experts with no to little knowledge of Python and SQL to establish and manage river databases. Only when code extensions or modifications of the data models (in the *models.pyscript*) are needed, knowledge of the Django Python library is required. The standard set of data models in River Analyst covers a wide variety of hydro-environmental features, with the goal of being generalizable. Other data attributes or a different conceptualization of hydro-environmental features may be required to meet other demands of data model simplicity or complexity (Morsy et al., 2017). For example, future measurement technology may allow for observations with higher spatial refinement and more dimensions (Pokorný, 2006), which can be added to the River Analyst database schema.

The use of relational databases in groundwater hydrology has improved the way how flow and solute transport are modeled (De Dreuzy et al., 2006), and they made the monitoring of groundwater recharge more efficient (Qiu et al., 2022). Similarly, water quality monitoring (Copp et al., 2010; Chini and Stillwell, 2017), and water distribution networks (Hernandez et al., 2016; Pacheco et al., 2021) were improved through structured hydrological databases. In contrast, database schemes for river data have only been established for sparse individual applications, such as for biochemical (Hartmann et al., 2014), discharge (Peucker-Ehrenbrink, 2009), or channel geometry (Andreadis et al., 2013) data. Yet, to leverage insights from multivariate surveys on rivers (Petersen et al., 2001), generally applicable software tools are required (Wohl et al., 2015), and River Analyst represents a viable pathway toward transparent and integrated river management. In addition, the featured PCAs addressing riverbed clogging and vertical connectivity did not yet exploit the full capacity of River Analyst to account for other biochemical and hydrodynamic parameters, such as flow velocity or turbulent kinetic energy. Still, because River Analyst and its database are a versatile open-source ecosystem, other statistical techniques, and parameters can easily be implemented and examined to augment the robustness of hydro-environmental analysis. The River Analyst database encompasses around twelve thousand hours of field and post-processing work and is made freely available for users worldwide in the spirit of open science.

7. Conclusions

A novel approach for managing hydro-environmental data with a specialized database application framework leverages valuable insights into hydro-environments that were previously unattainable through traditional dimensional analysis. The flexibility of this framework, notably River Analyst, allows for its adaptation to meet diverse research data requirements, making it an unprecedented tool for efficiently managing growing river datasets.

A principal component analysis (PCA) built on top of the new database management algorithms demonstrates the capacity of data-driven hydro-environmental assessments to inform river restoration. Specifically, the results of the PCA provide compelling evidence that riverbed clogging, a phenomenon that was previously unexplained by dimensionless parameters, can be attributed to environmental factors correlated with dominant grain sizes. The analysis provides first-time statistical evidence that clogging occurs primarily in coarse sediment environments when fine sediment (<2 mm) makes less than 50%–55% of the total sediment mass, and in particular, in anthropogenically forced low-energy environments. Conversely, sandy environments are little affected by riverbed clogging and exhibit characteristics of functional vertical connectivity. These insights suggest a paradigm shift for restoring the vertical connectivity of river systems, with the prospect of further data-informed advances for augmenting the knowledge of ecologically functional hydro-environments.

Dimensional analysis to compare findings across measurements from multiple sites, rivers, and scaled experiments, faces limitations

in characterizing dynamic river environments with manifold physical controls. In contrast, River Analyst, with its use of PCA and a centralized database management system, has been shown an effective tool for discerning key parameters driving environmental processes, as demonstrated in this study with the example of riverbed clogging.

Notation

Symbol	Name	Database name	Unit
IDO	Interstitial dissolved oxygen	idoc_mgl	mg L ⁻¹
h	Water depth	wl_m	m
k_f	Hydraulic conductivity	kf_ms	m s ⁻¹
Slurp. rate	Slurping rate	slurp_rate_avg_mls	mL s ⁻¹
d_m	Mean grain size	dm	mm
d_g	Geometric mean grain size	dg	mm
$d_{\%c}$	Grain size of which % _c is finer [†]	d10, d16, d50, d60, d84	mm
σ_d	Standard deviation of grain sizes	std_grain	–
σ_ϕ	Geometric standard deviation of grain sizes in the ϕ scale	geom_std_grain	–
FI	Fredle index	fi	–
S_0	Sorting coefficient	so	–
C_u	Uniformity coefficient	cu	–
C_c	Curvature coefficient	cc	–
FSF < 2 mm	Fine sediment fraction smaller than 2 mm	per-cent_finer_2mm	–
FSF < 0.5 mm	Fine sediment fraction smaller than 0.5 mm	per-cent_finer_0_5mm	–
η	Porosity	n	–
u	Streamwise flow velocity	v_x_ms	m s ⁻¹
v	Lateral flow velocity	v_y_ms	m s ⁻¹
w	Vertical flow velocity	v_z_ms	m s ⁻¹
k_t	Turbulent kinetic energy over u , v , and w	kt	m ² s ⁻²
$k_{t,2d}$	Turbulent kinetic energy over u and v	kt_2d	m ² s ⁻²

[†] For %_c ∈ {10, 16, 25, 30, 50, 60, 75, 84, 90}.

CRedit authorship contribution statement

Beatriz Negreiros: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Sebastian Schwandt:** Conceptualization, Investigation, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Federica Scolari:** Software, Investigation, Data curation, Writing – review & editing. **Ricardo Barros:** Software, Data curation, Writing – review & editing, Validation. **Alcides Aybar Galdos:** Investigation, Data curation, Writing – review & editing. **Markus Noack:** Investigation, Resources, Writing – review & editing, Funding acquisition. **Stefan Haun:** Investigation, Resources, Writing – review & editing, Funding acquisition. **Silke Wieprecht:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Beatriz Negreiros reports financial support was provided by the German Research Foundation. Stefan Haun reports financial support was provided by the Baden-Württemberg Foundation. Alcides Aybar Galdos and Markus Noack reports financial support was provided by the German Federal Environmental Foundation.

Data availability

The data and codes of this study are available at <https://osf.io/tk53d/>. The web app is available at <https://riveranalyst.github.io/app>.

Declaration of Generative AI and AI-assisted technologies in the writing process

Nothing to declare.

Acknowledgments

Beatriz Negreiros was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the framework of the China-NSFC-DFG 2020 program (GEPRIS GZ SCHW 2027/2-1, AOBJ: 673857, *Hydro-morphodynamic connectivity and ecosystem design in a changing environment*). Stefan Haun is indebted to the Baden-Württemberg Stiftung for the financial support of the Elite program for Postdocs. We acknowledge financial support of the German Federal Environmental Foundation (DBU, Deutsche Bundesstiftung Umwelt) for the project *Development of integrative guidelines for the assessment of clogging in rivers* (Grant 37315/01-33/2).

We thank the field crew involved with the data acquisition, including Maximilian Kunz, James Maxwell-Taylor, Ruslan Biserov, Romain Lacroix, Maria Ponce-Guzman, and in particular, Lydia Seitz for her pioneering MultiPAC surveys. We also thank Álvaro Negreiros for his key recommendations on the usage of the Django WAF functionalities.

The authors also thank three anonymous reviewers whose valuable feedback helped improving this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.envsoft.2023.105916>.

References

- Abbott, M.B., 1991. A central issue of teaching within the hydroinformatics paradigm. *Houille Blanche* 77 (3–4), 257–261. <http://dx.doi.org/10.1051/lhb/1991024>.
- Aggarwal, C.C., et al., 2015. *Data Mining: The Textbook*, Vol. 1. Springer, <http://dx.doi.org/10.1007/978-3-319-14142-8>.
- Andreadis, K.M., Schumann, G.J.-P., Pavelsky, T., 2013. A simple global river bankfull width and depth database. *Water Resour. Res.* 49 (10), 7164–7168. <http://dx.doi.org/10.1002/wrcr.20440>.
- Barenblatt, G.I., 1987. *Dimensional Analysis*. Gordon and Breach Science Publishers, New York.
- Barenblatt, G.I., 1996. *Scaling, Self-Similarity and Intermediate Asymptotics*. Dimensional Analysis and Intermediate Asymptotics. Cambridge University Press.
- Baumgartner, A., Liebscher, H.-J., 1996. *Allgemeine Hydrologie [General Hydrology]*, Auflage ed. In: Band 1. 2., Gebrüder Borntraeger, Berlin / Stuttgart, Germany.
- Benson, B.J., Bond, B.J., Hamilton, M.P., Monson, R.K., Han, R., 2010. Perspectives on next-generation technology for environmental sensor networks. *Front. Ecol. Environ.* 8 (4), 193–200. <http://dx.doi.org/10.1890/080130>.
- Bioco, J., Rocha, A., 2019. Web application for management of scientific conferences. In: *New Knowledge in Information Systems and Technologies: Volume 1*. Springer, pp. 765–774. http://dx.doi.org/10.1007/978-3-030-16181-1_72.
- Boggs, Jr., S., 2009. *Principles of Sedimentology and Stratigraphy*, fifth ed. Prentice Hall.
- Boulton, A.J., Findlay, S., Marmonier, P., Stanley, E.H., Valett, H.M., 1998. The functional significance of the hyporheic zone in streams and rivers. *Annu. Rev. Ecol. Syst.* 29 (1), 59–81. <http://dx.doi.org/10.1146/annurev.ecolsys.29.1.59>.

- Bovee, K.D., 1986. Development and Evaluation of Habitat Suitability Criteria for Use in the Instream Flow Incremental Methodology. Biological Report 21, National Ecology Center, U.S. Fish and Wildlife Service, Fort Collins, CO, USA, p. 235, URL: <https://digitalmedia.fws.gov/digital/api/collection/document/id/1770/download>.
- Bro, R., Smilde, A.K., 2014. Principal component analysis. *Anal. Methods* 6 (9), 2812–2831. <http://dx.doi.org/10.1039/C3AY41907J>.
- Brooking, C., Hunter, J., 2013. Providing online access to hydrological model simulations through interactive geospatial animations. *Environ. Model. Softw.* 43, 163–168. <http://dx.doi.org/10.1016/j.envsoft.2013.01.011>.
- Bunte, K., Abt, S., 2001. Sampling frame for improving pebble count accuracy in coarse gravel-bed streams. *J. Am. Water Resour. Assoc.* 37 (4), 1001–1014. <http://dx.doi.org/10.1111/j.1752-1688.2001.tb05528.x>.
- Carling, P., Reader, N., 1982. Structure, composition and bulk properties of upland stream gravels. *Earth Surf. Process. Landf.* 7 (4), 349–365. <http://dx.doi.org/10.1002/esp.3290070407>.
- Carman, P.C., 1937. Fluid flow through granular beds. *Trans. Inst. Chem. Eng.* 15, S32–S48. [http://dx.doi.org/10.1016/S0263-8762\(97\)80003-2](http://dx.doi.org/10.1016/S0263-8762(97)80003-2).
- Carman, P.C., 1956. *Flow of Gases Through Porous Media*. Butterworths Scientific Publications, London.
- Carrera-Hernández, J., Gaskin, S., 2008. The Basin of Mexico Hydrogeological Database (BMHDB): Implementation, queries and interaction with open source software. *Environ. Model. Softw.* 23 (10–11), 1271–1279. <http://dx.doi.org/10.1016/j.envsoft.2008.02.012>.
- Carrier III, W.D., 2003. Goodbye, hazen; hello, kozeny-carman. *J. Geotech. Geoenviron. Eng.* 129 (11), 1054–1056. [http://dx.doi.org/10.1061/\(ASCE\)1090-0241\(2003\)129:11\(1054\)](http://dx.doi.org/10.1061/(ASCE)1090-0241(2003)129:11(1054)).
- Chini, C.M., Stillwell, A.S., 2017. Where are all the data? The case for a comprehensive water and wastewater utility database. *J. Water Resour. Plan. Manag.* 143 (3), 01816005. [http://dx.doi.org/10.1061/\(ASCE\)WR.1943-5452.0000739](http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000739).
- Copp, J.B., Belia, E., Hübner, C., Thron, M., Vanrolleghem, P., Rieger, L., 2010. Towards the automation of water quality monitoring networks. In: 2010 IEEE International Conference on Automation Science and Engineering. IEEE, pp. 491–496. <http://dx.doi.org/10.1109/COASE.2010.5584638>.
- Cui, Y., Wooster, J.K., Baker, P.F., Dusterhoff, S.R., Sklar, L.S., Dietrich, W.E., 2008. Theory of fine sediment infiltration into immobile gravel bed. *J. Hydraul. Eng.* 134 (10), 1421–1429. [http://dx.doi.org/10.1061/\(ASCE\)0733-9429\(2008\)134:10\(1421\)](http://dx.doi.org/10.1061/(ASCE)0733-9429(2008)134:10(1421)), URL: [http://ascelibrary.org/doi/abs/10.1061/\(ASCE\)0733-9429\(2008\)134:10\(1421\)](http://ascelibrary.org/doi/abs/10.1061/(ASCE)0733-9429(2008)134:10(1421)).
- Dale, V.H., Beyeler, S.C., 2001. Challenges in the development and use of ecological indicators. *Ecol. Indic.* 1 (1), 3–10. [http://dx.doi.org/10.1016/S1470-160X\(01\)00003-6](http://dx.doi.org/10.1016/S1470-160X(01)00003-6).
- Dauzon, S., Bendoraitis, A., Ravindran, A., 2016. *Django: Web Development with Python*. Packt Publishing Ltd, ISBN: 978178712-1386.
- De Dreuzy, J.-R., Bodin, J., Le Grand, H., Davy, P., Boulanger, D., Battais, A., Bour, O., Gouze, P., Porel, G., 2006. General database for ground water site information. *Groundwater* 44 (5), 743–748. <http://dx.doi.org/10.1111/j.1745-6584.2006.00220.x>.
- Delipetrev, B., Jonoski, A., Solomatine, D.P., 2014. Development of a web application for water resources based on open source software. *Comput. Geosci.* 62, 35–42. <http://dx.doi.org/10.1016/j.cageo.2013.09.012>.
- DIN, 2006. DIN 18196: 2006-06, Erd-Und Grundbau–Bodenklassifikation FÜR Bautechnische Zwecke [Earth and Foundation Engineering Soil Classification for Civil Engineering Purposes]. Deutsches Institut für Normung e. V Beuth Verlag GmbH, Berlin.
- Dubuis, R., De Cesare, G., 2023. The clogging of riverbeds: A review of the physical processes. *Earth-Sci. Rev.* 239, 104374. <http://dx.doi.org/10.1016/j.earscirev.2023.104374>, URL: <https://www.sciencedirect.com/science/article/pii/S0012825223000636>.
- Eaton, B.C., Church, M., Millar, R.G., 2004. Rational regime model of alluvial channel morphology and response. *Earth Surf. Process. Landf.* 29 (4), 511–529. <http://dx.doi.org/10.1002/esp.1062>.
- Einstein, H.A., 1950. The bed-load function for sediment transport in open channel flows. In: Technical Bulletin of the USDA Soil Conservation Service, Vol. 1026. p. 71. <http://dx.doi.org/10.22004/ag.econ.156389>.
- Einstein, H.A., 1968. Deposition of suspended particles in a gravel bed. *J. Hydraul. Div.* 94 (5), 1197–1206. <http://dx.doi.org/10.1061/JYCEAJ.0001868>, URL: <https://ascelibrary.org/doi/10.1061/JYCEAJ.0001868>, Publisher: American Society of Civil Engineers.
- Fries, J.S., Trowbridge, J.H., 2003. Flume observations of enhanced fine-particle deposition to permeable sediment beds. *Limnol. Oceanogr.* 48 (2), 802–812. <http://dx.doi.org/10.4319/lo.2003.48.2.0802>, URL: <http://onlinelibrary.wiley.com/doi/10.4319/lo.2003.48.2.0802/abstract>.
- Frings, R.M., Schuettrumpf, H., Vollmer, S., 2011. Verification of porosity predictors for fluvial sand-gravel deposits. *Water Resour. Res.* 47, <http://dx.doi.org/10.1029/2010WR009690>.
- Goodrich, D.C., Guertin, D.P., Burns, I.S., Nearing, M.A., Stone, J.J., Wei, H., Heilman, P., Hernandez, M., Spaeth, K., Pierson, F., et al., 2011. AGWA: The automated geospatial watershed assessment tool to inform rangeland management. *Rangelands* 33 (4), 41–47. <http://dx.doi.org/10.2111/1551-501X-33.4.41>.
- Greig, S., Sear, D., Carling, P., 2007. A review of factors influencing the availability of dissolved oxygen to incubating salmonid embryos. *Hydrol. Process.: Int. J.* 21 (3), 323–334. <http://dx.doi.org/10.1002/hyp.6188>.
- Groff, J.R., Weinberg, P.N., Oppel, A.J., 2002. *SQL: The Complete Reference, Vol. 2*. McGraw-Hill/Osborne, ISBN: 0072118458.
- Gurnell, A.M., Bertoldi, W., 2022. The impact of plants on fine sediment storage within the active channels of gravel-bed rivers: A preliminary assessment. *Hydrol. Process.* 36 (7), e14637. <http://dx.doi.org/10.1002/hyp.14637>, arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.14637>, URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.14637>.
- Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al., 2020. Array programming with NumPy. *Nature* 585 (7825), 357–362. <http://dx.doi.org/10.1038/s41586-020-2649-2>.
- Hartmann, J., Lauerwald, R., Moosdorf, N., 2014. A brief overview of the GLOBal River Chemistry Database, GLORICH. *Procedia Earth Planet. Sci.* 10, 23–27. <http://dx.doi.org/10.1016/j.proeps.2014.08.005>.
- Hawley, N., Johengen, T.H., Rao, Y.R., Ruberg, S.A., Beletsky, D., Ludsin, S.A., Eadie, B.J., Schwab, D.J., Croley, T.E., Brandt, S.B., 2006. Lake Erie hypoxia prompts Canada-US study. *EOS Trans. Am. Geophys. Union* 87 (32), 313–319. <http://dx.doi.org/10.1029/2006EO320001>.
- Hernadez, E., Hoagland, S., Ormsbee, L., 2016. Water distribution database for research applications. In: World Environmental and Water Resources Congress 2016. pp. 465–474. <http://dx.doi.org/10.1061/9780784479865.049>.
- Hernandez, R.R., Mayerlik, M.S., Murphy-Mariscal, M.L., Allen, M.F., 2012. Advanced technologies and data management practices in environmental science: Lessons from academia. *BioScience* 62 (12), 1067–1076. <http://dx.doi.org/10.1525/bio.2012.62.12.8>.
- Hilty, L.M., Page, B., Hřebíček, J., 2006. Environmental informatics. *Environ. Model. Softw.* 21 (11), 1517–1518. <http://dx.doi.org/10.1016/j.envsoft.2006.05.016>.
- Holt, V., Ramage, M., Kear, K., Heap, N., 2015. The usage of best practices and procedures in the database community. *Inf. Syst.* 49, 163–181. <http://dx.doi.org/10.1016/j.is.2014.12.004>.
- Horsburgh, J.S., Morsy, M.M., Castronova, A.M., Goodall, J.L., Gan, T., Yi, H., Stealey, M.J., Tarboton, D.G., 2016. Hydrosare: Sharing diverse environmental data types and models as social objects with application to the hydrology domain. *JAWRA J. Am. Water Resour. Assoc.* 52 (4), 873–889. <http://dx.doi.org/10.1111/1752-1688.12363>.
- Huston, D.L., Fox, J.F., 2015. Clogging of fine sediment within gravel substrates: Dimensional analysis and macroanalysis of experiments in hydraulic flumes. *J. Hydraul. Eng.* 141 (8), 04015015, URL: <https://ascelibrary.org/doi/abs/10.1061/Publisher: American Society of Civil Engineers>.
- Jin, G., Chen, Y., Tang, H., Zhang, P., Li, L., Barry, D.A., 2019. Interplay of hyporheic exchange and fine particle deposition in a riverbed. *Adv. Water Resour.* 128, 145–157. <http://dx.doi.org/10.1016/j.advwatres.2019.04.014>.
- Kondolf, G.M., Boulton, A.J., O'Daniel, S., Poole, G.C., Rahel, F.J., Stanley, E.H., Wohl, E., Bång, A., Carlstrom, J., Cristoni, C., et al., 2006. Process-based ecological river restoration: Visualizing three-dimensional connectivity and dynamic vectors to recover lost linkages. *Ecol. Soc.* 11 (2).
- Kondolf, G.M., Wolman, M.G., 1993. The sizes of salmonid spawning gravels. *Water Resour. Res.* 29 (7), 2275–2285. <http://dx.doi.org/10.1029/93WR00402>.
- Kozeny, J., 1927. Über kapillare Leitung des Wassers im Boden [On the capillary conduction of water in soils]. In: Sitzungsberichte der Wiener Akademie der Wissenschaften. pp. 271–306.
- Krumbein, W.C., 1934. Size frequency distributions of sediments. *J. Sediment. Res.* 4 (2), 65–77. <http://dx.doi.org/10.1306/D4268EB9-2B26-11D7-8648000102C1865D>.
- Kundu, P., Cohen, I., 2008. *Fluid Mechanics*, fourth ed. Elsevier Inc., San Diego, California.
- Kunz, M., Negreiros, B., Schwindt, S., Haun, S., Noack, M., Wieprecht, S., 2021. Inn-ovative morphological restoration efforts. In: Proceedings of the International Symposium on Bedload Management. Interlaken, Switzerland, p. 5. <http://dx.doi.org/10.3929/ethz-b-000513098>.
- Lotspeich, F., Everest, F., 1981. A New Method for Reporting and Interpreting Textural Composition of Spawning Gravel, Vol. 369. Pacific Northwest Forest and Range Experiment Station, <http://dx.doi.org/10.1029/2010WR009690>.
- McKinney, W., 2010. Data structures for statistical computing in python. In: van der Walt, S., Millman, J. (Eds.), Proceedings of the 9th Python in Science Conference. pp. 51–56.
- Melton, J., 1996. SQL language summary. *ACM Comput. Surv.* 28 (1), 141–143. <http://dx.doi.org/10.1145/234313.234374>.
- Mooij, W.M., Brederveld, R.J., de Klein, J.J., DeAngelis, D.L., Downing, A.S., Faber, M., Gerla, D.J., Hipsey, M.R., Janse, J.H., Janssen, A.B., et al., 2014. Serving many at once: How a database approach can create unity in dynamical ecosystem modelling. *Environ. Model. Softw.* 61, 266–273. <http://dx.doi.org/10.1016/j.envsoft.2014.04.004>.
- Mooneyham, C., Strom, K., 2018. Deposition of suspended clay to open and sand-filled framework gravel beds in a laboratory flume. *Water Resour. Res.* 54 (1), 323–344. <http://dx.doi.org/10.1002/2017WR020748>, URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/2017WR020748>, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/2017WR020748>.

- Morsy, M.M., Goodall, J.L., Castronova, A.M., Dash, P., Merwade, V., Sadler, J.M., Rajib, M.A., Horsburgh, J.S., Tarboton, D.G., 2017. Design of a metadata framework for environmental models with an example hydrologic application in HydroShare. *Environ. Model. Softw.* 93, 13–28. <http://dx.doi.org/10.1016/j.envsoft.2017.02.028>.
- Negreiros, B., Galdos, A.A., Seitz, L., Noack, M., Schwindt, S., Wiprecht, S., Haun, S., 2023a. A multi-parameter approach to quantify riverbed clogging and vertical hyporheic connectivity. *River Res. Appl.* n/a (n/a). <http://dx.doi.org/10.1002/rra.4145>.
- Negreiros, B., Scolari, F., Barros, R., Schwindt, S., 2023b. River analyst. In: *Computer Software*. <http://dx.doi.org/10.17605/OSF.IO/KT53D>, URL: <https://osf.io/kt53d/>.
- Nikora, V.I., Goring, D.G., 1998. ADV measurements of turbulence: Can we improve their interpretation? *J. Hydraul. Eng.* 124 (6), 630–634. [http://dx.doi.org/10.1061/\(ASCE\)0733-9429\(1998\)124:6\(630\)](http://dx.doi.org/10.1061/(ASCE)0733-9429(1998)124:6(630)).
- Noack, M., Schneider, M., Wiprecht, S., 2013. The habitat modelling system CASiMR: a multivariate fuzzy approach and its applications. In: *Ecohydraulics: An Integrated Approach*. Wiley Online Library, pp. 75–91. <http://dx.doi.org/10.1002/9781118526576.ch4>.
- Orghidan, T., 1959. Ein neuer Lebensraum des unterirdischen Wassers: der hyporheische Biotop [A new habitat of subsurface waters: the hyporheic biotope]. *Arch. Hydrobiol.* 55 (3), 392–414.
- Pacheco, C.J.B., Brewer, J.C., Horsburgh, J.S., Caraballo, J., 2021. An open source cyberinfrastructure for collecting, processing, storing and accessing high temporal resolution residential water use data. *Environ. Model. Softw.* 144, 105137. <http://dx.doi.org/10.1016/j.envsoft.2021.105137>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Petersen, W., Bertino, L., Callies, U., Zorita, E., 2001. Process identification by principal component analysis of river water-quality data. *Ecol. Model.* 138 (1–3), 193–213. [http://dx.doi.org/10.1016/S0304-3800\(00\)00402-6](http://dx.doi.org/10.1016/S0304-3800(00)00402-6).
- Peucker-Ehrenbrink, B., 2009. Land2Sea database of river drainage basin sizes, annual water discharges, and suspended sediment fluxes. *Geochem. Geophys. Geosyst.* 10 (6), <http://dx.doi.org/10.1029/2008GC002356>.
- Plana, Q., Alferes, J., Fuks, K., Kraft, T., Maruéjols, T., Torfs, E., Vanrolleghem, P.A., 2019. Towards a water quality database for raw and validated data with emphasis on structured metadata. *Water Qual. Res. J. 54* (1), 1–9. <http://dx.doi.org/10.2166/wqrj.2018.013>.
- Plekhanova, J., 2009. Evaluating web development frameworks: Django, Ruby on Rails and CakePHP. In: *Institute for Business and Information Technology*, Vol. 20. p. 2009.
- Plotly Technologies Inc., 2022a. Dash python library. In: *Computer Software*. URL: <https://dash.plotly.com/>, Accessed: April 14, 2023.
- Plotly Technologies Inc., 2022b. Plotly python library. In: *Computer Software*. URL: <https://plotly.com/python/>, Accessed: April 14, 2023.
- Pokorný, J., 2006. Database architectures: Current trends and their relationships to environmental data management. *Environ. Model. Softw.* 21 (11), 1579–1586. <http://dx.doi.org/10.1016/j.envsoft.2006.05.004>.
- Python Software Foundation, 2023. Python software foundation. <http://python.org/about/>, Accessed: April 25 2023.
- Qiu, W., Ma, T., Wang, Y., Cheng, J., Su, C., Li, J., 2022. Review on status of groundwater database and application prospect in deep-time digital earth plan. *Geosci. Front.* 13 (4), 101383. <http://dx.doi.org/10.1016/j.gsf.2022.101383>.
- Rieger, L., Thomann, M., Joss, A., Gujer, W., Siegrist, H., 2004. Computer-aided monitoring and operation of continuous measuring devices. *Water Sci. Technol.* 50 (11), 31–39. <http://dx.doi.org/10.2166/wst.2004.0668>.
- River Analyst, 2023. River Analyst: User manual. URL: <https://riveranalyst.github.io/usermanual.pdf>, Accessed: June 26, 2023, Website.
- River Analyst, 2023. Using River Analyst for grain size analysis. URL: <https://www.youtube.com/@river-analyst/featured>, Accessed: June 26, 2023, Youtube Video.
- Roegner, G.C., Needoba, J.A., Baptista, A.M., 2011. Coastal upwelling supplies oxygen-depleted water to the Columbia River estuary. *PLoS One* 6 (4), e18672. <http://dx.doi.org/10.1371/journal.pone.0018672>.
- Schälchli, U., 1992. The clogging of coarse gravel river beds by fine sediment. *Hydrobiologia* 235, 189–197. <http://dx.doi.org/10.1007/BF00026211>.
- Schälchli, U., 1993. In: Vischer, D. (Ed.), *Die Kolmation Von Fließgewässersohlen: Prozesse Und Berechnungsgrundlagen [Riverbed Clogging: Processes and Calculation Methods]*. Mitteilung Nr. 124 der Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie (VAW) an der Eidgenössischen Technischen Hochschule Zürich, Zürich, Switzerland, URL: <https://ethz.ch/content/dam/ethz/special-interest/baug/vaw/vaw-dam/documents/das-institut/mitteilungen/1990-1999/124.pdf>.
- Schleiss, A.J., 2017. Better water infrastructures for a better world - The important role of water associations. *HydroLink* 3, 86–87.
- Schwindt, S., Pasternack, G.B., Bratovich, P.M., Rabone, G., Simodines, D., 2019. Hydro-morphological parameters generate lifespan maps for stream restoration management. *J. Environ. Manag.* 232, 475–489. <http://dx.doi.org/10.1016/j.jenvman.2018.11.010>, URL: <http://www.sciencedirect.com/science/article/pii/S0301479718312751>.
- Seitz, L., 2020. Development of New Methods to Apply a Multiparameter Approach - a First Step Towards the Determination of Colmation (Ph.D. thesis). University of Stuttgart, <http://dx.doi.org/10.18419/opus-11249>, Doctoral Dissertation.
- Severance, C., 2009. *Using Google App Engine: Building Web Applications*. O'Reilly Media, Inc., ISBN: 978-0-596-80069-7.
- Severance, C., 2014. Elizabeth Fong: Creating the SQL database standards. *Computer* 47 (8), 7–8. <http://dx.doi.org/10.1109/MC.2014.209>.
- Spiegel, M.R., 1990. *Schaum's Outline of Theory and Problems of Statistics*. McGraw-Hill Inc., ISBN: 0-07-060234-4.
- Stalnaker, C., Lamb, B.L., Henriksen, J., Bovee, K., Bartholow, J., 1995. In: Opler, P.A., Rockwell, E.D., Zuboy, J.R., Cox, J.D., Harris, D.K. (Eds.), *The Instream Flow Incremental Methodology: A Primer for IFIM*. In: Number 29 in Biological Report, National Biological Service, U.S. Department of the Interior, Washington, D.C., USA, p. 47. URL: <https://apps.dtic.mil/sti/citations/ADA322762>.
- Swain, N.R., Christensen, S.D., Snow, A.D., Dolder, H., Espinoza-Dávalos, G., Goharian, E., Jones, N.L., Nelson, E.J., Ames, D.P., Burian, S.J., 2016. A new open source platform for lowering the barrier for environmental web app development. *Environ. Model. Softw.* 85, 11–26. <http://dx.doi.org/10.1016/j.envsoft.2016.08.003>.
- Swain, N.R., Latu, K., Christensen, S.D., Jones, N.L., Nelson, E.J., Ames, D.P., Williams, G.P., 2015. A review of open source software solutions for developing water resources web applications. *Environ. Model. Softw.* 67, 108–117. <http://dx.doi.org/10.1016/j.envsoft.2015.01.014>.
- Tonina, D., Buffington, J.M., 2009. Hyporheic exchange in mountain rivers I: Mechanics and environmental effects. *Geogr. Compass* 3 (3), 1063–1086. <http://dx.doi.org/10.1111/j.1749-8198.2009.00226.x>.
- Vitolo, C., Elkhatib, Y., Reusser, D., Macleod, C.J.A., Buytaert, W., 2015. Web technologies for environmental Big Data. *Environ. Model. Softw.* 63, 185–198. <http://dx.doi.org/10.1016/j.envsoft.2014.10.007>.
- Voermans, J.J., Ghisalberti, M., Ivey, G.N., 2018. A model for mass transport across the sediment-water interface. *Water Resour. Res.* 54 (4), 2799–2812. <http://dx.doi.org/10.1002/2017WR022418>, URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/2017WR022418>, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/2017WR022418>.
- Webb, J.A., Miller, K.A., Stewardson, M.J., de Little, S.C., Nichols, S.J., Wealands, S.R., 2015. An online database and desktop assessment software to simplify systematic reviews in environmental science. *Environ. Model. Softw.* 64, 72–79. <http://dx.doi.org/10.1016/j.envsoft.2014.11.011>.
- Wentworth, C.K., 1922. A scale of grade and class terms for clastic sediments. *J. Geol.* 30 (5), 377–392. <http://dx.doi.org/10.1086/622910>.
- Wharton, G., Mohajeri, S.H., Righetti, M., 2017. The pernicious problem of streambed colmation: A multi-disciplinary reflection on the mechanisms, causes, impacts, and management challenges. *Wiley Interdiscip. Rev.: Water* 4 (5), e1231. <http://dx.doi.org/10.1002/wat2.1231>.
- Wohl, E.E., 2000. *Mountain Rivers*. In: Number 14 in Water Resources Monograph, American Geophysical Union, Washington, DC, USA, URL: <https://agupubs.onlinelibrary.wiley.com/doi/book/10.1029/WM014>.
- Wohl, E., 2019. Forgotten legacies: Understanding and mitigating historical human alterations of river corridors. *Water Resour. Res.* 55 (7), 5181–5201. <http://dx.doi.org/10.1029/2018WR024433>, URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR024433>.
- Wohl, E., Lane, S.N., Wilcox, A.C., 2015. The science and practice of river restoration. *Water Resour. Res.* 51 (8), 5974–5997. <http://dx.doi.org/10.1002/2014WR016874>.
- Wojda, P., Brouyere, S., Derouane, J., Dassargues, A., 2010. HydroCube: An entity-relationship hydrogeological data model. *Hydrogeol. J.* 18 (8), 1953–1962. <http://dx.doi.org/10.1007/s10040-010-0653-6>.
- Wooster, J.K., Dusterhoff, S.R., Cui, Y., Sklar, L.S., Dietrich, W.E., Malko, M., 2008. Sediment supply and relative size distribution effects on fine sediment infiltration into immobile gravels. *Water Resour. Res.* 44 (3), <http://dx.doi.org/10.1029/2006WR005815>.
- Wu, W., Wang, S.S.Y., 2006. Formulas for sediment porosity and settling velocity. *J. Hydraul. Eng.* 132 (8), 858–862. [http://dx.doi.org/10.1061/\(ASCE\)0733-9429\(2006\)132:8\(858\)](http://dx.doi.org/10.1061/(ASCE)0733-9429(2006)132:8(858)).
- Wuijts, S., Van Rijswijk, H.F., Driessen, P.P., Runhaar, H.A., 2023. Moving forward to achieve the ambitions of the European Water Framework Directive: Lessons learned from the Netherlands. *J. Environ. Manag.* 333, 117424. <http://dx.doi.org/10.1016/j.jenvman.2023.117424>, URL: <https://www.sciencedirect.com/science/article/pii/S0301479723002128>.
- Wyrick, J.R., Pasternack, G.B., 2014. Geospatial organization of fluvial landforms in a gravel-cobble river: Beyond the riffle-pool couplet. *Geomorphology* 213 (Supplement C), 48–65. <http://dx.doi.org/10.1016/j.geomorph.2013.12.040>, URL: <http://www.sciencedirect.com/science/article/pii/S0169555X14000099>.
- Yalin, M.S., 1971. *Theory of Hydraulic Models*. In: *Civil Engineering Hydraulics*, vol. 266, Macmillan, London.
- Yang, C.T., 1996. *Sediment Transport: Theory and Practice*. In: McGraw-Hill Series in Water Resources and Environmental Engineering, McGraw-Hill Companies, New York, NY, USA.
- Yi, H., Idaszak, R., Stealey, M., Calloway, C., Couch, A.L., Tarboton, D.G., 2018. Advancing distributed data management for the HydroShare hydrologic information system. *Environ. Model. Softw.* 102, 233–240. <http://dx.doi.org/10.1016/j.envsoft.2017.12.008>, URL: <https://www.sciencedirect.com/science/article/pii/S1364815217305856>.