Check for updates

Tocap: a web tool for ad-hoc campaign planning in terrestrial hydrology

Erik Nixdorf [a,*], Daniel Eggert [b], Peter Morstein[b], Thomas Kalbacher[a] and Doris Dransch[b]

[a] Helmholtz Centre for Environmental Research, Department of Environmental Informatics, Permoserstr.15, 04318, Leipzig, Germany
[b] GFZ German Research Centre for Geosciences, Telegrafenberg, 14473, Potsdam, Germany
*Corresponding author. E-mail: erik.nixdorf@ufz.de

EN, 0000-0001-5569-4597; DE, 0000-0003-0251-4390

ABSTRACT

Planning an event-based monitoring campaign on the regional scale is challenging, e.g. the timing and location of monitoring visits can dramatically impact monitoring efficacy and depends on the optimal environmental conditions required by the measurement parameters and the overarching monitoring goal. Therefore we developed a generic campaign planning approach utilizing interactive visualization methods and implemented this approach into the component-based web tool called Tocap: Tool for Campaign Planning. As a case study, we determine the most suitable time and location for event-driven, ad-hoc monitoring in hydrology using soil moisture measurements as our target variable. Our approach supports: (1) data acquisition from various digital data sources, (2) identification of the most suitable locations for measurements, (3) identification of the most suitable time for measurements at the selected locations, and (4) planning an optimized monitoring route.

Key words: geospatial visualization, monitoring campaign planning, soil moisture, visual-analytics, web-based interaction

HIGHLIGHTS

- Development of an open-source web-based environment to explore and process time-variant datasets.
- Providing target variable computation by multiple linear regression and filter functions.
- Determining monitoring locations by visualizing time series of hydro-meteorological forecast data.
- Computation and export of a monitoring route using OpenStreetMap data.
- Demonstration of the framework for planning a soil moisture monitoring campaign.

1. INTRODUCTION

A deeper understanding of the Earth system as a whole and its interacting sub-systems depends on both, accurate mathematical approximations of the physical processes and the availability of environmental data across time and spatial scales. Advanced numerical simulations (de Graaf *et al.* 2017; Zaherpour *et al.* 2018) and satellite-based remote sensing (Butler 2014; Simmons *et al.* 2016; Chen & Wang 2018) in conjunction with sophisticated algorithms such as machine learning tools (e.g. Mao *et al.* 2019; Sarafanov *et al.* 2020) can provide 4D environmental datasets with unprecedented resolution, coverage, and accuracy.

However, local and mesoscale monitoring continues to be the backbone in many disciplines such as population ecology (Rockwood 2015; Smith *et al.* 2017) and hydrology (Li *et al.* 2020). In general, a monitoring design comprises the number of locations and the monitoring frequency at each location. Both depend on factors such as the monitoring purpose, the available equipment, budgetary and quality requirements, the spatial-temporal characteristics of the study site, and the heterogeneity of parameters impacting the target variable (For an overview see e.g. Harmancioglu *et al.* 1999; De Gruijter *et al.* 2006; Brus & Knotters 2008). Furthermore, existing continuous monitoring networks are designed for objectives different from research interest, and establishing new stations is associated with high financial burdens and needs to be well prepared (Zhou *et al.* 2013; Bhat *et al.* 2015).

On the local scale, monitoring a single location may be sufficient, if spatial characteristics of relevant parameters can be assumed to be homogeneous. If not the case, environmental parameters can be monitored on a grid taking observations

at regular intervals across the study site (e.g. Joseph & Possingham 2008). However, on larger scales, monitoring at a grid with a resolution large enough to account for the scale of spatial heterogeneities is often limited by the availability of human and technical resources (Huston 1999). As an alternative, random, transect based as well as stratified monitoring strategies can be applied whereas the latter tries to capture the range of variability across parameters of interest which need to be available with sufficient spatial-temporal coverage and resolution (Li et al. 2017; Selmoni et al. 2020).

The selection of suitable monitoring frequency is influenced by technological constraints such as the incomplete automation of in-situ parameter estimation procedures as well as temporal characteristics of time-variant environmental factors such as habitat quality, temperature, and precipitation (Bonneau et al. 2018). This is even more important for ad-hoc monitoring campaigns of distinct dynamic events such as heavy precipitation events, floods, and droughts which require an adequate design that can evolve throughout the event and depending on the available data.

This calls for an integrated campaign planning approach that addresses the following main functional requirements:

1. Identify the most suitable location based on the state of environmental parameters, available historical measurement records as well as the target parameter and chosen measurement technique.
2. Identify the most suitable time to initialize the campaign depending on projected environmental conditions and synchronize with complementary measurements.
3. Optimize the route to reach all measurement areas within the campaign schedule.
4. Reprocess the workflow based on changing environmental conditions and retrieved datasets during an ongoing campaign.
5. Integrate monitoring results and environmental database updates into the data exploration framework without the need for any programming skills.

Most of the requirements above could be addressed with computational or visual methods. Since we are aiming at a generic solution, being as independent from the application domain and target parameter as possible, most existing computational methods and models analyzing spatio-temporal data require a lot of tailoring towards individual settings (Atluri et al. 2018). Therefore our approach will heavily rely on the expertise of the user in his research domain and only apply basic computational methods in the background to support the expert's assessment.

One helpful set of tools to address the shown requirements are classical Geographic Information System (GIS) tools, due to the spatial nature of the usually involved data. GIS tools work great with spatial data but often lack built-in possibilities for the analysis of spatio-temporal data or time series. For the visualization of gridded spatio-temporal data, specialized visualization software exists such as Panoply data viewer (National Aeronautics & Space Administration 2021) which is cross-platform and capable to produce either two-dimensional plots of geographically referenced data on a background map or time series curves of zonal averages as well as to export the output as images or comma-separated values. Although being great to visualize complex geospatial data stored in arrays, these viewers do not provide the functionality to manipulate and process the visualized data sets. On the other hand, data visualization and analytic platforms like Tableau (Hoelscher & Mortimer 2018) and Grafana (Grafana Labs 2018) are great to analyze big table-like data sets but lack built-in support for geospatial raster data. More general, versatile programming languages such as Python and R provide pre-packaged sets of functions to give the user maximum flexibility to process and analyze data sets in various formats, however, a certain level of programming skills is required.

Depending on the specific objectives, there have been previous initiatives and approaches for scientific data visualization in environmental science (for an overview see Lin et al. 2013; Chen et al. 2015). Exemplarily, Rink et al. (2013, 2020) developed 4D virtual geographic environment systems focusing on the animation of complex environmental datasets in a predefined static geographical setting employing the Unity engine. In contrast to tailor-made software implementations, Hunter et al. (2016) and Miao et al. (2017) present web-based solutions for their specific domains. They demonstrate the major advantages of web-based implementations, such as their support for simple accessibility via web browsers, cross-platform capability, and compatibility with common data formats and web services. Xu et al. (2019) developed a web-based geospatial platform for sedimentation at culverts in Iowa including mapping and analyzing functionality but focussing on the particular use-case.

However, approaches to address the specific requirements of event-driven analysis are less frequently developed. Goharian & Burian (2018) and Nikoloudi et al. (2021) developed interactive visual-analytical near-real-time decision-making tools for urban water management aiming to optimize predefined target functions (e.g operation costs) by manual interaction. In contrast, the web-based information system for landslides developed by Pumo et al. (2016), which integrates data from deployed sensors and rainfall forecast data, provides its product, a warning map, by an automatized algorithm without user interaction.

Zhai *et al.* (2016) and Ahamed & Bolten (2017) present a solution to visualize active floods and hydrological disasters using web service and sensor-web technologies but provide limited visual-analytical functionality.

On the other hand, web-distributed integrated environmental simulation and modeling systems allow linking, collaborative executing, and sharing environmental models but familiarisation with these systems and interpreting the results of heterogeneous models is demanding (Chen *et al.* 2020; Bayer *et al.* 2021; Qiao *et al.* 2021).

The dependence of required datasets on both, the type of event and the target parameter of the campaign raises significant data access and management challenges. Although standardization of data and metadata is progressing in environmental science led by initiatives such as the Open Geospatial Consortium and the GO FAIR Initiative, datasets from different disciplines and organizations are stored in disparate databases in heterogeneous formats and coordinate systems making a generic database access and integration approach difficult. Once an event occurs and develops, the challenge is to quickly evolve monitoring strategies in response based on existing heterogeneous data sets.

The proposed approach comprises a visual-analytical tool to provide an interactive web-based environment for a suitable representation and analysis of geographical datasets in the context of ad-hoc campaign planning. The remainder of the article is organized as follows. Section 2 describes our general approach, while section 3 outlines the web-based implementation of our approach, including the software architecture, the dataset and database integration, as well as the visual user interface. Section 4 provides an overview of potential application areas as well as a step-by-step application example from the field of event-driven hydrological monitoring to demonstrate the benefits of the proposed approach. Concluding remarks and plans for future extensions of Tocap are given in Section 5.

2. CONCEPT & METHODS

We propose a visual-analytical approach to address the defined requirements outlined in section 1. A UML (Object Management Group 2007) activity diagram (Figure 1) depicts the main steps and tasks of our approach. First, the relevant spatio-temporal data is acquired. Second, based on the imported data, the most suitable location for measurements is identified. Subsequently, the most suitable time for conducting the measurements in that area is identified. If viable a route can be planned for the identified measurement locations. The final output of the approach are measurement locations that satisfy the measurement preconditions, a time-optimized route to reach the proposed measurement locations, as well as the estimated driving time. Although the step-by-step order of the workflow components implies a closed-loop for simplicity's sake, it should be possible to switch between the individual processes at any time. Aiming to facilitate the planning and conduction of terrestrial monitoring campaigns, the platform and device-independent approach of Tocap gives both scientists and field technicians the flexibility to plan a campaign at the desktop computer first and to refine it later in the field using mobile devices. Using a visual-analytical approach with an app-like frontend visualization in a web browser, required user skills are reduced to normal IT competence that can be trained quickly. The following subsections will discuss each step and task, and provide details on how the individual requirements are addressed.

2.1. Data acquisition and visualisation

Our technical concept considers the integration of data from three major data sources into Tocap (Figure 2): data from the user's hard drive; data from standardized services such as Web Mapping Service (WMS), Web Coverage Service (WCS), and Web Feature Service (WFS) technologies; as well as data sent by the backend module of Tocap (see section 3.3), which
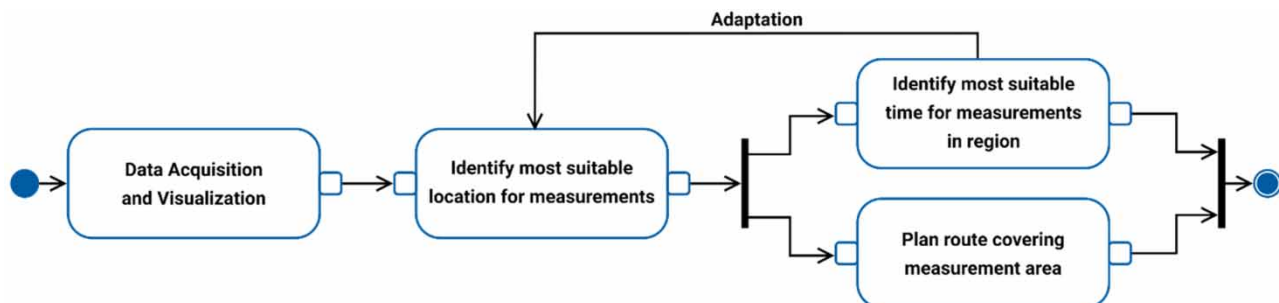

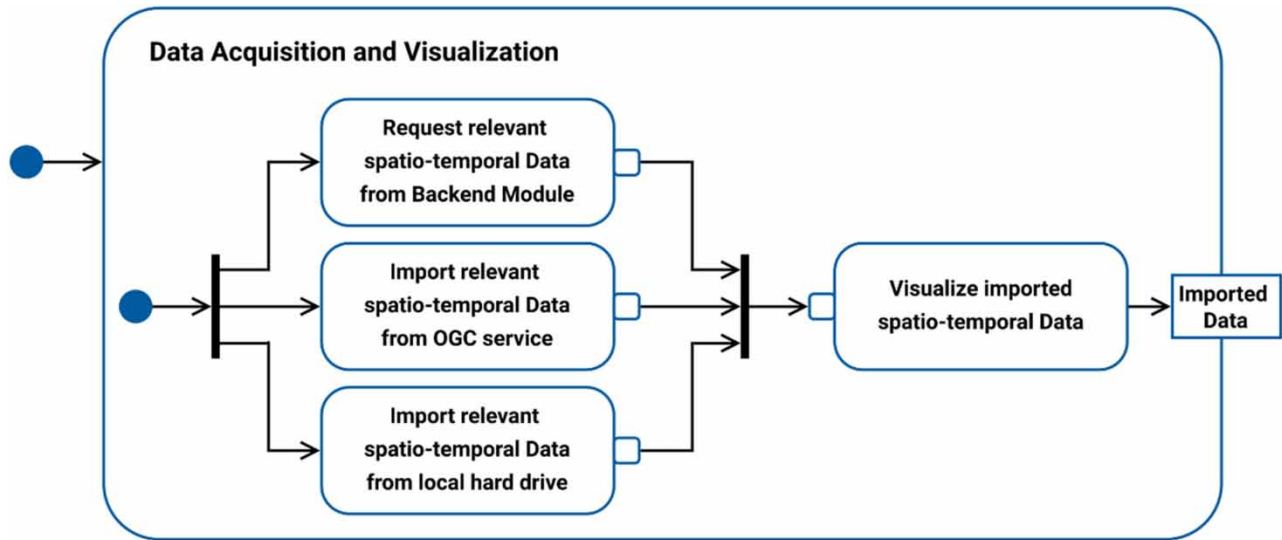**Figure 1** | UML activity diagram of the proposed approach.

Figure 2 | UML activity diagram of the 'Data Acquisition and Visualisation' step.

provides tailor-made solutions for data retrieval and processing from common environmental databases. The imported spatio-temporal data are immediately visualized to gain a first overview. This first step addresses, in particular, the requirement (e) 'Integrate monitoring results and environmental database updates into the data exploration framework without the need for any programming skills'. Its implementation is discussed in detail in section 3.3.

2.2. Identification of the most suitable location for measurements

Based on the acquired data, the most suitable location for the intended measurements can be identified within this step; it directly addresses requirement (a) 'Identify the most suitable location based on the state of environmental parameters, available historical measurement records as well as the target parameter and chosen measurement technique'. A generic way to support this shall be through the visual exploration of the spatial distribution of relevant parameters and the judgment by experts. In many cases, there will be multiple parameters with differing importance according to the case of application. Visually deriving a region where all parameters are optimal can be challenging, especially with multiple relevant parameters. We propose to identify the most suitable location through an assessment of the spatial distribution of individual parameters relevant for the intended monitoring campaign and as well their combinations.

This task involves three phases (Figure 3): First, cells containing relevant values can be selected from each dataset by applying a filter, known in image processing as 'binary threshold filter', over the entire spatial-temporal extension of the data layer. The result is treated as binary raster, meaning that a bit mask is derived where all pixels satisfying the constraints of the binary threshold filter are set to 1 and all non-satisfying pixels are set to 0. The threshold itself can be determined manually or by applying different statistical functions (the implemented options are described in section 3.2.2).

Second, the filtered cells of the selected data layers are merged into a single data product using linear superposition with user-defined weights for each data layer. Since the input data might use different resolutions and projections, it is reprojected and re-rasterized into a common raster with a unified projection (e.g. WGS 84/Pseudo-Mercator) and resolution. The spatial resolution of the common raster equals the highest available resolution of all selected input data layers. Once the data is reprocessed into a common grid, the individual grid cells/pixels are merged by treating each involved layer as a binary raster. Along with user-defined weights, the generated bit masks are merged into a single result. Assuming that all weights sum up to one, the resulting raster will contain a quasi-discrete set of values in the range from zero to one, with one indicating that this pixel satisfies all constraints and zero indicating that none of the constraints is satisfied. All values in between hint at locations where some constraints are satisfied but others are not. For time-dependent data, the linear superposition function uses the currently selected time step of each data layer as input data.

Finally, the resulting gridded dataset represents the combined distribution of all parameters considering their respective relevance via user-defined weights. Visualizing this distribution highlights areas with a high convergence of relevant
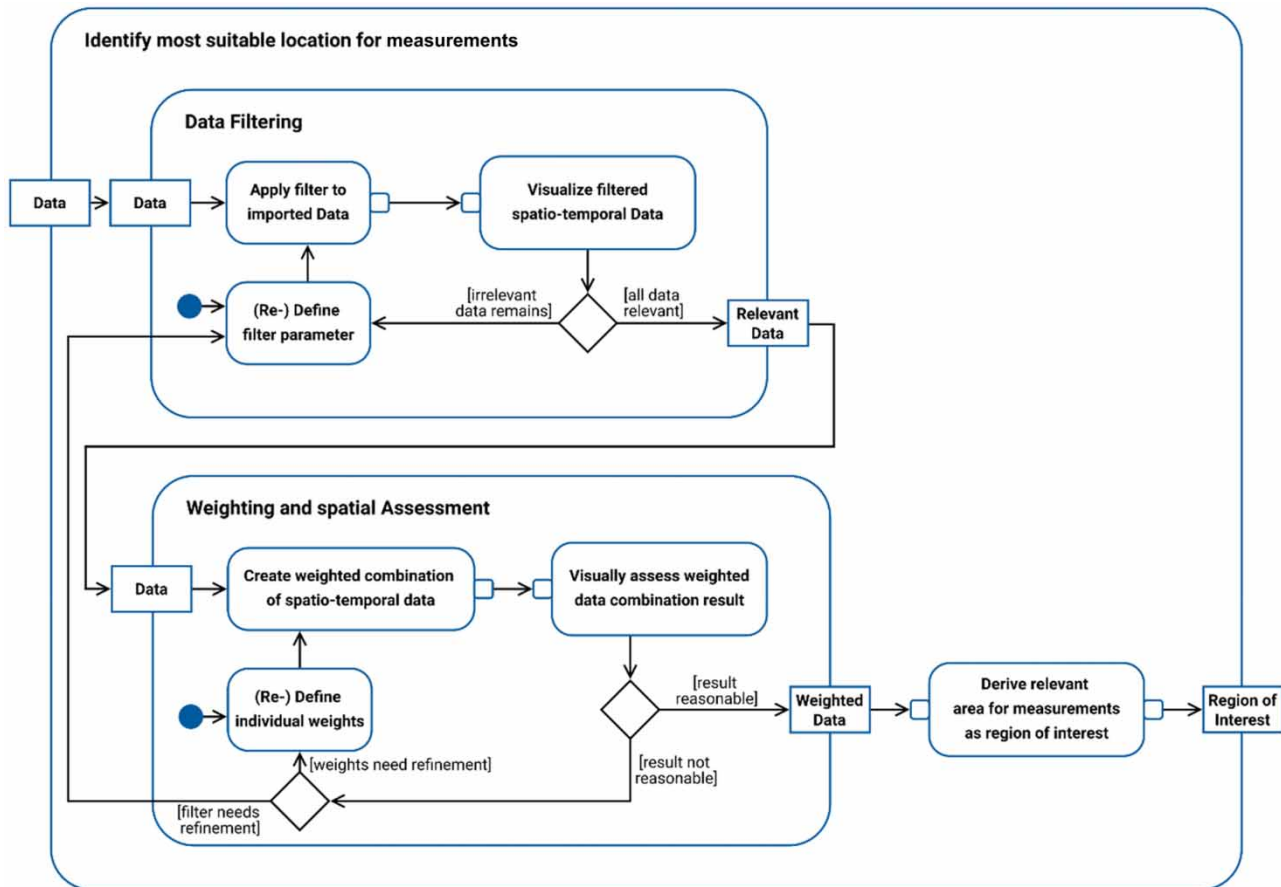
Figure 3 | UML activity diagram of the 'Identification of the most suitable location for measurements' step. The top-level activity takes (raster) data as input and generates the relevant area for measurements as a region of interest as output. To achieve this, the input data is processed via three sub-activities: 'Data filtering', 'Weighting and spatial assessment' as well as 'Derive relevant area for measurements'.

parameters and eventually supports the user in identifying the most suitable location for conducting the measurements. The filtering and weighting tasks both include visual assessment for their respective outputs to assert that the generated results are reasonable. In cases where for instance contradicting or misleading outputs have been generated the user can interactively refine the filter parameters and layer weights.

2.3. Identification of the most suitable time for measurements in a region

After selecting the relevant regions for measurements, the next task is to identify the most suitable time for conducting measurements (Figure 4), which addresses requirement (b) 'Identify the most suitable time to initialize the campaign depending on projected environmental conditions and synchronized with complementary measurements'. Hereby, a meaningful time frame for the upcoming campaign can be defined by the user when requesting the distributed available forecast data from the data catalog. The selection of the time frame depends on the forecast range of the dataset but also on the monitoring objective, organizational constraints as well as the dynamics of the event which induces the monitoring campaign. The user selects the relevant parameters from the available datasets. Subsequently, the corresponding forecast data for the selected region and time frame is acquired and subsequently visualized. Based on the time series visualization of chosen parameters, the user has the ability to determine a suitable time frame for the upcoming measurements.

2.4. Route planning for the measurement area

Following the identification of relevant measurement locations (see section 2.2), Tocap allows the user to request a route to reach these areas (Figure 5). Since the route calculation tries to determine the best route to reach all measurement locations in succession, the user has to provide a start location. Once the route is created the user can either visually explore or export it
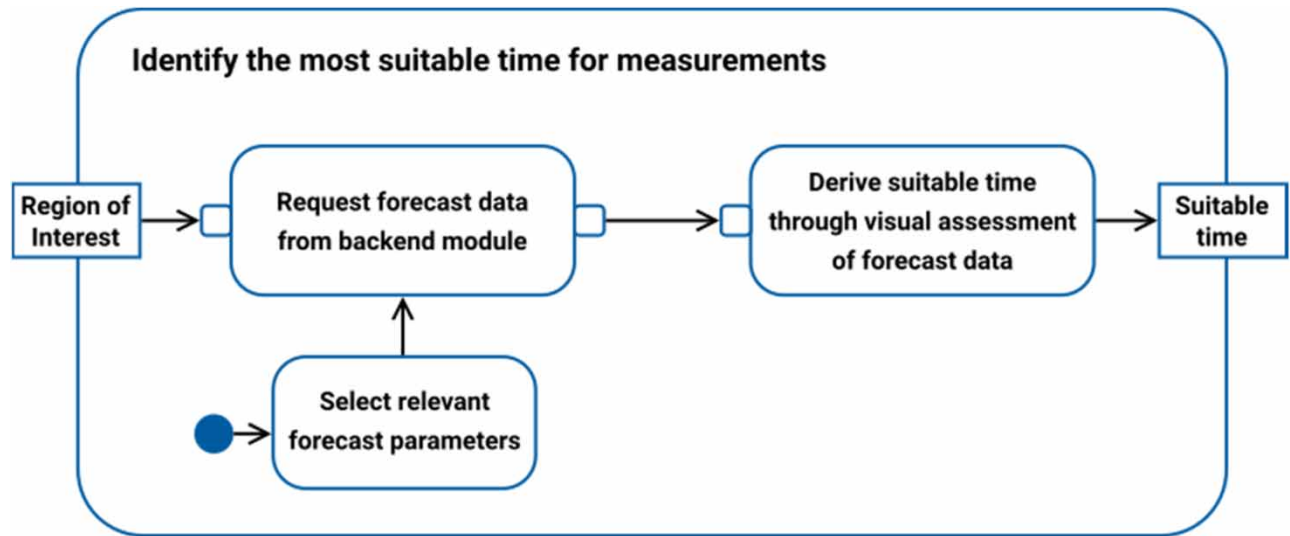
Figure 4 | UML activity diagram for the 'Identification of the most suitable time for measurements' step.
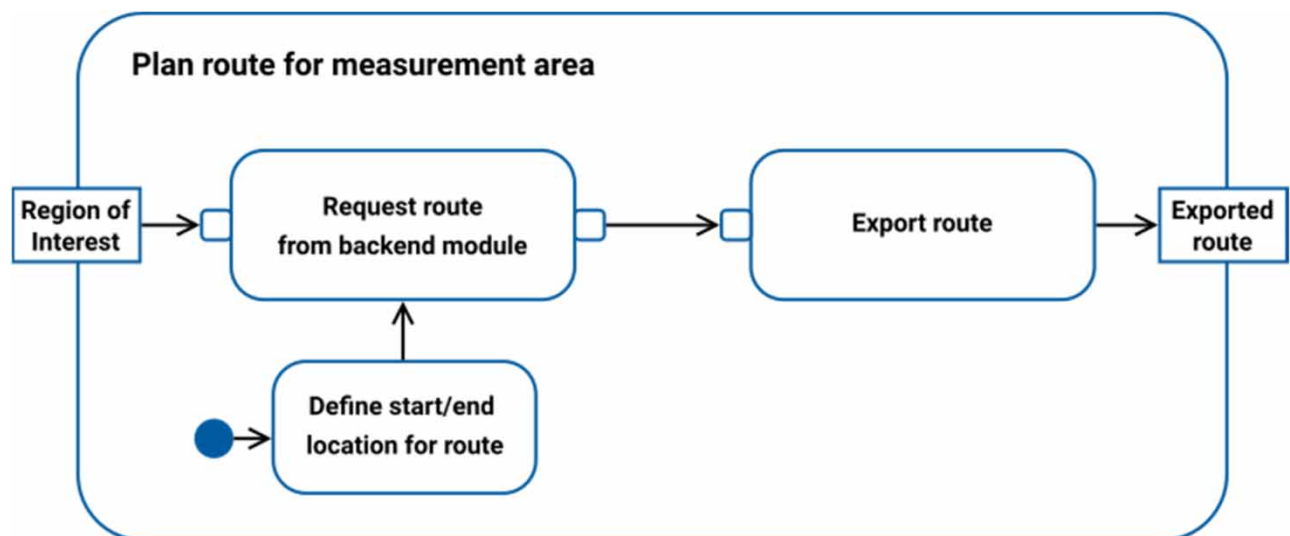

Figure 5 | UML activity diagram for the 'Route planning for the measurement area' step.

for further use. This addresses requirement (c) 'Optimize the route to reach all measurement areas within the campaign schedule'. Since route planning only relies on spatial data, the execution is independent of the identification of the most suitable time for measurements. For more information on the route calculation, see section 3.3.

3. IMPLEMENTATION

3.1. Application architecture

The campaign planning tool Tocap is implemented (Figure 6) based on the Data Analytics Software Framework (DASF) (Eggert & Dransch 2021). DASF allows building applications with a strong component-based design that operate in highly distributed environments, like multiple collaborating research centers. It is available under the Apache-2.0 license at https://git.geomar.de/digital-earth/dasf. A DASF-based application consists of three main components: a web application frontend to visualize data, a backend module to acquire and process the data, and a middleware connecting both. Tocap's
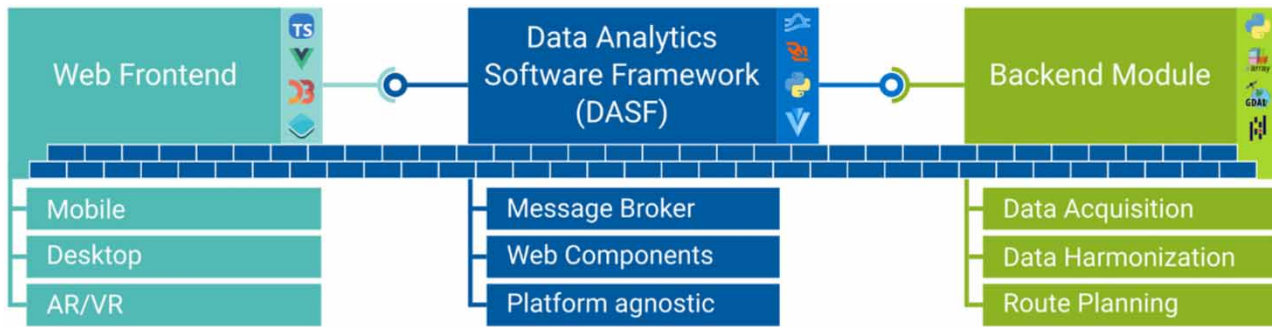
Figure 6 | Application architecture overview. The DASF serves as a foundation for the entire application.

backend (Eggert & Nixdorf 2022) is made freely available under 3-Clause BSD License in a public repository (https://git.geomar.de/digital-earth/de-smart-monitoring-backend-module) to enable wide usage, collaborative software development as well as to provide the user documentation of the problems tackled during the development process.

Tocap's web-frontend (Eggert & Nixdorf 2022) implementation uses the components provided by the DASF framework which are developed in Typescript and VueJs (Eggert *et al.* 2021). It is freely available under Apache-2.0 License in a public repository (https://git.geomar.de/digital-earth/flood-event-explorer/fee-smart-monitoring-workflow).

The framework's middle-ware is built around a publish-subscribe system with a central message broker instance, connecting all separated components. We use Apache Pulsar (https://pulsar.apache.org/) as the message broker implementation.

The backend module provides functionalities such as the retrieval and processing of datasets from external environmental data sources and the application of statistical functions and geometric operations on user-selected datasets.

3.2. Frontend

The frontend represents an integral and important component for all interactive visualization approaches. We decided to implement our approach as a web-based tool and platform-agnostic approach to lower the technological and overall computer skill requirements on the user side.

As a technological basis, we used the web application template and components provided by the DASF framework. Since all steps of the approach involve the visual assessment of spatio-temporal data a map view based on OpenLayers (OpenLayers 2021) is used as the main visual component. A collapsible interaction panel is used to provide the needed user functions (Figure 7). Furthermore, additional time-series data is visualized as times series charts based on D3 (Teller 2013).

3.2.1. Frontend for implementing 'data acquisition and overview'

The frontend provides access to local data via drag and drop onto the map component to support the data acquisition step (see section 2.1). Vector data needs to be provided in GeoJSON format describing polygon geometries and grid-based input data as NETCDF3 files. Aside from drag-and-drop capabilities, external databases can be accessed twofold. First, standard protocols developed by the Open Geospatial Consortium (OGC) (Reed 2011) for serving georeferenced maps (WMS), grids (WCS), and vector features (WFS) are supported. In this context, two open-access WebGIS projects (Haas *et al.* 2016) have been set up using the geodata infrastructure of the Alfred Wegener Institut Bremerhaven. Focusing on relevant datasets for applications in terrestrial hydrology, datasets stored in a PostgreSQL database include raster and vector products from state authorities, previous monitoring campaigns, and results of numerical models as well as environmental datasets that are freely available via online data repositories. All data is published as WebGIS services relying on OGC-standardized WMS and WFS technologies and accessible at https://maps.awi.de.

Apart from data access via standardized interfaces, the utilization of the backend module provides the flexibility to integrate a large number of additional data formats and data sources (see section 3.3). Selected datasets from the 'Layer from Catalog' panel are retrieved from the relevant server and homogenized automatically before being exported to the frontend for data visualization and exploration (Figure 7). To provide a first overview all imported data sets are rendered as map layers on top of a generic base map.
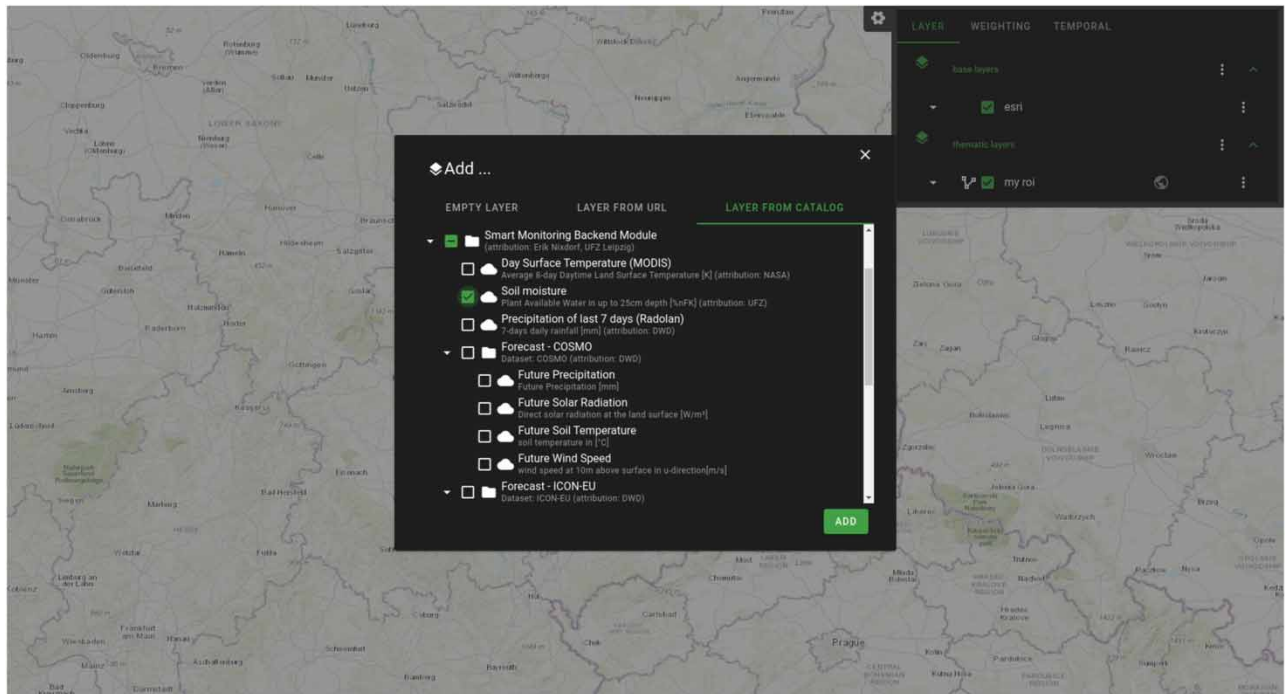
Figure 7 | The user interface of the web front-end contains the main map component with collapsible interaction panel and data acquisition component.

3.2.2. Frontend for implementing 'identification of the most suitable location for measurements'

The identification of the most suitable location demands certain filtering functionalities for each data layer as well as functionality to create a weighted combination of multiple input parameters (see section 2.2). The used framework provides a basic filtering mechanism (Figure 8 top). Filter thresholds can either be set manually utilizing a slider or derived from the percentile statistics of the parameter distribution. For time-dependent spatial data, a checkbox is used to define whether the statistic is derived from the parameter distribution of the selected time step only. Using a slider spares the manual input, which guarantees an error-free input. Additional spatio-temporal subsetting functions are added as a layer plugin exposed via the layer context menu. These functions are provided by the backend module, for details see the following section 3.3.

Since defining the weight for each layer in the individual layer components lacks the needed global overview, we provide an additional tab component (Figure 8 bottom). In this component, all visible layers are listed, while the user can define individual weights for each layer via an interactive slider component.

Following the fine-tuning of the weights, the weighted data merging is initialized via the corresponding button. The generated result is visualized as a heatmap on the base map. For a comprehensive assessment of the visualized result, the user can apply default pan & zoom operations, change the used color scale for the heatmap and even use the initially discussed filter functions. Finally, the result is compared with the original imported data layers by toggling the display of each layer. If the assessed result yields major shortcomings, the user can either refine the filter thresholds or change the weight distribution. Once the result is deemed acceptable, the area comprising all suitable measurement locations is visualized. For further processing, an empty layer can be added via the layer group context menu and the edit & draw feature can be used to create a polygon representing the derived region of interest.

3.2.3. Frontend for implementing 'identification of the most suitable time for measurements'

After promoting the layer containing the drawn polygon to the application-wide region of interest via the layer context menu, the temporal forecast data can be requested within the 'Temporal' tab. The user can select forecast parameters relevant for the intended measurement campaign which are provided by the backend module. Details on the provided data sources are given in section 3.3. The requested parameter, e.g. precipitation, as well as the visualization type (bar, line, area), can be selected from additional dropdown menus. Once the requested data is received, it is displayed in an interactive time series chart
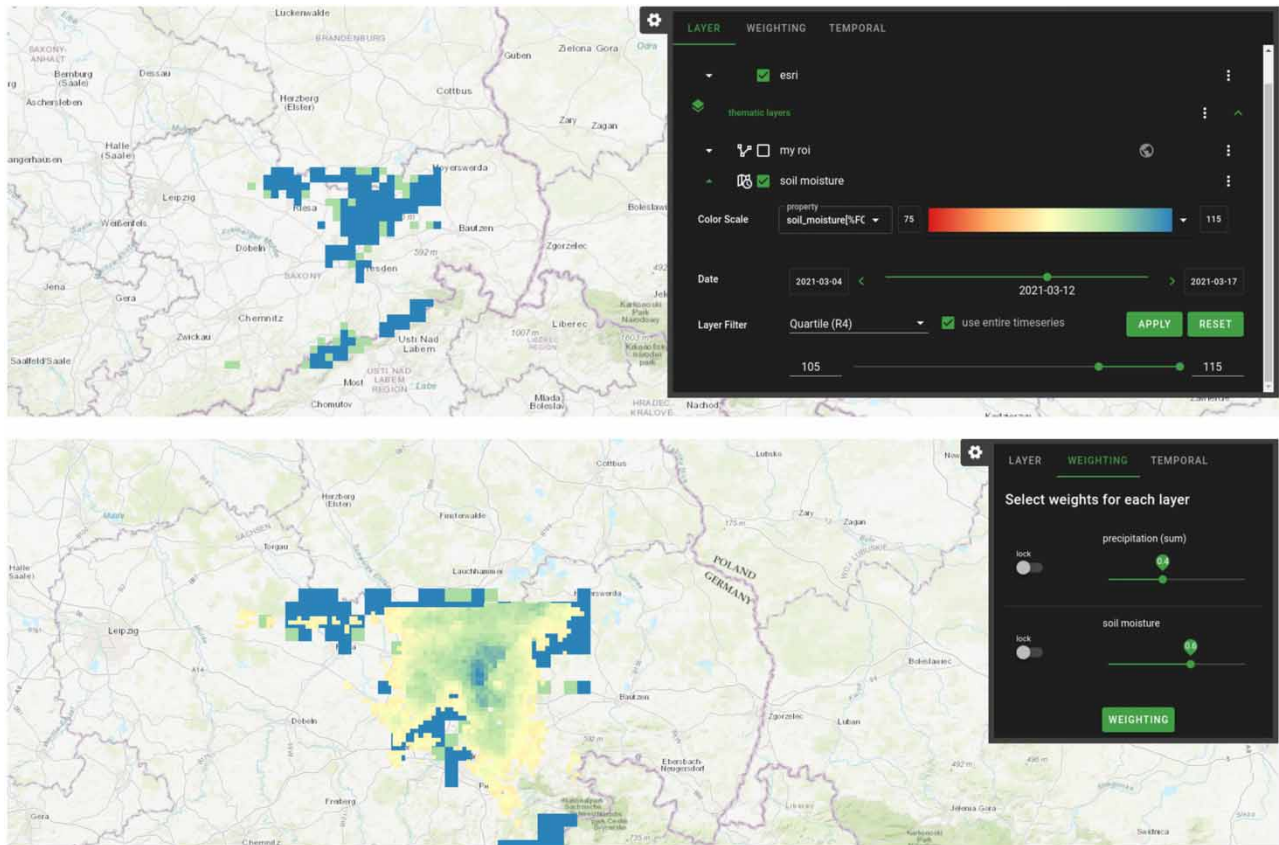
Figure 8 | Filter options via the layer menu (top) and weight distribution panel (bottom).

([Figure 9](#)). Forecast data is explored and assessed via the interactive chart, e.g. by zooming in and out on the temporal axis to derive the most suitable time frame for the intended measurement campaign.

3.2.4. Frontend for implementing 'plan route for the measurement area'

The route planning is based on an input raster layer, e.g. generated by the implemented weighting algorithm. Since the route planning is bound to a specific map layer, we decided to put the option to trigger the planning into the layer context menu. Next to the raster input layer, a starting point is needed, which can be entered in three different ways: First, the geographical coordinates can be manually typed in. Second, the device location can be requested via the web browser localization API. The third option lets the user click a point on the map to define the starting point. After receiving the calculated measurement route, it is displayed as a map layer and can be explored by the user. Finally, the user can download the route via the layers context menu.

3.3. Backend

The backend of Tocap comprises different modules embedded in a framework to address data layer and data logic operations. The modules of the backend are written platform-independent using Python 3.x scripting language ([Van Rossum & Drake 1995](#)). Python language is used due to its clear and readable syntax, its flexible data structure, the extensive standard library, and the many third-party libraries that are freely usable and distributable. Furthermore, its implementation in many GIS software packages and the wide application in Spatial Data Science provides high usability and simple extendibility.

The backend module uses a Python wrapper ([Eggert *et al.* 2021](#)) to generate results that serve as the content for the frontend visualization. In its current implementation, it consists of 8 sub-modules that are partly coupled with each other and with third-party python packages ([Figure 10](#)):

- The *downloader.py* module comprises all functions for the initialization of a class for data transfer between the communication module and the subordinate python modules addressing the communication with external databases. It determines
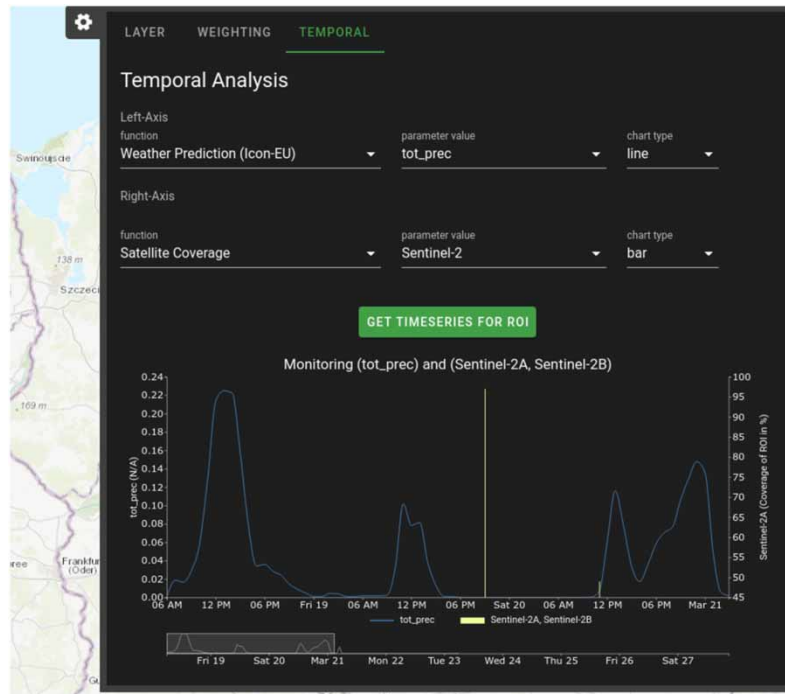
Figure 9 | Temporal data assessment panel enabling time-series data exploration via an interactive chart based on D3 library.

the spatial and temporal boundaries of the requested datasets by analyzing the user-defined region-of-interest (ROI) and date, respectively.

- The *satellites.py* module draws a rectangular boundary box around the ROI and requests information on upcoming satellite overpasses for more than 20 environmental earth observation satellites. Both, the swath width of different sensors deployed on the satellites as well as partial overpasses are considered in the calculation. The Python module creates a GeoJSON file containing information on partial and complete overpasses and sends it back to the frontend for visualization.
- The *rasterprocessing.py* module offers functionality and interfaces to external geospatial libraries to retrieve statistical metrics on raster datasets over space and time and to convert between different gridded data formats. It computes parameter statistics for each time step and filters the time step where a condition is met.
- The *rasterrouter.py* module contains functionality to calculate a time-optimized sampling route between the identified locations of interest within the ROI. It transforms the relevant areas defined by the frontend to polygon patches and assigns their location to the nearest nodes of the road network. The third-party Python library OSMnx (Boeing 2017) is used to interact with OpenStreetMap's API to access and download the road network data from the spatial database. A time-optimized monitoring route between all identified nodes and user-selected start and end positions is computed by employing the third-party Python library mlrose (Hayes 2019) which solves the Traveling Salesman Problem using genetic algorithms.
- The *weatherprediction.py* module provides functions to retrieve the most recent numerical weather model forecast runs from the German Weather Service (Deutscher Wetterdienst, (DWD)) FTP webserver. This includes extracting the binary data from compressed archives for each requested parameter and time step and converting the data into xarray datasets covering the entire run period by employing the xarray library (Hoyer & Hamman 2017).
- Similarly, the *precipitation.py* module accesses the DWD FTP webserver to retrieve a recent set of radar precipitation data covering the area of Germany in either hourly or daily resolution. Binary data of a user-defined period is extracted from single compressed archives and converted to xarray datasets employing the wradlib library (Heistermann *et al.* 2013). For areas outside of Germany, gridded data of global hourly precipitation is available after login by accessing the ERA5-Land database using the Python API provided by the European Centre for Medium-Range Weather Forecasts (ECMWF 2020).
- The *soilmoisture.py* module retrieves soil moisture datasets from two different databases. First, a file server hosted at the Helmholtz Centre for Environmental Research (UFZ) provides data for Germany generated from hydrological simulations on a river basin scale. Users need to request a secure shell key before getting access to the file server. Second, soil moisture
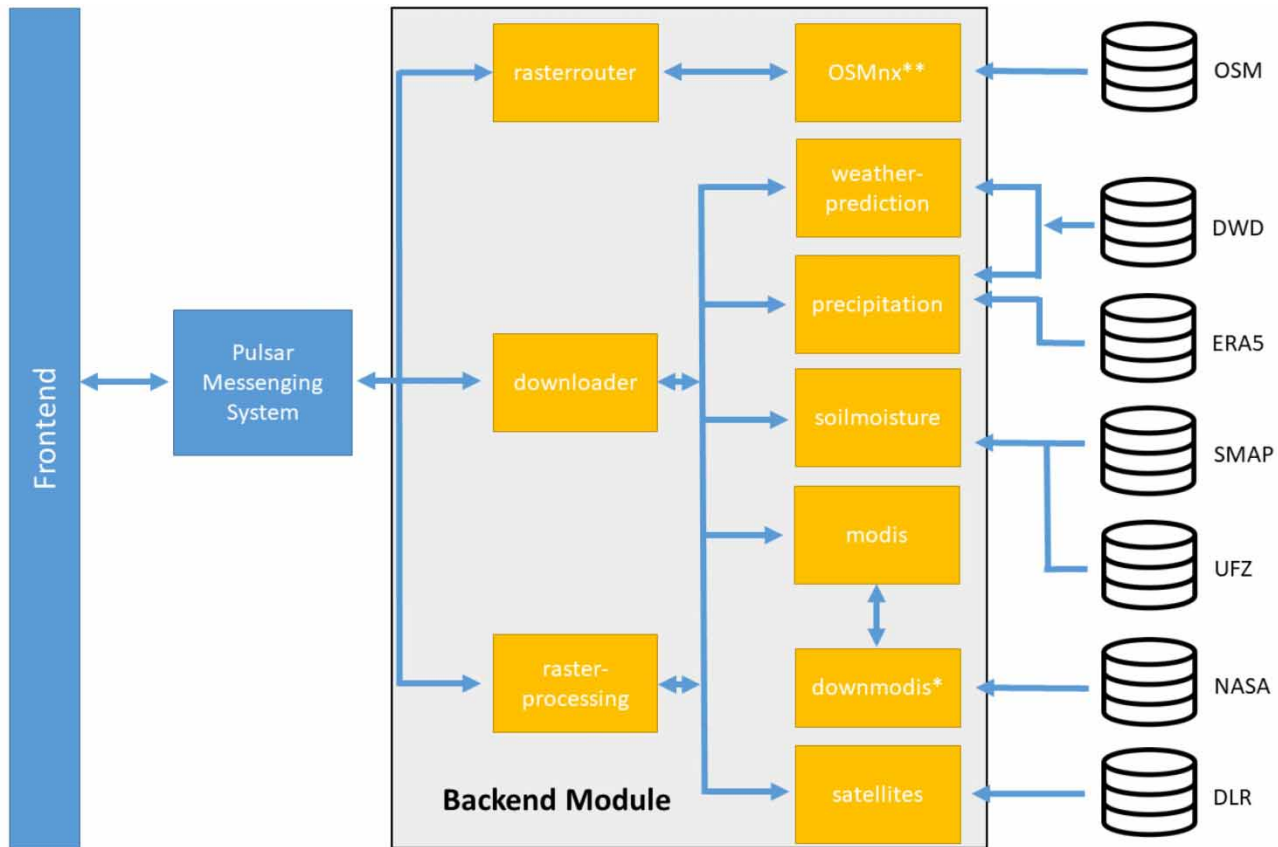
Figure 10 | The backend modeling framework with different modules, coupled with each other and with external databases. The downmodis module is marked with a * as it is a modified version of the script downmodis.py which is provided by the third-party PyModis Library. The OSMnx module is marked with two * to indicate that it is a third-party Python library.

estimates based on multiple satellite observations with global coverage (excluding the Polar Regions) can be retrieved from the NASA-hosted Earthdata database after login.

- The *modis.py* module allows retrieving MODIS data products available at a webserver maintained by NASA. After log-in with credentials, communication with the server is conducted by a modified version of the *downmodis.py* script provided by the PyModis library (Delucchi & Neteler 2013). The module downloads the most recent dataset available from the database, converts the data structure from MODIS HDF-EOS2 format to xarray datasets, and homogenizes the data using the data product-dependent fill and scale values.

The utilization of the backend module provides the flexibility to integrate a large number of additional data formats and data sources (Table 1). All retrieved grid-based datasets are converted into an in-memory representation of a NETCDF3 file with variables, coordinates, and attributes which together form a self-describing dataset. The in-memory NETCDF3 file contains a uniform structure by describing each data variable in the three dimensions latitude, longitude, and time as well as with meta-data information as attributes. All requested data is re-projected to World Geodetic System 1984 geographic reference system and clipped to the region of interest defined in the frontend by utilizing python libraries dealing with geo-spatial datasets such as GDAL (Warmerdam 2008) and Rasterio (Gillies & others, 2021).

We focus on integrating two types of hydro-climatological datasets into Tocap. First, for the applications presented in the paper, we implemented functionality to access databases containing gridded datasets of the target parameters covering Germany. However, if datasets do not cover areas outside of Germany, we provide access to an alternative dataset that provides data at a European or global scale. In addition, access to further datasets can be realized manually by adding the appropriate function to the Python downloader.py module in the backend architecture.

Table 1 | Characteristics of primary data sources accessible via data catalog option of the visual-analytical approach's frontend

Data type	Data source	Type	Spatial coverage/resolution	Temporal resolution
Precipitation	DWD RADOLAN Radar Network	Grid	Germany/1,000 m	1 hour
Precipitation	ECMWF ERA 5 Reanalysis data	Grid	Global/9 km	1 hour
Weather Forecast	DWD ICON-d2	Grid	Central Europe/2,100 m	3 hours
Weather Forecast	DWD ICON-EU	Grid	Europe/6,500 m	3 hours
Soil Moisture	UFZ Drought Monitor	Grid	Germany/4,000 m	1 day
Soil Moisture	SPL2SMAP product	Grid	Global (60S-60N, 180E-180 W)/3 km	6–12 days
Hydrological Forecast	UFZ mHM model	Grid	Mueglitz Basin/100 m	1 day
Selected MODIS Datasets	NASA	Grid	Global/500–1,000 m	Daily-weekly
Satellite Position Forecast	DLR /Heavens-Above	Vector	Satellite dependent	Satellite dependent
Street Network	OSM	Vector	–	–
Base Maps	OSM/ESRI	Grid	Scalable	–

Abbreviations are explained in the running text.

Gridded information on hourly precipitation in Germany with a 1-km resolution can be retrieved using the data from the RADOLAN Radar Network (Winterrath *et al.* 2012) provided by the DWD. Data is downloaded from the FTP server, re-projected to the World Geodetic System 1984, stacked and sent to the frontend. For regions outside of Germany, hourly precipitation estimates at a coarser scale of 0.1° are provided by the ERA 5 reanalysis dataset (Hersbach *et al.* 2020). Similarly, all climatological variables provided from the numerical weather prediction model for Germany, ICON-d2, and Europe, ICON-EU, (Reinert *et al.* 2015) can be integrated into the frontend via selection in the layer catalog.

Information on near-real-time soil moisture dynamics is provided by the German Drought Monitor (Zink *et al.* 2016) hosted at the UFZ. Soil moisture expressed in percentage of field capacity of uppermost 25 cm soil, is categorized in 12 distinct classes before being displayed in the frontend. The visual-analytical approach supports the integration of data provided by upcoming hydrological forecasting initiatives at the UFZ. Exemplarily, the backend module provides access to the hydrological forecast parameters computed for the Mueglitz River Basin in Saxony. Further, near-surface soil moisture on a 3 km grid is obtained from the SMAP/Sentinel-1 L2 soil moisture product SPL2SMAP (Das *et al.* 2019; Das *et al.* 2020) by selecting and downloading relevant tiles for each day available within the selected observation time. The files, provided in the Hierarchical Data Format, are converted into an in-memory NETCDF file before being stacked and sent to the frontend.

The backend module allows the integration of remote sensing data products derived from the MODIS earth observation program (Justice *et al.* 1998) which provides gridded information on environmental parameters such as evapotranspiration, emissivity, NDVI, and land surface temperature. After log-in at the NASA data server, the most recent raster of the selected dataset is downloaded employing the downmodis module (Delucchi & Neteler 2013) before being processed and sent to the frontend. Other remote sensing data products such as Sentinel 2 (Drusch *et al.* 2012) can be integrated directly via WMS/WCS interfaces (see section above) if available. In addition to data from remote sensing sources, information on location-dependent satellite swath coverage within the next two weeks is provided for all major earth observation satellites (Peat 2021). This allows practitioners to synchronize their measurement time and location with datasets from satellites available in the future.

Information on the existing global road network together with a topographic base map is embedded in the frontend visualization using existing interfaces to the OpenStreetMap database (OpenStreetMap contributors 2020). Alternatively, a topographic base map from Esri can be used as the foundation layer of the map layer.

4. APPLICATION

4.1. Areas of potential application

Due to its generic approach, Tocap has the potential to be applicable in multiple disciplines in terrestrial environmental science where the planning of monitoring and sampling campaigns depends on the spatial-temporal distribution of landscape parameters. Reflecting the dynamics in water catchments dynamics, short rainfall events are responsible for exporting the

majority of nutrients (Haraldsen & Stålnacke 2006; Puczko & Jekatierynczuk-Rudczyk 2020). Considering that technical options to measure many water quality properties with field-based continuous sensors are limited, field hydrologists need to rely on the short-term deployment of automatic samplers or manual grab sampling during storm events. If the objective is the correct estimation of pollutant loads, sampling activities have to continue during the rising and falling limb of a storm event (Marsh & Waters 2009). In this context, Tocap can be used to map the upcoming rainfall distribution and to plot the temporal distribution. This information can be combined with user-defined datasets on catchment-specific features and the location of surface water bodies to find optimal locations for the short-term deployment of samplers. Another application area is the detection of areas at risk of landslides in response to rainfall events. Aside from topographic characteristics like slope and soil texture, the occurrence of landslides is controlled by rainfall features (duration and intensity) as well as by the initial soil-moisture conditions (Comegna et al. 2016; De Vita et al. 2018). The filtering and weighting functionality of Tocap can be applied on required datasets to generate a map indicating areas with higher landslide risk for an upcoming rain event. Hence, a field campaign to determine the initital state of soil moisture in these areas can be initialized prior to an upcoming rainfall event, which can be determined using the temporal data assessment functionality of Tocap.

Another potential subject of the application is species sampling. For instance, Dusek et al. (2018) developed a Decision Tree based on 1,428 soil samples depicting simple implementable rules for the prevalence of Escherichia coli within soils depending on land cover, site characteristics and soil chemistry. After uploading all feature datasets into Tocap, its threshold and weighting functionality allows rebuilding each branch of the decision tree by applying the subsequent conditional statements and transferring the knowledge included in the decision tree to the user's region of interest. If target features of the species to be sampled depend on dynamic conditions, e.g precipitation events impacting the hygienic quality of blue mussels (Tryland et al. 2014), these data can be easily incorporated using the data catalog of Tocap to determine the optimal species sampling time.

4.2. Case study: planning a basin-scale ad-hoc cosmic ray rover campaign to monitor soil moisture with tocap

To demonstrate the capabilities of our developed visual-analytical tool Tocap, a typical data processing, and analysis workflow is illustrated to support the design of an ad-hoc campaign to measure soil moisture on the catchment scale from the practitioner viewpoint. The demonstration area is the Mueglitz River, which flows within an about 210 km^2 large catchment from the Eastern Ore Mountains in Germany into the Elbe River. The course of the Mueglitz River is partly characterized by narrow flood plains within steep-walled valleys and plateaus which are either used for forestry or agriculture. The shallow soils in the basin, the basin shape, and the occurrence of very intense precipitation events occasionally cause severe floods which, like the flood in August 2002, result in high financial losses (Walther & Pohl 2004). Subsequently, an efficient flood management strategy is of high importance in the Mueglitz River Basin which, among others, requires a comprehensive understanding of the hydrological processes leading to the formation of flash floods in the basin. In this context, intensive monitoring and modeling efforts are conducted within the MOSES initiative (https://www.ufz.de/moses/). MOSES is being developed to monitor the evolution of such weather extremes and their impacts on the affected regions by combining event-oriented measurements with stationary integrative monitoring programs and observatories. In the Mueglitz River Basin, it comprises mobile and modular sensor systems aiming to improve the understanding of the interaction between short-term storm events and the dynamics of the terrestrial water cycle.

One key state variable which controls run-off generation and flood hydrograph properties is soil moisture. However, the spatial-temporal variations of soil moisture distribution are highly variable making its observation challenging (Ochsner et al. 2013). Among the various existing soil moisture monitoring techniques, the application of vehicle-mounted Cosmic Ray Neutron Sensors (CRNS) allows monitoring soil moisture noninvasively by surveying larger regions within a reasonable time (Zreda et al. 2008; Schrön et al. 2018). CRNS-based soil moisture measurements can be validated with soil moisture estimates derived from space-borne sensors (Montzka et al. 2017), e.g. using data products of sensors deployed at the pair of Sentinel-1 satellites (Paloscia et al. 2013; Bauer-Marschallinger et al. 2018).

In the following, the showcase task for the proposed visual-analytical approach Tocap is to determine the optimal date and route to drive a CRNS equipped vehicle through the Mueglitz River Basin to areas where the highest values of soil moisture are expected within the next seven days. Statistical investigations on parameter sensitivity (e.g. Martö et al. 2001; Gill et al. 2006; Gwak & Kim 2017; Cai et al. 2019) have shown that initial soil moisture (SM), expected precipitation (P), topographic slope (SL), and clay content (CL) are among the most important variables to explain spatial and temporal trends in soil moisture distribution in a catchment. The first two parameters are functions of space and time but the slope and clay content is

spatially heterogeneous only. The latter can be derived from public databases such as SoilGrids (Hengl *et al.* 2017) and SRTM (Rabus *et al.* 2003). Tocap supports the CRNS campaign planning as follows. After manual format conversion, the prepared NETCDF files can be added by drag-and-drop (Figure 11). For initial soil moisture and expected precipitation, data from the UFZ soil moisture product as well as from the ICON-EU precipitation forecast estimate are added using the data catalog functionality. Information on past monitoring activities in the Mueglitz River Basin and about topographic and hydrological features can be integrated from the AWI WebGIS project 'MOSES/Digital Earth Mueglitz Campaign' by adding the WMS server address to the 'Add Layer from URL' function.

The implemented weighting functionality is utilized to find the areas with high expected soil moisture values by defining a target function:

$$f(\boldsymbol{x}, t) = w_1 * SM(\boldsymbol{x}, t) + w_2 * P(\boldsymbol{x}, t) + w_3 * SL(\boldsymbol{x}) + w_4 * CL(\boldsymbol{x}) \tag{1}$$

In contrast to soil moisture, for which the most recent data of the most recent time step should be used in eq.1, more than one significant precipitation event may occur within the upcoming seven days. Besides, the roving campaign should take place with some time lag to a storm event due to the adverse impact of raindrops and water accumulated in puddles on the CRNS measurements. The temporal analysis functionality assists to determine the optimal date for the CRNS rover campaign which is scheduled to be within 7 days after the 9th of April 2021. Adding both datasets to a time series chart with two y-Axes, the 15th of April 2021 is determined to be the optimal date for the campaign taking into account the conditions listed above (Figure 12). Subsequently, a new precipitation dataset comprising the accumulated future precipitation until the selected date is created by applying the '*accumulate to date*' filter in the layer panel.

In its current implementation, the weighting function does not consider the value of the pixel itself but follows a 'binary approach' when applying the target function (see section 2.2). Hence, filter functions are applied in advance to preselect pixels inside the user-defined data range. For the case study, we apply the '*above third quartile (Q4)*' filter on each dataset as our target function increases with increasing values of each parameter. For simplification, it is assumed that all parameters have the same weight, which is $w_n = 0.25$ in the case of 4 input parameters. The updated target function including 25% of the
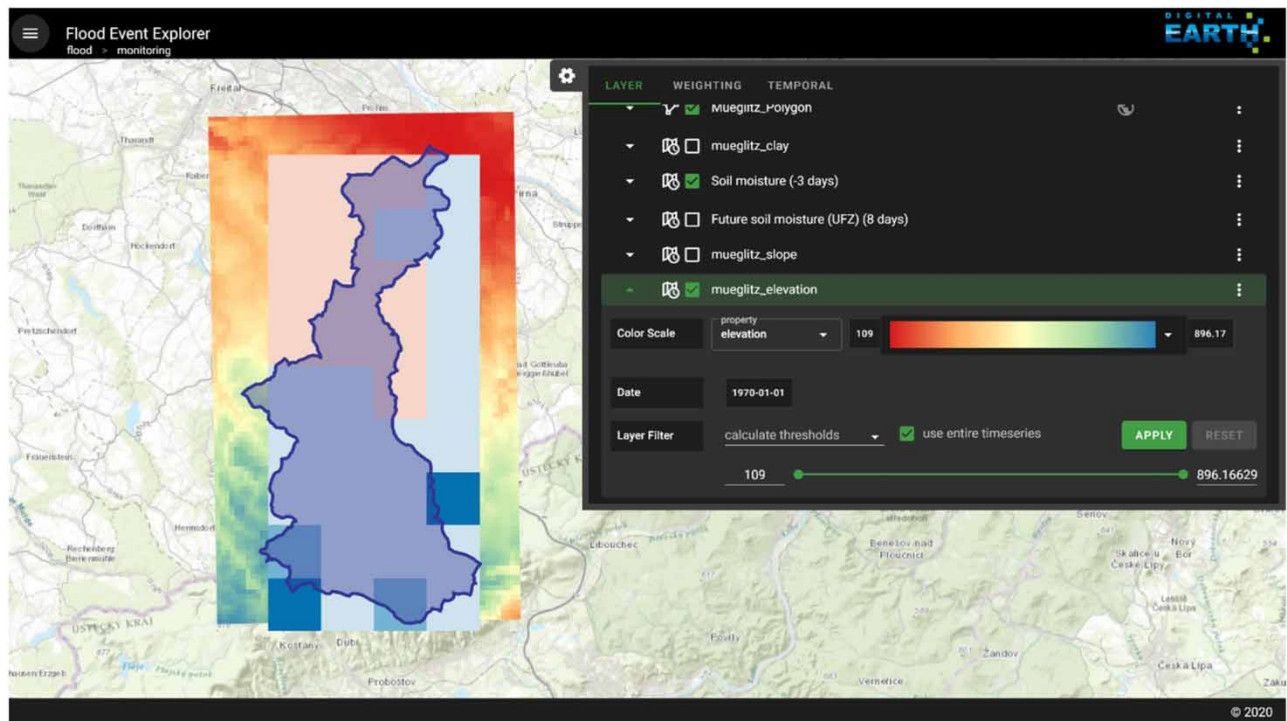

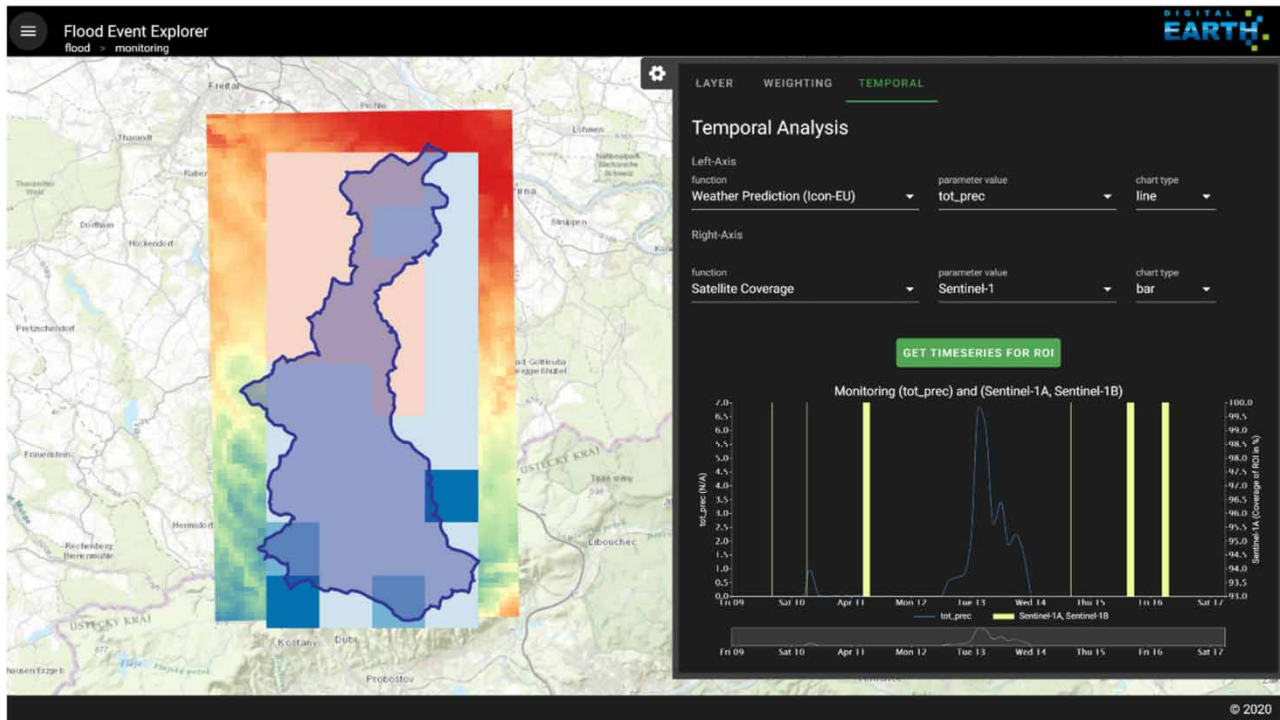**Figure 11** | Datasets of the case study integrated into the visualization frontend.

Figure 12 | Determining the optimal time to initialize the proposed monitoring campaign using the implemented temporal functionality. The 15th of April 2021 is the optimal date for the campaign because it is a day without forecasted rainfall following days of consecutive rain events and provides full spatial coverage by the sensors of the Sentinel-1 Earth observation satellite.

pixels of each input parameter is:

$$f(\boldsymbol{x}, t) = 0.25 * \left[SM_{Q4}(\boldsymbol{x}, t = t\_max) + \sum_{2021-04-09}^{2021-04-15} P(\boldsymbol{x}, t)_{Q4} + SL_{Q4}(\boldsymbol{x}) + CL_{Q4}(\boldsymbol{x}) \right] \qquad (2)$$

The application of Equation (2) reveals four areas where comparatively high soil moisture estimates are expected (Figure 13). Finally, the '*calculate path function*' computes the optimal driving route from the base of the user to all identified regions and back and provides an estimated travel time as supporting information whether the campaign can be conducted with one CRR rover within one working day. The suggested driving route is visualized using connected polylines and can be downloaded in GeoJSON-format readable by state-of-the-art automotive navigation systems.

4.3. Evaluation of results

The performance of Tocap to suggest areas in space and time where target parameter values are in the range of interest depends on multiple factors. A first factor is the quality and relevance of the input datasets selected by the user. For datasets accessible from the data catalog, information on data quality and areas of application are available from the references. Moreover, the chosen filters and layers included in the weighted linear regression affect the results significantly. Tocap allows a fast approach to evaluate the outcome visually and adapt the workflow parameters if required. Although Tocap provides the functionality, the performance of the outcome heavily relies on the user's expertise in selecting datasets and setting filter and weights, which best represent the target parameter dependencies.

The potential of using Tocap to improve campaign planning is demonstrated by applying Equation (2) and comparing the results with 29229 CRNS based soil moisture records obtained during 25 measurement campaigns in the Mueglitz River Basin between February 2019 and July 2020. Since the comparison covers a period in the past, legacy data of hourly ICON-EU forecasts with a horizon of +72 hours provided by the Fraunhofer Institute for Energy Economics and Energy System Technology was used instead of the DWD-based ICON-EU forecasts. Similarly, daily soil moisture estimates from
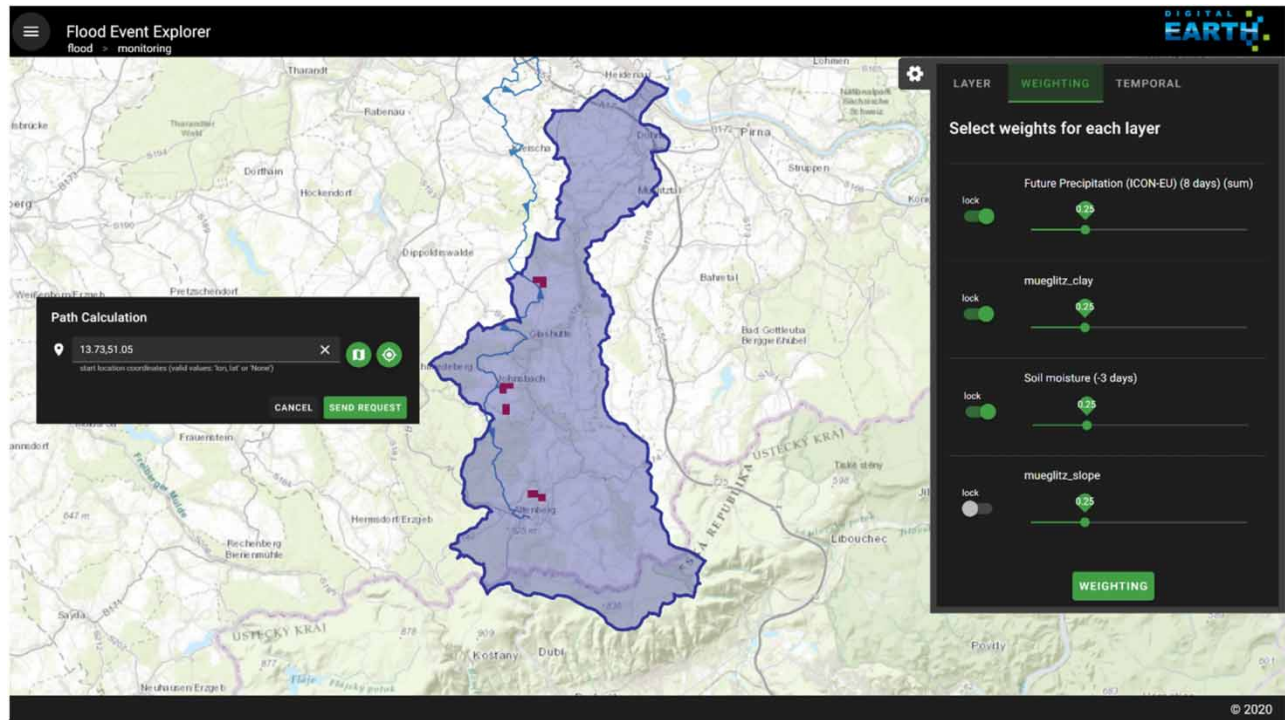
Figure 13 | Identified areas and suggested driving routes for the CRNS monitoring campaign.

the UFZ Drought Monitor are only available for the last 21 days. Subsequently, this data source is replaced with daily soil moisture data from 2019 to 2020 provided by a hydrological model for the Müglitz River Basin (Hannemann 2020). For comparison, the dates, which predict more than 5 mm daily rainfall for at least one day within the forecast horizon, are selected from all available ICON-EU forecasts. Hence, Tocap sums up the forecasted precipitation records until the date of the rain event and applies the algorithm described by Equation (2) to detect locations where higher soil moisture levels are expected. Subsequently, the distribution of CRNS based soil moisture records available within each forecast period and located nearest to a suggested measurement location within a maximum spatial distance of 1 km is compared with the distribution of all CRNS based soil moisture measurements (Figure 14). The results show that measured soil moisture at areas suggested by Tocap was at average 0.42 ± 0.19 kg/m^3, which was higher than the average of all measured records at the selected campaign dates with 0.32 ± 0.16 kg/m^3. These significant differences between both distributions are confirmed by a Kruskal-Wallis H Test (Kruskal & Wallis 1952), which results in a very low p-value of 2.05×10^{-13}. Summarizing, even with the default weighting scheme and only four input datasets, this comparison with measurement data demonstrates that Tocap has the ability to prioritize locations where target parameter values are in the range of interest.

5. CONCLUSIONS AND OUTLOOK

This paper introduces the development of the web-based visual-analytical tool Tocap to support ad-hoc campaign planning in terrestrial hydrology. It presents a comprehensive design for fast visualization and analysis of gridded environmental datasets. In contrast to standard GIS tools, it is platform and device-independent and tailored to meet the needs of processing time-variant distributed datasets by using handy statistical functions and filters.

The developed tool includes innovative features aiming to facilitate and optimize measurement and sampling location design for event-driven monitoring campaigns such as drag-drop functionality of datasets and the one-click integration of data from selected public sources without further user interaction. Furthermore, both, user-provided and accessed datasets are visualized together in space and time for a user-defined region of interest (e.g. a river basin). Threshold and statistical filters as well as weighting functions give the user the flexibility to identify regions of interest for his particular purposes. Time-series analysis graphs further assist to identify the most suitable time to measure and to synchronize the measurement plan with transit times of relevant earth observation satellites. Lastly, the workflow allows computing a time-optimized sampling
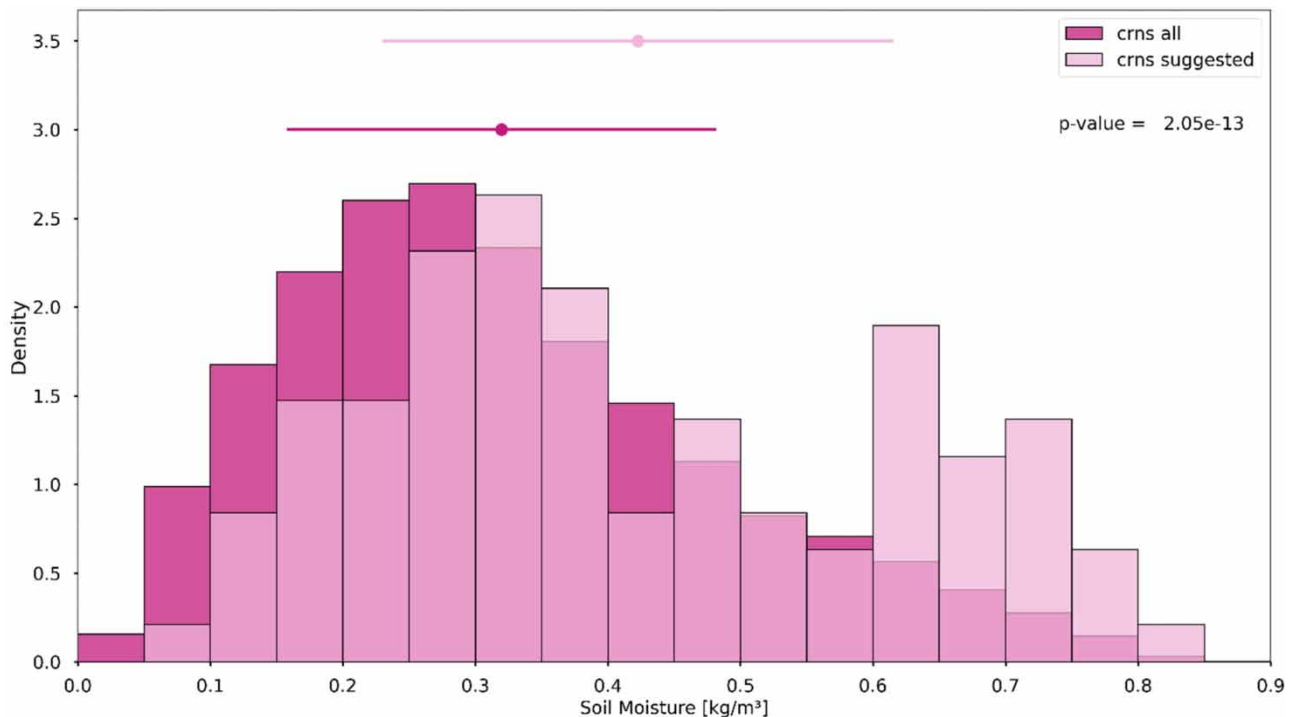
Figure 14 | Distribution of soil moisture estimates of all CRNS-based records within the selected campaign dates ('crns all') and of the records which are closest to the locations suggested by Tocap ('crns suggested'). The bin height is normalized such that the total area of each histogram equals one. The error bars above the histogram visualize the mean and standard deviation of each dataset.

route between identified regions of interest and the user location by accessing the OpenStreetMap road network. Both, the *open-source approach* and the usage of Python programming language in the backend module allow future users to modify the visual analytics tool by their needs.

Although the presented case study demonstrates the ability of Tocap at its current development state to support soil moisture monitoring campaigns, future work will focus on improving the capabilities of the system and applying it to other fields. Future releases may assist the user by including an automatized sensitivity and uncertainty analysis for the selected datasets and weighting schemes. One promising area of future application is population ecology, where human resources often limit species sampling. In this context, Tocap may assist to identify potential species habitats and optimizing sampling routing.

ACKNOWLEDGEMENTS

DISCLOSURE STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

FUNDING

DATA AVAILABILITY STATEMENT

All relevant data are available from the online repositories listed in section 3.1.

REFERENCES

Ahamed, A. & Bolten, J. D. 2017 A MODIS-based automated flood monitoring system for Southeast Asia. *International Journal of Applied Earth Observation and Geoinformation* **61**, 104–117. doi:10.1016/j.jag.2017.05.006.

Atluri, G., Karpatne, A. & Kumar, V. 2018 Spatio-temporal data mining: a survey of problems and methods. *ACM Computing Surveys (CSUR)* **51** (4), 1–41.

Bauer-Marschallinger, B., Freeman, V., Cao, S., Paulik, C., Schaufler, S., Stachl, T., Modanesi, S., Massari, C., Ciabatta, L., Brocca, L. & Wagner, W. 2018 Toward global soil moisture monitoring with Sentinel-1: harnessing assets and overcoming obstacles. *IEEE Transactions on Geoscience and Remote Sensing* **57** (1), 520–539.

Bayer, T., Ames, D. P. & Cleveland, T. G. 2021 Design and development of a web-based EPANET model catalogue and execution environment. *Annals of GIS* **27** (3), 247–260. doi:10.1080/19475683.2021.1936171.

Bhat, S., Motz, L. H., Pathak, C. & Kuebler, L. 2015 Geostatistics-based groundwater-level monitoring network design and its application to the Upper Floridan aquifer, USA. *Environmental Monitoring and Assessment* **187** (1). doi:10.1007/s10661-014-4183-x.

Boeing, G. 2017 OSMnx: a Python package to work with graph-theoretic OpenStreetMap street networks. *Journal of Open Source Software* **2** (12), 215. doi:10.21105/joss.00215.

Bonneau, M., Hauser, C. E., Williams, N. S. G. & Cousens, R. D. 2018 Optimal schedule for monitoring a plant incursion when detection and treatment success vary over time. *Biological Invasions* **20** (3), 741–756. doi:10.1007/s10530-017-1572-4.

Brus, D. J. & Knotters, M. 2008 Sampling design for compliance monitoring of surface water quality: a case study in a Polder area. *Water Resources Research* **44** (11), 1–10. doi:10.1029/2007WR006123.

Butler, D. 2014 Earth observation enters next phase. *Nature* **508** (7495), 160–161. doi:10.1038/508160a.

Cai, Y., Zheng, W., Zhang, X., Zhangzhong, L. & Xue, X. 2019 Research on soil moisture prediction model based on deep learning. *PLoS ONE* **14** (4), 1–19. doi:10.1371/journal.pone.0214508.

Chen, L. & Wang, L. 2018 Recent advance in earth observation big data for hydrology. *Big Earth Data* **2** (1), 86–107. doi:10.1080/20964471.2018.1435072.

Chen, M., Lin, H., Kolditz, O. & Chen, C. 2015 Developing dynamic virtual geographic environments (VGEs) for geographic research. *Environmental Earth Sciences* **10** (74), 6975–6980.

Chen, M., Voinov, A., Ames, D. P., Kettner, A. J., Goodall, J. L., Jakeman, A. J., Barton, M. C., Harpham, Q., Cuddy, S. M., DeLuca, C., Yue, S., Wang, J., Zhang, F., Wen, Y. & Lü, G. 2020 Position paper: open web-distributed integrated geographic modelling and simulation to enable broader participation and applications. *Earth-Science Reviews* **207** (June), 103223. doi:10.1016/j.earscirev.2020.103223.

Comegna, L., Damiano, E., Greco, R., Guida, A., Olivares, L. & Picarelli, L. 2016 Field hydrological monitoring of a sloping shallow pyroclastic deposit. *Canadian Geotechnical Journal* **53** (7), 1125–1137.

Das, N. N., Entekhabi, D., Dunbar, R. S., Chaubell, M. J., Colliander, A., Yueh, S., Jagdhuber, T., Chen, F., Crow, W., O'Neill, P. E., Walker, J. P., Berg, A., Bosch, D. D., Caldwell, T., Cosh, M. H., Collins, C. H., Lopez-Baeza, E. & Thibeault, M. 2019 The SMAP and Copernicus Sentinel 1A/B microwave active-passive high resolution surface soil moisture product. *Remote Sensing of Environment* **233**, 111380. doi:10.1016/j.rse.2019.111380.

Das, N. N., Entekhabi, D., Dunbar, R. S., Kim, S., Yueh, S., Colliander, A., O'Neill, P. E., Jackson, T., Jagdhuber, T., Chen, F., Crow, W., Walker, J. P., Berg, A., Bosch, D. D., Caldwell, T. & Cosh, M. H. 2020 SMAP/Sentinel-1 L2 Radiometer/Radar 30-Second Scene 3 km EASE-Grid Soil Moisture, Version 3. doi:10.5067/ASB0EQO2LYJV.

de Graaf, I. E. M., van Beek, R. L. P. H., Gleeson, T., Moosdorf, N., Schmitz, O., Sutanudjaja, E. H. & Bierkens, M. F. P. 2017 A global-scale two-layer transient groundwater model: development and application to groundwater depletion. *Advances in Water Resources* **102**, 53–67. doi:10.1016/j.advwatres.2017.01.011.

De Gruijter, J., Brus, D. J., Bierkens, M. F. P. & Knotters, M. 2006 *Sampling for Natural Resource Monitoring*. Springer, Berlin.

Delucchi, L. & Neteler, M. 2013 pyModis: the Python library for MODIS data. In: *FOSS4G* 2013.

De Vita, P., Fusco, F., Tufano, R. & Cusano, D. 2018 Seasonal and event-based hydrological and slope stability modeling of pyroclastic fall deposits covering slopes in Campania (Southern Italy). *Water (Switzerland)* **10** (9). doi:10.3390/w10091140.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F. & Bargellini, P. 2012 Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment* **120**, 25–36.

Dusek, N., Hewitt, A. J., Schmidt, K. N. & Bergholz, P. W. 2018 Landscape-scale factors affecting the prevalence of Escherichia coli in surface soil include land cover type, edge interactions, and soil pH. *Applied and Environmental Microbiology* **84** (10), e02714–17.

ECMWF 2020 cdsapi. *GitHub repository*

Eggert, D. & Dransch, D. 2021 DASF: A data analytics software framework for distributed environments. GFZ Data Services. https://doi.org/10.5880/GFZ.1.4.2021.004.

Eggert, D. & Nixdorf, E. 2022 The Smart Monitoring Workflow (Tocap) of the Flood Event Explorer: Determining the most suitable time and location for event-driven, ad-hoc monitoring. GFZ Data Services. https://doi.org/10.5880/GFZ.1.4.2022.004.

Eggert, D., Sommer, P. & Dransch, D. 2021 DASF: Messaging Python: A python RPC wrapper for the data analytics software framework. GFZ Data Services. https://doi.org/10.5880/GFZ.1.4.2021.005.

Gill, M. K., Asefa, T., Kemblowski, M. W. & McKee, M. 2006 Soil moisture prediction using support vector machines. *Journal of the American Water Resources Association* **42** (4), 1033–1046. doi:10.1111/j.1752-1688.2006.tb04512.x.

Gillies, S. & others 2021 *Rasterio: Geospatial Raster I/O for {Python} programmers*. https://github.com/rasterio/rasterio/blob/master/CITATION.txt.

Goharian, E. & Burian, S. J. 2018 Developing an integrated framework to build a decision support tool for urban water management. *Journal of Hydroinformatics* **20** (3), 708–727. doi:10.2166/hydro.2018.088.

Grafana Labs 2018 *Grafana Documentation*. Available from: https://grafana.com/docs/ (accessed 25 July 2019).

Gwak, Y. & Kim, S. 2017 Factors affecting soil moisture spatial variability for a humid forest hillslope. *Hydrological Processes* **31** (2), 431–445. doi:10.1002/hyp.11039.

Haas, A., Heim, B., Zubrzycki, S., Elger, K., Schäfer-Neth, C., Morgenstern, A. & Fedorova, I. 2016 Assembly of the CarboPerm WebGIS for the Laptev Sea Region, Arctic Siberia–data visualisation as a WebGIS service. *Polarforschung* **85** (2), 135–141.

Hannemann, M. 2020 *Investigating the Influence of Different Input Data Sets and Simulation Approaches on Runoff Dynamics in the Müglitz River Basin [German]*. University of Osnabrück, Osnabrück.

Haraldsen, T. K. & Stålnacke, P. 2006 Methods for water quality sampling and load estimation in monitoring of Norwegian agricultural catchments. *Nordic Hydrology* **37** (1), 81–92. doi:10.2166/nh.2005.032.

Harmancioglu, N. B., Fistikoglu, O., Ozkul, S. D., Singh, V. P. & Alpasan, M. N. 1999 Water quality monitoring network design. *Water Science and Technology Library* **33**, 1–290.

Hayes, G. 2019 Mlrose: Machine Learning, Randomized Optimization and SEarch package for Python.

Heistermann, M., Jacobi, S. & Pfaff, T. 2013 Technical note: an open source library for processing weather radar data (wradlib). *Hydrology and Earth System Sciences* **17** (2), 863–871. doi:10.5194/hess-17-863-2013.

Hengl, T., De Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S. & Kempen, B. 2017 *SoilGrids250 m: Global Gridded Soil Information Based on Machine Learning*. doi:10.1371/journal.pone.0169748.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. & Thépaut, J. N. 2020 The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146** (730), 1999–2049.

Hoelscher, J. & Mortimer, A. 2018 Using tableau to visualize data and drive decision-making. *Journal of Accounting Education* **44**, 49–59.

Hoyer, S. & Hamman, J. J. 2017 xarray: N-D labeled arrays and datasets in python. *Journal of Open Research Software* **5**, 1–6. doi:10.5334/jors.148.

Hunter, J., Brooking, C., Reading, L. & Vink, S. 2016 A Web-based system enabling the integration, analysis, and 3D sub-surface visualization of groundwater monitoring data and geological models. *International Journal of Digital Earth* **9** (2), 197–214. doi:10.1080/17538947.2014.1002866.

Huston, M. A. 1999 Local processes and regional patterns: appropriate scales for understanding variation in the diversity of plants and animals. *Oikos* **86** (3), 393. doi:10.2307/3546645.

Joseph, L. N. & Possingham, H. P. 2008 Grid-based monitoring methods for detecting population declines: sensitivity to spatial scale and consequences of scale correction. *Biological Conservation* **141** (7), 1868–1875. doi:10.1016/j.biocon.2008.05.003.

Justice, C. O., Vermote, E., Townshend, J. R. G., Defries, R., Roy, D. P., Hall, D. K., Salomonson, V. V., Privette, J. L., Riggs, G., Strahler, A., Lucht, W., Myneni, R. B., Knyazikhin, Y., Running, S. W., Nemani, R. R., Wan, Z., Huete, A. R., Van Leeuwen, W., Wolfe, R. E., Giglio, L., Muller, J. P., Lewis, P. & Barnsley, M. J. 1998 The moderate resolution imaging spectroradiometer (MODIS): land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing* **36** (4), 1228–1249. doi:10.1109/36.701075.

Kruskal, W. H. & Wallis, W. A. 1952 Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* **47** (260), 583–621. doi:10.1080/01621459.1952.10483441.

Li, Y., Zhang, X. X., Mao, R. L., Yang, J., Miao, C. Y., Li, Z. & Qiu, Y. X. 2017 Ten years of landscape genomics: challenges and opportunities. *Frontiers in Plant Science* **8** (December), 1–7. doi:10.3389/fpls.2017.02136.

Li, L., Sullivan, P. L., Benettin, P., Cirpka, O. A., Bishop, K., Brantley, S. L., Knapp, J. L. A., van Meerveld, I., Rinaldo, A., Seibert, J., Wen, H. & Kirchner, J. W. 2020 Toward catchment hydro-biogeochemical theories. *Wiley Interdisciplinary Reviews: Water*. (December). doi:10.1002/wat2.1495.

Lin, H., Chen, M. & Lu, G. 2013 Virtual geographic environment: a workspace for computer-aided geographic experiments. *Annals of the Association of American Geographers* **103** (3), 465–482. doi:10.1080/00045608.2012.689234.

Mao, H., Kathuria, D., Duffield, N. & Mohanty, B. P. 2019 Gap filling of high-Resolution soil moisture for SMAP/Sentinel-1: a two-layer machine learning-based framework. *Water Resources Research* **55** (8), 6986–7009. doi:10.1029/2019WR024902.

Marsh, N. & Waters, D. 2009 Comparison of load estimation methods and their associated error. In: *18th World IMACS Congress and MODSIM 2009 - International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, Proceedings*: 3322–3328.

Martõ, M., Albaladejo, J., Castillo, V. M. & Go, A. 2001 Factors regulating spatial distribution of soil water content in small semiarid catchments. *Journal of Hydrology* **253**, 211–226.

Miao, R., Song, J. & Zhu, Y. 2017 3D geographic scenes visualization based on WebGL. In: *2017 6th International Conference on Agro-Geoinformatics, Agro-Geoinformatics 2017*. doi:10.1109/Agro-Geoinformatics.2017.8046999.

Montzka, C., Bogena, H. R., Zreda, M., Monerris, A., Morrison, R., Muddu, S. & Vereecken, H. 2017 Validation of spaceborne and modelled surface soil moisture products with cosmic-Ray neutron probes. *Remote Sensing* 9 (2), 1–30. doi:10.3390/rs9020103.

National Aeronautics and Space Administration 2021 *Panoply netCDF, HDF and GRIB Data Viewer*. Available from: https://www.giss.nasa.gov/tools/panoply/

Nikoloudi, E., Romano, M., Memon, F. A. & Kapelan, Z. 2021 Interactive decision support methodology for near real-time response to failure events in a water distribution network. *Journal of Hydroinformatics* 23 (3), 483–499. doi:10.2166/HYDRO.2020.101.

Object Management Group 2007 *Unified Modeling Language: Superstructure Version 2.1. 1*. Available from: https://www.omg.org/spec/UML/2.1.1/Superstructure/PDF/

Ochsner, T. E., Cosh, M. H., Cuenca, R. H., Dorigo, W. A., Draper, C. S., Hagimoto, Y., Kerr, Y. H., Larson, K. M., Njoku, E. G., Small, E. E. & Zreda, M. 2013 State of the art in large-scale soil moisture monitoring. *Soil Science Society of America Journal* 77 (6), 1888–1919. doi:10.2136/sssaj2013.03.0093.

OpenLayers. 2021 *OpenLayers v6.X*. Available from: https://openlayers.org/ (accessed 19 April 2021).

OpenStreetMap contributors 2020 *Planet Dump*. Available from: https://planet.osm.org

Paloscia, S., Pettinato, S., Santi, E., Notarnicola, C., Pasolli, L. & Reppucci, A. 2013 Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation. *Remote Sensing of Environment* 134, 234–248.

Peat, C. 2021 *Heavens Above*. Available from: https://www.heavens-above.com/ (accessed 9 April 2021).

Puczko, K. & Jekatierynczuk-Rudczyk, E. 2020 Extreme hydro-meteorological events influence to water quality of small rivers in urban area: a case study in Northeast Poland. *Scientific Reports* 10 (1), 1–14. doi:10.1038/s41598-020-67190-4.

Pumo, D., Francipane, A., Lo Conti, F., Arnone, E., Bitonto, P., Viola, F., La Loggia, G. & Noto, L. V. 2016 The SESAMO early warning system for rainfall-triggered landslides. *Journal of Hydroinformatics* 18 (2), 256–276. doi:10.2166/hydro.2015.060.

Qiao, X., Li, Z., Zhang, F., Ames, D. P., Chen, M., James Nelson, E. & Khattar, R. 2021 A container-based approach for sharing environmental models as web services. *International Journal of Digital Earth* 14 (8), 1067–1086. doi:10.1080/17538947.2021.1925758.

Rabus, B., Eineder, M., Roth, A. & Bamler, R. 2003 The shuttle radar topography mission – a new class of digital elevation models acquired by spaceborne radar. *ISPRS Journal of Photogrammetry and Remote Sensing* 57 (4), 241–262. doi:10.1016/S0924-2716(02)00124-7.

Reed, C. N. 2011 The open geospatial consortium and web services standards. In: *Geospatial Web Services: Advances in Information Interoperability* IGI Global; 1–16.

Reinert, D., Frank, H. & Prill, F. 2015 *ICON Database Reference Manual*. Deutscher Wetterdienst.

Rink, K., Fischer, T., Selle, B. & Kolditz, O. 2013 A data exploration framework for validation and setup of hydrological models. *Environmental Earth Sciences* 69 (2), 469–477. doi:10.1007/s12665-012-2030-3.

Rink, K., Nixdorf, E., Zhou, C., Hillmann, M. & Bilke, L. 2020 A virtual geographic environment for multi-compartment water and solute dynamics in large catchments. *Journal of Hydrology* 582, 124507.

Rockwood, L. L. 2015 *Introduction to Population Ecology*. John Wiley & Sons, Oxford.

Sarafanov, M., Kazakov, E., Nikitin, N. O. & Kalyuzhnaya A, V. 2020 A machine learning approach for remote sensing data gap-filling with open-source implementation: an example regarding land surface temperature, surface Albedo and NDVI. *Remote Sensing* 12 (23), 3865. doi:10.3390/rs12233865.

Schrön, M., Rosolem, R., Köhli, M., Piussi, L., Schröter, I., Iwema, J., Kögler, S., Oswald, S. E., Wollschläger, U., Samaniego, L., Dietrich, P. & Zacharias, S. 2018 Cosmic-ray neutron rover surveys of field soil moisture and the influence of roads. *Water Resources Research* 54 (9), 6441–6459. doi:10.1029/2017WR021719.

Selmoni, O., Vajana, E., Guillaume, A., Rochat, E. & Joost, S. 2020 Sampling strategy optimization to increase statistical power in landscape genomics: a simulation-based approach. *Molecular Ecology Resources* 20 (1), 154–169. doi:10.1111/1755-0998.13095.

Simmons, A., Fellous, J. L., Ramaswamy, V., Trenberth, K., Asrar, G., Balmaseda, M., Burrows, J. P., Ciais, P., Drinkwater, M., Friedlingstein, P., Gobron, N., Guilyardi, E., Halpern, D., Heimann, M., Johannessen, J., Levelt, P. F., Lopez-Baeza, E., Penner, J., Scholes, R. & Shepherd, T. 2016 Observation and integrated earth-system science: a roadmap for 2016–2025. *Advances in Space Research* 57 (10), 2037–2103. doi:10.1016/j.asr.2016.03.008.

Smith, A. N. H., Anderson, M. J. & Pawley, M. D. M. 2017 Could ecologists be more random? straightforward alternatives to haphazard spatial sampling. *Ecography* 40 (11), 1251–1255. doi:10.1111/ecog.02821.

Teller, S. 2013 *Data Visualization with d3. js*. Packt Publishing Ltd, Birmingham.

Tryland, I., Myrmel, M., Østensvik, Ø., Wennberg, A. C. & Robertson, L. J. 2014 Impact of rainfall on the hygienic quality of blue mussels and water in urban areas in the Inner Oslofjord, Norway. *Marine Pollution Bulletin* 85 (1), 42–49. doi:10.1016/j.marpolbul.2014.06.028.

Van Rossum, G. & Drake Jr., F. L. 1995 *Python Reference Manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Walther, P. & Pohl, R. 2004 Hochwasserrisiken im osterzgebirge–eine historische retrospektive. *Risiken bei der Bemessung und Bewirtschaftung von Fließgewässern und Stauanlagen Dresdner Wasserbauliche Mitteilungen* 27, 383–400.

Warmerdam, F. 2008 The geospatial data abstraction library. In: B. Hall & M. G. Leahy, eds. *Open Source Approaches in Spatial Data Handling*. Springer, Berlin, pp. 87–104.

Winterrath, T., Rosenow, W. & Weigl, E. 2012 On the DWD quantitative precipitation analysis and nowcasting system for real-time application in German flood risk management. *IAHS-AISH Publication* 351 (April 2011), 323–329.

Xu, H., Muste, M. & Demir, I. 2019 Web-based geospatial platform for the analysis and forecasting of sedimentation at culverts. *Journal of Hydroinformatics* **21** (6), 1064–1081. doi:10.2166/hydro.2019.068.

Zaherpour, J., Gosling, S.N., Mount, N., Schmied, H.M., Veldkamp, T.I.E., Dankers, R., Eisner, S., Gerten, D., Gudmundsson, L., Haddeland, I., Hanasaki, N., Kim, H., Leng, G., Liu, J., Masaki, Y., Oki, T., Pokhrel, Y., Satoh, Y., Schewe, J. & Wada, Y. 2018 Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. *Environmental Research Letters* **13** (6). doi:10.1088/1748-9326/aac547.

Zhai, X., Yue, P. & Zhang, M. 2016 A sensor web and web service-based approach for active hydrological disaster monitoring. *ISPRS International Journal of Geo-Information* **5**, 10. doi:10.3390/ijgi5100171.

Zhou, Y., Dong, D., Liu, J. & Li, W. 2013 Upgrading a regional groundwater level monitoring network for Beijing Plain, China. *Geoscience Frontiers* **4** (1), 127–138. doi:10.1016/j.gsf.2012.03.008.

Zink, M., Samaniego, L., Kumar, R., Thober, S., Mai, J., Schafer, D. & Marx, A. 2016 The German drought monitor. *Environmental Research Letters* **11** (7). doi:10.1088/1748-9326/11/7/074002.

Zreda, M., Desilets, D., Ferré, T. P. A. & Scott, R. L. 2008 Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. *Geophysical Research Letters* **35** (21), 1–5. doi:10.1029/2008GL035655.