LegacyMine_HydroGeo: Dataset on geochemical and hydrological dynamics in a historic mine system

(https://doi.org/10.23728/b2share.32958b4c93284b739182774a18756fdb)

Anita Alexandra Sanchez[1,2]*, Maximilian P. Lau[1,2], Sean Adam[1,3], Karl-Julius Haas[1,3], Sabrina Hedrich[1,4], & Conrad Jackisch[1,3]

[1]*Interdisciplinary Environmental Research Centre, Technische Universität Bergakademie Freiberg*
[2]*Institute of Mineralogy, TU Bergakademie Freiberg, Brennhausgasse 14, 09599 Freiberg, Germany*
[3]*Institute of Drilling Technology and Fluid Mining, TU Bergakademie Freiberg, Germany*
[4]*Institute of Biosciences, TU Bergakademie Freiberg, Germany*
**Anita.Sanchez@mineral.tu-freiberg.de*

Abstract

This dataset presents a comprehensive approximately two-year continuous monitoring effort of subsurface mine water dynamics in the Reiche Zeche mine, a decommissioned silver mining site in the Ore Mountains of Central Europe. This dataset integrates high-resolution geochemical sampling, hydrological in situ sensing, and UV-Vis spectrometry complemented by machine learning modeling to characterize hotspots and hot moments of metal mobilization across stratified and mixed hydrological regimes. Within 42 sampling campaigns (February 2022 - May 2024), water samples were collected from 26 water flowing and dripping sites, with detailed hydrological and automated spectroscopic monitoring at four distinct locations (sites 1, 2, 3A, and 3B). Dissolved metal(loid) and organic carbon concentrations, and stable water isotopes ($\delta^{2}H$ and $\delta^{18}O$) were recorded alongside hourly flow rates and optical absorbance spectra. A cubist-based modeling framework was applied to trace metal concentrations from spectral data, with models validated using autosampler measurements. This curated dataset supports the study of contaminant transport under variable flow conditions in legacy mining environments and is intended to facilitate data reuse for hydrogeochemical modeling, machine learning applications, and environmental monitoring research.

Table of contents

1. Licence

2. Citation

When using the data please cite:

Sanchez, A. A., Lau, M. P., Adam, S., Haas, K. J., Hedrich, S., and Jackisch, C. (2025). LegacyMine_HydroGeo: Dataset on geochemical and hydrological dynamics in a historic mine system. CDI B2SHARE. DOI: 10.23728/b2share.32958b4c93284b739182774a18756fdbSanc

3. Data Description

Our study site, “Reiche Zeche”, situated in the Ore Mountains of Central Europe, is located at 50.928° N latitude and 13.357°E longitude. This site was active in extracting high-grade minerals, specifically silver ore, and processing mine waste rock since the 12th century up until 1969. Subsequently, the lower sections of the adit system, which extend down to 1300 meters, became inundated with water up to the level of the central adit “Rothschönberger Stollen” (RSS) – a depth of about 230 meters at the Reiche Zeche mine shaft (Zhiteneva et al., 2016; Mischo et al., 2021). This study focuses on a single slanted vertical extraction structure which spans over three levels before reaching the central drainage adit. These levels are: Level 1 (located 103 meters below the surface); Level 2 (149 meters below the surface); and Level 3 (191 meters below the surface). We collected geochemical and isotopic data at all sites, obtained hydrological monitoring data for 4 sites, and at one site on level 2, we collected online uv-vis spectrometer data and performed matching learning modeling using ‘Cubist’ to further validate our findings of metal(loid) concentration fluctuations during different hydrological phases.

3.1. Sampling sites

A total of 26 sites were selected for ongoing sampling with 14 sites exhibiting high flow, 11 sites presenting lower, drip flow rates, and 1 site displaying alternating very stagnant (i.e., stratified) and non-stagnant waters. However, four specific sites positioned with water level loggers (C4, A12, B4, and B3) will be focused here. The 4 locations, renamed as sites 1, 2, 3A, and 3B, were selected due to the presence of continuous and ample amounts of flowing water in comparison to the rest of the locations which were not as great in volume of flowing water. In three-week intervals, we conducted 42 sample

campaigns to all sites from February 3rd, 2022 to May 31st, 2024. Figure 1 displays a representation of sampling sites within the mine.

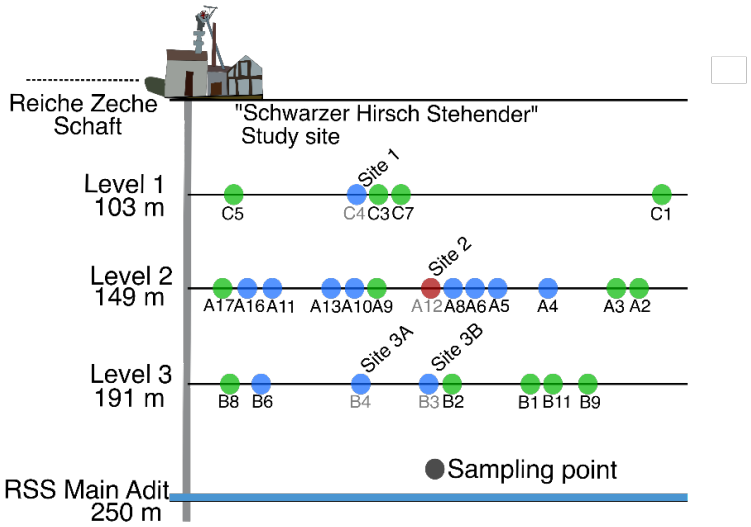

Figure 1: Schematic representation of study site "Schwarzer Hirsch Stehender" at Reiche Zeiche, a former lead-zinc-silver mine in Germany. Samples were collected at 26 sites of running water across the three levels and central drainage adit. Labelled sites 1, 2, 3A, and 3B (previously named C4, A12, B4, B3) positioned with water loggers are specifically discussed in this study. The color of each site corresponds to water flow dynamic (blue - flowing, green - dripping, red - stratified).

3.2. Geochemical data

Acid-washed HDPE bottles, pre-rinsed with deionized water, were used to collect water samples. For all samples, pH and conductivity were measured (pH 340 and Cond 3310 sensors, WTW Weilhem). Prior to conducting analyses for dissolved organic carbon (DOC), dissolved inorganic carbon (DIC), and metal(loid)s, we filtered the samples using polyethersulfone filters with 0.45 µm pores (Filtropur S, Sarstedt Nümbrecht). The DIC and DOC concentrations were measured in triplicate for each sample using a total organic carbon analyzer (TOC-L series, Shimadzu Duisburg). For DOC measurements, we employed a high temperature combustion method, categorizing it as non-purgeable organic carbon (NPOC). This involved acidifying and then purging the samples with oxygen to expel inorganic carbon before the analysis. The precision of our data was validated by computing the standard deviation of the triple measurements, ensuring data reliability within the instrument's precision range (coefficient of variation < 2% and standard error < 0.1). We quantified metal(loid) concentrations using inductively coupled plasma optical emission spectroscopy (ICP-OES Optima 5300 DV Spectrometer, PerkinElmer Rodgau). For metal(loid) analysis, we prepared the samples with an addition of 1 mL of 2% nitric acid and included the following metal(loid)s in our analysis: iron (Fe), zinc (Zn), arsenic (As), copper (Cu), cadmium (Cd), lead (Pb), aluminum (Al), nickel (Ni), and manganese (Mn).

We further calculated the Pollution Load Index (PLI) to assess the cumulative level of metal(loid)n contamination across the four flow monitored locations. The PLI provides an aggregated measure of contamination by integration the contamination factors (CFs) of individual metal(loid)s, calculated as a ratio of observed metal(loid) concentrations to their respective background reference values (Jahan & Strezov, 2018):

$$PLI = (CF1 \times CF2 \times CF3 \ldots\ldots CFn)^{1/n} \quad (1)$$

where CF= $C_{metal}/C_{background}$, and n is the number of metal(loid)s considered. A PLI > 1 indicates pollution, whereas PLI < 1 implies no contamination (Tomlinson et al., 1980). Reference concentrations (for all metal(loid)s except Al) were derived from average values in the Elbe river at the Magdeburg station, located near the midpoint of the river, for the year 2022, obtained from FGG Elbe Data Portal (FGG Elbe, 2025). This evaluation allowed us to assess relative contamination levels at specific locations in the mine against a representative background from a major regional river.

3.3. Online UV-Vis Spectrometer data and modeling results

An autosampler (6712 full-size portable sampler, ISCO Nebraska) and an online UV-Vis spectrometer probe (spectro::lyser V3, s::can GmbH Vienna; in the following simply termed spectrolyzer) were positioned at site 2 as an approach to avoid missing unseen aspects in the temporal dynamics of mine drainage water quality. The autosampler was calibrated to take a sample daily. 21 out of 24 1 L autosampler bottles were each filled with 10 mL of 2 M HCl prior to each start of the autosampler run to perform adequate metal(loid) measurements in the laboratory, while three autosampler bottles (one every seven days) were unacidified to record accurate pH and electrical conductivity measurements. From May 16th, 2022 to February 14th, 2023 the autosampler was filled every three weeks and 250 mL samples were collected from each 1 L bottle in the machine, while the remaining was discarded. Samples were filtered in the lab and prepped for doc, dic, and metal(loid) analyses. Prior to each new campaign, all 250 ml autosampler bottles were cleaned in a dishwasher and rinsed with deionized water. To complement this daily automated sampling, we submersed the spectrolyzer to record hourly absorbance measurements from May 16th, 2022 to May 23rd, 2023 over a wavelength range of 200 to 720 nm at 2.5 nm increments.

To analyze and compare the spectral data obtained from the spectrolyzer with the metal(loid) concentration data collected by the autosampler at site 2 over time, we employed Cubist modeling. Cubist, a rule-based machine learning method, combines decision trees with linear models, allowing for the prediction of continuous numerical variables. Chemometric models were calibrated from UV-Vis scans (i.e, measurements from the spectrolyzer) and corresponding analytical data (i.e., measurements from the Autosampler) to predict trace element concentrations.Raw spectra were cleaned by removing implausible scans (absorbance < 0 or absorbance > 4.5). Spectra were processed using the prospectr package Stevens and Ramirez-Lopez (2022). First, spectra were smoothed using a Savitzky Golay filter with a polynomial degree of 3 and a window size of 11 to remove noise. Then, a Standard Normal Variate (SNV) correction was applied. Both raw, smoothed and smoothed + snv-corrected spectra were used for model calibration. Chemometric models were calibrated using the Cubist package Kuhn and Quinlan (2023) in the caret framework Kuhn (2008) for the following dissolved metal(loid)s: Mn, Ni, As, Cd, Pb, Fe, Al, Cu, and Zn. Models were calibrated both for untransformed and log1p-transformed variables, and using the three differently processed spectral sets (raw, smoothed, smoothed+snv). Cubist models were tuned for 1, 2, 5, 10, 20, and 50 committees and 0-9 neighbours using a 10-fold cross-validation. 75 % of the dataset was used for calibration and 25 % was held back for testing. The dataset was split using random, percentile-binned sampling. Model accuracy was evaluated using the test set, using a modified version of the evaluate_model() function found in the simplerspec package Baumann (2020) to calculate RMSE, R2, RPD, and Lin's concordance correlation coefficient among other valuation statistics.

3.4. Hydro-meteorological monitoring data

Weather data is recorded at an automated weather station at Reiche Zeche. We aggregated the 5-minute records to daily values for air temperature, actual atmospheric vapor pressure, wind velocity, incoming radiation and precipitation. Moreover, we calculated the FAO 56 Penman-Monteith reference evapotranspiration (using pyeto, https://github.com/woodcrafty/PyETo/) and the biweekly Self-Calibrating Palmer Drought Severity Index (Palmer, 1965; Wells et al., 2004; using scPDSI, https://github.com/Sibada/scPDSI/). In addition we added monthly values of the UFZ Drought Monitor

/ Helmholtz Centre for Environmental Research https://www.ufz.de/index.php?en=37937 (Boeing et al., 2022) for the closest gridcell as monthly data for the total soil column. Both landscape water state measures are interpolated to the daily meteorologic records.

The hydrological dynamics were observed with automated water level loggers (Levelogger 5, Solinst Canada Ltd.) at sites 1, 2, 3A, and 3B. The loggers at levels 1 and 2 recorded measurements from March 2022 to approximately January 2024 and the loggers at levels 3A and 3B recorded measurements from February 2022 to March 2024. For precise water level-discharge relationships sites 1 and 2 were equipped with plastic V-weirs. Sites 3A and 3B were equipped with V-shaped spillways in the given rock material.
To establish the water level-discharge relationship we manually measured both values during the sampling campaigns and fitted the Henderson (1996) equation for triangular measuring weirs to the data:

$$Q = \frac{8}{15} * C_d * \sqrt[2]{2 * g} * tan\left(\frac{\alpha}{2}\right) * h^{\frac{5}{2}} \qquad (2)$$

where Q is the flow rate (in m^3/s], C_d is the discharge coefficient, g is the gravitational constant (9.81 m/s^2), α is the angle of the lower weir edge (deg) and h is the water level above the outlet (m). The established water level-discharge relationship was applied, bias corrected and smoothed with a Savitzky-Golay filter. Using the flow rate data, we identified and assigned different hydrological phases (flush, declining flow, low flow).

3.5. Isotopic data

Water stable isotopes (δ^2H and $\delta^{18}O$) were analyzed via cavity ring-down spectroscopy (L2130-i, Picarro Inc.) to trace water sources. All samples have been analysed 8-fold to either determine the mean of the last 5 samples when no asymptotic approach to a true value can be detected (mostly when the repeated measurements do not drift at all), or their asymptotic value is used (Hachgenei et al., 2022) as true measurement. These values are reported as 2H_m and 18O_m (measured). The standard deviation of the fit is reported as 2H_err and 18O_err (error/uncertainty).
Batches of ±12 samples were referenced to several standards before and after each batch. The respective drift is accounted for as two individual linear regressions for the standards and its weighted application to the measured values. The respective weight is a linear function of the order of measurements with the first analysis weighing heavily to the antecedent reference and the last analysis weighing mostly to the posterior reference measurements. Moreover and since the measured standards can deviate substantially from the range of measured samples, an amplification for the standards close to the measured samples is applied. The respective calibrated values are reported as 2H_cal and 18O_cal (calibrated).

3.6. Hysteresis Analysis

To capture dynamic transport mechanisms, C-Q relationships and hysteresis behavior were analyzed using hysteresis index (HI) methods developed by Lloyd et al. (2016) and Zuecco et al (2016). The following phases and hysteresis classes were derived based on a point-to-point analysis of the C-Q patterns for each solute and site: Loading with Q↓ C↑, flushing with Q↑ C↑, dilution with Q↑ C↓, chemostatic with varying Q and stable C, recession with Q↓ C↓ and variable sources with stable Q and varying C. The hysteresis phase classification was based on the direction of the C-Q loop as counter-clockwise (distant sources, delayed response) with negative HI and clockwise (near-stream sources, rapid mobilization) with positive HI.

4. File description

In this section the details to the provided files of the LegacyMine_HydroGeo dataset are provided.

4.1. File inventory

The geochemical, hydrological, and spectrolyzer data, and Python scripts for the figures presented in the manuscript are stored in the following files:

File	Content
'Fulldat_PC.csv'	The dataset includes the physicochemical parameters (PC) measured for all the samples manually collected from Reiche Zeche.
'Site2_autosampler_PC.csv'	The dataset includes the physicochemical parameters (PC) measured for all the samples from the autosampler at site 2.
'Site2_stratifed_sampling_PC.csv'	The dataset includes the physicochemical parameters (PC) measured for all the samples manually collected from site 2's top and bottom waters during stratification.
'Meteo.csv'	The meteorologic conditions measured at the weather station at the position of the Reiche Zeche mine access are included here.
'Flow.csv'	The water discharge data measured from the four sites with the hydrological phase identified are included here.
'Isotopes.csv'	The isotope values for all grab samples from Reiche Zeche and from precipitation are included here.
'Hysteresis_analysis.csv'	The hysteresis index values from the four sites with water loggers are included here.
'Site2_spectrolyzer_autosampler_prediction.csv'	We present the input concentration data (spectrolyzer and autosampler data) to perform the cubist modeling procedure. The output predicted results are included here too.
'Site2_spectrolyzer_raw.csv'	We present the compilation of all the raw data obtained from the spectrolyzer over a wavelength from 200 nm to 720 nm here.
'Site2_Cdspec_Qlogg_hourly.csv'	We present the matched timestamps of cadmium concentration values from cubist modeling and the water discharge values measured from the water logger.
'LegacyMine_ReicheZeche.ipynb' and 'hysteresis_analysis.py'	The jupyter notebook code used to develop the plots in the manuscript and the package with the hysteresis analysis.

4.2. Description of data tables

4.2.1. 'Fulldat_PC.csv'

This file reports the physicochemical parameters measured for all grab samples.

Column header	Unit	Description
Sample_ID		Sample site and event identifier
campaign		Sampling event identifier

site_id		Sample site identifier
EC	µs/cm	Electrical conductivity
pH		pH value of water sample
level_num		Level in the mine where sample was collected
doc_mgL	mg/L	Dissolved organic carbon concentration
dic_mgL	mg/L	Dissolved inorganic carbon concentration
date	YYYY-MM-DD	Date of sample collection in year (Y), month (M), and day (D)
Mn_mgL	mg/L	Dissolved manganese concentration
Ni_mgL	mg/L	Dissolved nickel concentration
As_mgL	mg/L	Dissolved arsenic concentration
Cd_mgL	mg/L	Dissolved cadmium concentration
Pb_mgL	mg/L	Dissolved lead concentration
Fe_mgL	mg/L	Dissolved iron concentration
Al_mgL	mg/L	Dissolved aluminium concentration
Cu_mgL	mg/L	Dissolved copper concentration
Zn_mgL	mg/L	Dissolved zinc concentration
Q_mLs	mL/s	Flow rate extracted from mine drainage waters
PLI		Pollution load index value for each specific sample
Event		Flow phase identified during the time and year

4.2.2. 'Site2_autosampler_PC.csv'

This file reports the physicochemical parameters measured for all samples collected from the Autosampler positioned at site 2.

Column header	Unit	Description
Sample_ID		Sample site and event identifier
campaign		Sampling event identifier
site_id		Sample site identifier
EC	µs/cm	Electrical conductivity
pH		pH value of water sample

Group		Month in which sample was collected
doc_mgL	mg/L	Dissolved organic carbon concentration
dic_mgL	mg/L	Dissolved inorganic carbon concentration
date	YYYY-MM-DD	Date of sample collection in year (Y), month (M), and day (D)
Mn_mgL	mg/L	Dissolved manganese concentration
Ni_mgL	mg/L	Dissolved nickel concentration
As_mgL	mg/L	Dissolved arsenic concentration
Cd_mgL	mg/L	Dissolved cadmium concentration
Pb_mgL	mg/L	Dissolved lead concentration
Fe_mgL	mg/L	Dissolved iron concentration
Al_mgL	mg/L	Dissolved aluminium concentration
Cu_mgL	mg/L	Dissolved copper concentration
Zn_mgL	mg/L	Dissolved zinc concentration
Q_mLs	mL/s	Flow rate extracted from mine drainage waters

4.2.3. 'Site2_stratified_sampling_PC.csv'

This file reports the physicochemical parameters measured for samples collected manually from the top and bottom stratified waters of site 2.

Column header	Unit	Description
Sample_ID		Sample site and event identifier
campaign		Sampling event identifier
site_id		Sample site identifier
EC	µs/cm	Electrical conductivity
pH		pH value of water sample
level_num		Level in the mine where sample was collected
doc_mgL	mg/L	Dissolved organic carbon concentration
dic_mgL	mg/L	Dissolved inorganic carbon concentration
date	YYYY-MM-DD	Date of sample collection in year (Y), month (M), and day (D)
Mn_mgL	mg/L	Dissolved manganese concentration

Ni_mgL	mg/L	Dissolved nickel concentration
As_mgL	mg/L	Dissolved arsenic concentration
Cd_mgL	mg/L	Dissolved cadmium concentration
Pb_mgL	mg/L	Dissolved lead concentration
Fe_mgL	mg/L	Dissolved iron concentration
Al_mgL	mg/L	Dissolved aluminium concentration
Cu_mgL	mg/L	Dissolved copper concentration
Zn_mgL	mg/L	Dissolved zinc concentration
Event		Flow phase identified during the time and year
Group		Distinction between which layer sample was collected
Q_mLs	mL/s	Flow rate extracted from mine drainage waters (only collected from top waters)

4.2.4. 'Meteo.csv'

This file reports the meteorologic conditions measured at the weather station at the position of the Reiche Zeche mine access.

Column header	Unit	Description
date	YYYY-MM-DD	Date of sample collection in year (Y), month (M), day (D)
rsds	MJ/m^{2}day	Incoming shortwave radiation
tas	°C	Air temperature 2 m above ground
sfcWind	m/s	Wind velocity 10 m above ground
vabar	bar	Vapor pressure deficit calculated from daily max/min air temperature and max/min relative humidity
Prec	mm	Precipitation
EToPM	mm	Potential evapotranspiration calculated from rsds, tas, sfcWind, vabar using FAO56 Penman-Monteith
scPDSI	-	Self-Calibrating Palmer Drought Severity Index calculated based on 14d sums of Prec and EToPM [<-4 extreme drought, -1..1 normal .. +4 very wet] (interpolated to daily timesteps)
SMI	-	Soil Moisture Index monthly data (interpolated to daily timesteps) from UFZ Drought Monitor / Helmholtz Centre for Environmental Research https://www.ufz.de/index.php?en=37937 (Boeing et al.,

		2022) [<0.3 abnormally dry, <0.1 severe drought, <0.02 exceptional drought]

4.2.5. 'Flow.csv'

This file reports the water flow rate measured at the four locations with water loggers.

Column header	Unit	Description
time	YYYY-MM-DD HH:mm:SS	Time of sample collection in year (Y), month (M), day (D), hour (H), minute (m), and second (S), tz=UTC
Site 1	mL/s	Water flow rate measured hourly at renamed site 1
Site 2	mL/s	Water flow rate measured hourly at renamed site 2
Site 3B	mL/s	Water flow rate measured hourly at renamed site 3B
Site 3A	mL/s	Water flow rate measured hourly at renamed site 3A
HydPhase		Flow phase identified during the time

4.2.6. 'Isotopes.csv'

This file reports the stable water isotope (δ^2H and $\delta^{18}O$) values measured across all sites within the mine and from precipitation at the surface

Column header	Unit	Description
P		Sample Run Identifier
2H_m	‰	Mean δ^2H (deuterium) value (ref. VSMOW)
2H_cal	‰	Calibrated δ^2H value (ref. VSMOW)
2H_err		Error/uncertainty of asymptotic δ^2H fit (std)
18O_m	‰	Mean $\delta^{18}O$ (oxygen-18) value (ref. VSMOW)
18O_cal	‰	Calibrated $\delta^{18}O$ value (ref. VSMOW)
18O_err		Error/uncertainty of asymptotic $\delta^{18}O$ fit (std)
kx		Processing batch index during calibration
date	YYYY-MM-DD	Date of sample collection in year (Y), month (M), and day (D)
campaign		Sampling event identifier (only for mine water samples)
position		Sample site identifier (only for mine water samples)
dev_med		Deviation from median (only for mine water samples)
dev_med_pos		Positive deviation from median (only for mine water samples)

Identifier_mine1		Sample site and event identifier
Identifier_mine2		Sample event identifier
Identifier_precip1	EventDDMMYY	Rain or snow event and date in day (D), month (M), and year (Y)
Identifier_precip2		Sampling event (for mine sample collection) closest to the date of collection of precipitation sample
Group		Distinction between whether sample was collected from the mine water or precipitation from the surface

4.2.7. 'Hysteresis_analysis.csv'

This file reports the summary of the hysteresis analysis for the four main sites. Please see the respective code 'hysteresis_analysis.py' for details.

Column header	Unit	Description
site_id		Sample site identifier
compound		Specific element being analyzed
segment_id		Number of samples collected from start to end date at site id
start_date	YYYY-MM-DD	Start date and time in year (Y), month (M), and day (D)
end_date	YYYY-MM-DD	End date and time in year (Y), month (M), and day (D)
start_flow	mL/s	Water flow rate measured at start date
end_flow	mL/s	Water flow rate measured at end date
start_conc	mg/L	Dissolved concentration of compound at start date
end_conc	mg/L	Dissolved concentration of compound at end date
flow_difference		Difference in start and end flow
conc_difference		Difference in start and end concentration
behavior		Pattern based off of concentration-discharge relationships
hysteresis_index_lloyd		Hysteresis index value - method using a linear difference with concentration and discharge normalization
hysteresis_index_zuecco		Hysteresis index value - method using a geometric method with vector angle calculated between concentration and discharge normalized values

4.2.8. 'Site2_spectrolyzer_autosampler_prediction.csv'

This file reports the input measured data from the autosampler and output predicted results from cubist machine learning modeling used from the spectrolyzer data at site 2.

Column header	Unit	Description
date	YYYY-MM-DD	Date and time in year (Y), month (M), and day (D)
Fe_pred	mg/L	Dissolved iron concentration cubist model prediction
Zn_pred	mg/L	Dissolved zinc concentration cubist model prediction
Cd_pred	mg/L	Dissolved cadmium concentration cubist model prediction
Cu_pred	mg/L	Dissolved copper concentration cubist model prediction
Al_pred	mg/L	Dissolved aluminium concentration cubist model prediction
Pb_pred	mg/L	Dissolved lead concentration cubist model prediction
Mn_pred	mg/L	Dissolved manganese concentration cubist model prediction
Ni_pred	mg/L	Dissolved nickel concentration cubist model prediction
As_pred	mg/L	Dissolved arsenic concentration cubist model prediction
Sample_ID		Sampling bottle and event identifier
campaign		Sampling event identifier
site_id		Sample bottle identifier
EC	µs/cm	Electrical conductivity
pH		pH value of water sample
Group		Month in which sample was collected
doc_mgL	mg/L	Dissolved organic carbon concentration
dic_mgL	mg/L	Dissolved inorganic carbon concentration
Mn_mgL	mg/L	Dissolved manganese concentration (input for model)
Ni_mgL	mg/L	Dissolved nickel concentration (input for model)
As_mgL	mg/L	Dissolved arsenic concentration (input for model)
Cd_mgL	mg/L	Dissolved cadmium concentration (input for model)
Pb_mgL	mg/L	Dissolved lead concentration (input for model)
Fe_mgL	mg/L	Dissolved iron concentration (input for model)
Al_mgL	mg/L	Dissolved aluminium concentration (input for model)

Cu_mgL	mg/L	Dissolved copper concentration (input for model)
Zn_mgL	mg/L	Dissolved zinc concentration (input for model)
Q_mLs	mL/s	Water flow rate measured from top waters

4.2.9. 'Site2_spectrolyzer_raw.csv'

This file reports the compiled spectrolyzer data collected from waters at site 2.

Column header	Unit	Description
Date_Time	YYYY-MM-DDTHH:mm:SSZ	Date and time in year (Y), month (M), day (D), delimiter (T), hour (H), minute (m), second (S), and time zone designator (Z)
200.0 - 720.0	nm	Wavelength at which absorbance was measured
Serial_No		Spectrometer device serial number
DOCeq	eq	Dissolved organic carbon measured as equivalent weights
Flags_DOCeq		Notes of any source of error from DOCeq
TOCeq	eq	Total organic carbon measured as equivalent weights
Flags_TOCeq		Notes of any source of error from TOCeq
Turbidity	NTU	Measures haziness of fluid sample
Flags_Turbidity		Notes of any source of error from turbidity
Temperature	℃	Temperature of water
Flags_Temperature		Notes of any source of error from temperature
identifier	DD.MM.YY-HH:mm:SS Serial-No	Date and time in day (D), month (M), year (Y), hour (H), minute (m), second (S), and serial number of device

4.2.10. 'Site2_Cdspec_Qlogg_hourly.csv'

This file reports matched timestamps of Cadmium concentration values from cubist modeling and water flow rate values from the logger collected from waters at site 2.

Column header	Unit	Description
date	YYYY-MM-DDTHH:mm:SSZ	Date and time in year (Y), month (M), day (D), delimiter (T), hour (H), minute (m), second (S), and time zone designator (Z)
Q_mLs	mL/s	Flow rate extracted from mine water at site 2

Cd_pred	mg/L	Dissolved cadmium concentration cubist model prediction
time_index		Time points for each measurement from start to end

4.2.11. Scripts 'LegacyMine_ReicheZeche.ipynb' and 'hysteresis_analysis.py'

This file LegacyMine_ReicheZeche.ipynb contains a jupyter notebook to reproduce each plot in the manuscript with the data given in this repository. The hysteresis analysis is given as a separate package hysteresis_analysis.py. The software is developed and tested with Python 3.12, NumPy 2.0, Pandas 2.2.3 and Plotly 5.24.1.

5. References

Baumann, P. (2020), Simplerspec: Soil and Plant Spectroscopic Model Building and Prediction. https://github.com/philipp-baumann/simplerspec.

Boeing, F., Rakovec, O., Kumar, R., Samaniego, L., Schrön, M., Hildebrandt, A., Rebmann, C., Thober, S., Müller, S., Zacharias, S., Bogena, H., Schneider, K., Kiese, R., Attinger, S., and Marx, A. (2022), High-resolution drought simulations and comparison to soil moisture observations in Germany, Hydrol. Earth Syst. Sci., 26, 5137–5161, https://doi.org/10.5194/hess-26-5137-2022.

FGG Elbe. *Datenportal der Flussgebietsgemeinschaft Elbe – Schadstoffe in Wasserphase*. https://www.elbe-datenportal.de/FisFggElbe/content/auswertung/UntersuchungsbereichChemWas_export_tabelle.action (accessed 11 July 2025).

Hachgenei, N., Vaury, V., Nord, G., Spadini, L., and Duwig, C. (2022), Faster and more precise isotopic water analysis of discrete samples by predicting the repetitions' asymptote instead of averaging last values, MethodsX, 9, 101656, https://doi.org/10.1016/j.mex.2022.101656.

Henderson, F. M. (1996), Open channel flow, New York: Macmillan Company.

Jahan, S. and Strezov, V. (2018). Comparison of pollution indices for the assessment of heavy metals in the sediments of seaports of NSW, Australia. Mar. Pollut. Bull. 128, 295–306. https://doi.org/10.1016/j.marpolbul.2018.01.036.

Kuhn, M. and Quinlan, R. (2023), Cubist: Rule- and Instance-Based Regression Modeling. https://CRAN.R-project.org/package=Cubist.

Kuhn, M. (2008), Building Predictive Models in r Using the Caret Package. *Journal of Statistical Software* 28 (5). https://doi.org/10.18637/jss.v028.i05.

Lloyd, C. E. M., Freer, J. E., Johnes, P. J. and Collins, A. L. (2016), Technical Note: Testing an improved index for analysing storm discharge–concentration hysteresis. Hydrol. Earth Syst. Sci. 20, 625–632.

Mischo, U.-P. D.-I. H., Eng, P., Barakos, D.-I. G., Szkliniarz, K. and Kisiel, J. (2021), Site Description and data of the Reiche Zeche Site Services, Characteristics and Data. http://tu-freiberg.de/lfbw.

Palmer, W.C. (1965), Meteorological Drought. U.S. Weather Bureau Research Paper No. 5.

Tomlinson, D.L., Wilson, J.G., Harris, C.R., and Jeffrey, D.W. (1980). Problems in the assessment of heavy-metal levels in estuaries and the formation of a pollution index. Helgoländer Meeresunters. 33, 566–575. https://doi.org/10.1007/bf02414780.

Wells, N., Goddard, S., and Hayes, M. J. (2004), A Self-Calibrating Palmer Drought Severity Index, J. Clim., 17, 2335–2351, https://doi.org/10.1175/1520-0442(2004)017<2335:aspdsi>2.0.co;2.

Zhiteneva, V., Brune, J., Mischo, H., Weyer, J., Simon, A., and Lipson, D. (2016), Reiche Zeche Mine Water Geochemistry. Proceedings IMWA 2016, Freiberg/Germany.

Zuecco, G., Penna, D., Borga, M. & Meerveld, H. J. van. (2016), A versatile index to characterize hysteresis between hydrological variables at the runoff event timescale. Hydrol. Process. 30, 1449–1466.