

Title

Spatial Study 2021: Sample-Based Surface Water Chemistry and Organic Matter Characterization across Watersheds in the Yakima River Basin, Washington, USA (v3)

Summary

This dataset supports a broader study examining the drivers of spatial variability in sediment respiration rates in the Yakima River Basin. The dataset provides geochemistry and organic matter characterization data generated from samples collected during the same two-week period at 47 sites within multiple rivers throughout the Yakima River Basin in Washington, USA. Related sensor data are published at <https://data.ess-dive.lbl.gov/datasets/doi:10.15485/1892054>.

This data package was originally published September 2022. It was updated May 2023 (modified files) and November 2024 (new and modified files). See the change history section below for more details.

We thank the United States Forest Service, Washington Department of Natural Resources, Washington Department of Fish and Wildlife, Washington State Parks, Confederated Tribes and Bands of the Yakama Nation, and Cowiche Canyon Conservancy for access to field locations where these samples were collected. We also thank the Yakama Nation Tribal Council and Yakama Nation Fisheries for working with us to facilitate sample collection and optimization of data usage according to their values and worldview.

Critical Details

- 1) The SPE aliquot used for the BPCA analysis is the same sample used in Roebuck et al. (2023, <https://data.ess-dive.lbl.gov/datasets/doi:10.15485/1995543>)
- 2) In version 3 of this data package, previous ion data generated from water samples was removed due to concerns about instrument quality. In its place, a reduced amount of ion data is published in version 3 of the data package. Only one replicate was analyzed for ions of interest (nitrate, sulfate, and chloride). In most cases, replicate number 1 was used, but if that was unavailable, replicate number 2 was used (-1 or -2 appended to sample name, respectively). The water samples were collected in 2021 and were analyzed in 2024 for this new ion data, which surpasses the typical hold time (6 months- 1 year) and the samples have undergone multiple freeze-thaw cycles, which we typically try to avoid (see the methods codes for more information). However, we found that the magnitude of the nitrate values were similar when compared to the original run.
- 3) Formularity, the software used to assign molecular formulas to peaks in FTICR-MS data, has changed its name to Formultitude, and can be referenced via <https://github.com/PNNL-Comp-Mass-Spec/Formultitude>. This documentation within this data package has not been updated to reflect the new name.

Brief Overview of Methods

Surface water samples were collected at 47 sites by 2-person teams in multiple rivers within in the Yakima River Basin in late August through mid-September 2021. Site metadata, including general environmental information, were manually recorded in the field. For complete information regarding data collection, see the Sampling Protocol for details on water sample and site metadata collection. Surface water samples were processed and analyzed in the laboratory after field collection. For details regarding laboratory methods, see the alphanumeric methods codes located in the header rows of the chemistry data csv file and their associated definitions in SPS_Methods_Codes.csv.

Data Package Structure

This dataset is comprised of one main data folder containing (1) file-level metadata; (2) data dictionary; (3) field metadata; (4) dissolved inorganic carbon (DIC), dissolved organic carbon (DOC; reported as non-purgeable organic carbon; NPOC), total nitrogen (TN), total suspended solids (TSS), ions, and benzene polycarboxylic acid (BPCA) concentration and stable isotope data; (5) averaged values from water chemistry data; (6) surface water sampling protocol; (7) sensor protocol (8) readme; (9) methods codes; (10) international generic-sample number (IGSN) mapping file; and (11) folder of high resolution characterization of organic matter via 12 Tesla Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) through the Environmental Molecular Sciences Laboratory (EMSL; <https://www.pnnl.gov/environmental-molecular-sciences-laboratory>). This folder contains two subfolders, one containing the .xml data files and the other containing instructions for using Formultitude (<https://github.com/PNNL-Comp-Mass-Spec/Formultitude>) and an R script to process the data based on the user's specific needs. All files are .csv, .pdf, .R, .ref, or .xml.

Acknowledgements

This research was supported by the U.S. Department of Energy (DOE) Biological and Environmental Research (BER) Environmental System Science (ESS) program (<https://ess.science.energy.gov/>) through the Pacific Northwest National Laboratory River Corridor Science Focus Area (SFA). PNNL is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract No. DE-AC05-76RL01830.

FTICR-MS data were analyzed on User Proposal 60221 at the Environmental Molecular Sciences Laboratory, a DOE Office of Science User Facility sponsored by the Biological and Environmental Research program under Contract No. DE-AC05-76RL01830 (EMSL; <https://ror.org/04rc0xn13>).

Contact

James Stegen, james.stegen@pnnl.gov

Change History

Approach to change history and versioning:

Updates to **data package** version: When any file within a data package is updated, the data package version number is updated. The data package version number is indicated in the title of the data package, the data package folder name, and in the change history table below. You can access previous versions of the data package by sending a request to ESS-DIVE.

Updates to **individual file** versions: As files are changed, the file version number is also updated. The file version number is indicated in the file name, file level metadata (flmd) file, and the change history table below. The version number on an individual file may not match the version number of the data package. For example, v3 of a data package may include v2 of an individual file.

The change history below describes each file revised during versioning. If you are interested in seeing the exact cells within a file that have changed, you can utilize the daff package in R (<https://github.com/edwindj/daff>) to compare a previously downloaded file to a newly downloaded file.

In the change history table below, the sub-headers and bullets indicate the type of change in each file:

- New files: Describes new files added that were not present in previous data package versions
- Bulk changes to files: Describes a change to many files within the data package. The indicated superscript will be added to each file name that the change applies to.

- Modified files:
 - Corrected: Describes existing information modified or removed to prevent sharing of incorrect information
 - Added: Describes new information inserted into an existing file (e.g., appending new columns/rows)
 - Updated: Describes modifying existing information to maintain accuracy though version changes. (e.g., changing version number to new version number)

Change history:

Data Package Version	Changes
Version 1 <i>September 2022</i>	Original data package publication
Version 2 <i>May 2023</i>	<p><u>MODIFIED FILES</u></p> <p>SPS_DIC.csv (v2)</p> <ul style="list-style-type: none"> • Corrected values to remove original QAQC approach (blank correction) <p>SPS_Ions.csv (v2)</p> <ul style="list-style-type: none"> • Corrected alphanumeric methods codes to indicate that the values are not blank corrected. The ion data (original v1 and current v2) were not blank corrected. <p>SPS_NPOC_TN.csv (v2)</p> <ul style="list-style-type: none"> • Corrected values to remove original QAQC approach (blank correction) <p>SPS_NPOC_TN_DIC_TSS_Ions_Summary.csv (v2)</p> <ul style="list-style-type: none"> • Corrected values to remove original QAQC approach (blank correction) <p>SPS_Methods_Codes (v2)</p> <ul style="list-style-type: none"> • Corrected ion method alphanumeric codes to remove indication of blank corrections. The ion data (original v1 and current v2) were not blank corrected. <p>SPS_Sample_flgmd.csv (v2)</p> <ul style="list-style-type: none"> • Updated version number of modified files. <p>readme_SPS_Spatial_Sample.pdf (v2)</p> <ul style="list-style-type: none"> • Updated change history • Updated version number in data package title.
Version 3 <i>November 2024</i>	<p><u>NEW FILES</u></p> <ul style="list-style-type: none"> • SPS_Water_B5CA_B6CA_Conc_Isotopes.csv <p><u>BULK CHANGES TO FILES</u></p> <p>¹ indicates the sample material type was added to the file name. ² indicates a column was added with international generic sample numbers (IGSN)</p> <p><u>MODIFIED FILES</u></p> <p>SPS_Water_Ions.csv (v3)^{1, 2}</p> <ul style="list-style-type: none"> • Corrected all values. See critical details and methods code for more information. <p>SPS_Water_NPOC_TN.csv (v3)^{1, 2}</p> <ul style="list-style-type: none"> • Corrected the Methods_Deviations. • Added outlier Methods_Deviation. • Updated the data flag.

SPS_Water_DIC.csv (v3)^{1,2}

- Added data from additional samples.

SPS_Water_Sample_Data_Summary.csv (v3)^{1,2}

- Updated file name. File previously named SPS_NPOC_TN_DIC_TSS_Ions_Summary.csv.
- Corrected summarized values to correctly exclude any outliers.
- Added summarized values for data added.

SPS_Methods_Codes (v3)

- Updated to reflect changes.

SPS_Sample_flgnd.csv (v3)

- Updated version number of modified files.
- Updated file descriptions.

SPS_Sample_dd.csv (v2)

- Updated to reflect changes.

readme_SPS_Spatial_Sample.pdf (v3)

- Updated change history
- Updated version number in data package title.

SPS_Water_FTICR_Methods.csv (v2)^{1,2}

- Updated the file name. File previously named SPS_FTICR.csv

SPS_Water_TSS.csv (v2)^{1,2}