Title

Spatial Study 2022: Water Column, Sediment, and Total Ecosystem Respiration Rates across the Yakima River Basin, Washington, USA (v2)

Summary

This dataset supports a broader study examining the drivers of spatial variability in sediment respiration rates in the Yakima River Basin and is associated with the manuscript "Sediment-associated processes account for most of the spatial variation in ecosystem respiration in the Yakima River basin" submitted to Nature Communications Earth & Environment (Garayburu-Caruso et al., in review). The dataset provides ecosystem metabolism estimates generated from streamMetabolizer (Appling et al.; 2018) using data collected during the same five-week period at 48 sites within multiple rivers throughout the Yakima River Basin in Washington, USA. Additionally, it includes the scripts used for the analysis and producing the figures in the manuscript. The contents include streamMetabolizer inputs and outputs and additional relevant data needed to generate the main manuscript results. The data included are: total ecosystem respiration, water respiration, calculated sediment-associated respiration, gross primary production outputs from the river corridor model for the Yakima River Basin, median grain size (d50), depth, dissolved oxygen, water temperature, pressure, and annual oxygen consumption. The associated GitHub repository can be found at https://github.com/river-corridors-sfa/SSS_metabolism.

Samples collected during this study were labeled as "Second Spatial Study" or "SSS." Raw time series sensor data, total suspended solids, and depth data from SSS were published at https://data.ess-dive.lbl.gov/datasets/doi:10.15485/1969566. A subset of data from the SSS samples were published in the contiguous United States (CONUS)-Scale Model-Sample (CM) study data package available at https://data.ess-dive.lbl.gov/view/doi:10.15485/1923689 that presents data from across the CONUS. They include dissolved organic carbon (DOC, measured as non-purgeable organic carbon, NPOC), total nitrogen (TN), grain size, aerobic sediment respiration, dissolved oxygen (DO), and temperature. Parent IDs and Site IDs are consistent between the SSS and CM data packages, and they can be mapped directly so data across packages can be used together. Field metadata for the samples in this data package can be found at either link.

Brief Overview of Methods

In situ dissolved oxygen, temperature, and water pressure sensors were deployed at 48 sites by 2-person teams in multiple rivers within the Yakima River Basin during the week of July 25-28, 2022. Barometric pressure dataloggers were deployed at a subset of the sites at that time. Triplicate 2-hour dark-bottle stream water incubations were carried out and the datalogging sensors were checked and cleaned at the same 48 sites by 2-person teams during the week of August 8-12, 2022. At most sites, reach-average depth data was also collected during the week of August 8-12, 2022 by wading the stream, taking multiple manual measurements along multiple transects. When this wading depth method was not feasible, depth was measured via sonar from a kayak or jetboat, or calculated using public models or gauge data. In situ sensors from July deployments were retrieved by 2-person teams during the week of August 29-September 1, 2022. As a result, the logged deployments consist of approximately one month. The inverse model streamMetabolizer (Appling et al.; 2018) was used to estimate a deployment-period-average ecosystem respiration rate, gross primary production, and gas exchange velocity (K600) for each site. D50 was estimated as described in Gomez-Velez et al. (2015). Model-predicted HZ respiration values used in this manuscript were previously published in Son et al., (2023), and the data is publicly available on ESS-DIVE at https://data.ess-dive.lbl.gov/datasets/doi:10.15485/. Sediment-associated respiration was calculated as the difference

between total ecosystem respiration and water column respiration. To explore how physical variables covary within sediment-associated respiration at different sites, a LASSO (Least Absolute Shrinkage and Selection Operator) regression model was built using physical, chemical and environmental variables as inputs, and sediment-associated respiration as the target variable. See the associated manuscript for more details.

Critical Details

1 – Each physical site has a "Site_ID" and each sampling event in time at that Site has a "Parent_ID" (i.e., in 2021, Site_ID T07 was sampled and the resulting Parent_ID of the samples was SPS_0053. In 2022, Site_ID T07 was sampled again and the resulting Parent_ID of the samples was SSS013). The files indicate both the Site_ID and the Parent_ID to indicate at which Site the sensor was deployed (Site_ID) and to which sampling event the data is most relevant (Parent_ID). The Parent_ID can be used to match sample data from the CM and SSS data packages to the data presented in this data package. The identifier most relevant to the data user will depend on the specific analyses being done. If the user is interested in other data from the same sites collected at other times, they can access at https://data.ess-dive.lbl.gov/datasets/doi:10.15485/1898914, https://data.ess-dive.lbl.gov/datasets/doi:10.15485/1892052, https://data.ess-dive.lbl.gov/datasets/doi:10.15485/1898912, and https://data.ess-dive.lbl.gov/datasets/doi:10.15485/1892054

2 – The manual chamber data used to calculate water column respiration, the depth data used in the stream metabolizer inputs, and input data for the LASSO are from https://data.ess-dive.lbl.gov/datasets/doi:10.15485/1969566 (Delgado et al., 2023). Additional input data for the LASSO are from https://data.ess-dive.lbl.gov/view/doi:10.15485/1923689 (Forbes et al., 2023).

3 – The annual oxygen consumption was calculated using data from https://data.ess-dive.lbl.gov/datasets/doi:10.15485/1962818 (Son et al., 2023). See "create_O2_consumption_rates_based_model_outputs.R" for more details.

Data Package Structure

This dataset is comprised of one main data folder with four subfolders. The main data folder contains of (1) file-level metadata; (2) data dictionary; (3) total/water column/sediment respiration; (4) gross primary production (GPP); (5) median grain size (d50); and (6) annual oxygen consumption. The "Figures" subfolder contains the figures used in the paper and all intermediate files (including geospatial files). The "Published_Data" contains a readme directing the user to download the public data to reproduce analyses and figures. The "Scripts" folder contains all scripts used in the analyses that were not part of running StreamMetabolizer. Lastly, the "Stream_Metabolizer" folder contains all files associated with running StreamMetabolizer including (1) model input files, (2) model output files, (3) processing scripts, (4) histogram plots of the outputs, and (5) an R project. All files are .csv, .pdf, .R, .Rmd, .Rproj, .html, .png, .txt, .qgz, .cpg, .dbf, .prj, .shp, .shp.ea.iso.xml, .shp.iso.xml, .shx, .sbn,

Citations, Acknowledgements, and License

Acknowledgements

We acknowledge the Yakama Nation as owners and caretakers of the lands where we collected these data. We thank the Confederated Tribes and Bands of 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.

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.

Citations

- Kaufman M H ; Barnes M ; Chen X ; Forbes B ; Garayburu-Caruso V A ; Goldman A E ; Myers-Pigg A ; Powers-McCormack B ; Stegen J C ; Scheibe T D (2023): Geospatial Information, Metadata, and Maps for Global River Corridor Science Focus Area Sites (v3). River Corridor and Watershed Biogeochemistry SFA, ESS-DIVE repository. Dataset. doi:10.15485/1971251
- Son K ; Fang Y ; Gomez-Velez J D ; Chen X (2023): Model Inputs, Outputs, and Scripts associated with: "Spatial microbial respiration variations in the hyporheic zones within the Columbia River Basin". River Corridor and Watershed Biogeochemistry SFA, ESS-DIVE repository. Dataset. doi:10.15485/1962818 accessed via https://data.ess-dive.lbl.gov/datasets/doi:10.15485/1962818
- Forbes B ; Barnes M ; Boehnke B T ; Bowden M E ; Chen X ; Cornwell K ; Crawford M ; Delgado D ; Fulton S G ; Garayburu-Caruso V A ; Gary S ; Goldman A E ; Gonzalez B I ; Grieger S ; Hammond G E ; Jiang P ; Kaufman M H ; Laan M ; Li B ; Li Z ; McKever S A ; Mudunuru M K ; Muller K A ; Myers-Pigg A ; Ocejo J A ; Otenburg O ; Pelly A ; Peta K ; Powers-McCormack B ; Regier P ; Renteria L ; Roebuck A ; Scheibe T D ; Son K ; Tfaily M M ; Torgeson J M ; Stegen J C ; WHONDRS Consortium T (2023): WHONDRS River Corridor Sediment and Water Geochemistry and In Situ Sensor Data from Machine-Learning-Informed Sites across the Contiguous United States (v4). River Corridor and Watershed Biogeochemistry SFA, ESS-DIVE repository. Dataset. doi:10.15485/1923689
- Delgado D ; Barnes M ; Boehnke B T ; Chen X ; Chen Y ; Cornwell K ; Forbes B ; Fulton S G ; Garayburu-Caruso V A ; Goldman A E ; Gonzalez B I ; Grieger S ; Hammond G E ; Jiang P ; Kaufman M H ; Laan M ; Li B ; Li Z ; Lin X ; McKever S A ; Mudunuru M K ; Muller K A ; Myers-Pigg A ; Otenburg O ; Pelly A ; Peta K ; Powers-McCormack B ; Regier P ; Renteria L ; Roebuck A ; Scheibe T D ; Son K ; Torgeson J M ; Zheng J ; Stegen J C (2023): Spatial Study 2022: Surface Water Samples, Cotton Strip Degradation, and Hydrologic Sensor Data across the Yakima River Basin, Washington, USA (v3). River Corridor and Watershed Biogeochemistry SFA, ESS-DIVE repository. Dataset. doi:10.15485/1969566

License

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 with 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 *June 2023*	Original data package publication
Version 2 *October 2025*	**NEW FILES** • *_Temp_DO_Press_Depth_Plot.html • histogram_*.png • Stream_Metabolizer.Rproj • Figure*.pdf • Figure*.png • Figure6-Cluster_Map.qgz • cb_2018_us_state_20m.* • CB_catchments_clean_joined.* • Columbia_Basin_Watershed_Boundary.* • Yakima_River_Basin_Flowlines.* • Yakima_River_Basin.* • YRB_Cluster.* • ERwc_ERtot_Scatter.pdf • ERsed_ERtot_Scatter.pdf • ERwc_Contribution.pdf • ERsed_Map.pdf • ERsed_Map_Zscores.pdf • ERhz_Map_Zscores.pdf

- ERsed_Contribution.pdf
- ERwc_Map.pdf
- ERhz_Map.pdf
- ERtot_Map.pdf
- readme_Published_Data.txt
- Calculate_ERwc_ERsed.R
- create_O2_consumption_rates_based_model_outputs.R
- Create_SM_Input_Files.R
- ERsed_Lasso.R
- ERsed_Manuscript_Figures.R
- ERsed_Manuscript_Maps.R

BULK CHANGES TO FILES

[1] indicates files were modified due to the corrections and updates to streamMetabolizer inputs and outputs. Changes include:

1. correcting the dissolved oxygen (DO; mg/L) data by recalculating the offset that is applied to it
2. correcting the DO percent saturation data by using the streamMetabolizer::calc_DO_sat() function
3. correcting K600 by dividing by depth to convert to the correct units
4. correcting site IDs for parent IDs SSS001, SSS004, SSS006, and SSS022
5. updating model parameters
6. files were moved within the Stream_Metabolizer folder

MODIFIED FILES

SSS_ER_Installation_Methods.csv (v2)
- Updated methods information

readme_SSS_ER.pdf (v2)
- Updated version number in data package title
- Updated data package structure to include new files
- Added new versioning information.

SSS_ER_flmd.csv (v2)
- Added rows for new files.
- Added version number to updated files.

SSS_ER_dd.csv (v2)
- Added rows for new column headers from new files

***_Temp_DO_Press_Depth.csv** (v2) [1]
SSS_K600.csv (v2) [1]
SSS_Slope_Discharge_Velocity.csv (v2) [1]
SSS_combined_SM_results.csv (v2) [1]
***_SM_final_daily_prediction_results.csv** (v2) [1]
***_SM_final_full_prediction_results.csv** (v2) [1]
***_SM_final_instant_fit_results.csv** (v2) [1]
***_SM_final_overall_fit_results.csv** (v2) [1]
***_SM_output.HTML** (v2) [1]
K_estimation_final.Rmd (v2) [1]
SM_results_analysis_final.Rmd (v2) [1]

	SSS_SM_final_template.Rmd (v2) [1] **SSS_SM_loop.R** (v2) [1]

Note: An asterisks (*) in a file name indicates multiple files with the same file and file naming structure. See the flmd for more details about specific files and the meaning of the asterisks.