Title

Data and Scripts Associated with the Manuscript "Yakima River Basin Water Column Respiration is a Minor Component of River Ecosystem Respiration"

Summary

This data package is associated with the publication "Yakima River Basin Water Column Respiration is a Minor Component of River Ecosystem Respiration" submitted to EGU Biogeochemistry (Fulton et al. 2024). In this research, water column respiration (ERwc) data, surface water chemistry data, organic matter (OM) chemistry data, and publicly available geospatial data were used in a multiple linear regression model to evaluate the drivers of spatial variability in ERwc at 47 sites across the Yakima River basin in Washington, USA.

Brief Overview of Methods

The sensor data and surface water chemistry data are associated with the previously published 2021 spatial study data packages (Fulton et al. 2022; Grieger et al. 2022); the raw data and associated metadata (including geographic locations, methods information, and protocols) can be found within each associated data package. For full references for these data packages, refer to the Citations and Acknowledgment section below. Water column respiration (ERwc) data and surface water chemistry samples were collected in triplicate at 47 sites in the Yakima River basin distributed across Strahler stream orders 2 through 7 and different hydrological and biophysical settings during summer baseflow conditions in 2021. To test the transferability of study results to catchments throughout the Columbia River basin, we used cluster analysis to group catchments in the Columbia River basin into six classes sharing similar landscape characteristics using key biophysical and hydrological attributes selected from readily available spatial datasets. The results were used to guide site selection. Watershed characteristics and surface water chemistry data were then input to a multiple linear regression model to determine which explanatory variables best explained spatial variation in ERwc across the Yakima River basin. Please refer to Methods section in the publication for more information on the data and analyses described in the publication.

Critical Details

1) This data package is associated with the public GitHub repository "Yakima River Basin Water Column Respiration" found at https://github.com/river-corridors-sfa/YRB_Water_Column_Respiration. The repository contains all of the code (R scripts) and data to run the analyses, interpret the results, and plot the figures (excluding Figure 2) as presented in the publication. Please see the Methods section in the publication for more information on the data processing steps used to run the statistical analyses for this research.

2) The sensor data collection field protocol and dissolved oxygen sensor dataset used to generate water column respiration rates (ERwc) for the Yakima River basin (this study) are from the published dataset described in Fulton et al. (2022) (https://data.ess-dive.lbl.gov/view/doi:10.15485/1892052).

3) The surface water chemistry sample collection field protocol and water chemistry dataset (including organic matter (OM) chemistry data) used in the multiple linear regression analysis to evaluate the primary drivers of spatial variation in water column respiration rates (ERwc) across the Yakima River

basin (this study) are from the published dataset described in Grieger et al. (2022) (https://data.ess-dive.lbl.gov/view/doi:10.15485/1898914).

4) Please see Methods sections 2.4 and 2.5 of the publication for detailed information on the analytical methods used to characterize the surface water chemistry and OM chemistry samples, respectively.

5) The OM biogeochemical transformation analysis was performed following methods described in Garayburu-Caruso et al. (2020) (https://doi.org/10.3390/metabo10120518).

6) Please see the Methods section in the publication for more information on the sources for the publicly available data that were used in the cluster analysis to group Columbia River basin catchments into classes sharing key biophysical and hydrological characteristics.

7) The Appling et al. (2018b and 2018c) stream metabolism dataset was used to generate reach-scale ecosystem respiration (ERtot) data in volumetric units to compare to the water column respiration rate. Data from this research can be found at https://www.sciencebase.gov/catalog/item/59bff507e4b091459a5e0982.

8) The Bernhardt et al. (2022) StreamPULSE dataset was used to subset out 208 sites from the Appling et al. (2018c) that had been through a rigorous quality assurance/quality control process. The StreamPULSE dataset can found at https://data.streampulse.org/download_bulk.

Data Package Structure

The data package includes the data inputs, and outputs, and R scripts to calculate descriptive statistics, run the multiple linear regression models, analyze and interpret the results, and create manuscript figures. The data package is comprised of three main folders (Code, Data, and Figures). The Code folder is comprised of five analysis-specific subfolders that contain the R scripts to perform the analyses described in the publication and create publication figures. The Data folder is comprised of six subfolders that contain R script-specific data input and output files. Related code and data subfolders use the same naming convention to easily identify the data input files associated with each R script. The Figures folder includes three map figures from the manuscript in ".pdf" and ".png" format. This data package also includes a file-level metadata (flmd) csv, and a data dictionary (dd) csv, and readme pdf. Please see "Fulton_2024_Water_Column_Respiration_Data_Package_flmd.csv" for a list of all files contained in the data package, along with descriptions for each for each file. The data dictionary (Fulton_2024_Water_Column_Respiration_Data_Package_dd.csv) describes the column headers within any tabular data file (e.g., csv). The readme (Fulton_2024_Water_Column_Respiration_Data_Package_flmd.csv) describes the data package, methods, critical details, etc. This data package is associated with a GitHub repository which can be found at https://github.com/river-corridors-sfa/YRB_Water_Column_Respiration/.

Citations and Acknowledgements

Cite this data package with the appropriate DOI. Cite the associated manuscript in any work that that uses analyses or conclusions presented in the manuscript. To cite the paper:

Fulton, S. G., Barnes, M., Borton, M. A., Chen, X., Farris, Y., Forbes, B., Garayburu-Caruso, V. A., Goldman, A. E., Grieger, S., Hall Jr., R., Kaufman, M. H., Lin, X., McCann, E., McKever, S. A., Myers-Pigg, A., Otenburg, O. C., Pelly, A. C., Ren, H., Renteria, L., Scheibe, T. D., Son, K., Tagestad, J., Torgeson, J. M., and Stegen, J. C.: Yakima River Basin Water Column Respiration is a Minor Component of River Ecosystem Respiration, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-3038, 2024.

Citations

- Agarwal, D., Cholia, S., Hendrix, V. C., Crystal-Ornelas, R., Snavely, C., Damerow, J., & Varadharajan. (2022). ESS-DIVE Reporting Format for Dataset Metadata. Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE), ESS-DIVE repository. https://doi.org/10.15485/1866026
- Appling, A. P., Hall, R. O., Yackulic, C. B., and Arroita, M., 2018a. Overcoming Equifinality: Leveraging Long Time Series for Stream Metabolism Estimation, Journal of Geophysical Research: Biogeosciences, 123, 624-645. https://doi.org/10.1002/2017jg004140
- Appling, A. P., Read, J. S., Winslow, L. A., Arroita, M., Bernhardt, E. S., Griffiths, N. A., Hall, R. O., Jr., Harvey, J. W., Heffernan, J. B., Stanley, E. H., Stets, E. G., and Yackulic, C. B., 2018b. The metabolic

regimes of 356 rivers in the United States, Scientific Data, 5, 180292.
https://doi.org/10.1038/sdata.2018.292

- Appling, A. P., Read, J. S., Winslow, L. A., Arroita, M., Bernhardt, E. S., Griffiths, N. A., Hall, R. O., Jr., Harvey, J. W., Heffernan, J. B., Stanley, E. H., Stets, E. G., and Yackulic, C. B., 2018c. Metabolism estimates for 356 U.S. rivers (2007-2017): U.S. Geological Survey data release [dataset]. https://doi.org/10.5066/F70864KX

- Bernhardt, E. S., Savoy, P., Vlah, M. J., Appling, A. P., Koenig, L. E., Hall, R. O., Arroita, M., Blaszczak, J. R., Carter, A. M., Cohen, M., Harvey, J. W., Heffernan, J. B., Helton, A. M., Hosen, J. D., Kirk, L., McDowell, W. H., Stanley, E. H., Yackulic, C. B., and Grimm, N. B., 2022. Light and flow regimes regulate the metabolism of rivers, Proceedings of the National Academy of Sciences, 119, e2121976119. https://doi.org/10.1073/pnas.2121976119

- Fulton, S. G., Barnes, M., Borton, M. A., Chen, X., Farris, Y., Forbes, B., Garayburu-Caruso, V. A., Goldman, A. E., Grieger, S., Hall Jr., R., Kaufman, M. H., Lin, X., McCann, E., McKever, S. A., Myers-Pigg, A., Otenburg, O. C., Pelly, A. C., Ren, H., Renteria, L., Scheibe, T. D., Son, K., Tagestad, J., Torgeson, J. M., and Stegen, J. C.: Yakima River Basin Water Column Respiration is a Minor Component of River Ecosystem Respiration, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-3038, 2024

- Fulton, S. G., Barnes, M., Borton, M. A., Chen, X., Farris, Y., Forbes, B., Garayburu-Caruso, V. A., Goldman, A. E., Grieger, S., Kaufman, M. H., Lin, X., McKever, S. A., Myers-Pigg, A., Otenburg, O., Pelly, A., Ren, H., Renteria, L., Scheibe, T. D., Son, K., Torgeson, J. M., and Stegen, J. C., 2022. Spatial Study 2021: Sensor-Based Time Series of Surface Water Temperature, Specific Conductance, Total Dissolved Solids, Turbidity, pH, and Dissolved Oxygen from across Multiple Watersheds in the Yakima River Basin, Washington, USA (v2) [dataset]. https://doi.org/10.15485/1892052

- Garayburu-Caruso, V. A., Danczak, R. E., Stegen, J. C., Renteria, L., McCall, M., Goldman, A. E., Chu, R. K., Toyoda, J., Resch, C. T., Torgeson, J. M., Wells, J., Fansler, S., Kumar, S., and Graham, E. B., 2020. Using Community Science to Reveal the Global Chemogeography of River Metabolomes, Metabolites, 10, 518. https://doi.org/10.3390/metabo10120518

- Grieger, S., Barnes, M., Borton, M. A., Chen, X., Chu, R., Farris, Y., Forbes, B., Fulton, S. G., Garayburu-Caruso, V. A., Goldman, A. E., Gonzalez, B. I., Kaufman, M. H., McKever, S. A., Myers-Pigg, A., Otenburg, O., Pelly, A., Renteria, L., Scheibe, T. D., Son, K., Torgeson, J. M., Toyoda, J. G., and Stegen, J. C., 2022. Spatial Study 2021: Sample-Based Surface Water Chemistry and Organic Matter Characterization across Watersheds in the Yakima River Basin, Washington, USA (v2) [dataset]. https://doi.org/10.15485/1898914

- Kaufman, M. H., Barnes, M., Chen, X., Forbes, B., Garayburu-Caruso, V. A., Goldman, A. E., Stegen, J. C., Myers-Pigg, A., and Scheibe, T. D., 2023. Geospatial Information, Metadata, and Maps for Global River Corridor Science Focus Area Sites (v2) [dataset]. https://doi.org/10.15485/1971251

- McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., and Rea, A., 2012. NHDPlus Version 2: User Guide.

- Velliquette, T., Welch, J., Crow, M., Devarakonda, R., Heinz, S., Crystal-Ornelas, R., 2021. ESS-DIVE Reporting Format for Comma-separated Values (CSV) File Structure. Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE), ESS-DIVE Repository. https://doi.org/10.15485/1734841

- Velliquette, T., Welch, J., Crow, M., Devarakonda, R., Heinz, S., Crystal-Ornelas, R., 2021. ESS-DIVE Reporting Format for File-level Metadata. Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE), ESS-DIVE Repository. https://doi.org/10.15485/1734840
- Wieczorek, M. E., Jackson, S. E., and Schwarz, G. E., 2018. Select Attributes for NHDPlus Version 2.1 Reach Catchments and Modified Network Routed Upstream Watersheds for the Conterminous United States (ver. 3.0, January 2021), https://doi.org/10.5066/F7765D7V

Contact

James Stegen, james.stegen@pnnl.gov

Change History

Data Package Version	Changes
Version 1 *January 2024*	Original data package publication