Supplementary Information

Wildfires drive multi-year water quality degradation over the western United States

Carli P. Brucker[1,2,3], Ben Livneh[1,2], Fernando L. Rosario-Ortiz[1,4], Fangfang Yao[5], A. Park Williams[6], William C. Becker[4,7], Stephanie K. Kampf[8], and Balaji Rajagopalan[1]*

[1]Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder, Boulder, CO, 80309, United States

[2]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, 80309, United States

[3]Carollo Engineers, 390 Interlocken Crescent # 800, Broomfield, CO 80021, United States

[4]Environmental Engineering Program, University of Colorado Boulder, Boulder, CO, 80309, United States

[5]Department of Environmental Science, University of Virginia, Charlottesville, VA, 22904, United States

[6]Department of Geography, University of California Los Angeles, Los Angeles, CA, 90095, United States

[7]Hazen and Sawyer, 498 Seventh Avenue, 11th Floor, New York, NY, 10018, United States

[8]Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO, 80523, United States

*Corresponding author email: carli.brucker@gmail.com

This file includes:

Supplementary Figures

Supplementary Tables

Supplementary Methods

Data sources and pre-processing

Empirical water quality data, as well as remotely sensed geospatial data, climate variables, and wildfire information were first compiled across the western United States from 1974-2022. Data was filtered for sites within 11 states: Arizona, California, Colorado, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington and Wyoming. These data were downloaded, filtered, and pre-processed to prepare them for overlaying and site identification processes, as well as for use as covariates in model building. Pre-processing steps were completed using R Statistical Software[1] and Python[2] programming languages, as well as QGIS[3] and the Climate Data Operators (CDO)[4] command-line tool. The following paragraphs describe data sources and the pre-processing steps involved.

Water quality observations

The Water Quality Portal (WQP) created by the National Water Quality Monitoring Council[5] provides empirical water quality data for sites across the western United States. The database includes publicly available data from the United States Geological Survey (USGS), the Environmental Protection Agency (EPA), and over 400 state, federal, tribal, and local agencies. Sediment, DOM, nutrient, and turbidity data from the WQP were filtered for the temporal and spatial scope of this study's analyses, minimized anthropogenic contaminants, and maximized consistency in site and data types. Date ranges from 01-01-1974 to 01-01-2023 were chosen to span the ignition dates of wildfires available through the MTBS database, with 10 years of buffer on the front end to construct and fit pre-wildfire models. Sample media was limited to "Water" and "water" (different terms were used for various agencies) and monitoring station types were limited to "streams" and "rivers" for increased consistency in site types.

As measurement types and descriptions of sediment, DOM, nutrients, and turbidity varied widely across agencies, a frequency analysis was completed to identify search queries for each constituent—selecting the ones most common across agencies. Search term frequencies were first calculated for water quality data across the western United States. Thus, "Organic Carbon" was the chosen search term for organic carbon;

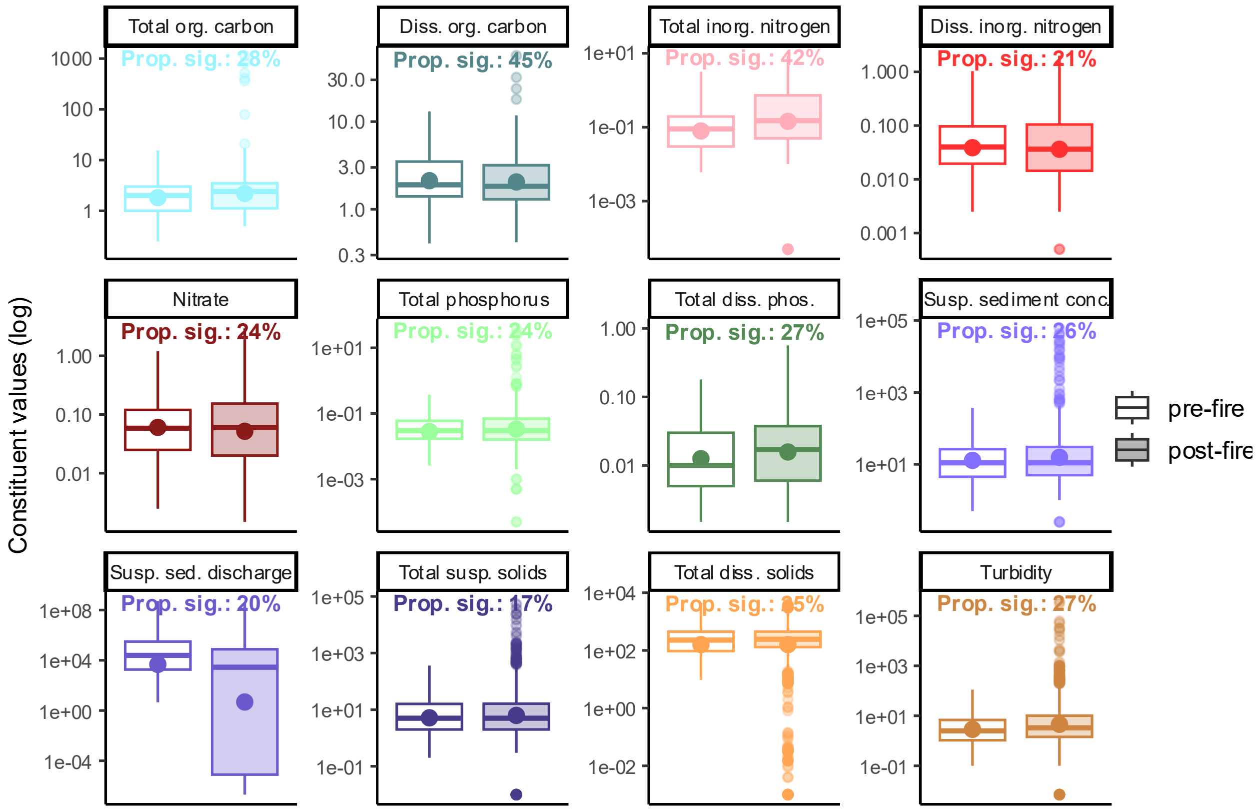
“Inorganic nitrogen (nitrate and nitrite)”, “Inorganic nitrogen (nitrate and nitrite) as N”, and “Nitrate + Nitrite” were used for inorganic nitrogen; “Nitrate”, “Nitrate as N”, and “Nitrate-Nitrogen” were used for nitrate; “Phosphorus”, “Phosphorus as P”, and “Total Phosphorus, mixed forms” were used for phosphorus; “Sediment”, “Total suspended solids”, “Total Suspended Particulate Matter”, “Suspended Sediment Concentration (SSC)”, “Suspended Sediment Discharge”, and “Total dissolved solids” were used for sediment and dissolved organic matter; and “Turbidity” and “Turbidity Field” were used for turbidity. Total nitrogen, organic nitrogen, and nitrite constituents were also downloaded, but ultimately left out of the analysis due to small sample sizes. Note, although it had lower data availability, organic carbon was included due to being a common analytical measurement associated with dissolved organic matter. Once these data were downloaded, their characteristics were used to designate which measurements were dissolved or total (dissolved + particulate), as well as adjust data points to maximize consistency in units and speciation methods within constituent types.

The collection time and date of each measurement were recorded on the granularity of one second, but the actual time interval between successive measurements was irregular, ranging from several minutes to years. Values were resampled to a daily timescale (i.e., total daily values for flux measurements with units of “kg” and mean daily values for concentrations with units of “mg l^{-1}”) for increased consistency, but high enough temporal resolution to capture rapid changes in concentrations. In total, 61,002 water quality stations were available for analysis and ~3.5 million water quality data points.

Supplementary Table 1 displays a breakdown of the total number of available data points for each constituent, as well as the average post-fire constituent changes observed in individual basins and the proportion where changes were significant. Data compiled from all basins for the regression-based regional analysis are shown in Supplementary Fig. 1, illustrating overall changes in raw data pre- and post-fire.

Supplementary Table 1 | Water quality constituents' availability and change after fires. Water quality constituents' availability and change after fires. Median and mean constituent percent differences in basins between eight years of pre-fire data and data within two years of a wildfire event—averaged across all burned basins. The percent of all basins exhibiting a significant change, tested using Mann-Whitney U-tests, is also displayed. Total number of datapoints is shown for all available data across burned basins for each constituent.

Constituent	Total number of datapoints	Average median change (%)	Average mean change (%)	Percent of basins significant (%)
TOC	5,228	2	0	28
DOC	2,211	46	58	45
TIN	1,580	46	57	42
DIN	9,767	84	40	21
NO_3^-	3,395	180	89	24
TP	11,443	49	166	24
TDP	2,321	210	121	27
SSC	10,044	217	320	26
SSD	6,534	287	124	20
TSS	18,574	79	236	17
TDS	18,523	19	9	25
TURB	17,690	381	1361	27



Supplementary Fig. 1 | Post-fire changes in water quality data aggregated across the western United States. Boxplots of pre- and post-fire water quality data aggregated from wildfire affected basins on a log scale. Data shown are from eight years pre-fire (clear boxes), as well as two years following wildfire events (shaded boxes), and different colors represent different constituents. The upper and lower bounds of the boxes represent interquartile ranges, the vertical whiskers represent 1.5 times the interquartile ranges, the horizontal lines in the middle of the boxes represent medians, the shaded points represent outliers, and the solid points represent means. The percentage of individual basins which exhibited significant post-fire changes, or “proportion significant”, is also displayed for each constituent.

Fire data

The United States Forest Service’s Monitoring Trends in Burn Severity (MTBS) database[6] supplied data for wildfires in the western United States occurring from 1984-2022 which were greater in size than 1,000 ac (404.7 ha). This satellite-derived data has a 30-m spatial resolution of both burn intensity and burn perimeter delineations. National wildfire perimeter and centroid datasets were used for initial modeling site identification and calculation of burn extent in each basin, with 11,400 wildfires available for analysis after filtering to the study’s temporal and spatial scope.

Physiographic data

For elevation data, the 1-Arc second Global Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM)[7] was used—a near-global land elevation dataset acquired from radar data by the National Aeronautics and Space Administration (NASA) and the National Geospatial-Intelligence Agency (NGA). This dataset covers over 80% of the Earth's land surface between 60° north and 56° south latitude, with datapoints from every 1 arc-second (~30 m). The SRTM Void Filled elevation data were used in this analysis, which has additional processing to address areas of missing data, or voids where initial processing did not meet quality specifications. 1-degree x 1-degree SRTM DEM rasters were first tiled together for the extent of the western United States, or latitudes from 31°N to 49°N and longitudes from 125°W to 102°W.

Outlines from the Watershed Boundary Dataset (WBD)[8] were then used to create more manageable file sizes. The WBD is a comprehensive aggregated collection of hydrologic unit data, created by the USGS, United States Department of Agriculture – Natural Resource Conservation Service (USDA NRCS), and the EPA. It defines the extent of surface water drainage to a certain point, delineating watershed boundaries of varying sizes. With these polygons, the SRTM DEM was then cropped to 10 HUC 2 watershed extents, then those cropped to 77 HUC 4 watershed extents. These pre-existing watershed boundaries were used to ensure that DEM subdivisions would not cut short natural flow paths during the basin delineation process. State boundaries from the USGS National Boundary Dataset (NBD)[9] were used for figure creation.

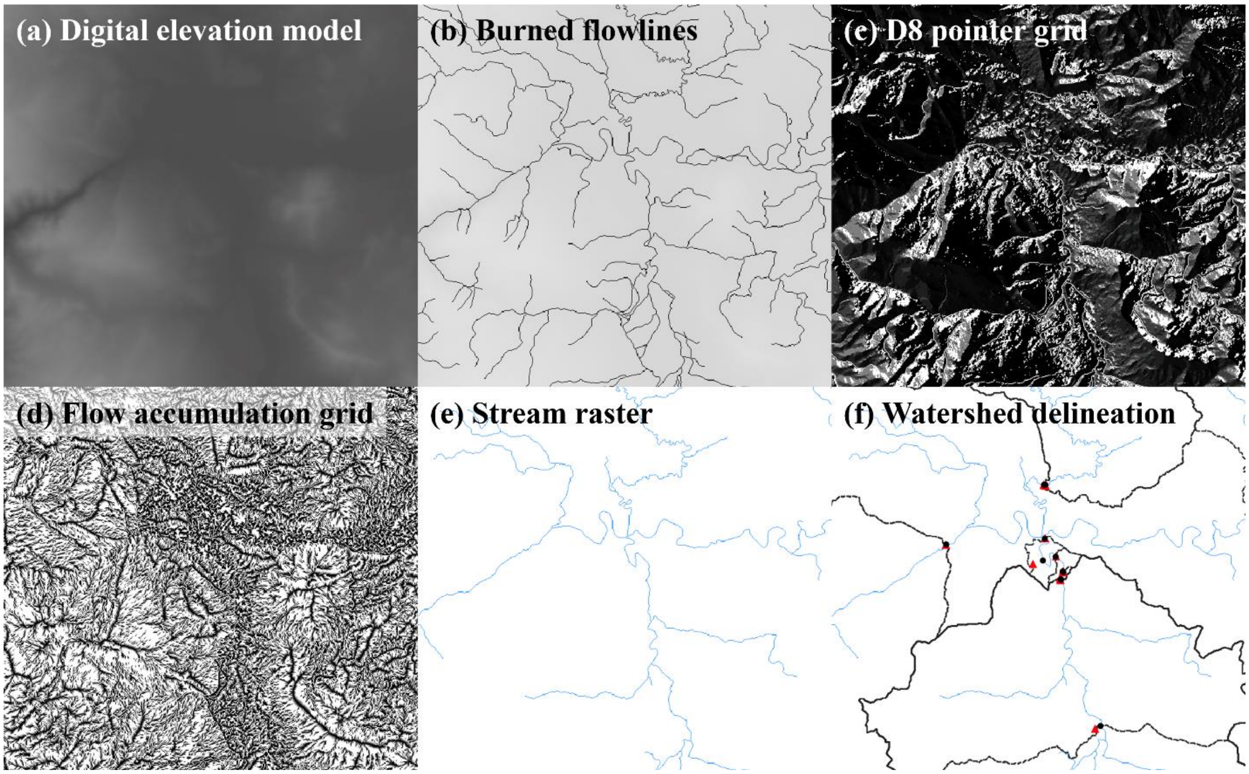
The National Hydrography Dataset (NHD)[10] was used to "burn" streams into DEM segments during the hydrologic conditioning process. This dataset represents the water drainage network of the United States, defining rivers, streams, canals, lakes, ponds, dams, and stream gages. Mapped at a 1:24,000 or larger scale, these data are updated and maintained through Stewardship partnerships and other collaborative bodies. Finally, to assess land cover characteristics in each basin, the National Land Cover Database (NLCD)[11] was used. The NLCD is a land cover database for the United States with 28 different land cover products characterizing land cover and land cover changes across 8 epochs from 2001-2019. Median land cover data from across these available epochs were calculated for each basin.

Hydroclimatic observations

The European Environment Agency ERA5-Land[12] datasets provided climate data for this study. These climate data are a reanalysis dataset which provides information on changing land variables over several decades. At a 0.1° x 0.1° spatial resolution and hourly timestep, this dataset combines model data with observations across the world to form a globally complete and consistent dataset using the laws of physics. The temperature of air at 2 m above the surface of the Earth, potential evaporation calculated through the surface energy balance, surface runoff calculated as the total amount of water accumulated during the forecast duration, and total precipitation, or the sum of accumulated liquid and frozen water, were used in this analysis. This data was aggregated to a daily timescale to match water quality and streamflow data temporal resolutions. Total daily values were calculated for precipitation, potential evaporation, and runoff, and max daily values calculated for temperatures.

Hydrologic conditioning and basin delineation

Watersheds were delineated using the coordinates of each of the 61,002 water quality monitoring stations as a pourpoint, or outlet of a contributing drainage area—resulting in a total subset of 51,101 successfully delineated watersheds. As shown in Supplementary Fig. 2, DEM segments were first prepared through hydrologic conditioning, then flow information extracted, before finally moving pourpoints to be coincident with streams and executing watershed delineation algorithms.



Supplementary Fig. 2 | Watershed delineation process. Images representing each step involved in delineating watersheds from pourpoints: **a,** A digital elevation model (DEM) was tiled for the western United States and segmented to HUC 4 subdivisions, **b,** NHD flowline streams were "burned" into the DEM segments, **c,** D8 pointer grids were extracted from conditioned DEMs, **d,** flow accumulation was calculated for each grid cell, **e,** stream grid cells were designated where flow accumulation exceeded a threshold of 5 km^2, and **f,** water quality station pourpoints (red triangles) were snapped (black points) to stream grid cells, then a watershed delineation algorithm applied to each pourpoint (black outlines).

Hydrologic conditioning algorithms prepare DEMs for hydrologic analyses by smoothing out depressions and correcting disrupted flow paths[13]. These processing steps were completed using the *whitebox* package in R[14]. Here, two subsequent processes were applied to the HUC 4 DEM segments: stream burning and filling. Stream burning is a flow enforcement technique which corrects surface drainage patterns in DEMs by lowering grid cells underneath overlaid stream vector shapefiles[15]. 1:100,000 resolution flowline vectors from the NHD were trimmed to the extent of each of the 77 HUC 4 segments, overlaid on the associated DEM subdivision, then the flow paths burned in. This was executed using the Whitebox Tools *FillBurn* function, interfaced in R using the *whitebox* package. Next, depressions in the DEMs' topography which were disruptive to flow paths were identified and "filled", i.e., grid cell elevations raised, using a method described in Wang and Liu, 2006. Implemented with the *FillDepressionsWangAndLiu* Whitebox tool, this algorithm was chosen due to its computational efficiency, as it simultaneously determines flow paths and spatial partitions of watersheds with one pass of processing[16].

Once DEM segments were conditioned, flow direction, flow accumulation, and stream raster grids were extracted to execute the basin delineation algorithm. Flow direction, or pointer, grids contain information about the direction of flow in each grid cell, calculated based on the slope and aspect of surrounding grid cells. A D8 algorithm, which assumes 8 possible flow directions from each grid cell[17], was implemented for each conditioned HUC 4 DEM segment using the *D8Pointer* tool. These pointer grids were then input into a flow accumulation algorithm—the *D8FlowAccumulation* tool—which determines the contributing number of cells, or area of drainage, to each grid cell. Finally, stream network rasters were then extracted from the flow accumulation grids using the *ExtractStreams* tool, which designated stream grid cells at high flow accumulation locations. A threshold contributing area of 5 km^2 was determined by testing several values and assessing the resulting stream raster's similarity to the 1:100,000 resolution NHD flowlines.

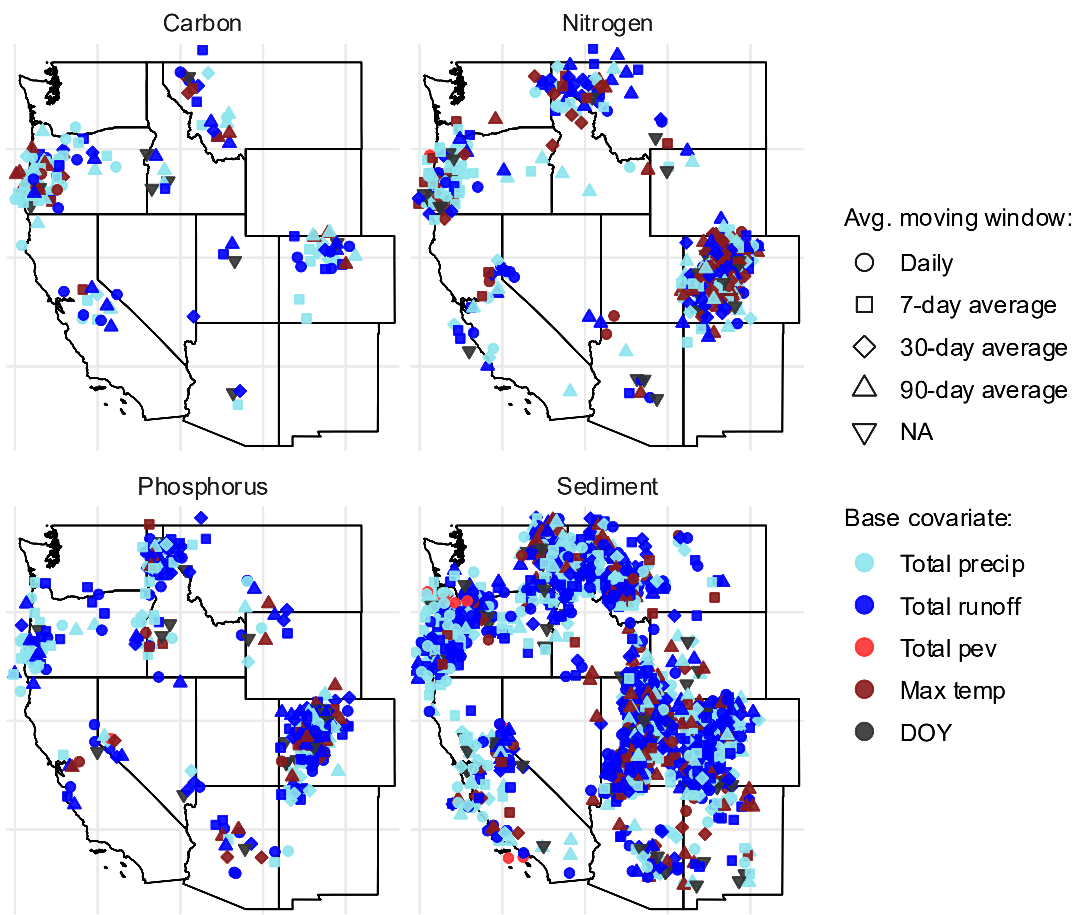
Once stream rasters were created for each HUC 4 segment, pourpoints which fell within the extent of each segment were snapped or moved to be coincident with a stream raster grid cell. This step ensured limited truncation in delineations due to slight discrepancies in the pourpoint coordinates and the location of flow paths defined through the DEM conditioning process. The Jenson algorithm was used, which snaps pourpoints to the nearest stream grid cell[18]—commonly preferred over less sophisticated algorithms which snap pourpoints to the largest flow accumulation grid cells within a certain radius[13]. This was implemented using the *JensonSnapPourPoints* tool, with a 5 km snap radius threshold.

The final step in the basin delineation process was to execute a search algorithm from each snapped pourpoint which looked upslope and determined the contributing area. Water quality monitoring station coordinates which snapped to the same 30 m grid cell, thus resulting in the same delineated watershed, were assigned a unique basin ID, and their available data combined. In total, 51,101 basins were successfully delineated from 61,002 water quality monitoring station locations. Stations within the same 30 x 30 m grid cell were merged to form a single pourpoint and their data combined. 299 water quality stations were screened from the analysis due to lack of proximity to stream raster grid cells. This was likely due to slight inaccuracies or low resolution of recorded water quality monitoring station coordinates, or inaccuracies in the extracted stream raster to natural flowlines. USGS water quality monitoring stations which had equivalent basin delineations in the GAGES II dataset were overlaid and visually compared to available GAGES II delineations to validate the implemented basin delineation process.

Supplementary Discussion

Model Covariate Regional Variability

Dominant covariates selected in the model building process for each basin and constituent varied regionally across the western United States. Supplementary Fig. 3 shows variation in influence of driver types and moving average window sizes on different constituent types across a spatial map of the western United States. These results are discussed further in the *Inter-basin variability* section of the *Results*.



Supplementary Fig. 3 | Regional variation in covariates' influence on water quality. Distribution of dominant covariates selected through the linear model building process for each constituent in each basin in the West Coast, Rocky Mountains, and Southwest regions of the western United States. Different covariates are represented by different colors, with temperature-related variables in red colors and water-related variables in blue colors. "DOY" represents the day of the water year of the water quality measurement. Evaluated constituents are combined into carbon, nitrogen, phosphorus, and sediment categories.

Model Performance Metrics and Filtering Criteria

Models built for each basin-constituent pair were evaluated using a leave-one-out cross-validation (LOOCV) method over pre-fire data. Several metrics were then calculated using the resulting predicted data to characterize model performance: Nash-Sutcliffe efficiency (NSE), percent bias (PBIAS), and the ratio of the root mean squared error to standard deviation (RSR). These metrics were calculated using the following equations (Supplementary Eqn. 1, Supplementary Eqn. 2, and Supplementary Eqn. 3, respectively):

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{N}(S_i - O_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2} \tag{1}$$

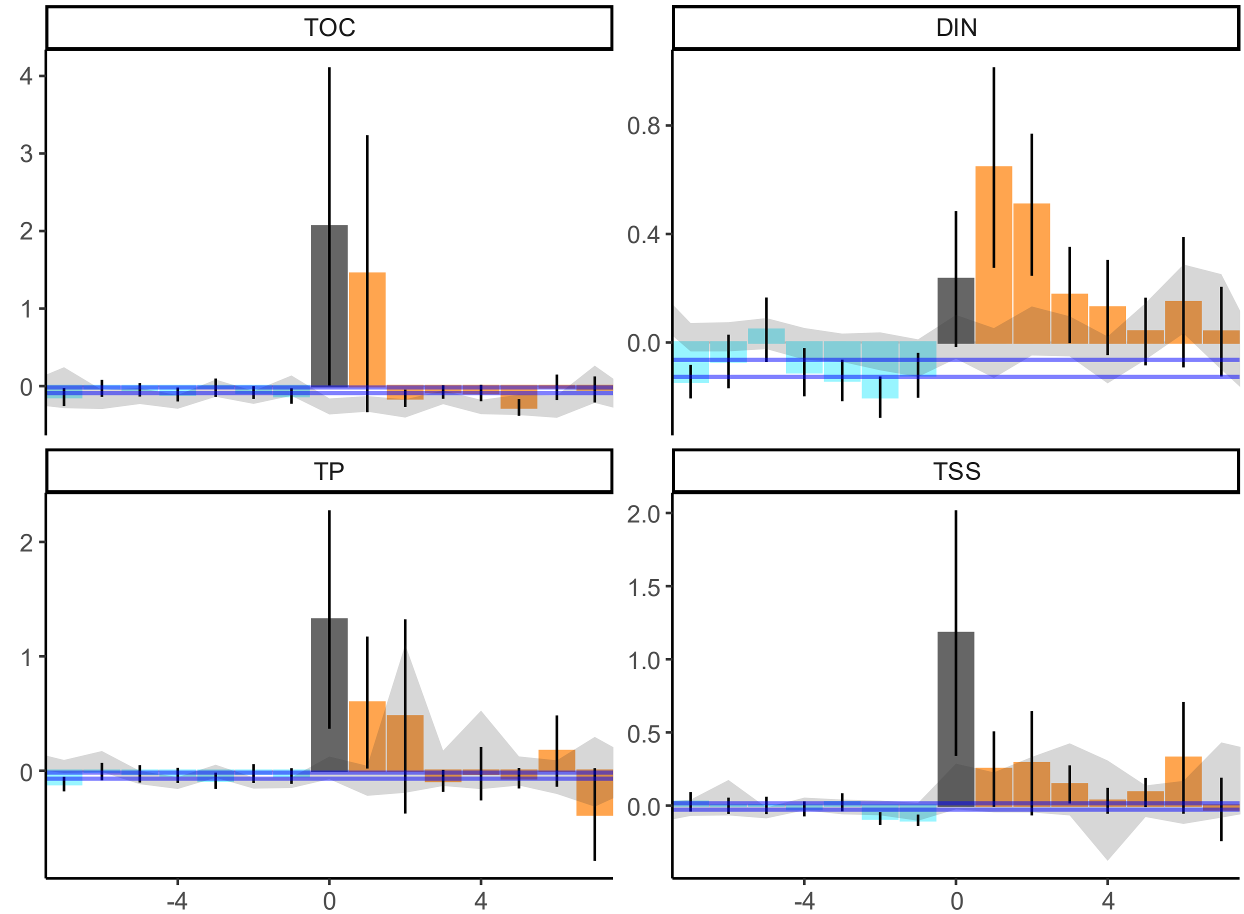
$$\text{PBIAS} = 100 \times \frac{1}{N}\sum_{i=1}^{N}\left(\frac{O_i - S_i}{|O_i|}\right) \tag{2}$$

$$\text{RSR} = \frac{1}{N}\sum_{i=1}^{N}(S_i - O_i)^2 \tag{3}$$

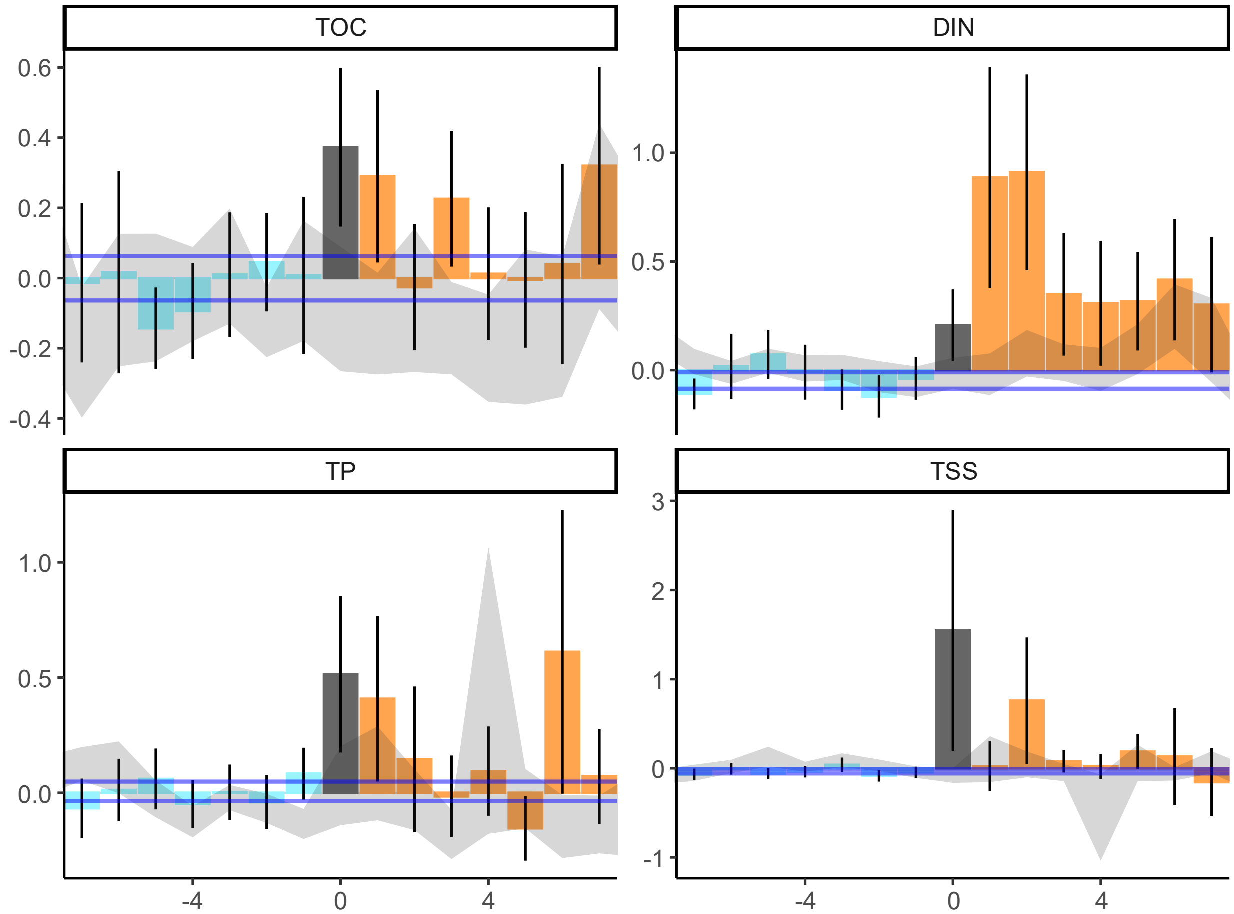
In these equations, *N* is number of pre-fire data used in the LOOCV analysis, *O* represents each observed datapoints, and *S* represents each simulated datapoint, or those predicted using the model being tested.

A major challenge of this study was to balance data availability limitations with obtaining a large enough sample size. Therefore, the thresholds for “acceptable” model performance metrics in this analysis were dictated in part to maximize the number of residual datapoints in the results. Thus, while performance criteria were applied to models used in the final results (or RSR scores > 1, as discussed in the *Regression-based response analysis* section in the *Methods*), stricter criteria were not applied to allow for a robust post-fire residual sample size for assessment of response significance.

To further investigate the impact of this decision, results from the current analysis were compared to results generated using stricter filtering criteria: RSR > 1, NSE > 0.25, and PBIAS > -15% and < 15%. These filtering criteria reduced the total number of burned basins to 122 and unburned basins to 167, and excluded DOC, TIN, and TDP from the analysis, due to lack of data availability (i.e., under the stricter criteria, results for these constituents did not meet the minimum requirement of including data from at least one burned and unburned basin for each pre- and post-fire year). However, despite the stricter filtering and lower residual sample sizes, the qualitative findings and general features of the analysis were consistent with the original results. This is shown for four constituents in Supplementary Fig. 4 (original results) and Supplementary Fig. 5 (results with more strict filtering criteria).



Supplementary Fig. 4 | Constituent responses with original model performance criteria. Residuals from models with < 1 ratio of root mean squared error scores for four key constituents representing the main water quality categories in this study: total organic carbon (TOC), dissolved inorganic nitrogen (DIN), total phosphorus (TP), and total suspended solids (TSS). For burned basins (n = 245), mean residuals for each year are shown for seven years pre-fire (light blue bars), the year following wildfire events (grey bars), as well as the following seven years post-fire (orange bars). The black vertical lines on each bar represent 90% confidence intervals of burned basins' residuals for each year. Horizontal blue lines represent the overall 90% confidence interval bounds for all seven pre-fire years together, extended through the post-fire years to assess the significance of post-fire responses. The gray ribbons represent the residuals from unburned basins (n = 293), showing their 90% confidence intervals for each pre- and post-fire year.



Supplementary Fig. 5 | Constituent responses with more strict model performance criteria. Residuals from models with < 1 ratio of root mean squared error scores, Nash-Sutcliffe efficiency scores > 0.25, and percent bias scores in between -15% and 15%. Residuals for four key constituents representing the main water quality categories in this study: total organic carbon (TOC), dissolved inorganic nitrogen (DIN), total phosphorus (TP), and total suspended solids (TSS). For burned basins (n = 122), mean residuals for each year are shown for seven years pre-fire (light blue bars), the year following wildfire events (grey bars), as well as the following seven years post-fire (orange bars). The black vertical lines on each bar represent 90% confidence intervals of burned basins' residuals for each year. Horizontal blue lines represent the overall 90% confidence interval bounds for all seven pre-fire years together, extended through the post-fire years to assess the significance of post-fire responses. The gray ribbons represent the residuals from unburned basins (n = 167), showing their 90% confidence intervals for each pre- and post-fire year.

Due to the similarities in qualitative findings, the original filtering criteria was kept. This more closely aligned with the goals of this study—maximizing the number of basins and datapoints used in the final results to more fully capture regional variability in post-fire responses across the western United States.

Supplementary References

1. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria, 2022).
2. Van Rossum, G. & Drake, F. L. *Python 3 Reference Manual*. (CreateSpace, Scotts Valley, CA, 2009).
3. QGIS Development Team. *QGIS Geographic Information System*. (QGIS Association, 2023).
4. Schulzweida, U. *CDO User Guide*. (Zenodo, 2023). doi:10.5281/zenodo.10020800.
5. National Water Quality Monitoring Council, Environmental Protection Agency (EPA), & United States Geological Survey (USGS). Water Quality Portal. U.S. Geological Survey https://doi.org/10.5066/P9QRKUVJ (2021).
6. MTBS Project (USDA Forest Service/U.S. Geological Survey). Fire Level Geospatial Data. MTBS Data Access (2022).
7. Earth Resources Observation And Science (EROS) Center. Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global. U.S. Geological Survey https://doi.org/10.5066/F7PR7TFT (2017).
8. U.S. Geological Survey, National Geospatial Technical Operations Center. Watershed Boundary Dataset (WBD) - USGS National Map Downloadable Data Collection. U.S. Geological Survey (2023).
9. U.S. Geological Survey. USGS National Boundary Dataset (NBD). (2020).
10. USGS - National Geospatial Technical Operations Center (NGTOC). National Hydrography Dataset (NHD) - USGS National Map Downloadable Data Collection. U.S. Geological Survey (2023).

11. Dewitz, J. & U.S. Geological Survey. National Land Cover Database (NLCD) 2019 Products (ver. 2.0, June 2021). U.S. Geological Survey data release https://doi.org/10.5066/P9KZCM54 (2021).

12. Muñoz-Sabater, J. ERA5-Land hourly data from 2001 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) https://doi.org/10.24381/CDS.E2161BAC (2019).

13. Lindsay, J. B., Rothwell, J. J. & Davies, H. Mapping outlet points used for watershed delineation onto DEM-derived stream networks. *Water Resour. Res.* **44**, (2008).

14. Wu, Q. & Brown, A. whitebox: 'WhiteboxTools' R Frontend. (2023).

15. Lindsay, J. B. The practice of DEM stream burning revisited. *Earth Surf. Process. Landf.* **41**, 658–668 (2016).

16. Wang, L. & Liu, H. An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. *Int. J. Geogr. Inf. Sci.* **20**, 193–213 (2006).

17. O'Callaghan, J. F. & Mark, D. M. The extraction of drainage networks from digital elevation data. *Comput. Vis. Graph. Image Process.* **28**, 323–344 (1984).

18. Jenson, S. K. Applications of hydrologic information automatically extracted from digital elevation models. *Hydrol. Process.* **5**, 31–44 (1991).