Analysis of Anthropogenic, Climatological, and Morphological Influences on Dissolved Organic Matter in Rocky Mountain Streams

Nicolás Rodríguez-Jeangros [1], Amanda S. Hering [2],* and John E. McCray [1,3]

[1] Civil and Environmental Engineering Department, Colorado School of Mines, 1500 Illinois St, Golden, CO 80401, USA; nicolairj@gmail.com (N.R.J.); jmccray@mines.edu (J.E.M.)

[2] Department of Statistical Science, Baylor University, One Bear Place #97140, Waco, TX 76798, USA

[3] Hydrologic Science and Engineering Program, Colorado School of Mines, 1500 Illinois St, Golden, CO 80401, USA

* Correspondence: mandy_hering@baylor.edu; Tel.: +1-254-710-1120

check for updates

Abstract: In recent decades, the Rocky Mountains (RM) have undergone significant changes associated with anthropogenic activities and natural disturbances. These changes have the potential to alter primary productivity and biomass carbon storage. In particular, dissolved organic carbon (DOC) in RM streams can affect heterotrophic processes, act as a source for the nutrient cycle, absorb sunlight radiation, alter metal transport, and can promote the production of carcinogenic byproducts during water treatment. Recent studies have focused on the relationship between bark beetle infestations and stream organic matter but have reached conflicting conclusions. Consequently, here we compile and process multiple datasets representing features of the RM for the period 1983–2012 with the purpose of assessing their relative influence on stream DOC concentrations using spatial statistical modeling. Features representing climate, land cover, forest disturbances, topography, soil types, and anthropogenic activities are included. We focus on DOC during base-flow conditions in RM streams because base-flow concentrations are more representative of the longer-term (annual to decadal) impacts and are less dependent on episodic, short-term storm and runoff/erosion events. To predict DOC throughout the network, we use a stream network model in a 56,550 km^2 area to address the intrinsic connectivity and hydrologic directionality of the stream network. Natural forest disturbances are positively correlated with increased DOC concentrations; however, the effect of urbanization is far greater. Similarly, higher maximum temperatures, which can be exacerbated by climate change, are also associated with elevated DOC concentrations. Overall, DOC concentrations present an increasing trend over time in the RM region.

Keywords: water quality; stream networks; forest disturbances; climate change; urbanization; large-scale analysis; mountain pine beetle; wildfires

1. Introduction

Anthropogenic activities have highly influenced ecosystems in the RM. Urbanization and forest logging are clear examples of direct impact (e.g., [1–5]). Furthermore, climate change has provided optimal conditions for fires and mountain pine beetle (MPB) outbreaks, which can lead to hydrologic disturbances. The MPB is an autochthonous bark beetle species from the RM, and it has shifted its impacts from periodic and localized outbreaks to widespread infestations (e.g., [6–9]). Additionally, fire management policies have resulted in an overall increase in and homogenization of the forest age, making pine forests more vulnerable to bark beetles [10,11]. During the last two decades, bark beetle infestations have affected more than 5 million ha of forest in the western U.S. and British

Columbia [6,12]. Potential impacts associated with MPB disturbances include modifications of the hydrologic processes and alterations of the biogeochemical cycle [13–15]. For example, water storage and supply can be affected by changes in snow accumulation and melt derived from variations in the surface energy fluxes [16], and carbon and nitrogen soil concentrations can increase in areas impacted by the MPB [17–19].

Although evidence exists that surface water contributes as a regulator of the carbon cycle and that the temporal change of these contributions is associated with a changing environment [20,21], little research has been published regarding the effects of MPB infestations on the carbon concentrations of RM streams. Understanding how MPB infestations affect carbon concentrations in streams is relevant because in addition to being a source for the nutrient cycle, dissolved organic matter (DOM) plays a central role in other surface water quality processes [22]. First, it absorbs sunlight radiation, influencing biochemical processes that depend on ultraviolet radiation and temperature [23]. Second, it can increase the solubility of metals by acting as a ligand, which has implications for the restoration of polluted ecosystems and the mitigation of the impact of metal pollutants [24]. Third, high concentrations of DOM during water treatment represent a risk because DOM can react during disinfection processes to form carcinogenic byproducts [25].

DOM in soils and water originates from the degradation of organic matter from plants and microorganisms, and therefore, processes altering forest and vegetation life cycles may have an indirect effect on surface water quality [26]. At the hillslope scale, increases in DOM and nitrogen have been observed in MPB-impacted areas [17,19]. Mikkelson et al. observed increased concentrations of water disinfection by-products associated with DOM in water treatment facilities where the contributing watershed had suffered MPB disturbances [27]. Increased particulate and metal transport have also been observed in impacted watersheds [13]. However, the biogeochemical dynamics of the soil and the streams of the RM cannot be directly attributed to MPB disturbances without also accounting for other factors that contribute to a dynamic environment, such as climate change (e.g., temporal trends in precipitation and temperature) and anthropogenic disturbances (e.g., land cover changes associated with urbanization). Examples include MPB disturbances and forest fires: As MPB outbreaks have increased, fire activity has also increased in the western U.S., suggesting a possible link. Nonetheless, recent research suggests that increased fire activity does not depend on MPB infestations; instead, both are consequences of climate change and other factors [28–30]. Consequently, although a relationship has been observed between MPB infestations and DOM alterations at the hillslope and watershed scales, very little is understood about the causal relationship between MPB infestations and these environmental alterations because they are both highly influenced by multiple factors, including climate change and anthropogenic disturbances.

Further evidence of the difference in impact between climate change and local factors on DOM is revealed by conflicting reports at different locations and scales. Trahan et al. reported reduced DOM and nutrients in soils with high tree mortality during the first 4 years after MPB disturbance [19]. Contrasting results were obtained by Norton et al. in an analog study, where nutrient concentration increases were observed [15]. Furthermore, Beudert et al. reported no long-term effects in a 28-year period study on water quality associated with bark beetle infestations in forests of southeastern Germany [31]. Anthropogenic disturbances and other processes associated with climate change were not accounted for in these three studies. Furthermore, the spatial domains in these studies corresponded to one or two small watersheds with areas of less than 40 km^2.

DOM change in streams is potentially one of the most practically relevant impacts of bark beetle and forest disturbances. Thus, we present an unprecedented large-scale analysis in space and time of the effects of beetle infestations, along with relevant climatological, forest disturbance, anthropogenic, and morphological processes on DOM in the RM streams. Our 56,550 km^2 study area includes highly heterogeneous watersheds with respect to scale, attributes, and hydrological processes during a 30-year period. We account for these heterogeneous features and other potentially contributing factors associated with other forest disturbances, anthropogenic activities, and the morphology of the

region by selecting and processing extensive spatial–temporal datasets and products with appropriate attributes for the scale and resolution of the analysis. In short, we employ a statistical approach to model the relationship between DOM and many features, and this model incorporates the physical connectivity of the stream network in the region of interest for analysis and prediction. We then interpret the statistical significance and relative importance of each anthropogenic, climatological, and morphological feature for predicting DOM in the RM streams and investigate the possible existence of a temporal trend in DOM concentrations.

The remainder of this article is organized as follows: We describe the region of interest, the compilation and preparation of DOM stream data, and the selection and processing of the datasets and products representing the predictive features in Section 2. Section 3 presents the necessary considerations for a proper analysis of the data's correlated features and spatial dependence in the network and introduces the associated statistical models appropriate for these considerations. Finally, the results and analysis regarding importance of the different features for DOM in the RM streams and the prediction of DOM in space and time are presented in Section 4. We conclude in Section 5.

2. Materials and Methods

In this section, we describe the region of interest and the compilation of the observed DOM dataset and the different potential predictors representing anthropogenic, climatological, and morphological features and processes in the RM.

2.1. Region of Interest

The region of interest consists of the section of the Rocky Mountains in the U.S. that has been most impacted by forest disturbances and anthropogenic development, as shown in Figure 1. More than 6500 km^2 in the region have been affected by bark beetle outbreaks (~12% of the region of interest), wildfires have affected more than 1750 km^2 in the last decades (~3% of the region of interest), and approximately 1500 km^2 have been urbanized in the 1983–2012 period (~2.7% of the region of interest) [1,32–34]. Figure 1 also presents the stream network, consisting of sixteen separate basins and covering an area of 56,550 km^2. The network was developed by the National Stream Internet Project by reconditioning the NHD-Plus Version 2 (Rocky Mountain Research Station, Fort Collins, CO, USA) hydrography, and it has a scale of 1:100,000 [35]. We corrected some topology issues in the network such as isolated watersheds. The stream network consists of 17,166 stream segments with a total stream length of 3,598,413 km. Specific features of the region are described in detail in Sections 2.2 and 2.3.

2.2. Dissolved Organic Matter Data

A statistical analysis on the strength of the dependence of DOM on anthropogenic, climatological, and morphological features of the RM involves accounting for possible correlation among the predictors, expected temporal changes, and potential spatial dependence in the response associated with a stream network setting. Therefore, a comprehensive database representing the response and potential predictors in space and time is needed. Experimentally, DOM is measured as dissolved organic carbon (DOC) [36]. Therefore, we first compile DOC data for the study area from the USGS Water-Quality Data for the Nation because it provides a considerable spatial and temporal coverage and uses certified sampling and processing standards [37]. The data complied corresponds to 12,407 DOC samples in streams for the study area during the period 1983–2012.
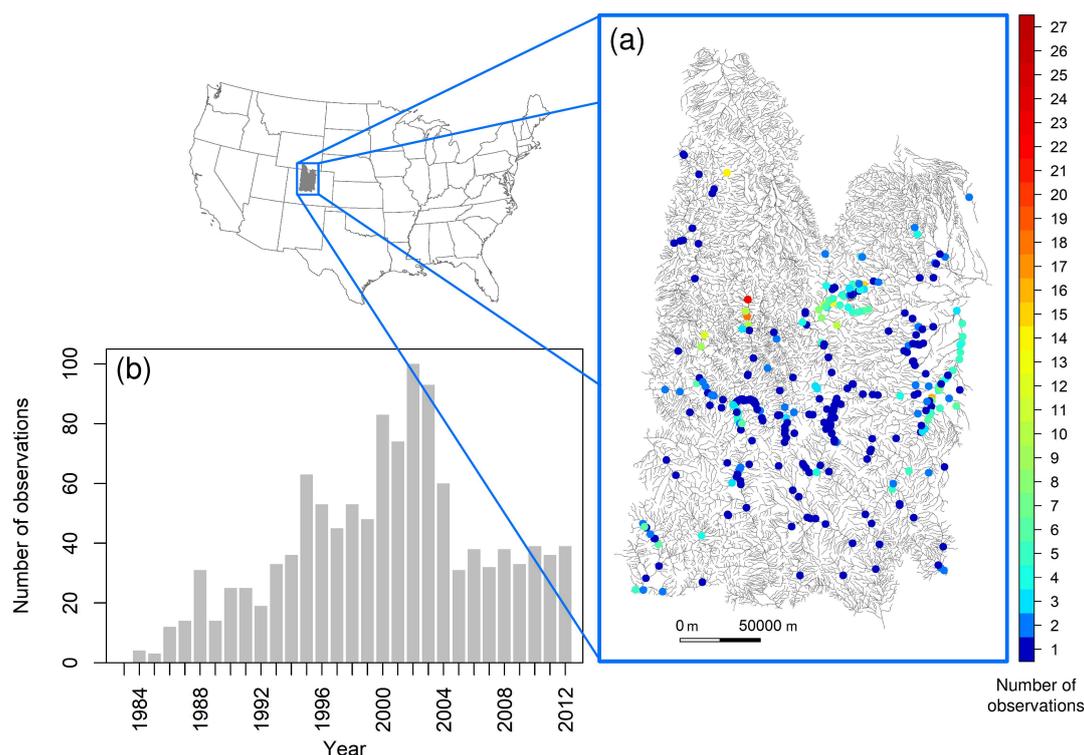
Figure 1. Study area located in the section of the Rocky Mountains that has been affected by anthropogenic and natural disturbances in the 1983–2012 period (**a**), and number of observations available per year once the dissolved organic matter (DOM) data have been transformed to yearly averages for the corresponding base-flow periods as described in Section 2.2 (**b**). In (**a**), the stream network is represented by the gray line segments, and the 371 points in the stream network represent the locations with DOM measurements. The color scale represents the number of observations as in (**b**) but at each location. Basins located near the northeast corner of the study area are not included due to a lack of observations.

For this analysis, base-flow conditions are preferred because they are more representative of the longer-term (annual to decadal) impacts from forest disturbances and climate change, rather than short-term impacts of precipitation, runoff events, and snowmelt, which can occur at an hourly to monthly temporal scale. The base-flow months used here are August, September, October, and November. Carbon concentrations from late base-flow months (December to March) are not included because more significant fluctuations in the concentrations have been observed during these late months, possibly because of larger snow-melt contributions [38]. Although an analysis of DOC loads during runoff events would be of interest, the instances when stream-flow and DOC are sampled rarely match in the historical records. As shown in Figure 2, the selected months have considerably lower peak flow and fewer runoff events than summer months. Once the DOC data are filtered based on the selected base-flow months, 3680 samples remain.

We construct a representative DOC value of each base-flow period for each sampling location and year. Because this value is constructed based on observations that are unequally spaced in time during the base-flow period, a simple average of these observations may be suboptimal because it can be biased due to redundant observations that are close in time. We tested three approaches for producing a single yearly value, but they result in negligible differences in constructed base-flow values, and the statistical analysis presented below (Sections 3 and 4) produces practically the same results regardless of the averaging approach. Thus, we compute a monthly weighted average where the weight corresponds to the number of samples in each day; then the overall average is the weighted average of the monthly averages where the weight is the number of days with samples within each

month. We then have 1174 observations representing base-flow DOC concentrations in 371 locations of the stream network. Figure 1a presents the spatial distribution of the 371 locations, and the color scale shows the number of observations at each location. The majority of locations have observations for 5 or fewer different years, while locations near urban areas (such as the front-range on the eastern limit of the stream network) tend to have more observations. Figure 1b displays the number of observations for each year during the 1983–2012 study period. Recent years tend to have more observations, and the period between 2000 and 2004 has the most observations.
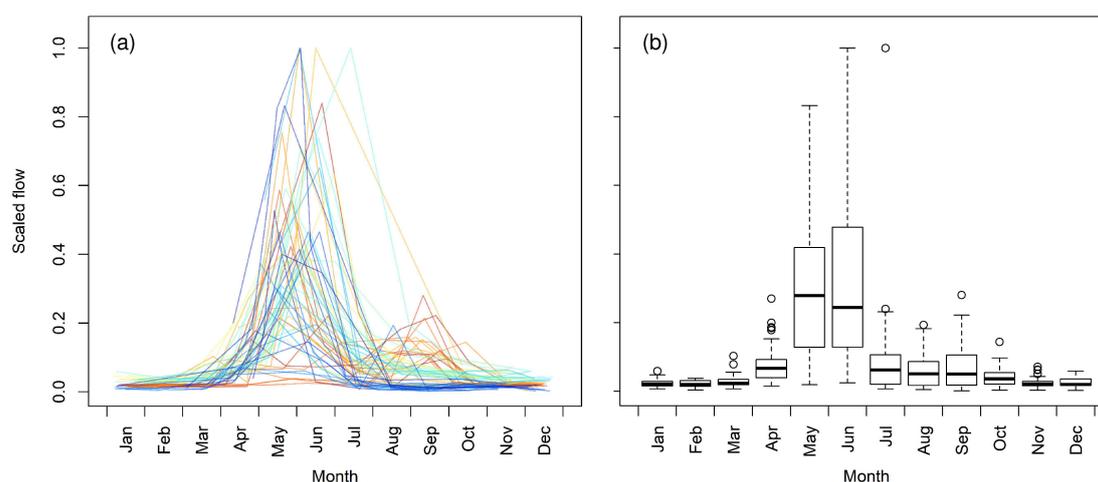
Figure 2. Yearly stream-flow data for 64 randomly selected locations in the study area. The locations were selected from a set in which flow gaging occurred consistently during the year with gaps between measurements of less than 4 months. The year corresponding to each location varies in the period 1983–2012 because gaging campaigns were not consistent. For this same reason, flow gaging observations are not equally spaced in time. The abscissa represents the sampling date, and the ordinate represents a scaled flow, which corresponds to the ratio of streamflow to maximum streamflow recorded at the location. The scaling is performed to improve visualization. Continuous lines connect the streamflow values at a given location and year (**a**), and the box plots represent the distribution of flows in each month (**b**), where the thick horizontal line is the median, the box limits are the first and third quartiles, the whiskers are 1.5 times the interquartile range beyond the quartiles, and the open circles are potential outliers. The data were compiled from the USGS Water-Quality Data for the Nation [37].

2.3. Potential Predictors from Datasets and Products

As described in Section 1, climate change can influence stream water quality through multiple processes, and other factors such as the morphology and natural processes of a region are also strong determinants of surface water quality. This section summarizes the different features that represent stream-water quality determinants in the region of interest and that are used as potential predictors. Appendix A provides a detailed description of the compilation and processing of each of the features. Table 1 summarizes and groups the predictors into four main categories: Forest Disturbance, Climate, Anthropogenic, and Morphological.

Table 1. Features potentially influencing carbon concentrations in the region of interest. Although Land Cover appears in the Anthropogenic category, it can also represent Forest Disturbance, as described in Appendix A. The potential predictors are used to represent the features in the model described in Section 3, in which each coefficient represents the strength of the relationship between the response variable (dissolved organic carbon (DOC)) and the predictor.

Category	Feature	Number of Coefficients	Potential Predictors
Forest Disturbance	Fires	1	Accumulated Relative differenced Normalized Burn Ratio (*RdNBR*)
	Mountain pine beetle (MPB) Infestations	3	Fraction of affected area for the green, red, and gray phases
Morphological	Topography	9	Mean elevation and predominant aspect among 9 classes [a]
	Hydrologic Soil Types	3	Fraction covered by soil types A, B, and C [b]
Climate	Precipitation	3	Mean monthly minimum and maximum temperatures for the base-flow period, non-base-flow period, and the entire year
	Temperature	6	Mean monthly minimum and maximum temperatures for the base-flow period, non-base-flow period, and the entire year
	Snow Cover	3	Mean monthly snow water equivalent (SWE) for the base-flow period, non-base-flow period, and the entire year
Anthropogenic	Land Cover	3	Fraction of area covered by developed, agricultural, and forest land cover types, individually
	Wastewater Point Sources	1	Number of wastewater point sources

[a] The predominant aspect is a categorical value with nine different classes: North, Northeast, East, Southeast, South, Southwest, West, Northwest, and Flat, and thus it is included in the model using eight dummy variables, leaving one of the classes out to avoid collinearity (see Section Appendix A for details). [b] There are four types of soil representing the similarity of physical and runoff soil characteristics, but type D is left out since the sum of the area fraction covered by all classes is always 1, and therefore including type D would produce collinearity (see Section Appendix A for details).

For the case of fires, we use data from the Monitoring Trends in Burn Severity project for extent and severity [33,39]. Because water quality responses last longer than the duration of a fire, we include the severity of each fire in all years posterior to its ignition with a decay over time as vegetation and water quality recover. See Appendix A for details. The fire severity metric that we choose to use is the Relative differenced Normalized Burn Ratio (*RdNBR*) because it can characterize fire severity on low biomass sites and enhances inter-fire comparability at larger ecological scales [40]. We use and process MPB infestation data from the United States Forest Service, which provides a yearly geospatial inventory of tree mortality and the associated cause, i.e., type of insect or disease [41]. For topography, elevation and aspect data are included as potential predictors due to their influence on watershed size, biogeochemical reactions, and solar exposure [42–47] using information from the National Elevation Dataset [48,49]. Hydrologic soil types are included as possible predictors from the U.S. Department of Agriculture because this categorization is based on the similarity of physical and runoff soil characteristics [50]. For the representation of climatological features, we use the hydrometeorological dataset developed by Livneh et al. [51] because it is the only product that covers the entire spatial–temporal domain and includes temperature, precipitation, and snow water equivalent (SWE) variables. This approach minimizes transboundary discontinuities while maintaining a consistent gridding methodology in space and time, which are important for analyzing large-scale hydrometeorological phenomena [52]. Land cover features are included as predictors by using the SCaMF-RM product, because it provides high resolution maps for the region of interest with a yearly frequency during the entire period 1983–2012 [53]. Finally, the locations of wastewater point sources in the region of interest are collected from the environmental authorities in Colorado and Wyoming [54,55].

3. Water Quality Model

Here, we model the effects of climate change, anthropogenic processes, and the morphology of the RM on water quality with a linear model. Linear models can be represented by Equation (1),

$$y = X\beta + \epsilon, \tag{1}$$

where y is a vector that represents the response variable of the model (DOC in this case), X is a matrix that represents the predictors (Table 1), β is a vector of parameters that represents the strength and direction of the linear relationship between y and X, and ϵ represents the error.

The DOC value at each observation location corresponds to a concentration that has been influenced by water flowing and mixing from different sections of the watershed for which the observation location is the collection point. These different portions of water have been affected by biogeophysical processes influenced by many different potential predictors (e.g., see Table 1). Additionally, the most important value of each predictor at each location is not the value of the predictor at that specific location, but instead, the average value of the predictor across its associated watershed [56]. Hence, the accumulated or average value of each predictor in the watershed associated with each observation location is used in X (see Appendix A for details about using an accumulated or averaged value, depending on the predictor). Obtaining these values requires (a) the delineation of the watersheds in thousands of locations for both the locations with observations and the locations in which the model is used to predict the response variable; and (b) the accumulation or averaging of each potential predictor in thousands of watersheds. Consequently, we use the Cheyenne supercomputer for such processing-intensive computations by parallelizing each independent task [57].

There are 26 primary predictors in the model; however, for the case of soil types, only three types are included as predictors because the sum of the area fraction covered by all four soil types in a given watershed is always 1. Therefore, including all four types would produce collinearity. Similarly, the predominant aspect in the watershed associated with a location is a categorical value with nine different possible classes. Consequently, eight different dummy binary variables are created to represent each of these possible categories, receiving a value of 1 if they represent the predominant aspect for the watershed associated with a specific location, and 0 otherwise. The dummy variable associated with a Flat predominant aspect is not included as a potential predictor and represents a base scenario in the regression. Although several of the predictors are temporally dynamic, they may not explain all of the temporal variability in the base-flow DOC data, and thus the year associated with each observation is also included as a potential predictor to represent time. Consequently, the X matrix has 1174 rows, corresponding to the observations (see Section 2.2), and 34 columns, corresponding to 33 potential predictors and a column of ones for the intercept coefficient. Similarly, β is a vector with 34 entries corresponding to the coefficients for the 33 potential predictors and the intercept (see Table 1).

In a simple linear model, the values of β can be estimated by minimizing the sum of the squared errors, i.e., $\sum \epsilon^2$. This approach produces the best linear unbiased estimators, which means that the estimator for β has the smallest variance among unbiased estimators and an expected value equal to the real value of β. Nevertheless, when this approach is used with highly correlated predictors, the resulting variance of $\hat{\beta}$ can be large, meaning that the estimated coefficient values from different samples of the same population may differ substantially. As shown in Figure 3, this is the case here, where several predictors are strongly correlated. For example, it is expected that greater yearly precipitation will translate to greater snow depth. Similarly, the areal fraction of watersheds affected by MPB infestations is limited by the forest coverage and host tree species.
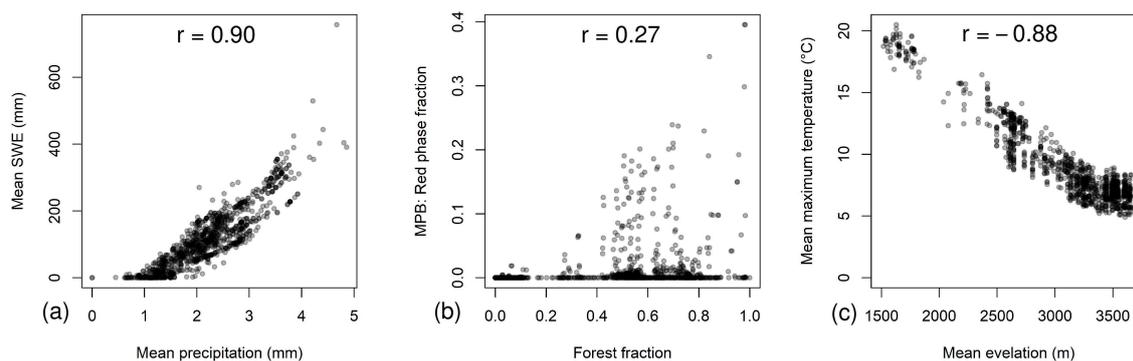
Figure 3. Pairwise comparison of potential predictor values for the 1174 DOC data points and their associated Pearson correlation: (**a**) yearly mean monthly precipitation and yearly mean monthly SWE; (**b**) forest fraction and MPB-Red phase fraction; and (**c**) mean elevation and yearly mean monthly maximum temperature.

We use the Least Absolute Shrinkage and Selection Operator (LASSO) regression method to reduce the variance of the estimated coefficients, which can be enlarged by the correlation in the predictors, and to filter and select a subset of potential predictors that are most important for explaining the response variable [58,59]. With LASSO, the estimated coefficients, $\hat{\beta}$, are not unbiased, but the resulting variance is considerably smaller when the predictors are correlated, meaning that the estimated coefficient values from different samples of the same population tend to be similar. LASSO achieves this by solving $\min_{\beta}\{\|y - X\beta\|_2^2\}$ subject to $\|\beta\|_1 \leq c$, which can be re-written as $\min_{\beta}\{\|y - X\beta\|_2^2 + \lambda\|\beta\|_1\}$ in order to use the method of Lagrange multipliers to restrict the domain in which the coefficients are estimated. In LASSO, the λ parameter controls the space of the coefficient domain. Larger values of λ produce small estimates in the $\hat{\beta}$ coefficients by creating a more restrictive constraint for the coefficient domain, forcing some of the regression coefficients to zero and automatically making the associated predictors insignificant. Therefore, LASSO helps to identify the predictors that have a significant impact on the response while dealing with their correlation, achieving both variable selection and coefficient estimation simultaneously. The LASSO model is fit using penalized maximum likelihood via coordinate descent [60,61].

Although we standardize all predictors within the LASSO model to ensure that the penalization scheme, i.e., the $\|\beta\|_1 \leq c$ constraint, is fair to all regressors [62], the reported coefficients are based on the following centering and scaling changes in the predictors. These operations are performed to facilitate the interpretation of the coefficient magnitudes.

Predictors centered to their mean: The year of observation, predictors related to climate (i.e., maximum and minimum temperatures, precipitation, and SWE), the fraction of area covered by soil types (A, B, and C), the mean elevation, and the number of wastewater points sources are the 18 potential predictors that are centered to their mean. That is, the mean value of a feature across all 1174 data points is subtracted from each of the respective 1174 values. After transformation, the intercept coefficient from the regression represents the expected DOC base-flow concentration when all these 18 predictors have a value equal to their mean. All other predictors are not centered because they have an upper and/or lower limit with a physically interpretable meaning. For example, a value of zero in the fraction of area affected by MPB infestations in the different phases can be understood as a reference value. In other words, the intercept coefficient represents the expected DOC base-flow concentration for undisturbed watersheds.

Predictors scaled by their standard deviation: The accumulated *RdNBR*, representing fire severity, is the only predictor scaled by its standard deviation. This predictor is scaled because an accumulated value of *RdNBR* over a watershed is not easy to interpret intuitively. On the other

hand, when scaled, the associated regression coefficient can be interpreted as the average increase or decrease in the DOC base-flow concentration when the accumulated *RdNBR* changes by one standard deviation for a given watershed. All other predictors are not scaled because a one unit difference in their magnitudes is intuitive, and therefore, it would facilitate the interpretation of its associated coefficient. For example, the coefficient associated with a temperature predictor would represent the average increase or decrease in the DOC base-flow concentration when temperature changes one Celsius degree.

In summary, the intercept coefficient in the regression represents the DOC base-flow concentration when a watershed has not been disturbed by fires or MPB infestations; does not contain urbanized, agricultural, or forested areas; has a Flat predominant slope; and all other predictors have their mean value.

Although general linear models usually assume the residual, ϵ, to be random and independent (i.e., $var(\epsilon) = \sigma^2 I$ with I being the identity matrix), in this case it may be necessary to account for the spatial correlation in the residuals by using a covariance matrix, i.e., $var(\epsilon) = \Sigma$, because of the inherent connectivity of a stream network where upstream processes tend to highly affect downstream states [63]. Consequently, in case of observed spatial dependence, we model the residuals from the LASSO model as $\epsilon = z + \eta$, where z is a random effect with a covariance matrix, Σ, that accounts for spatial dependence, and η is the residual from the spatial model with mean zero and $\sigma^2 I$ variance. With this, the linear model from Equation (1) becomes

$$y = X\beta + z + \eta. \tag{2}$$

It is not feasible to estimate all entries in Σ because the matrix usually has more entries than the number of data points. Consequently, models with a small number of parameters are usually fit to the observed covariance. Commonly, these models are a function of the Euclidean distance between points, h, where closer locations tend to be more correlated than locations farther apart (see Appendix B for an example). When the covariance is estimated using this approach, the statistical method of Kriging is usually implemented to optimally predict (i.e., make inference on a random quantity of interest) at unobserved locations and/or times [63]. However, our covariance model should also be a function of stream distance because the spatial dependence between locations can be highly affected by their connectedness in the network and by the direction of the stream flow [64]. Stream distance, h_{sn}, is defined as the shortest distance between two locations, where distance is calculated along the stream network [65].

Standard spatial covariance models developed for Euclidean distances may not be valid when using stream distance (i.e., the covariance matrix may not be positive definite) [56,63,65]. Consequently, Ver Hoef et al. and Cressie et al. proposed moving average constructions (also called kernel convolutions) to develop valid models for stream networks [63,65]. The moving average is a standard construction for a single line segment that is continuous from $-\infty$ to ∞ on the real line. However, for stream networks, the line segments split into two at confluence points. Therefore, the moving average is also split into two parts. Each part is weighted to incorporate the flow rate of each segment at the confluence. In the absence of flow data for all confluences, because it is often difficult to measure flow for each stream segment, a proxy variable such as the area of each watershed is used [65]. This approach produces a valid covariance model for a stream network where the dependence of two flow-connected locations is non-zero while the covariance is zero for all pairs that are not flow-connected [56]. Here, we test for spatial dependence in ϵ, and if present, we use the stream network covariance models proposed by Ver Hoef et al. to model the residuals spatially [65].

Additionally, it is likely that the set of potential predictors included in the LASSO model is not comprehensive of all possible covariates that might influence DOC concentrations. These omitted covariates are likely to present spatial dependence due to their geospatial character [64]. Therefore, including a component of traditional covariance models for Euclidean distances may be appropriate if

ϵ exhibits spatial dependence between locations that are not flow-connected. In such a case, we would model the covariance of the random effects with Equation (3),

$$\Sigma = C_{sn}(h_{sn}|\boldsymbol{\theta}_{sn}) + C_{euc}(h|\boldsymbol{\theta}), \tag{3}$$

where $C_{sn}(h_{sn}|\boldsymbol{\theta}_{sn})$ is a spatial stream network covariance model, $C_{euc}(h|\boldsymbol{\theta})$ is a traditional covariance model based on Euclidean distance, and $\boldsymbol{\theta}_{sn}$ and $\boldsymbol{\theta}$ are vectors of parameters associated with each model. Common implementations of models based on Euclidean distance are linear, spherical, and exponential structures. We base the selection of the structure on the features of spatial dependence observed in the data and on the optimal structure associated with the root-mean-square error (RMSE) from leave-one-out cross-validation and the Akaike information criterion (AIC) [65]. See Appendix B for the mathematical construction of these covariance models.

4. Results and Discussion

The first step in the implementation of a LASSO regression is the selection of λ, which is the parameter that determines how strong the penalty is for the model coefficients. We compared three different approaches for the selection of λ: (a) leave-one-out cross-validation; (b) 10-fold cross-validation; and (c) generalized cross-validation. We repeated approach (b) for three different partitions of the data into ten random sub-groups. Multiple approaches are compared because each has its limitations. Leave-one-out cross-validation tends to overfit the data for big sample sizes, as is the case here; 10-fold cross-validation produces results that may vary depending on the partitioning of the sample; and generalized cross-validation does not withhold any of the data while fitting but instead includes a penalty to compensate for the fact that the same observations are used in fitting and predicting [66–68]. For approaches (a) and (b), we use the mean squared error (MSE) as the objective function, while for approach (c) we use a penalized version of the MSE (PMSE) given by Equation (4) [69]:

$$\text{PMSE} = \text{MSE}\left(\frac{n}{n-df}\right)^2 = \frac{n\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{(n-df)^2}, \tag{4}$$

where n is the sample size (1174 DOC base-flow concentrations); df is the number of degrees of freedom in the model, corresponding to the sample size minus the number of non-zero regression coefficients; y_i is the ith entry of $\boldsymbol{y}$ in Equation (2); and $\hat{y}_i$ is the predicted value of y_i from the LASSO regression model.

Figure 4 presents the value of the different objective functions for different values of λ, ranging from a non-restrictive value where all regression coefficients have non-zero values to a restrictive value where there is only one non-zero regression coefficient. The values of λ at which each approach reaches an optimum are variable; nevertheless, all approaches seem to produce stable values near their optimum for $\log(\lambda) \leq -3$. Consequently, we select the λ that produces the minimum number of predictors with non-zero regression coefficients while the change in the objective function is negligible with respect to its optimal value. Therefore, $\log(\lambda) = -3$ or $\lambda = 0.0498$ is selected.
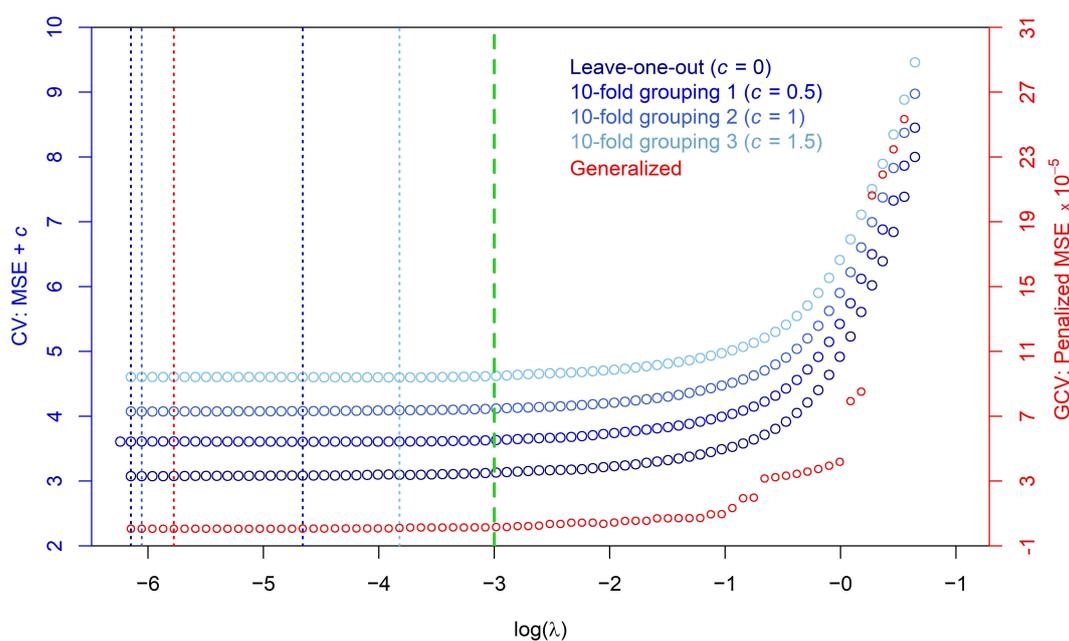
Figure 4. Cross-validation approaches for the selection of λ in the Least Absolute Shrinkage and Selection Operator (LASSO) regression. The horizontal axis represents different values of λ. The left vertical axis represents the mean squared error (MSE) objective function for the leave-one-out and 10-fold cross-validation (CV) approaches, where a constant c is added to avoid overlapping and to facilitate visualization. The right vertical axis represents the penalized MSE for the generalized cross-validation (GCV). The vertical dotted lines with colors corresponding to the curves represent the values of λ at which the corresponding objective functions reach their minimum value for each approach. The green dotted line represents the value of λ of maximum regularization (least number of potential predictors with non-zero regression coefficients) for which changes in the objective functions are negligible with respect to the optimal values.

Table 2 summarizes basic statistics, the transformations, and the estimated LASSO regression coefficients for all potential predictors. From the 33 potential predictors, only 17 have non-zero coefficients and are therefore significant. Only one of the three soil type coefficients is different from zero, and this implies that soil type as a predictor is significant. Similarly, six of the eight predominant aspect coefficients are different from zero, and this implies that aspect as a predictor is significant. The intercept coefficient has a value of 2.52 mg/L, which can be interpreted as the average DOC concentration during the months between August and November for a watershed during the end of the 1980s that has not been disturbed by fires, urbanization, or agriculture; may have been affected by MPB outbreaks but the trees have not transitioned to the gray phase yet; has an average maximum monthly temperature of 12 °C and an average monthly precipitation of 1.9 mm from August to November; has an average monthly precipitation of 2.3 mm for the rest of the months; has a 37% coverage of moderately permeable and transmissible soil (type B); has a mean elevation of 3000 m; has four or five wastewater point sources; and has an aspect that is predominantly flat or faces in the south or southeast directions. The average DOC concentration of 2.52 mg/L at control conditions is within the range of 1 to 3 mg/L, representative of an oligotrophic water body in which the low availability of nutrients limits primary productivity, leading to growth of algae. The low availability of organic matter reduces oxygen consumption, thereby increasing dissolved oxygen concentrations and supporting the growth of different fish species [70].

Table 2. Basic statistics, transformations, and LASSO coefficients when $\log(\lambda) = -3$ for all potential predictors. Basic statistics are reported for the untransformed features because the mean for the centered features would be zero, and the standard deviation for the scaled features would be one otherwise. The predictors with a coefficient significantly different from zero are highlighted. The accumulated *RdNBR* used corresponds to a decay of 90% in 5 years.

Covariate	Mean	Standard Deviation	Minimum	Maximum	Transformation	Coefficient Value
Intercept (mg/L)					None	2.5194
Time (year)	2000.32	6.47	1984	2012	Center to mean	0.0004
Accumulated *RdNBR*	3676.94	29,963.35	0	714,890.71	Scale by st. dev.	0.1257
MPB: Green phase fraction	0.00	0.02	0	0.30	None	0
MPB: Red phase fraction	0.01	0.04	0	0.40	None	0
MPB: Gray phase fraction	0.01	0.04	0	0.49	None	0.0218
Developed fraction	0.06	0.16	0	0.99	None	2.9290
Agricultural fraction	0.01	0.06	0	0.65	None	−0.1682
Forest fraction	0.40	0.30	0	1	None	0
Base-flow mean minimum monthly temperature (°C)	−4.45	3.16	−8.78	5.97	Center to mean	0
Non-base-flow mean minimum monthly temperature (°C)	−8.94	3.41	−14.15	2.18	Center to mean	0
Yearly mean minimum monthly temperature (°C)	−7.44	3.30	−12.15	3.33	Center to mean	0
Base-flow mean maximum monthly temperature (°C)	12.64	3.27	7.63	24.00	Center to mean	0.0999
Non-base-flow mean maximum monthly temperature (°C)	7.43	3.51	2.69	19.29	Center to mean	0
Yearly mean maximum monthly temperature (°C)	9.17	3.38	4.90	20.50	Center to mean	0
Base-flow mean monthly precipitation (mm)	1.89	0.65	0.54	4.63	Center to mean	0.1238
Non-base-flow mean monthly precipitation (mm)	2.26	0.94	0.19	5.35	Center to mean	−0.1299
Yearly mean monthly precipitation (mm)	2.14	0.77	0.45	4.86	Center to mean	0
Base-flow mean monthly SWE (mm)	11.91	12.33	0	74.05	Center to mean	0
Non-base-flow mean monthly SWE (mm)	161.37	134.08	0.05	1111.40	Center to mean	0
Yearly mean monthly SWE (mm)	111.55	91.99	0.07	757.00	Center to mean	0
Type A soil fraction	0.46	0.41	0	1	Center to mean	0
Type B soil fraction	0.37	0.31	0	1	Center to mean	0.6371
Type C soil fraction	0.09	0.20	0	1	Center to mean	0
Mean elevation (m)	3071.74	530.74	1517.78	3680.83	Center to mean	−0.0014
Number of wastewater point sources	4.42	11.99	0	55	Center to mean	0.0630
North predominant aspect	0.09	0.28	0	1	None	0.1195
Northeast predominant aspect	0.05	0.21	0	1	None	−0.1727
East predominant aspect	0.25	0.43	0	1	None	0.9548
Southeast predominant aspect	0.05	0.23	0	1	None	0
South predominant aspect	0.08	0.26	0	1	None	0
Southwest predominant aspect	0.14	0.34	0	1	None	0.0775
West predominant aspect	0.07	0.26	0	1	None	−0.4632
Northwest predominant aspect	0.27	0.45	0	1	None	−0.1511

The accumulated *RdNBR* values used in Table 2 correspond to a severity decay of 90% in 5 years. We repeated the LASSO regression using 10 and 15 years for the 90% severity decay (see Appendix A) and obtained the exact same significant predictors and negligible differences in the goodness of fit. Nevertheless, the coefficient corresponding to the accumulated *RdNBR* is reduced by 21% for a 10-year decay and 30% for a 15-year decay. The rest of the coefficients remain approximately the same except for the ones associated with the base-flow mean maximum monthly temperature, which increased by a factor of 2.3 for 10-year and 15-year decays; the mean elevation, which increased by a factor of 5.3 for 10-year and 15-year decays; and time, which increased by 26% for a 10-year decay and 41% for a 15-year decay.

No temporal trend is evident in the residuals from the LASSO regression; however, Figure 5 shows there is spatial dependence between pairs of nearest neighbors connected through water flow in the network. Figure 5a compares all network-connected nearest neighbors regardless of the year, and Figure 5b only compares network-connected nearest neighbors with observations in the same year. The correlation between nearest neighbors indicates that the value of the residual at a given location in the network tends to be similar to the value of its nearest upstream neighbor. Consequently, we use the moving average constructions proposed by Ver Hoef et al. [65] to model the covariance in the LASSO residuals. To do so, we use empirical semivariograms to determine the form of the covariance model [71]. In this case, the distance between locations used to estimate the semivariograms corresponds to the stream-distance, and the empirical semivariogram is called a torgegram [72].
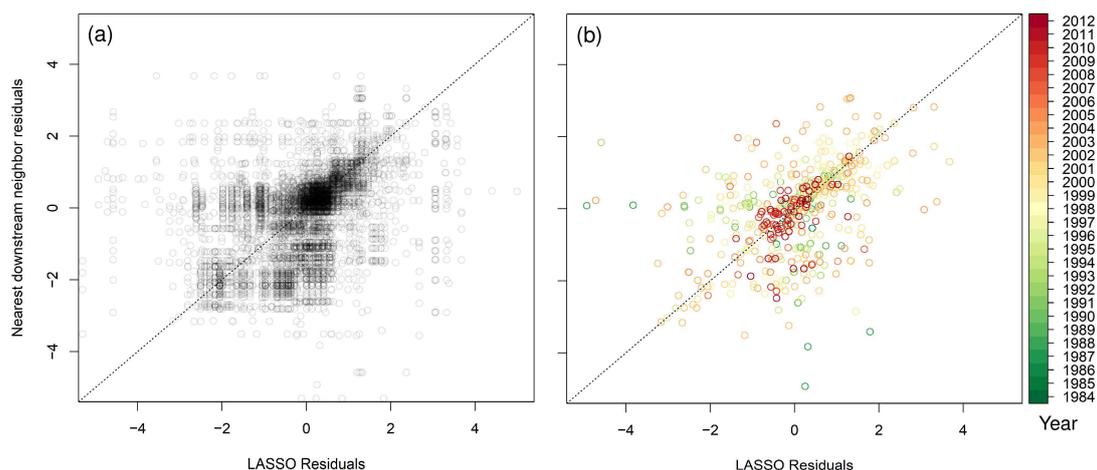
Figure 5. Comparison of the LASSO residuals with respect to the residual of the corresponding nearest downstream neighbors. The units of the residuals are mg/L. Panel (**a**) includes all neighboring pairs regardless of the year while panel (**b**) only includes neighboring pairs from the same year, and the color bar represents the corresponding year.

Figure 6 presents the empirical torgegram computed from all 1174 LASSO residuals. A low value plotted here indicates a high value of covariance and vice versa. Additionally, the distance at which the points level off represents the distance at which spatial dependence ceases, and it is called the range [71]. As depicted in Figure 6, flow-connected locations are more similar when they are close to each other, and this dependence decreases as distance increases with a range of approximately 50 km (measured as a stream-distance). Additionally, Figure 6 also presents the empirical semivariance for flow-unconnected locations, and it is clear that this spatial dependence is also present with a similar range as the one for flow-connected locations, indicating that the set of potential predictors included in the LASSO model is not comprehensive of all possible covariates with an effect on DOC concentrations and a geospatial trend. Consequently, we model the spatial dependence of the LASSO residuals using the covariance model from Equation (3). We used a single spatial model across time for all residuals because (1) they did not present a temporal trend; (2) the relationship observed in Figure 5a has a

similar pattern to the one in Figure 5b; and (3) the covariance model cannot be fit separately for each year because the number of observations in some years is very limited.
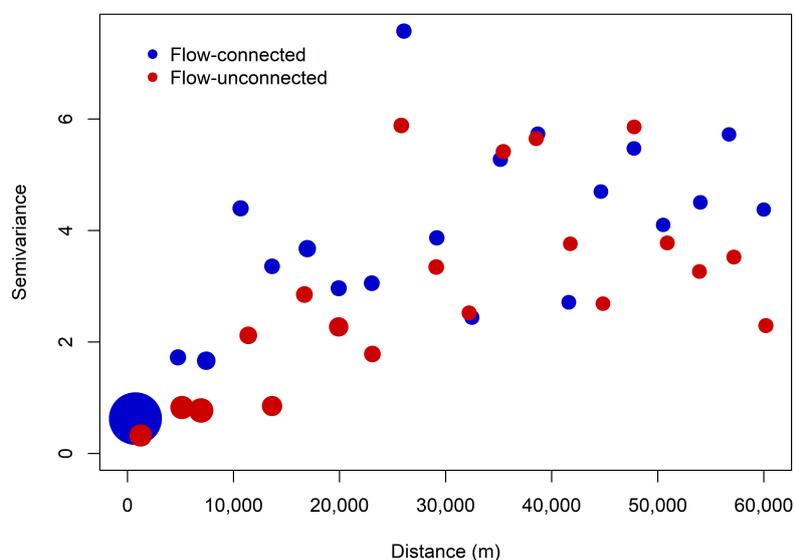

Figure 6. Empirical torgegram for flow-connected pairs (blue) and semivariogram for flow-unconnected pairs (red) in the network. The size of each point represents the number of pairs used to estimate the semivariance at each distance lag. The distance between flow-connected pairs corresponds to the stream distance, which is the shortest path between them following the network segments.

We compared different model types to represent the spatial dependence in Equation (3). For the stream network covariance, we fit linear, Mariah, spherical, and exponential models; while for the Euclidean covariance, we fit Gaussian, Cauchy, spherical, and exponential models (16 different combinations). The AIC value only varied by 0.14% across all model type combinations while the RMSE varied 0.77% with respect to their corresponding optimal values. The optimal combination varied depending on the criterion (AIC or RMSE). We select the exponential type of model for both the flow-connected spatial dependence and the Euclidean spatial dependence models because (1) it had the best AIC and the third best RMSE, showing consistency across both criteria, and (2) the torgegram shown in Figure 6 presents an exponential-like behavior in the decay of the covariance as distance increases for flow-connected and flow-unconnected locations. The RMSE of the LASSO regression is 1.141 mg/L, and it is reduced to zero when the spatial model is included because ordinary kriging is an exact interpolator that honors the data. However, if we perform a leave-one-out cross-validation with the spatial model, we obtain an improvement over the LASSO regression RMSE of 24.8%. This fitted model then allows us to make predictions of DOC at all other locations within the stream network.

4.1. Discussion of Static Significant Predictors

There are four significant predictors from the LASSO regression that do not change over time: number of wastewater point sources, soil type, elevation, and predominant aspect. Note that seven significant predictors from the LASSO regression do change over time (accumulated *RdNBR*, MPB-gray phase fraction, developed and agricultural land cover fractions, base-flow mean maximum monthly temperature, and base-flow and non-base-flow mean monthly precipitation) and will be discussed in Section 4.2. Figure 7 presents a qualitative comparison of the four static predictors with the predicted base-flow DOC concentrations for four years distributed across the 30-year period of interest. Inspection of Figure 7 indicates a strong positive correlation between the number of wastewater point sources and base-flow DOC concentrations. This is expected because the increased disposal of

organic waste into wastewater is associated with increased population densities in urbanized areas [73]. Nevertheless, the LASSO regression coefficient associated with wastewater point sources evidences a weaker correlation wherein the appearance of a new wastewater point source in a watershed would produce an average increase in the base-flow DOC concentrations of 0.063 mg/L. This is of course conditional on all other predictors remaining unchanged, which is unlikely in reality, but it is a convenient way to understand and measure the importance of a feature. For example, the appearance of a wastewater point source is usually accompanied with increased urbanization, which produces additional changes in DOC concentrations. This weak coefficient can be attributed to (1) the fact that the location of wastewater point sources does not account for the flow rate and associated pollution load present at each location; (2) there are non-point pollution sources associated with urbanization and industrialization that may be more significant contributors to DOC concentrations in surface water streams; and (3) although the location of a pollution source may not change over time, its pollution loads can be dynamic and highly affected by the urban changes over time associated with the discharge. Consequently, a predictor that accounts for the dynamics of urban development can be a more important predictor, as is the case with the land cover feature.
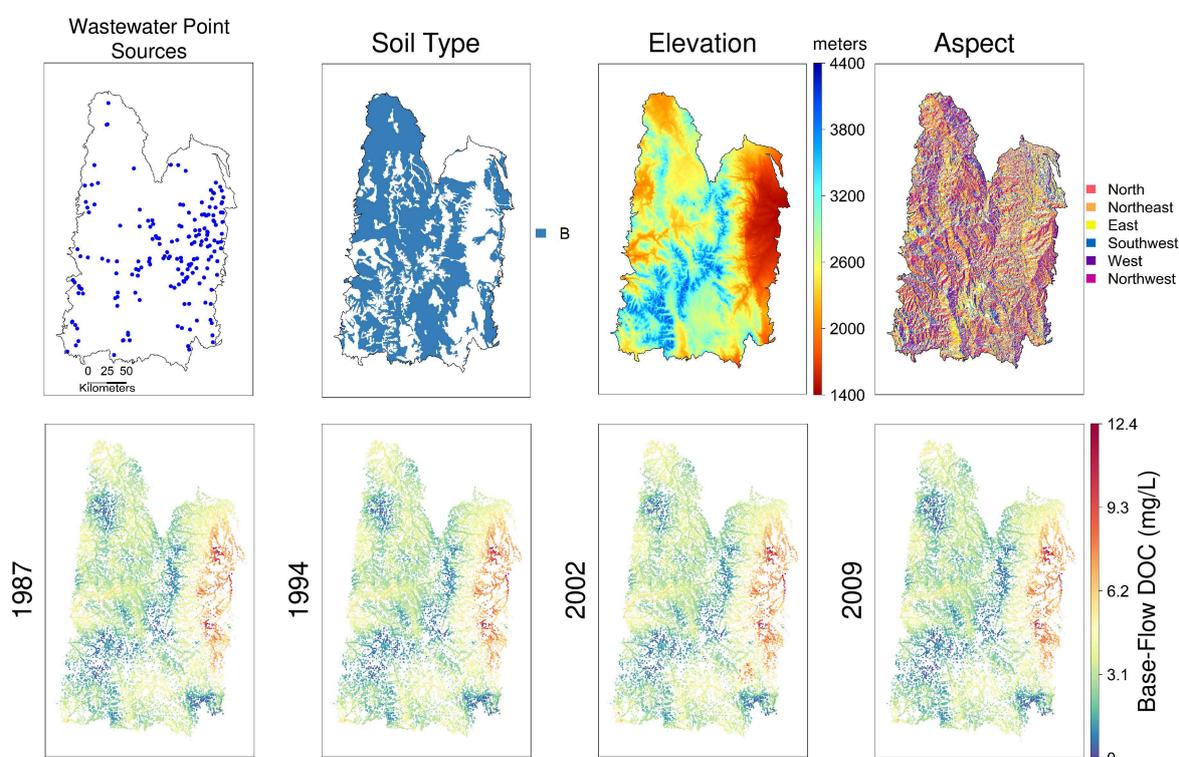
Figure 7. Comparison of the four time invariant significant predictors from the LASSO regression (**top row**) and predicted base-flow DOC concentrations in the region of interest for the years 1987, 1994, 2002, and 2009 (**bottom row**). For categorical predictors, only the significant classes are depicted.

Although Figure 7 does not depict a clear qualitative relationship between soil type B and base-flow DOC concentrations, the LASSO regression reveals a positive correlation between the two. A watershed fully covered by soil type B can be expected to have a 0.43 mg/L increase in average base-flow DOC concentrations with respect to a watershed with a typical distribution of soil types, which would have only 37% coverage of type B soil. This positive correlation is expected because type B soils present a moderate to high transmissivity, allowing a reasonable exchange between streams and groundwater during base-flow, while providing longer residence times than type A soils, and therefore enabling more dissolution of organic matter from the soil. Additionally, soils with an intermediate

transmissivity and a high water capacity, such as type B soils, tend to have greater concentrations of organic matter [74,75].

Elevation shows a very strong qualitative negative relationship with base-flow DOC concentrations in Figure 7. In general, low elevations tend to have higher DOC concentrations. The associated LASSO regression coefficient also reveals a negative correlation; however, its magnitude reveals a weaker dependence. A reduction of 500 m in the mean elevation of a watershed would represent a 0.71 mg/L increase in its average base-flow DOC concentration. The smaller LASSO coefficient is likely due to the fact that elevation is also correlated to other predictors that may have a more direct causal effect on DOC concentrations. For example, the majority of the urbanized and agricultural areas tend to occur below 2000 m in elevation (see Appendix A). Similarly, temperature has an almost linear inverse proportionality with elevation, and it has a substantial effect on the rate at which microorganisms hydrolyze and decompose organic matter [76]. Nevertheless, despite its correlation with other features, elevation is still a significant predictor. This is possibly accounted for by the relationship between watershed size and elevation. Watersheds with elevated pour/outlet/collection points tend to be smaller, and therefore, have shorter travel times for the water particles in the soil and the surface, where the organic matter is usually assimilated. Likewise, elevated watersheds tend to have less vegetation and thus fewer sources of organic matter. The opposite holds for watersheds with collection points at lower elevations.

Unlike for elevation, there is no clear qualitative relationship between aspect and base-flow DOC concentrations in Figure 7. Except for the Southeast and South predominant aspects, all other aspect orientations produce significant differences in the DOC concentrations of a watershed with respect to the Flat control aspect class (see Table 2). Although the aspects oriented to the Southwest are expected to have different DOC concentrations because they receive more solar radiation and thus have higher temperature and more MPB outbreaks, this effect is expected to be explained by the temperature and MPB predictors [77]. Of all of the significant aspect orientations, only two, the East and West predominant aspects, produce a change of more than 10% in the control base-flow DOC concentration (intercept value in Table 2). While a watershed facing in the East direction is expected to have an average increase of 0.955 mg/L in base-flow DOC concentrations with respect to a predominantly Flat watershed, a watershed facing in the west direction is expected to have an average decrease of 0.463 mg/L. East oriented watersheds tend to occur on the east side of the continental divide, where there is also more urbanization and industrial activities, causing increased organic matter contributions to surface water. Nevertheless, we expect this to be explained by the land cover features of the region of interest. Another likely reason for the significance of the aspect as a predictor despite its relationship with other features is that it is the main determinant on the exposure of a stream to solar radiation. Solar radiation can produce photochemical oxidation of DOM to inorganic compounds and alter the microbial metabolism associated with organic matter [78].

4.2. Discussion of Dynamic Significant Predictors

As mentioned earlier, time is included as a potential predictor because although the set of all other potential predictors incorporate several dynamic features, it may not be comprehensive of all possible covariates that represent biogeochemical processes over time in the study area with an effect on DOC concentrations. Consequently, time being a significant predictor in the LASSO regression indicates that there is a process affecting DOC concentrations beyond the included predictors. When the period of decay in fire severity is increased, the significance of time increases while the significance of fire severity decreases. This indicates that short-term effects of fire severity in DOC concentrations are more relevant, and therefore shorter decay periods produce better predictors. This confirms that the recovery of a watershed from a water quality standpoint tends to be much shorter than from a forestry/biomass standpoint, and it is approximately 5 years [79–86]. All subsequent results are therefore presented for a 5-year 90% decay of fire severity.

Although there are positive linear correlations between accumulated *RdNBR* and mean elevation and maximum monthly temperature, they are not strong correlations in the analyzed dataset. However, from a physical perspective, there is a logical reason why these two predictors become more significant as the fire severity predictor starts to give more importance to previous wildfires. The ignition of a wildfire is highly affected by the presence of combustible materials and an ignition source [87]. Elevation is an important determinant of the abundance of fuel because forested areas in the region of interest tend to occur above front-range elevations and below the tree-line (see Appendix A). Similarly, maximum temperatures are strongly associated with ignition, because as they increase, the magnitude needed for an external source of ignition, such as a flame or a spark, decreases [88]. Although their relationship may not be linear, elevation and maximum temperatures are thus expected to be linked to new wildfires. As accumulated wildfires become less significant predictors when past fires receive more importance, elevation and maximum temperature become more important as well.

As shown in Table 2, the LASSO regression coefficient corresponding to fire severity has a positive value indicating that base-flow DOC concentrations tend to increase with wildfire occurrences. This is expected because most overstorey and understorey surface organic matter (i.e., plants and roots) is burned [85], and wildfires tend to reduce the particle and aggregate sizes for organic matter, facilitating its dissolution while the transport of particulate and coarse organic matter as suspended material is reduced [80]. Consequently, higher concentrations of burned organic matter have been observed during low-flow periods [84]. However, it is important to note that the amount of fine and dissolvable organic matter that reaches stream base-flow can also be highly influenced by the effects of the fire on soil permeability and pore-sizes [83]. Similarly, fire reduces evapotranspiration, and thus may increase flow during dry periods, which may dilute dissolved compounds. The trade-off of this effect with increased fine organic matter will determine the response of each watershed individually [89].

Based on the LASSO regression results, an increase in one standard deviation of accumulated *RdNBR* in a watershed, which would mean a watershed with an area of 0.054 km^2 (5.4 ha) burned to a fire severity of 500 (a value commonly exceeded in wildfire affected areas) would produce an increase of 0.126 mg/L in average base-flow DOC concentrations. Although this seems to be a small effect, it is more substantial when we consider that the area of interest had more than 800 km^2 affected by wildfires during 2002 (see Appendix A). Specifically, Figure 8 compares the predicted base-flow DOC concentrations with the accumulated fire severity for four years distributed across the 30-year period of interest. Although small DOC changes are not easily identifiable due to the magnitude of differences between the fire-affected areas and the domain area, the year with largest fires in the period of interest is 2002, affecting 0.00015% of the domain area, and noticeable changes in DOC concentrations are produced. For instance, the largest fire in the 30-year period also ignited in 2002 (see Appendix A), affecting an area of 524 km^2 and increasing base-flow DOC concentrations by approximately 5 mg/L. The year 2012 represents the second year with the largest wildfire impacted areas in the study region and the period of interest (see Appendix A), affecting 0.00009% of the domain area. Although the fractional area affected by wildfires in the years 2002 and 2012 is small, they both produced an estimated increase of 0.01% in the average base-flow DOC concentrations across the entire domain area in each specific year with respect to the average base-flow DOC concentration across the entire domain area and the entire period of interest. The average base-flow DOC concentration corresponds to 3.4 mg/L, which is obtained from averaging the predicted values over the 17,091 prediction locations and all years in the 1983–2012 period.
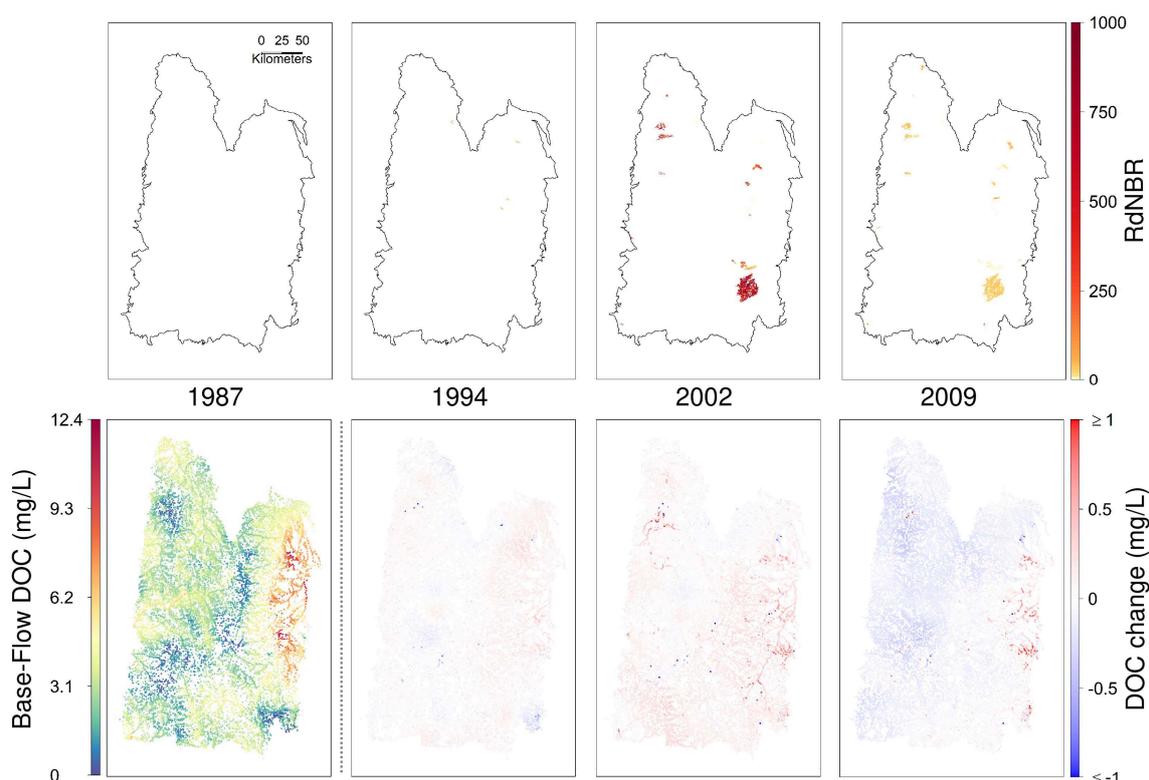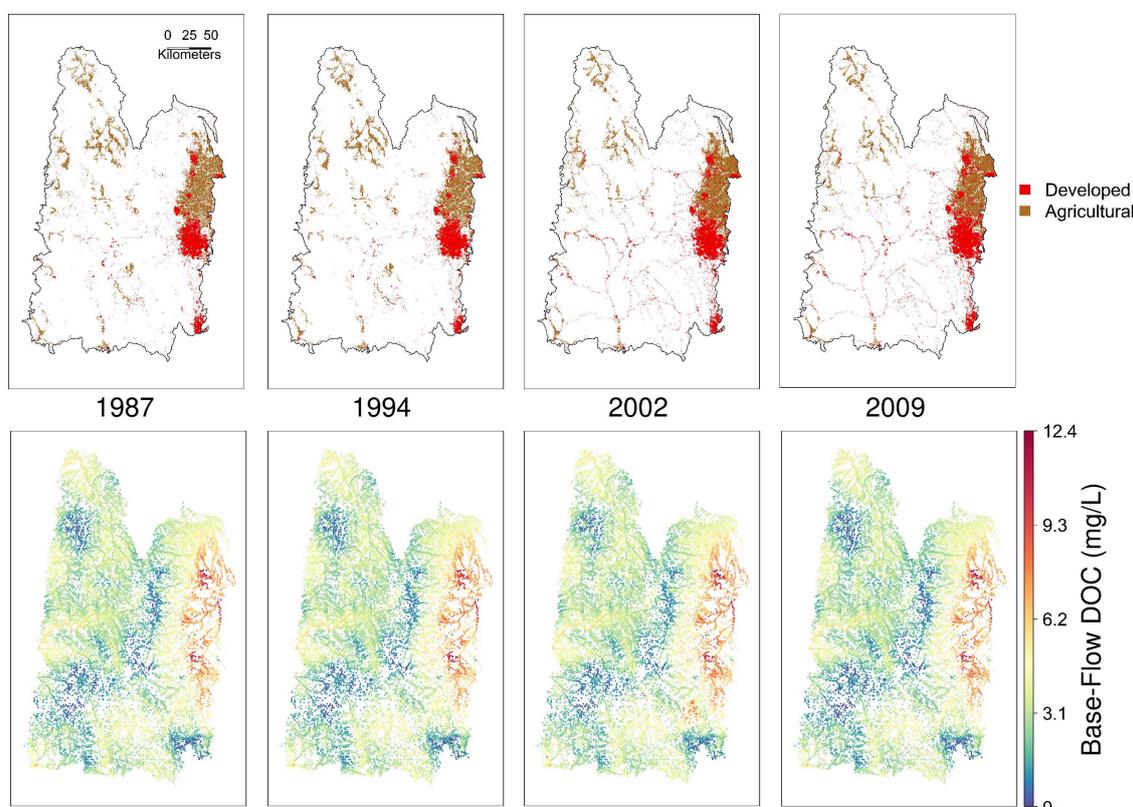
Figure 8. Comparison of fire severity and predicted base-flow DOC concentrations in the region of interest for the years 1987, 1994, 2002, and 2009. Instead of presenting the DOC concentrations for 1994, 2002, and 2009, the difference between their concentrations and the concentrations in 1987 are displayed to facilitate visualization. Although localized points reach *RdNBR* values of up to 13,000, these are not common, so the color scale used for *RdNBR* has an upper bound of 1,000 to facilitate visualization.

From the three different stages followed by a forest after an MPB outbreak, only the gray phase is a significant predictor in the LASSO regression. When beetle-killed trees transition to the gray phase, they lose their needles, increasing the input of degradable organic matter to the soil surface [38]. Consequently, the observed positive relationship is expected, but it is quite weak. A watershed fully covered by a gray-phase forest is expected to have an increase in average base-flow DOC concentrations of 0.022 mg/L with respect to an undisturbed watershed, which is less than 1% of the control base-flow DOC concentration (intercept value in Table 2). Similarly, a qualitative comparison of the gray-phase areas affected by the MPB with predicted DOC concentrations in the region of interest does not reveal any strong pattern (see Appendix C–Figures A10 and A11). A stronger correlation between MPB-infested areas and DOM has been observed during months with significant precipitation and individual runoff events [25,27,38,90], but we note that these studies did not control for the additional potential predictors considered in our study. Moreover, there may exist an important lag between the time at which the needles fall and the time by which their decomposition alters DOC concentrations, largely due to their recalcitrant chemical composition, especially in high-altitude and low-temperature climates [91,92].

The strong relationship between the land cover anthropogenic classes (developed and agricultural) and base-flow DOC concentrations is evidenced in Figure 9 and in their significance as predictors in the LASSO regression. The developed land cover class is the most significant predictor of base-flow DOC concentrations, producing an average increase of 2.93 mg/L when a watershed transitions from completely rural to fully developed. This is more than a 100% increase in the control base-flow DOC concentration (intercept value in Table 2). As mentioned earlier, this correlation is expected because

developed areas involve higher population densities and more concentrated pollutant and chemical loads, including DOC, to the receiving water bodies [42].

Figure 9. Comparison of developed and agricultural land cover with predicted base-flow DOC concentrations in the region of interest for the years 1987, 1994, 2002, and 2009.

As the developed land cover increases from 3 to 6% in the study area during 1983–2012 (see Appendix A), the maximum predicted DOC concentrations also increase from 11.5 to 12.3 mg/L. The front-range (east side of the study area) presents the largest developed zones including the Denver Metropolitan area, and the highest predicted DOC concentrations also occur in this region, exceeding 7 mg/L. Nevertheless, although a positive correlation between the agricultural land cover class and DOC concentrations is also expected from qualitative inspection of Figure 9, the LASSO regression coefficient reveals a negative correlation. Various studies have suggested that DOC in streams tends to increase with agricultural activities [93,94]. However, this increase has always been associated with higher concentrations of microbial-processed DOC and with lower concentrations of plant-derived DOC [95,96]. Consequently, the negative correlation observed here could be associated with a more significant reduction in plant-derived DOC than the increase in microbial-processed DOC. The reduction in plant-derived DOC is associated with the removal of native vegetation and harvesting, which in turn can have a negative effect on instream heterotrophic processes due to the reduced availability of organic matter [93,95]. On the other hand, our hypothesized less significant increase in microbial processed DOC could enhance the reactivity of catchment DOC emissions associated with higher carbon dioxide outgassing [93–95].

The three climatic features that are significant predictors in the LASSO regression are mean base-flow monthly precipitation, mean non-base-flow monthly precipitation, and mean base-flow monthly maximum temperature. Based on the regression results, an increase of 1 °C in the mean base-flow monthly maximum temperature of a watershed would represent a 0.099 mg/L increase in its average base-flow DOC concentration. Previous studies have found that DOC concentrations

increase with higher temperatures due to an elevated production and release of microbial-processed DOC [96]. This can also explain why minimum temperatures are not significant while maximum temperatures are as follows: the microbial maximum rate of metabolism occurs during the warmest periods of the day, which typically correspond to maximum temperatures [97,98]. Additionally, elevated maximum temperatures can increase evapotranspiration rates, reducing the base-flow rates, and therefore concentrating DOC for a fixed carbon input [99].

Precipitation is more important than maximum temperature as a DOC predictor. An increase of only 1 mm in the mean base-flow monthly precipitation of a watershed represents a 0.124 mg/L increase in its average base-flow DOC concentration. On the other hand, the same increase in the mean non-base-flow monthly precipitation of a watershed would have a negative effect of approximately the same magnitude, reducing the average base-flow DOC concentration by 0.130 mg/L. The base-flow period used here presents low precipitation with rare substantial runoff events. However, when such a runoff event occurs during the base-flow period, it has the potential of producing a storm flow flushing of DOC, explaining the observed positive correlation [100]. Similarly, small precipitation events during base-flow can enhance infiltration and thus the transport of subsurface DOC to streams. In other words, years with above average precipitation during base-flow months also tend to have higher DOC concentrations due to the flushing produced by the associated runoff events. On the other hand, although flushing can also occur during the non-base-flow period, it is not detected in base-flow DOC concentrations. Additionally, increased precipitation during the non-base-flow period can increase groundwater levels and thus produce higher base-flow rates, which would dilute the same contributions of organic matter and reduce DOC concentrations [101,102]. For example, the year 2002 had high base-flow precipitation and low non-base-flow precipitation. As seen in Figure 8, the year 2002 presents increased base-flow DOC concentrations that go beyond the acute effects of localized wildfires. Conversely, the year 2009 had low base-flow precipitation and high non-base-flow. As seen in Figure 8, the year 2009 presents decreased base-flow DOC concentrations over most of the study area. A qualitative comparison of these three significant climatological features with predicted DOC concentrations does not evidence a connection for the precipitation due to its spatial heterogeneity (see Appendix C), and the qualitative relationship between temperatures and predicted DOC concentrations is analogous to the one between elevation and DOC concentrations (see Figure 7) due to the strong link between temperature and elevation.

As mentioned earlier, time is a significant predictor in the LASSO regression, suggesting that the set of potential predictors used here is not comprehensive of all possible covariates that represent temporally dynamic geophysical processes with an effect on DOC concentrations. Although the coefficient for time is positive in Table 2, this does not indicate that base-flow DOC concentrations have been increasing over time in the region of interest for the period 1983–2012 because other dynamic predictors may be causing a decrease in base-flow DOC concentrations over time. However, if we re-fit the regression using ordinary least squares with time as the only predictor, then we obtain a model with an intercept of 2.87 mg/L and a coefficient associated with time that indicates an increase of 0.057 mg/L in base-flow DOC concentrations per year. Time is a significant predictor in this linear model with a *p*-value of 6.1×10^{-6}. Consequently, ignoring all other predictors, base-flow DOC concentrations present a temporal trend in the region of interest with an average increase of 1.7 mg/L from the beginning to the end of the 1983–2012 period.

When analyzing the significant predictors, the microbiology of the soils, a factor that could not be included in the LASSO regression, frequently appeared to be a possible link between some of the predictors and the response of the stream DOC concentrations. For example, microbial metabolism is highly affected by temperature, and microbial communities tend to vary across soil types. Additionally, because the set of potential predictors used here may not be comprehensive, as evidenced by time being a significant predictor, we hypothesize that important predictors not included in the LASSO regression are mainly associated with the abundance, diversity, spatial distribution, and dynamics of the microbiology in the soils of the Rocky Mountains. For example, soil microbiology has been found

to respond to forest disturbances, and, especially for gradual forest changes, this response can act as a buffer for water quality changes, possibly explaining why the MPB infestations, which develop over multiple years, are some of the weakest significant predictors for DOC in the streams of the Rocky Mountains [38,103,104] (contrasted to fires, which occur over weeks). Additionally, the spatial heterogeneity of microbial community composition in the Rocky Mountains is highly complex [105] and has been shown to change over time [106]. Finally, microbial communities may not only be a significant predictor of DOC concentrations, but also of the character of DOC, which highly determines its water quality relevance, such as the potential to enhance metal transport [13,107].

The relevance of microbiology goes beyond the abundance of DOM in the streams of the Rocky Mountains because the source of DOM in soils and water is the degradation of organic matter from plants and microbes, and its character is also determined by these sources. Typically, the main components of DOM in land ecosystems are humic substances, which are derived from the decomposition of plants, and is called allochthonous. Conversely, DOM sources resulting from the metabolism and degradation products of microbes are referred to as autochthonous. Disturbances such as fires, bark beetle outbreaks, and agriculture have shifted DOM abundances by increasing autochthonous organic matter in soils. Allochthonous DOM has a high concentration of aromatic species, which are correlated with metal–ligand exchange and the absorbance of ultra-violet radiation. Consequently, the character of the DOM is relevant for its effects on water quality, and future work includes analyzing how the abundance and diversity of microbial communities in the Rocky Mountains are linked to the concentration and character of DOM in its streams.

5. Conclusions

DOC is a relevant water quality constituent of lotic water bodies because it affects heterotrophic processes, acts as a source for the nutrient cycle, absorbs sunlight radiation, alters the transport of metals, and can promote the appearance of carcinogenic byproducts during water treatment. The Rocky Mountain region has suffered significant changes over recent decades, including anthropogenic development and natural disturbances such as wildfires and bark beetle infestations. These changes can modify the carbon cycle by altering primary productivity and carbon storage, and thus, it is insightful to determine if DOC concentrations have shifted in the region and to identify factors that contribute to these changes.

Some recent research has analyzed the effects of MPB infestations on carbon concentrations in the soils and streams of affected watersheds, but while some studies have identified a relationship between the two, others have not found evidence of a relationship [15,19,31]. These analyses typically involved single watersheds with small contributing areas and omitted other features that can also affect carbon concentrations, which may explain the conflicting results. Consequently, we have performed this large-scale analysis to account for numerous features that can affect DOC concentrations in streams, such as urbanization, wildfires, MPB infestations, and precipitation, among others.

We use a statistical model to assess the effect of climatological, morphological, anthropogenic, and forest disturbance features on base-flow DOC concentrations in the streams of a section of the Rocky Mountains with an area of 56,550 km^2 for the period 1983–2012. LASSO regression is used to account for the possible correlation among the predictors and to select predictors at the same time that the regression coefficients are estimated. Then, a spatial stream network model is fit to account for the spatial dependence in the residuals from the LASSO model, taking into account how the dependences are affected by the directionality of the water flow and the connectivity of the network. Taken together, the fitted LASSO and spatial dependence models are used to predict DOC throughout the RM stream network.

Wildfires, MPB infestations, urban development, agricultural development, monthly precipitation, maximum monthly temperature, hydrologic soil type, wastewater point sources, elevation, and aspect were found to be significant predictors of DOC. As expected, beetle-induced forest disturbances and wildfires appear to be related to increases in stream DOC because they release the carbon stored in

the biomass of the forest. Similarly, urbanization is also linked to increased DOC concentrations, likely due to higher population densities, and therefore, more organic pollutants in stormwater runoff and domestic/industrial wastewater. Surprisingly, the spread of agricultural activities seemed to produce a slight decline in DOC concentrations. Based on previous work, this effect may be associated with a strong decrease in plant-derived organic matter (because vegetation is harvested from crops) that is not offset by an increase in microbial-processed organic matter. Maximum temperatures are associated with higher DOC concentrations, most likely because higher temperatures are associated with increased microbial metabolism, and possibly because they produce lower base-flow rates, therefore concentrating the DOC. While increased precipitation during the base-flow periods seems to increase DOC concentrations, possibly because of increased infiltration and associated soil flushing, precipitation during the wet season seems to have the opposite effect by increasing base-flow rates and diluting organic matter. The reason is that base-flow rates depend on groundwater levels, which in turn are dictated by the depth of precipitation during the non-base-flow period. Watersheds with moderately permeable and transmissible soil types also appear to have higher DOC concentrations, likely because of higher organic matter content of the soils, and relatively higher infiltration rates that enable infiltration to efficiently flush DOC to groundwater, a primary contributor to base-flow. The composition of these moderately permeable and transmissible soil types is highly variable, and thus further analyses should utilize a more detailed soil classification such as the USDA textural soil classification that may better distinguish types of soil with abundant and soluble organic matter. The topography of the Rocky Mountains also influences DOC concentrations, perhaps due to its relationship to watershed size and solar radiation exposure.

Overall, DOC concentrations present an increasing trend over time for the region of interest, owing largely to the rapid anthropogenic development of the region. Urban development doubled during the analyzed 30-year period and was the most significant predictor of DOC in the statistical model. Additionally, wildfires can also be a significant contributor to DOC changes over time because recent decades have had more severe and extensive fires. Wildfires reduce the particle and aggregate sizes for organic matter by burning overstorey and understorey vegetation, facilitating the dissolution of organic matter. Finally, the microbiology of the soils provides a clear theoretical linkage between the processes that connect the causality of the important predictors with DOC in the streams. The significant predictors accounted for 62% of the variability of DOC concentrations. Therefore, microbial dynamics represent a feature of the region of interest that can potentially explain more of the temporal changes in DOC concentrations. Although related products, such as more detailed soil classifications, could have been included as a proxy for the microbiology and the biogeochemistry of the region, the temporal dynamics of these two features are not present in the data. Unfortunately, data quantifying the spatial or temporal abundance and diversity of microbial communities in the study area over the analyzed time period are unavailable and thus could not be included as a potential predictor. Microbial community data could be used at a smaller spatial–temporal scale to test this hypothesis in future work. Similarly, groundwater levels could be included since higher levels are usually associated with higher flow rates during the base-flow period, potentially diluting DOC concentrations. The opposite holds for low groundwater levels. However, data quantifying groundwater levels in the study area over the analyzed time period are unavailable.

The datasets and products associated with the predictors of DOC used in this study involved considerable work to select, compile, and process. These datasets represent a valuable input for other studies in the region associated with other water quality determinants in streams, the hydrologic behavior of the network, and the ecological composition and dynamics of communities, among others. Because the development of these datasets involved other organizations and scientists, the processed data is not publicly available, but interested parties can contact the authors of this work to receive authorized access. Because of the complex relationships among the predictors and DOC and the heterogeneous spatial domain, future work could also include the implementation of a mixture of

regression models to partition the observations into spatial subsets with similar relationships among explanatory variables and DOC concentration [108].

Even though our results indicate that MPB infestations are not as crucial to predicting DOC in RM streams compared to other relevant features, acute and localized effects and correlations with other water quality constituents have been reported previously. Moreover, the effects of the MPB infestations reach beyond water quality. For example, the timber and real-estate industries, tourism and recreation, and biodiversity support can also be positively and/or negatively impacted by MPB infestations [109,110]. Consequently, efficient forest management should integrate and account for all of these environmental and societal impacts when planning in the short and long term [111]; otherwise, unexpected side effects may arise. The MPB infestations are an example themselves of an unexpected consequence since previous management approaches, focused on short-term maximization of timber production, provided a connected and homogenous ecosystem that fostered the spread and expansion of MPB outbreaks [109]. Integrative management frameworks should be informed by combining the best available science and data [110], including results of studies such as this one. Human perceptions can vary and may be inaccurate, particularly with respect to the effects of MPB infestations [112]. For instance, social factors, such as value orientation, culture, and visual perception, have highly influenced public opinion and management practices regarding MPB infestations [111–113]. To conclude, we recommend that efforts in objectively informed and integrative management frameworks for understanding the tradeoffs among ecosystem services and appropriate scientific formulations, such as the ones proposed by Maguire et al. and Morris et al. [109,110], respectively, continue to be developed.

Author Contributions: Nicolás Rodríguez-Jeangros proposed the study; carried out the high-performance computing, coding, and analysis; and is the primary author on the manuscript. Amanda S. Hering contributed to conceptualizing the project, developing the methodology, and writing the manuscript. John E. McCray contributed to conceptualizing the project, interpreting the processes impacting DOC, and proofreading the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

As described in Section 1, climate change can influence stream water quality through multiple processes, and other factors such as the morphology and natural processes of a region are also strong determinants of surface water quality. This section describes in detail the different features that represent stream-water quality determinants in the region of interest and that are used as potential predictors.

Appendix A.1 Fires

The forested regions of the Rocky Mountains are one of the most susceptible ecosystems to wildfires, and this is especially relevant from a hydrological standpoint because the impact of fires on a forest can potentially produce drastic alterations on water quality and quantity processes. The reason is that the loss of vegetation and the soil transformations can change the geochemical cycle [81] and

the flow pattern [89]. Analyzing the temporal sequence of recovery in a given forest affected by a fire, termed forest succession, is challenging because it would be necessary to fund and implement a project with a time span on the order of a century. Hence, most succession studies are based on chrono-sequencing, an approach in which the regrowth is followed in multiple plots where original forests have been disturbed at different times [114]. Fires play a central role in the ecological cycling of a forest because, for example, they can release cone seeds that have been in the soil for several years [115].

Although the majority of the vegetation recovers within the first five to ten years following a wildfire, the time taken by climax species (mainly pine trees in the Rocky Mountains) to return to full height can range from forty years to a century [115–117]. Specifically, studies on local regions have reported that the biotic cover takes at least four years to recover to pre-fire levels [118]. For example, Turner et al. and Kinoshita and Hogue analyzed hundreds of areas affected by fires, finding that biomass reached a stable state after approximately five years and that it equaled pre-fire biomass by the end of the seventh post-fire year [89,119]. Similarly, Savage and Mast reported that the increase in abundance of trees in pine forests stabilizes between three and twenty years after a fire [120].

Nevertheless, it is important to recognize that although a fire-recovered forest may have approximately the same biomass and tree abundance as the pre-fire forest, the recovered forest would not be geochemically and hydrologically identical to the pre-fire forest. This is due to the influence of random effects in its development, such as weather patterns, and in some cases, grazing [114,115].

Although the complete recovery of a watershed from a forestry standpoint may span multiple decades, the recovery associated with water quality seems to be shorter. Statistically significant differences in post-fire water quality have a duration that ranges from 2 months to 10 years, and in the Rocky Mountains, most studies have found that water quality is fully recovered after approximately the fifth post-fire year at which time it is not statistically different from the pre-fire state [79–86].

Finally, it is important to recognize that the severity of a wildfire is spatially variable, and the severity of a fire is highly correlated to the succession pattern followed by the forest [79,117]. For example, Baker et al. remark that low-severity fires in the Rocky Mountain National Park and south-western Colorado may only take 10–20 years to regenerate [121]. However, few studies have quantified wildfire severity and extent effects on watershed responses in the Rocky Mountains [79].

Consequently, here we use the data from the Monitoring Trends in Burn Severity (MTBS) project, which maps the extent, size, and severity of all fires that occurred in our region of interest between 1984 and 2012 [33,39]. The fire severity metric from MTBS that we choose to use is the Relative differenced Normalized Burn Ratio ($RdNBR$) because it can characterize fire severity on low biomass sites and enhances inter-fire comparability at larger ecological scales [40]. $RdNBR$ is based on a normalization of the Differenced Normalized Burn Ratio ($dNBR$), which is the difference between the post-fire Normalized Burn Ratio (NBR) and the pre-fire NBR. NBR is computed with Landsat TM/ETM bands 4 (near-infrared—NIR) and 7 (shortwave-infrared—SWIR), as shown in Equation (A1).

$$NBR = \frac{NIR - SWIR}{NIR + SWIR}. \tag{A1}$$

This normalized index emphasizes the spectral response of fire-affected vegetation by contrasting photosynthetically healthy and burned vegetation. Because it is a normalized ratio, NBR ranges between -1 and 1. However, to simplify its manipulation, MTBS truncates NBR to three decimal places and multiplies it by 1000. Thus, NBR ranges between -1000 and 1000, and it only takes integer values. Hence, $dNBR$, calculated as $NBR_{post-fire} - NBR_{pre-fire}$, ranges between -2000 and 2000. Finally, $RdNBR$ is calculated using Equation (A2):

$$RdNBR = \frac{dNBR}{\sqrt{|NBR_{pre-fire}/1000|}}, \tag{A2}$$

where $NBR_{pre-fire}$ represents the pre-fire value of NBR. This normalization removes the biasing effect of pre-fire conditions.

Before the computation of NBR, MTBS pre-processes the pre-fire and post-fire Landsat imagery to generate top-of-atmosphere reflectance images that include geometric (including terrain correction) and radiometric correction. Pre-fire and post-fire images are also inspected for co-registration accuracy and are corrected if spatial differences are present [39]. Ideally, fires that occurred before 1983 should also be included in this analysis since, as described previously, their effects may extend for years. However, MTBS represents the most comprehensive and detailed dataset currently available that not only delineates the fire-affected area, but also estimates a spatially variable fire severity in a region that encompasses our entire study area during several years. Unfortunately, it only includes fires that ignited after 1983. Figure A1 presents the number of fires and the total affected area for the region of interest. The years in which the majority of fires occur are 1988, 1989, 2000, 2002, 2003, 2011, and 2012, each with at least three fires; however, years 2002 and 2012 have the greatest burned areas. Here, we integrate all $RdNBR$ fire images in a given year into a single raster map. Each fire is not only included as a potential predictor in the raster map corresponding to its year of occurrence, but also in subsequent yearly maps, since, as described earlier, a fire can affect water quality for a decade and also forestry for multiple decades.
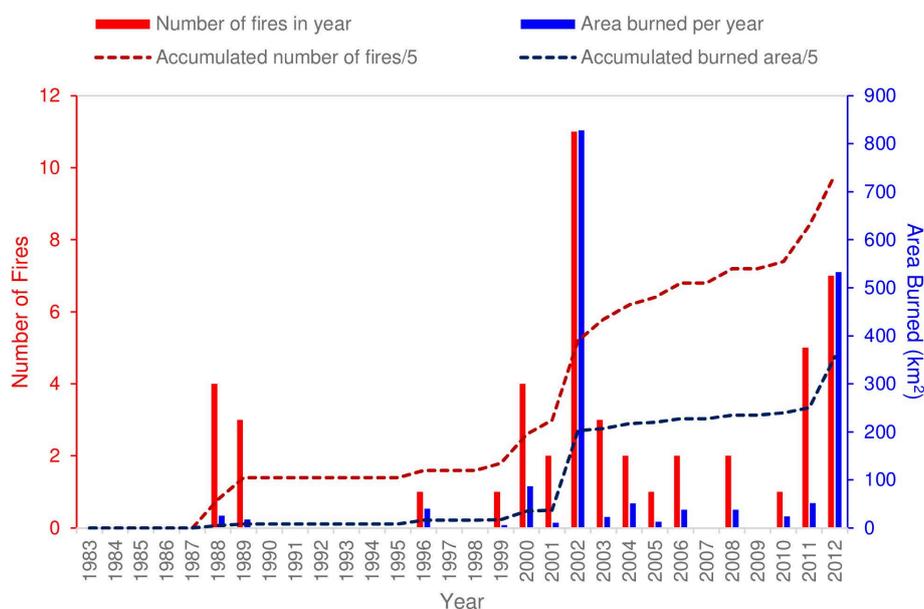
Figure A1. Number of fires in red and burned area per year in blue for the region of interest in the period 1983–2012. The dotted lines represent the accumulated number of fires and burned area scaled by a factor of 0.5 to facilitate visualization [33,39].

To confirm the post-fire recovery time-window, we select five fires in the region of interest that ignited during the time period 1983–2012 and had different sizes, severities, and locations. Figure A2 presents yearly $NBRs$ and Normalized Difference Vegetation Indices ($NDVIs$) averaged spatially for the area of each of these five fires (Figure A2) and for a control scenario corresponding to a region with no fires prior to 2012 (Figure A2). The NBR and $NDVI$ metrics are based on composites created from Landsat 5 TM orthorectified raw scenes [122,123]. These composites are created with images from the months of June, July, and August for each year to avoid seasonal effects and to focus on the most vegetated time of the year. Within each year, images with a cloud score index exceeding 10 are discarded, and each pixel in the composite is assigned with the median value from the remaining

images. *NDVI* quantifies the live green vegetation and is also a normalized ratio computed from bands 4 (near-infrared—*NIR*) and 3 (*RED*) [124], as shown in Equation (A3).

$$NDVI = \frac{NIR - RED}{NIR + RED}. \tag{A3}$$

Regions with scarce vegetation, such as barren rock, sand, or snow usually show low *NDVI* values. As vegetation becomes denser, *NDVI* values increase, and in densely vegetated areas, such as a tropical forest, or crops, it reaches values near 1. On the other hand, *NBR* does not have a direct conceptual interpretation, but *dNBR* does, with values of approximately 0 for unburned areas, slightly negative values for areas where there is strong post-fire vegetation re-growth, and positive values for burned vegetation, usually exceeding 0.6 with high burn severity.

Figure A2 shows how *NBR* and *NDVI* values reach an approximately stable state in all cases in 5 to 15 years (with values that are between 50% and 90% of the pre-fire value). The recovery follows an approximately exponential behavior (rapid initial recovery with an asymptotic stabilization), and *NBR* (Normalized Burn Ratio) and *NDVI* do not recover to the pre-fire values because, as mentioned above, the recovered forest is never exactly the same as the original due to random effects.
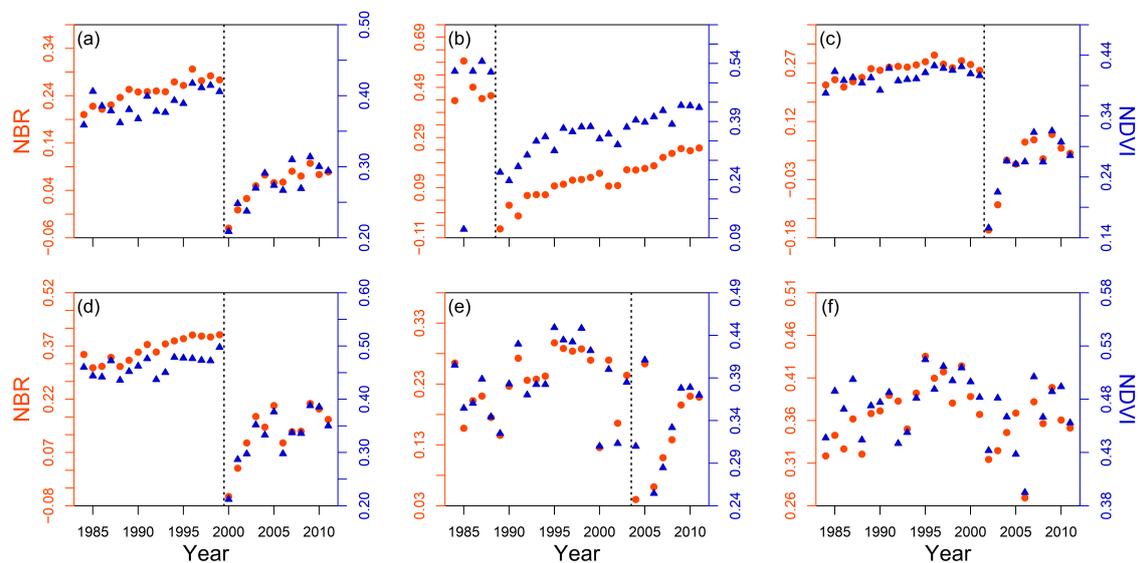

Figure A2. Normalized Burn Ratio (*NBR*) on left axis in red and Normalized Difference Vegetation Index (*NDVI*) on right axis in blue for five fires (**a**–**e**) that occurred at different locations in the region of interest between the period 1984–2011, with different severities and sizes ranging from mostly low to mostly high burn severity and from 5.9 to 523.7 km^2. The vertical dotted line represents the occurrence date of each fire. A control area is also included (**f**) corresponding to an area with no fires during the period depicted in the abscissa. See Figure A3 for fire locations.

Consequently, the *RdNBR* values associated with a fire from the MTBS project are included in the raster maps corresponding to subsequent years and are dampened using Equation (A4), where $RdNBR_{s_i,y_0}$ is the severity of the fire at location s_i in the year of its occurrence, y_0; $RdNBR_{s_i,y_j}$ is the dampened severity in a subsequent year, y_j (i.e., $y_j > y_0$); and θ is a parameter that represents the strength of the exponential decay.

$$RdNBR_{s_i,y_j} = RdNBR_{s_i,y_0} \cdot exp\left(-\frac{y_j - y_0}{\theta}\right). \tag{A4}$$

We choose an exponential decay model so that fire severity has a greater relevance in earlier post-fire years and asymptotically approaches a severity of zero. The *RdNBR* never reaches zero since, as described above, the nature of the recovered forest is never identical to the pre-fire forest due to random effects [114,115]. We test three different values for θ (2.17, 4.34, and 5.51) because these values produce a reduction of approximately 90% in the severity after 5, 10, and 15 years after the fire occurrence, respectively. These values are chosen because water quality tends to be completely recovered, and the majority of the biomass is also recovered within a period of 5–15 years [79–86,89,119]. Additionally, yearly *NBR* and *NDVI* values for fires in the region of interest in Figure A2 show recovery and stabilization within this period of time. Each of these cases is tested separately as a possible predictor for water quality.

Figure A3 presents an example of maps accumulating *RdNBR* values from previous years for 2002, 2006, and 2012 for the fire in Figure A2 and contrasts the exponential decay to 90% in 5 and 15 years. When $\theta = 5.51$ years, wildfire severity is reduced by approximately one order of magnitude after ten years. On the other hand, $\theta = 2.17$ years produces a reduction in the *RdNBR* values of approximately two orders of magnitude in ten years. It is also worth noting that although the maximum *RdNBR* in the region of interest for these three years exceeds 10,000, this is a localized and rare case, and the majority of fire affected areas have *RdNBR* values less than 4000.
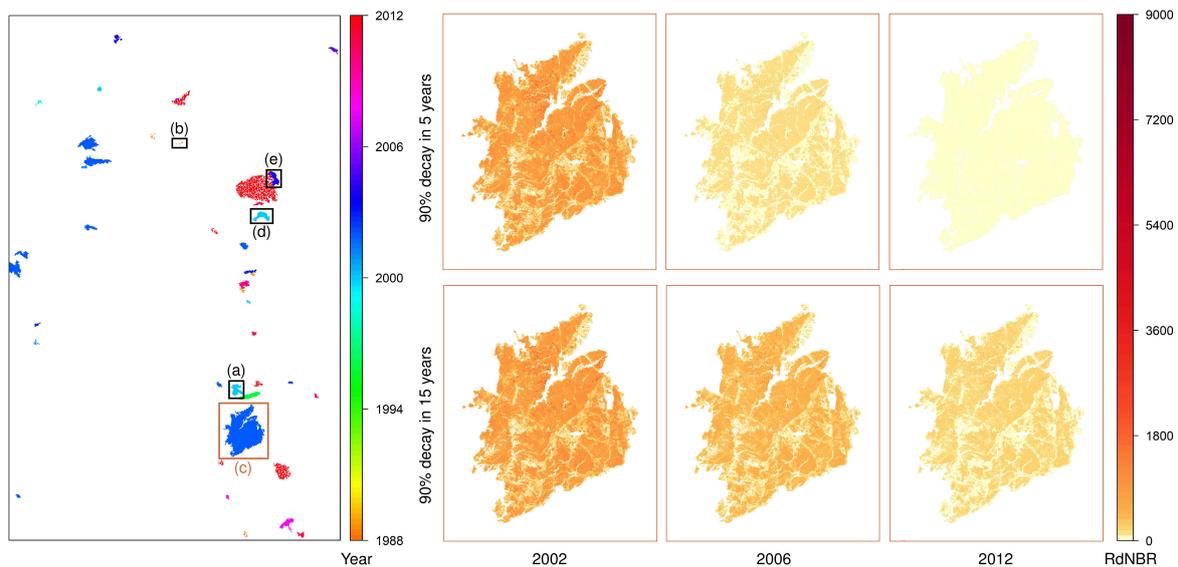
Figure A3. Fire affected areas in the region of interest with color indicating the year of ignition (left panel). The fire areas surrounded by rectangles in the left panel correspond to the fires in Figure A2, and the naming convention (**a**–**e**) matches the one used in Figure A2. *RdNBR* values for fire (**c**) for the years 2002, 2006, and 2012 are shown in the six panels on the right. The top row represents the case when the *RdNBR* values are accumulated over time with $\theta = 2.17$, and the bottom row represents the case when $\theta = 5.51$.

Appendix A.2 Mountain Pine Beetle

The region of interest has experienced significant tree mortality induced by MPB infestations possibly caused by warmer temperatures, more persistent droughts, and fire suppression [125,126]. Altered hydrology and biochemistry in the streams are potential consequences of these infestations. Although little research has been published regarding the effects of MPB infestations on water quality, increases in carbon and nitrogen concentrations have been observed in MPB-impacted areas at the hillslope scale [17,19]. Mikkelson et al. observed increased concentrations of water disinfection by-products associated with carbon compounds in water treatment facilities where the contributing watershed had suffered MPB disturbances [27]. Increased particulate and metal transport have also

been observed in impacted watersheds [13]. Consequently, we include MPB infestations as a potential predictor of water quality. MPB-affected forests usually follow three mortality phases. Initially, the trees remain green for approximately one year, and their metabolism slows down during this period. Next, the needles turn red, and the metabolic processes cease completely. Finally, the needles and twigs drop to the forest floor after approximately three years of being red, and the trees turn gray. These phases are usually referred to as the green, red, and gray phases due to the forest appearance.

We use the MPB infestation data previously processed by Brouillard et al. for the western United States [127]. These data are based on aerial surveys for forest insect and disease conditions from the United States Forest Service, which provides a yearly geospatial inventory of tree mortality and the associated cause, i.e., type of insect or disease [41]. Brouillard et al. filtered these data to only identify MPB as the damage causal agent, which is the greatest source of disease in trees in the region of interest by multiple orders of magnitude [127–130]. The aerial surveys are based on optical detection, and therefore, the first detection of an infestation is usually associated with the phase transition from green to red [28]. Consequently, Brouillard et al. considered this first detection to be the beginning of the red phase and accumulated the infested areas over time, transitioning them to gray phase after three years [127]. The aerial survey data has an accuracy of approximately 70% when assessed with respect to control ground references [128]. This accuracy increases to 80% when a 50-m tolerance is allowed [128], which is approximately the resolution at which the data is analyzed here. To further minimize errors, Brouillard et al. only included infestation data in regions where host tree species were present [127]. Although this dataset is not completely accurate, it represents the most spatially and temporally complete and precise product regarding MPB infestations in the region of interest, and it has been utilized in several studies that analyzed the consequences of the MPB infestations [6,126,130–132]. Finally, we add the green phase infestation period by including the locations with a first detection (transition from green to red phase) in the previous year map.

Figure A4 presents the temporal evolution of the fraction of the stream network area that has been affected by MPB infestations and shows three examples of the temporal distribution of the infestations in the region of interest for the years 2001, 2006, and 2011 that approximately correspond to the beginning, middle, and peak of the infestations. The area affected by MPB outbreaks is negligible before the year 1999. During the years 2011 and 2012, the green phase is also negligible, and therefore the total affected area stabilizes, reaching a maximum of 13% of the stream network. Approximately 2% of the area remains red and does not transition to gray by 2012. The fraction of each watershed affected by each of the three infestation phases in each year is included as a potential predictor of water quality.

Appendix A.3 Topography

Topography controls several drivers of water quality processes [46,47]. For example, higher elevations are associated with lower temperatures, which determine the speed of some biogeochemical reactions in the streams [45]. Similarly, the aspect and altitude of the terrain determine its exposure to ultraviolet light and solar radiation, which are also important determinants for biogeochemical reactions in streams [42–44]. Consequently, we include the average elevation of a watershed and its predominant aspect as possible predictors. These topographic features are calculated from the National Elevation Dataset [48,49]. Here, the aspect corresponds to a categorical variable with nine different classes: North, Northeast, East, Southeast, South, Southwest, West, Northwest, and Flat. Figure A5 depicts elevation and aspect and their corresponding frequencies for the stream network region of interest. The most predominant altitudes are between 2400 and 2900 meters, and although the East aspects are most prevalent, all eight directions have similar abundances. Only 0.4% of the region has Flat aspects, corresponding mainly to lentic water bodies.
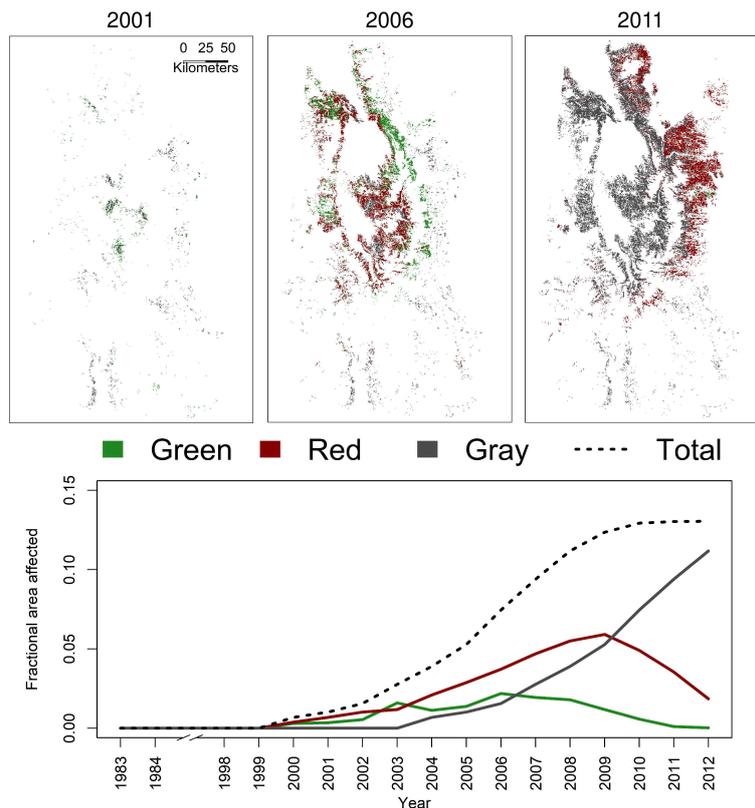
Figure A4. MPB infestation maps in the region of interest for the years 2001, 2006, and 2011 (**top**). This 10-year period depicts the beginning, middle, and peak of the infestations in the study area differentiating the different infestation phases. Temporal evolution of the fraction of the stream network of interest affected by the infestations is shown during the period 1983–2012 (**bottom**). The years between 1984 and 1998 are not depicted since the fraction of area affected is negligible.

Appendix A.4 Soil Types

Because soil types not only determine runoff patterns, but also affect biogeochemical processes that control water quality, we include soil type as a possible predictor [133–137]. The soil classification used here corresponds to the hydrologic soil groups defined by the Natural Resources Conservation Service division of the U.S. Department of Agriculture because this categorization is based on the similarity of physical and runoff soil characteristics [50]. These characteristics include depth to a restrictive layer or to the water table, water transmissivity, texture, structure, and degree of swelling when saturated, among others. There are four hydrologic groups:

- Group A includes pervious soils, with high infiltration rates and low runoff potential, and therefore a high transmissivity. These soils include drained sands, drained gravels, loamy sands, and sandy loams.
- Group B includes soils with moderate infiltration rates. These soils include moderately drained materials with textures that are not extremely fine or coarse.
- Group C includes soils with low infiltration rates with a moderately fine texture. These soils are mainly sandy clay loams.
- Group D includes highly impervious soils with high runoff potential, and therefore a low transmissivity. These soils are typically located in regions with high water tables or shallow impervious materials and include clay loams, silty clay loams, sandy clays, silty clays, and clays.
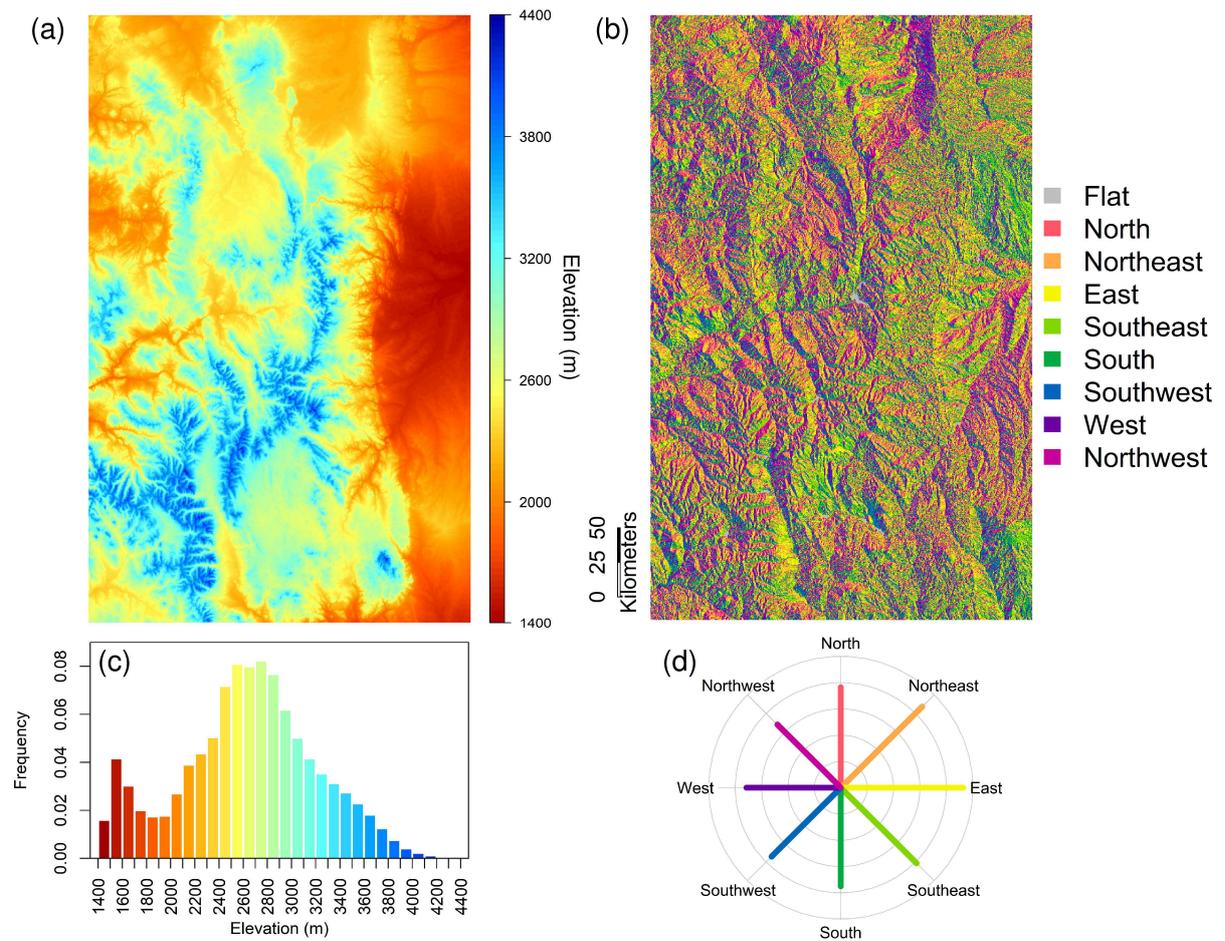
Figure A5. Digital elevation model for the region of interest (**a**), and the corresponding aspect categorized as North, Northeast, East, Southeast, South, Southwest, West, Northwest, and Flat (**b**). Frequency of elevations (**c**) and relative frequency of aspects (**d**) for the stream network of interest. The frequency of the flat aspect is not depicted, and it corresponds to only 0.4% of the region.

Figure A6 presents the compiled hydrologic soil group data for the region of interest from the U.S. General Soil Map, provided by the U.S. Department of Agriculture and manipulated with the Soil Data Viewer tool [138,139]. We use this product to compute four potential predictors: percentage of a watershed covered with soils in groups A, B, C, and D.
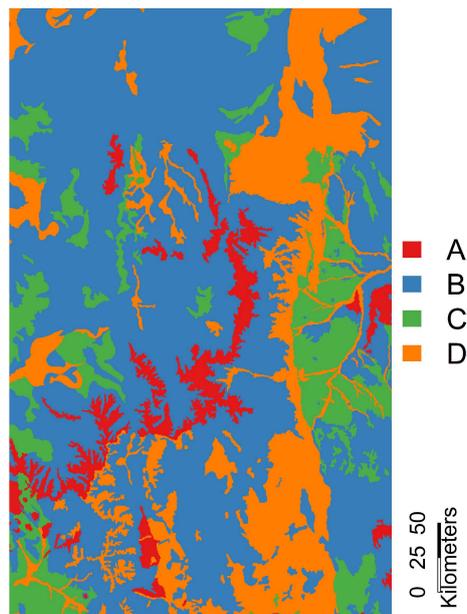
Figure A6. Hydrologic soil groups for the region of interest where A represents the highest transmissivities and infiltration rates, and these decrease in groups B and C, reaching highly impervious soils in group D.

Appendix A.5 Precipitation, Temperature, and Snow Cover

Meteorological factors such as temperature and precipitation can be two of the most relevant factors affecting stream hydrology [46,47]. Temperature controls water quality in the streams by altering biogeochemical cycles and reactions; for instance, the maximum dissolved oxygen concentration in a stream is highly and inversely correlated to temperature [140–142]. Flow rates and patterns in a stream are determined by the amount of precipitation that occurs and how much of this precipitation becomes runoff. Infiltration, evapotranspiration, and storage are the main processes determining the conversion of precipitation into runoff [143]. The hydrologic soil types described earlier serve as a proxy for the potential of precipitation to infiltrate depending on the soil features [50]. Similarly, temperature, elevation, aspect, and land cover are highly correlated to evapotranspiration [129,144–148]. Finally, the amount of water accumulated in the snow cover is the main determinant of water storage in the Rocky Mountains [149]. Consequently, we use temperature, precipitation, and snow water equivalent (SWE) as the potential predictors associated with hydrometeorology.

Minimizing transboundary discontinuities while maintaining a consistent gridding methodology in space and time are important features in a product for it to be suitable for analyzing large-scale hydrometeorological phenomena, as is the case here [52]. Therefore, the use of a single product for the entire space-time region of interest that includes temperature, precipitation, and SWE is ideal. However, most products do not include all three of these meteorological factors and/or do not cover the period 1983–2012. Accordingly, we use the hydrometeorological dataset developed by Livneh et al., referred to as Livneh-Climate hereafter [51,150,151]. Monthly means for the period 1983–2012 are obtained from this product for precipitation, maximum and minimum temperature, and SWE. The data are gridded to a 1/16° (~6 km) resolution. It is possible to find other data products with a finer resolution (~1 km) for hydrometeorological data; however, the Livneh-Climate product is chosen here because the products with a higher spatial resolution do not cover the period of interest, the study area, and/or do not have all three hydrometeorological factors of interest (temperature, precipitation, and SWE). If a product that meets our requirements were not available, it would be necessary to use multiple products and to predict their values at times and in locations with no data,

which would decrease the consistency across hydrometeorological factors and increase transboundary discontinuities and uncertainty [51,52].

The Livneh-Climate product is an observation-based meteorological dataset that offers insights into changes to hydro-climatic systems based on the detection of spatial–temporal characteristics. It is derived from daily temperature, precipitation, and wind observations from approximately 20,000 NOAA Cooperative Observer stations. These data were used to derive humidity and down-welling solar and infrared radiation by using algorithms that index these quantities to the daily mean temperature, temperature range, and precipitation. The precipitation data also followed an orographic adjustment. Livneh et al. used the Variable Infiltration Capacity model to produce estimates of soil moisture, SWE, discharge, and surface heat fluxes [51]. Finally, the data were aggregated into daily and monthly averages for the period 1950–2013 in the conterminous United States. The final product reduces transboundary discontinuities when compared to other commonly used reanalysis products because of its consistent gridding methodology [51,52,150,151].

We aggregate the monthly averages for the period 1983–2012 in the stream network of interest into three yearly averages: (i) a yearly average that only includes base-flow months, i.e., August–November (BF); (ii) a yearly average that only includes all the months that are not included in the base-flow case (nBF); and (iii) a yearly average that includes all months of the year (YR). Figure A7 presents an example of the yearly averaged hydrometeorological factors of interest in the study area for the year 2001, and it also displays the temporal evolution of these factors in the stream network of interest.

As seen in the top panels from Figure A7, higher elevations receive the majority of the precipitation, have lower temperatures, and accumulate most of the snow. Although the water quality data analyzed here only includes measurements during the base-flow period, we included the three different cases for the hydrometeorological yearly averages (base-flow months, non-base-flow months, and all months) as possible predictors because there can be a lag time between hydrometeorological changes and responses in the water quality. For example, the snow-melt can percolate to the groundwater table, and later on, during the base-flow period, groundwater is the main source of stream-flow during the months between late August and November. It is likely that the yearly averages in these three periods are redundant predictors in some cases, since, for example, temperature is highly correlated across these three times periods. However, temporal temperature fluctuations are stronger during the base-flow period, while the snow cover depth is negligible when compared to months outside of the base-flow period. In these cases, one of the three yearly average cases may be a better predictor of water quality. High precipitation during the end of the 1990s is associated with the El Niño Southern Oscillation event [152–155].

Appendix A.6 Land Cover

The water cycle, biogeochemical cycling, hydrologic processes, soil erosion, ecological community composition, ecosystem productivity, and other natural environmentally relevant processes are highly affected by land cover [129,144–148]. In some cases, the impact on the environment produced by land cover changes is equal to or greater than that produced by climate change [146,156,157]. Consequently, we use the SCaMF-RM land cover product to construct three predictors associated with land cover: percentage of a watershed covered with Developed, Forest, and Agricultural land cover classes [53]. One of the main advantages of SCaMF-RM is that it provides high resolution maps for the region of interest with a yearly frequency during the entire period 1983–2012. Figure A8 presents an example of the land cover coverage from these three classes during three years in the period of interest. Additionally, Figure A8 presents the temporal evolution of the fraction of the stream network covered by these three classes, where we observe that the majority of the region is covered by Forest, with an abundance that oscillates around 52% and is highly influenced by climatological events. For example, the Forest coverage increased substantially during 1997 and 1998, and these two years also presented above average precipitation due to a strong El Niño Southern Oscillation [152–155]. The Agricultural coverage is about 7% and is relatively stable during the period of interest. Finally,

urbanization is widespread in the stream network region, with the Developed class coverage changing from approximately 3 to 6% in 30 years [1–5].
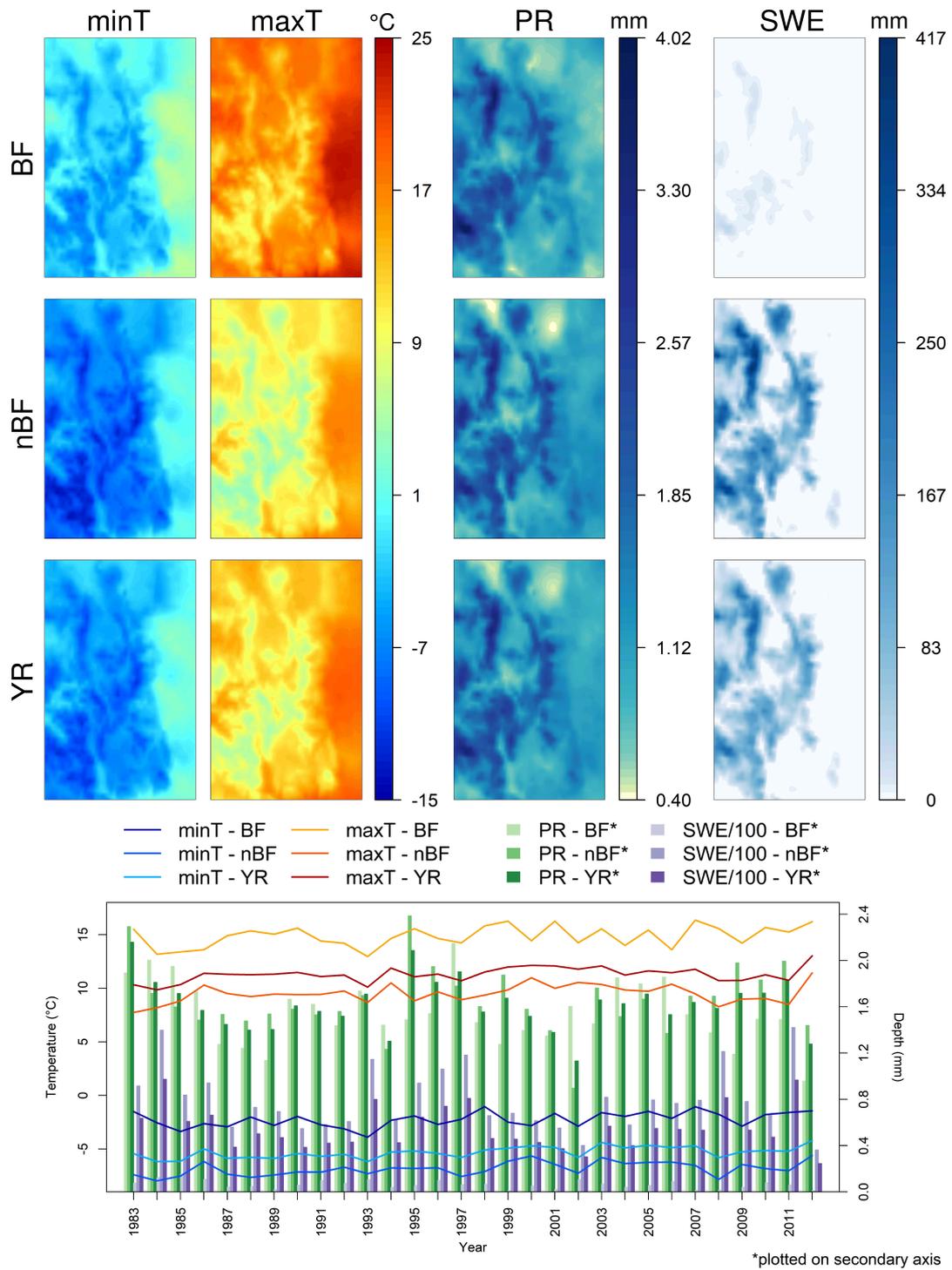
Figure A7. Minimum temperature (minT), maximum temperature (maxT), precipitation (PR), and SWE yearly averages for the cases including only base-flow months (BF), all other months (nBF), and all months of the year (YR). The top 12 panels present these averages for the year 2001 in the region of interest, while the bottom panel presents the evolution of the mean of these yearly averages in the stream network of interest for the period 1983–2012. SWE is scaled by two orders of magnitude in the bottom panel for visualization purposes.
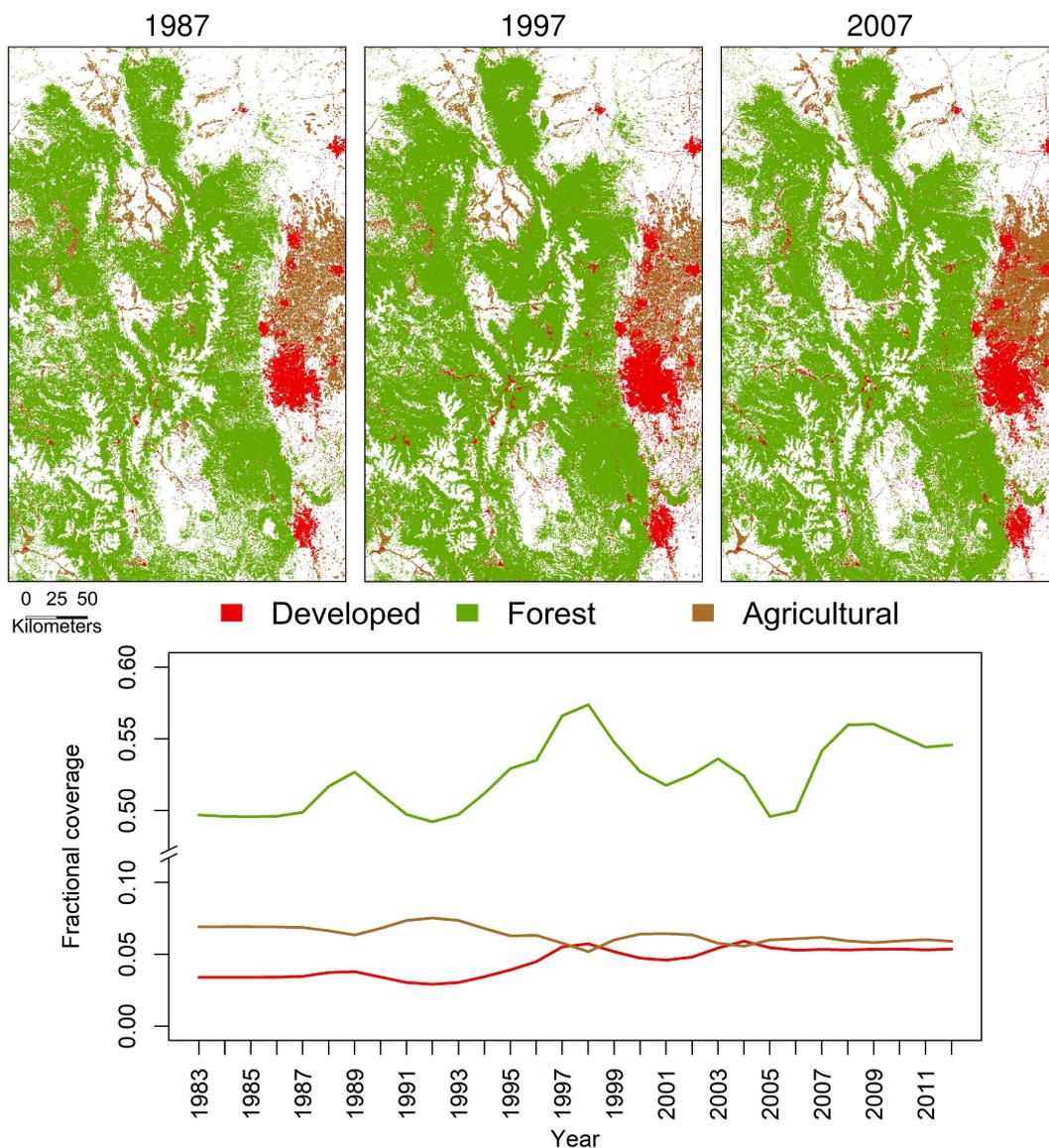
Figure A8. Land cover maps in the region of interest for the classes Developed, Forest, and Agricultural and the years 1987, 1997, and 2007 (**top**), and temporal evolution of the fraction of the stream network of interest covered by these three classes during the period 1983–2012 (**bottom**).

Appendix A.7 Wastewater Point Sources

The Rocky Mountains have been affected, as mentioned earlier, by anthropogenic activities such as urbanization and logging [1–5]. Urbanization not only alters the physical and chemical patterns of runoff, but it also produces wastewater sources of organic matter that can alter carbon concentrations and chemistry in surface waters [73,158–160]. The urbanization process in the region of interest is already represented by the Developed land cover class from the SCaMF-RM product, as described in detail in Section Appendix A.6. However, this representation does not capture the specific locations of point sources for untreated or treated wastewater.

We compile information regarding the location of wastewater treatment plants in the states of Colorado and Wyoming (the two states that intersect the region of interest) from their corresponding environmental authorities, i.e., the Department of Public Health and Environment-Water Quality Control Commission and the Department of Environmental Quality-Water Quality Division, respectively [54,55]. Ideally, this information would include a temporal component that differentiates

the wastewater sources between treated and untreated. Nevertheless, it is not required by any State Agency to keep track of this date by regulation [54], and therefore, these meta-data are not available. Thus, only the number of point sources of wastewater in each watershed are included as possible predictors. Figure A9 plots the 174 wastewater point sources compiled for the stream network of interest.
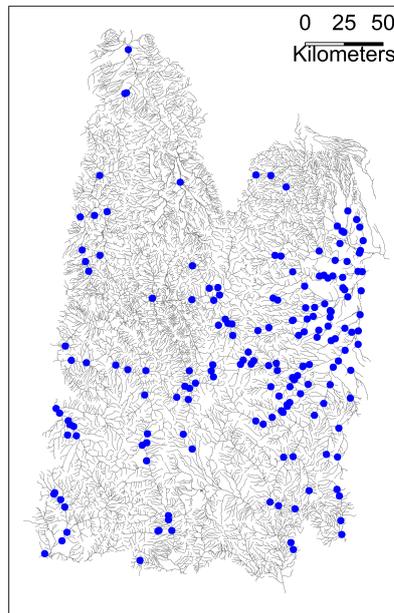
Figure A9. Wastewater point source locations (blue dots) for the stream network of interest (gray segments).

Appendix B

In a stream network, there are locations at stream junctions that are very close spatially, but their values for the variable of interest may be quite different because they are on two stream segments that do not share flow. In comparison, values in flow-connected locations tend to be more similar, as are values within a stream segment [65].

As an example, Equation (A5) presents a traditional linear bounded covariance model used for spatial prediction when distances are Euclidean.

$$C(h|\boldsymbol{\theta}) = \begin{cases} \theta_n + \theta_s & if\ h = 0 \\ \theta_s\left(1 - \frac{h}{\theta_r}\right) & if\ 0 < h < \theta_r \\ 0 & if\ \theta_r \le h \end{cases}, \tag{A5}$$

where $C(h|\boldsymbol{\theta})$ is the covariance between two locations separated by Euclidean distance h, and the model parameters $\boldsymbol{\theta} = (\theta_n, \theta_s, \theta_r)$ represent the nugget, the sill, and the range of dependence of the model. However, this model does not produce a positive-definite matrix for covariance if h is a stream network distance. Ver Hoef et al. proposed the model in Equation (A6) to obtain a positive definite covariance matrix when stream distance, h_{sn}, is used and when downstream locations depend spatially on upstream locations, but not the opposite [65]. Because of the asymmetric spatial dependence, these models are called tail-up covariance models.

$$C_1(h_{sn}|\boldsymbol{\theta}) = \int_{h_{sn}}^{\infty} g(x|\boldsymbol{\theta})g(x - h_{sn}|\boldsymbol{\theta})dx, \tag{A6}$$

where h_{sn} is the stream distance between two locations, and $g(x|\boldsymbol{\theta})$ is a moving-average function. As an example, this function takes the form given by Equation (A7) for a linear model with range parameter θ_r:

$$g(x|\boldsymbol{\theta}) = \vartheta \cdot \mathbb{I}\left(0 \leq \frac{x}{\theta_r} \leq 1\right), \tag{A7}$$

where ϑ is a parameter related to the sill, θ_s, as demonstrated later on; $\mathbb{I}$ is an indicator function that returns a value of 1 if the clause it evaluates is true and 0 otherwise; and the nugget effect is introduced in Equation (A9).

Because the range, θ_r, is a positive number, which can be practically understood as the distance when spatial dependence ceases between two locations, Equation (A7) can be re-written as:

$$g(x|\boldsymbol{\theta}) = \theta_s \cdot \mathbb{I}(0 \leq x \leq \theta_r), \tag{A8}$$

where $\theta_s = \vartheta \cdot \theta_r$.

Inserting Equation (A8) into Equation (A6) simplifies the covariance structure as follows:
If $h_{sn} = 0$:

$$C_1(0|\boldsymbol{\theta}) = \int_0^{\theta_r} (g(x|\boldsymbol{\theta}))^2 dx + \int_{\theta_r}^\infty 0 dx = \vartheta^2 \left(0 + x \,\Big|\, \begin{matrix} \theta_r \\ 0 \end{matrix}\right) + 0 = \vartheta^2 \theta_r.$$

If $h_{sn} > \theta_r$, we have that $x - h_{sn} < 0 \ \forall \ x : 0 \leq x \leq \theta_r$, and therefore $g(x - h_{sn}|\boldsymbol{\theta}) = 0$. Similarly, $g(x|\boldsymbol{\theta}) = 0$ for all other values of x. Therefore, $C(h_{sn}|\boldsymbol{\theta}) = 0$.
If $0 < h_{sn} \leq \theta_r$, the covariance is given by:

$$C_1(h_{sn}|\boldsymbol{\theta}) = \int_{h_{sn}}^{\theta_r} \vartheta g(x - h_{sn}|\boldsymbol{\theta}) dx + \int_{\theta_r}^\infty 0 g(x - h_{sn}|\boldsymbol{\theta}) dx = \vartheta \int_0^{\theta_r} g(x - h_{sn}|\boldsymbol{\theta}) dx.$$

To evaluate the integral, we can do a simple substitution with $u = x - h_{sn} \Rightarrow du = dx$:

$$C_1(h_{sn}|\boldsymbol{\theta}) = \vartheta \int_{u=h_{sn}-h_{sn}}^{u=\theta_r-h_{sn}} g(u|\boldsymbol{\theta}) du = \vartheta \int_{u=0}^{u=\theta_r-h_{sn}} \vartheta du = \vartheta^2 \left(0 + u \,\Big|\, \begin{matrix} \theta_r - h_{sn} \\ 0 \end{matrix}\right) = \vartheta^2(\theta_r - h_{sn}).$$

Thus, the covariance model is simplified to:

$$C_1(h_{sn}|\boldsymbol{\theta}) = \begin{cases} \vartheta^2(\theta_r - h_{sn}) & if \ 0 \leq h_{sn} \leq \theta_r \\ 0 & if \ h_{sn} > \theta_r \end{cases}, \tag{A9}$$

Equation (A9) can be reorganized to a more standard form, setting $\theta_s = \vartheta^2 \theta_r$:

$$C_1(h_{sn}|\boldsymbol{\theta}) = \theta_s \left(1 - \frac{h_{sn}}{\theta_r}\right) \mathbb{I}\left(0 \leq \frac{h_{sn}}{\theta_r} \leq 1\right). \tag{A10}$$

Ver Hoef and Peterson remark that any moving average function can be chosen for $g(x|\boldsymbol{\theta})$ as long as it has finite volume in order to create a stationary process, and preferably, it should be centered at 0 where most of its mass occurs [64].

The covariance construction in Equation (A6) is defined for a single line segment that is continuous from $-\infty$ to ∞ on the real line. However, for stream networks, the line segments split into two. Therefore, the moving average is split into two parts as well, and each part is weighted using a scheme that maintains stationarity. The proportion of the area for each sub-watershed in the confluence can

be used as a weight. This allows the construction of a valid covariance model for the entire stream network, given in Equation (A11).

$$C(s_i, t_j | \boldsymbol{\theta}) = \begin{cases} 0 & if\ s_i\ and\ t_j\ are\ not\ flow-connected \\ C_1(0|\boldsymbol{\theta}) + \theta_n & if\ s_i = t_j \\ \prod_{k \in B_{s_i, t_j}} \sqrt{w_k} C_1(h_{sn}|\boldsymbol{\theta}) & otherwise \end{cases}, \qquad (A11)$$

where s_i represents a location that belongs to the stream segment i; t_j represents a location that belongs to the stream segment j; h_{sn} is the stream distance between s_i and t_j; θ_n is the nugget of the model; B_{s_i, t_j} is the set of stream segments between the locations s_i and t_j excluding the lowermost location; and w_k is the weight of stream segment k.

Assuming $\theta_n = 0$ and using the linear model for $C_1(h_{sn})$ developed earlier with unitary sill and range parameters, i.e., $\boldsymbol{\theta} = (0, 1, 1)$, the covariance model is reduced to Equation (A12) [56]:

$$C(s_i, t_j | \boldsymbol{\theta}) = \begin{cases} 0 & if\ s_i\ and\ t_j\ are\ not\ flow-connected \\ \prod_{k \in B_{s_i, t_j}} \sqrt{w_k}(1 - h_{sn}) & otherwise \end{cases}. \qquad (A12)$$

Appendix C
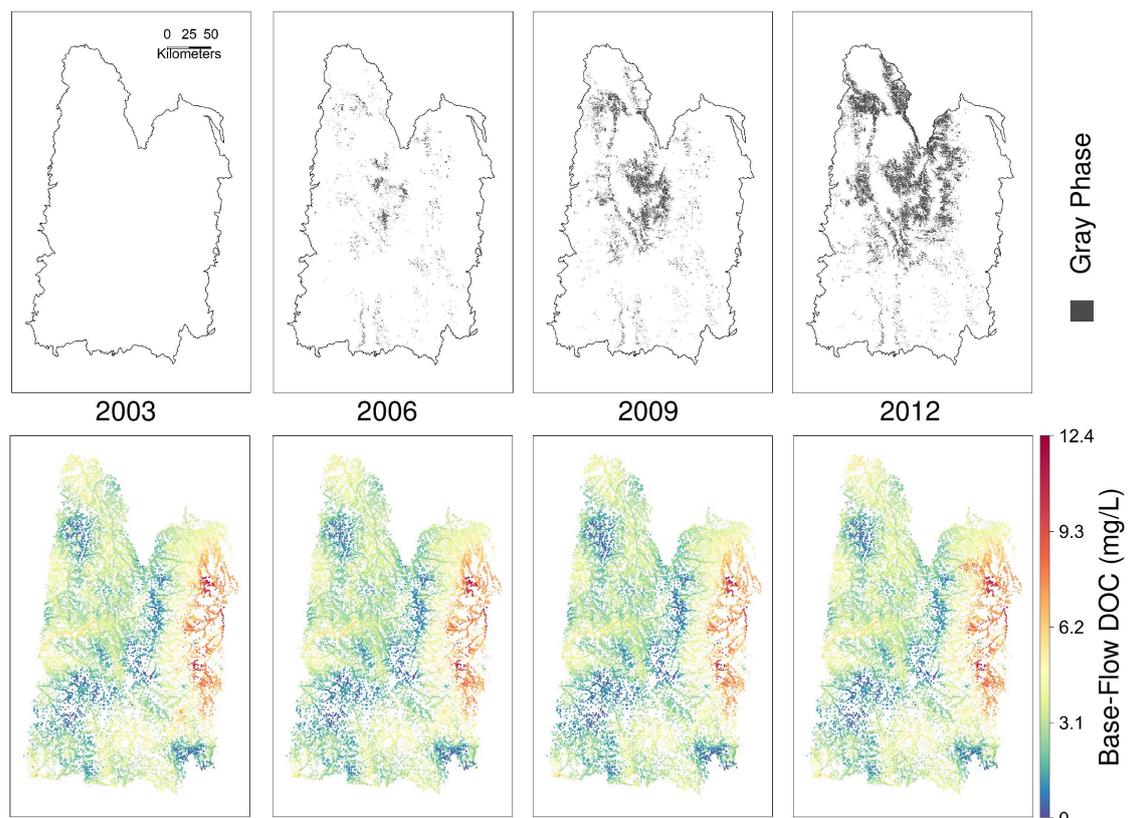
Figure A10. Comparison of gray-phase areas affected by the MPB and predicted base-flow DOC concentrations in the region of interest for the years 2003, 2006, 2009, and 2012.
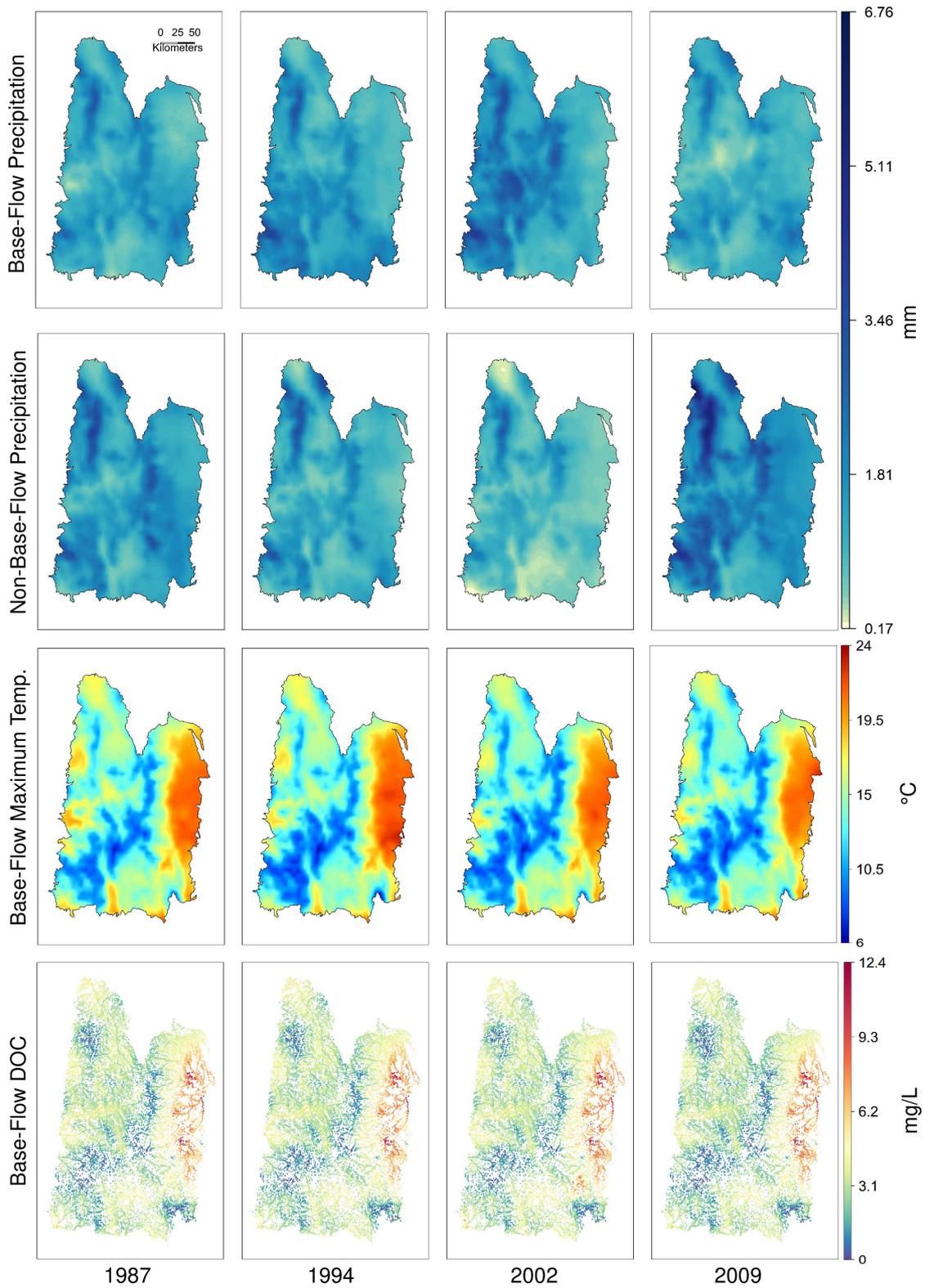
Figure A11. Comparison of significant climatological predictors (mean base-flow monthly precipitation, mean non-base-flow monthly precipitation, and mean base-flow monthly maximum temperature) and predicted base-flow DOC concentrations in the region of interest for the years 1987, 1994, 2002, and 2009.

References

1. Golubiewski, N.E. Urbanization Increases Grassland Carbon Pools: Effects of Landscaping in Colorado's Front Range. *Ecol. Appl.* **2006**, *16*, 555–571. [CrossRef]
2. Manfredo, M.J.; Zinn, H.C. Population change and its implications for wildlife management in the New West: A case study of Colorado. *Hum. Dimens. Wildl.* **1996**, *1*, 62–74. [CrossRef]
3. Pepin, N.; Losleben, M. Climate change in the Colorado Rocky Mountains: Free air versus surface temperature trends. *Int. J. Climatol.* **2002**, *22*, 311–329. [CrossRef]
4. Riebsame, W.E.; Gosnell, H.; Theobald, D.M. Land Use and Landscape Change in the Colorado Mountains I: Theory, Scale, and Pattern. *Mt. Res. Dev.* **1996**, *16*, 395–405. [CrossRef]
5. Smutny, G. Legislative Support for Growth Management in the Rocky Mountains: An Exploration of Attitudes in Idaho. *J. Am. Plan. Assoc.* **1998**, *64*, 311–323. [CrossRef]
6. Mikkelson, K.M.; Maxwell, R.M.; Ferguson, I.; Stednick, J.D.; McCray, J.E.; Sharp, J.O. Mountain pine beetle infestation impacts: Modeling water and energy budgets at the hill-slope scale. *Ecohydrology* **2013**, *6*, 64–72. [CrossRef]
7. Moore, M.L.; Six, D.L. Effects of Temperature on Growth, Sporulation, and Competition of Mountain Pine Beetle Fungal Symbionts. *Microb. Ecol.* **2015**, 1–12. [CrossRef] [PubMed]
8. Negron, J.F.; Fettig, C.J. Mountain pine beetle, a major disturbance agent in US Western coniferous forests: A synthesis of the state of knowledge [Research In Review]. *J. For.* **2014**, *112*, 257. [CrossRef]
9. Zhang, Z. A review of the Effects of Climate and Weather on Mountain Pine Beetle Population Dynamics and Impacts of Climate Change on Range Expansion in Canada. Bachelor's Thesis, University of Vancouver, Vancouver, BC, Canada, 2014.
10. Helie, J.F.; Peters, D.L.; Tattrie, K.R.; Gibson, J.J. *Review and Synthesis of Potential Hydrologic Impacts of Mountain Pine Beetle and Related Harvesting Activities in British Columbia*; Natural Resources Canada: Victoria, BC, Canada, 2005.
11. Pfeifer, E.M.; Hicke, J.A.; Meddens, A.J.H. Observations and modeling of aboveground tree carbon stocks and fluxes following a bark beetle outbreak in the western United States. *Glob. Chang. Biol.* **2011**, *17*, 339–350. [CrossRef]
12. Liang, L.; Hawbaker, T.J.; Chen, Y.; Zhu, Z.; Gong, P. Characterizing recent and projecting future potential patterns of mountain pine beetle outbreaks in the Southern Rocky Mountains. *Appl. Geogr.* **2014**, *55*, 165–175. [CrossRef]
13. Bearup, L.A.; Mikkelson, K.M.; Wiley, J.F.; Navarre-Sitchler, A.K.; Maxwell, R.M.; Sharp, J.O.; McCray, J.E. Metal fate and partitioning in soils under bark beetle-killed trees. *Sci. Total Environ.* **2014**, *496*, 348–357. [CrossRef] [PubMed]
14. Biederman, J.A.; Harpold, A.A.; Gochis, D.J.; Ewers, B.E.; Reed, D.E.; Papuga, S.A.; Brooks, P.D. Increased evaporation following widespread tree mortality limits streamflow response. *Water Resour. Res.* **2014**, *50*, 5395–5409. [CrossRef]
15. Norton, U.; Ewers, B.E.; Borkhuu, B.; Brown, N.R.; Pendall, E. Soil Nitrogen Five Years after Bark Beetle Infestation in Lodgepole Pine Forests. *Soil Sci. Soc. Am. J.* **2015**, *79*, 282–293. [CrossRef]
16. Winkler, R.; Boon, S.; Zimonick, B.; Baleshta, K. Assessing the effects of post-pine beetle forest litter on snow albedo. *Hydrol. Process.* **2010**, *24*, 803–812. [CrossRef]
17. Clow, D.W.; Rhoades, C.; Briggs, J.; Caldwell, M.; Lewis, W.M., Jr. Responses of soil and water chemistry to mountain pine beetle induced tree mortality in Grand County, Colorado, USA. *Appl. Geochem.* **2011**, *26*, S174–S178. [CrossRef]
18. Mikkelson, K.M.; Bearup, L.A.; Maxwell, R.M.; Stednick, J.D.; McCray, J.E.; Sharp, J.O. Bark beetle infestation impacts on nutrient cycling, water quality and interdependent hydrological effects. *Biogeochemistry* **2013**, *115*, 1–21. [CrossRef]
19. Trahan, N.A.; Dynes, E.L.; Pugh, E.; Moore, D.J.P.; Monson, R.K. Changes in soil biogeochemistry following disturbance by girdling and mountain pine beetles in subalpine forests. *Oecologia* **2015**, *177*, 981–995. [CrossRef] [PubMed]
20. Tranvik, L.J.; Downing, J.A.; Cotner, J.B.; Loiselle, S.A.; Striegl, R.G.; Ballatore, T.J.; Dillon, P.; Finlay, K.; Fortino, K.; Knoll, L.B.; et al. Lakes and reservoirs as regulators of carbon cycling and climate. *Limnol. Oceanogr.* **2009**, *54*, 2298–2314. [CrossRef]

21. Boyer, E.W.; Hornberger, G.M.; Bencala, K.E.; McKnight, D. Overview of a simple model describing variation of dissolved organic carbon in an upland catchment. *Ecol. Model.* **1996**, *86*, 183–188. [CrossRef]

22. Wu, F.C.; Kothawala, D.N.; Evans, R.D.; Dillon, P.J.; Cai, Y.R. Relationships between DOC concentration, molecular size and fluorescence properties of DOM in a stream. *Appl. Geochem.* **2007**, *22*, 1659–1667. [CrossRef]

23. McKnight, D.M.; Boyer, E.W.; Westerhoff, P.K.; Doran, P.T.; Kulbe, T.; Andersen, D.T. Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *Limnol. Oceanogr.* **2001**, *46*, 38–48. [CrossRef]

24. Aiken, G.R.; Hsu-Kim, H.; Ryan, J.N. Influence of Dissolved Organic Matter on the Environmental Fate of Metals, Nanoparticles, and Colloids. *Environ. Sci. Technol.* **2011**, *45*, 3196–3201. [CrossRef] [PubMed]

25. Beggs, K.M.H.; Summers, R.S. Character and Chlorine Reactivity of Dissolved Organic Matter from a Mountain Pine Beetle Impacted Watershed. *Environ. Sci. Technol.* **2011**, *45*, 5717–5724. [CrossRef] [PubMed]

26. Cory, R.M.; Miller, M.P.; McKnight, D.M.; Guerard, J.J.; Miller, P.L. Effect of instrument-specific response on the analysis of fulvic acid fluorescence spectra. *Limnol. Oceanogr. Methods* **2010**, *8*, 67–78. [CrossRef]

27. Mikkelson, K.M.; Dickenson, E.R.V.; Maxwell, R.M.; McCray, J.E.; Sharp, J.O. Water-quality impacts from climate-induced forest die-off. *Nat. Clim. Chang.* **2013**, *3*, 218–222. [CrossRef]

28. Hart, S.J.; Schoennagel, T.; Veblen, T.T.; Chapman, T.B. Area burned in the western United States is unaffected by recent mountain pine beetle outbreaks. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 4375–4380. [CrossRef] [PubMed]

29. Harvey, B.J.; Donato, D.C.; Turner, M.G. Recent mountain pine beetle outbreaks, wildfire severity, and postfire tree regeneration in the US Northern Rockies. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 15120–15125. [CrossRef] [PubMed]

30. Wooten, G. Relationships between Bark Beetle Outbreaks and Subsequent Fire Severity in the 2006 Tripod Fire. Available online: http://www.okanogan1.com/ecology/webfire/research/Wooten-tripod-fire-bark-beetles-2014-10-01.pdf (accessed on 15 April 2018).

31. Beudert, B.; Bässler, C.; Thorn, S.; Noss, R.; Schröder, B.; Dieffenbach-Fries, H.; Foullois, N.; Müller, J. Bark Beetles Increase Biodiversity While Maintaining Drinking Water Quality. *Conserv. Lett.* **2015**. [CrossRef]

32. Bentz, B.J.; Régnière, J.; Fettig, C.J.; Hansen, E.M.; Hayes, J.L.; Hicke, J.A.; Kelsey, R.G.; Negrón, J.F.; Seybold, S.J. Climate Change and Bark Beetles of the Western United States and Canada: Direct and Indirect Effects. *BioScience* **2010**, *60*, 602–613. [CrossRef]

33. Finko, M.; Quayle, B.; Zhang, Y.; Lecker, J.; Megown, K.A.; Brewer, C.K. Monitoring Trends and Burn Severity (MTBS): Monitoring wildfire activity for the past quarter century using landsat data. In *Moving from Status to Trends: Forest Inventory and Analysis (FIA) Symposium*; Morin, R.S., Liknes, G.C., Eds.; U.S. Department of Agriculture, Forest Service, Northern Research Station: Baltimore, MD, USA, 2012; pp. 222–228.

34. Rodríguez-Jeangros, N.; Hering, A.S.; Kaiser, T.; McCray, J.E. SCaMF–RM: A Fused High-Resolution Land Cover Product of the Rocky Mountains. *Remote Sens.* **2017**, *9*, 1015. [CrossRef]

35. Nagel, D.; Wollrab, S.; Parkes-Payne, S.; Peterson, E.; Isaak, D.; Ver Hoef, J. *National Stream Internet Hydrography Datasets for Spatial-Stream-Network (SSN) Analysis*; Rocky Mountain Research Station: Fort Collins, CO, USA, 2017.

36. McKnight, D.M.; Bencala, K.E.; Zellweger, G.W.; Aiken, G.R.; Feder, G.L.; Thorn, K.A. Sorption of dissolved organic carbon by hydrous aluminum and iron oxides occurring at the confluence of Deer Creek with the Snake River, Summit County, Colorado. *Environ. Sci. Technol.* **1992**, *26*, 1388–1396. [CrossRef]

37. U.S. Geological Survey. Water Data for the Nation: The National Water Information System. Available online: https://www.usgs.gov/water-data-nation-national-water-information-system (accessed on 15 April 2018).

38. Brouillard, B.M.; Dickenson, E.R.V.; Mikkelson, K.M.; Sharp, J.O. Water quality following extensive beetle-induced tree mortality: Interplay of aromatic carbon loading, disinfection byproducts, and hydrologic drivers. *Sci. Total Environ.* **2016**, *572*, 649–659. [CrossRef] [PubMed]

39. Eidenshink, J.C.; Schwind, B.; Brewer, K.; Zhu, Z.-L.; Quayle, B.; Howard, S.M. A project for monitoring trends in burn severity. *Fire Ecol.* **2007**, *3*, 321. [CrossRef]

40. Miller, J.D.; Knapp, E.E.; Key, C.H.; Skinner, C.N.; Isbell, C.J.; Creasy, R.M.; Sherlock, J.W. Calibration and validation of the relative differenced Normalized Burn Ratio (RdNBR) to three measures of fire severity in the Sierra Nevada and Klamath Mountains, California, USA. *Remote Sens. Environ.* **2009**, *113*, 645–656. [CrossRef]

41. USDA Forest Service. *Rocky Mountain Region, Forest Health Management Annual Aerial Detection Overview Survey—Vector Digital Data of Forest Insect and Disease Conditions*; USDA Forest Service: Washington, DC, USA, 1950.

42. Wear, D.N.; Turner, M.G.; Naiman, R.J. Land Cover Along an Urban–Rural Gradient: Implications for Water Quality. *Ecol. Appl.* **1998**, *8*, 619–630. [CrossRef]

43. Mitasova, H.; Hofierka, J.; Zlocha, M.; Iverson, L.R. Modelling topographic potential for erosion and deposition using GIS. *Int. J. Geogr. Inf. Syst.* **1996**, *10*, 629–641. [CrossRef]

44. Oates, P.M.; Shanahan, P.; Polz, M.F. Solar disinfection (SODIS): Simulation of solar radiation for global assessment and application for point-of-use water treatment in Haiti. *Water Res.* **2003**, *37*, 47–54. [CrossRef]

45. Chapra, S.C. Reaction Kinetics. In *Surface Water-Quality Modeling*; Waveland Press: Long Grove, IL, USA, 2008; pp. 24–42, ISBN 978-1-4786-0830-1.

46. Pratt, B.; Chang, H. Effects of land cover, topography, and built structure on seasonal water quality at multiple spatial scales. *J. Hazard. Mater.* **2012**, *209*, 48–58. [CrossRef] [PubMed]

47. Ye, L.; Cai, Q.; Liu, R.; Cao, M. The influence of topography and land use on water quality of Xiangxi River in Three Gorges Reservoir region. *Environ. Geol.* **2009**, *58*, 937–942. [CrossRef]

48. Gesch, D.B.; Oimoen, M.J.; Greenlee, S.K.; Nelson, C.A.; Steuck, M.J.; Tyler, D.J. The national elevation data set. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 5–31.

49. U.S. Geological Survey. *National Elevation Dataset (NED) 1/9 Arc-Second Downloadable Data Collection*; U.S. Geological Survey: Reston, VA, USA, 2014.

50. United States Department of Agriculture—Natural Resources Conservation Service Hydrologic Soil Groups. *Hydrology National Engineering Handbook*; United States Department of Agriculture—Natural Resources Conservation Service Hydrologic Soil Groups: Washington, DC, USA, 2017.

51. Livneh, B.; Rosenberg, E.A.; Lin, C.; Nijssen, B.; Mishra, V.; Andreadis, K.M.; Maurer, E.P.; Lettenmaier, D.P. A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions. *J. Clim.* **2013**, *26*, 9384–9392. [CrossRef]

52. Livneh, B.; Bohn, T.J.; Pierce, D.W.; Munoz-Arriola, F.; Nijssen, B.; Vose, R.; Cayan, D.R.; Brekke, L. A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013. *Sci. Data* **2015**, *2*, 150042. [CrossRef] [PubMed]

53. Rodríguez-Jeangros, N.; Hering, A.S.; Kaiser, T.; McCray, J. Fusing multiple existing space-time land cover products. *Environmetrics* **2017**, *28*, e2429. [CrossRef]

54. McGowan, S. *Surface Wastewater Treatment Plant Locations in Colorado, United States*; GIS Coordinator—Watershed Section; Colorado Department of Public Health and Environment-Water Quality Control Commission: Denver, CO, USA, 2017.

55. Lohman, P. *Surface Wastewater Treatment Plant Locations in Wyoming, United States*; Southeast District Engineer; Wyoming Department of Environmental Quality-Water Quality Division: Cheyenne, WY, USA, 2017.

56. Asadi, P.; Davison, A.C.; Engelke, S. Extremes on river networks. *Ann. Appl. Stat.* **2015**, *9*, 2023–2050. [CrossRef]

57. Computational and Information Systems Laboratory Cheyenne: SGI ICE XA System (Climate Simulation Laboratory). *Boulder CO Natl. Cent. Atmospheric Res. NCAR* **2017**. [CrossRef]

58. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B* **1994**, *58*, 267–288.

59. Breiman, L. Better Subset Regression Using the Nonnegative Garrote. *Technometrics* **1995**, *37*, 373–384. [CrossRef]

60. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [CrossRef] [PubMed]

61. Simon, N.; Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J. Stat. Softw.* **2011**, *39*, 1–13. [CrossRef] [PubMed]

62. Tibshirani, R. The Lasso Method for Variable Selection in the Cox Model. *Stat. Med.* **1997**, *16*, 385–395. [CrossRef]

63. Cressie, N.; Frey, J.; Harch, B.; Smith, M. Spatial prediction on a river network. *J. Agric. Biol. Environ. Stat.* **2006**, *11*, 127. [CrossRef]

64. Ver Hoef, J.M.; Peterson, E.E. A Moving Average Approach for Spatial Statistical Models of Stream Networks. *J. Am. Stat. Assoc.* **2010**, *105*, 6–18. [CrossRef]

65. Ver Hoef, J.M.; Peterson, E.; Theobald, D. Spatial statistical models that use flow and stream distance. *Environ. Ecol. Stat.* **2006**, *13*, 449–464. [CrossRef]

66. Golub, G.H.; Heath, M.; Wahba, G. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics* **1979**, *21*, 215–223. [CrossRef]

67. Rao, R.; Fung, G.; Rosales, R. On the Dangers of Cross-Validation. An Experimental Evaluation. In Proceedings of the 2008 SIAM International Conference on Data Mining, Atlanta, Georgia, 24–26 April 2008; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2008; pp. 588–596.

68. Smirnova, A.; Martcheva, M.; Liu, H. On generalized cross validation for stable parameter selection in disease models. *J. Inverse Ill-Posed Probl.* **2015**, *23*, 451–464. [CrossRef]

69. Craven, P.; Wahba, G. Smoothing noisy data with spline functions. *Numer. Math.* **1978**, *31*, 377–403. [CrossRef]

70. Thurman, E.M. Amount of Organic Carbon in Natural Waters. In *Organic Geochemistry of Natural Waters*; Developments in Biogeochemistry; Springer: Dordrecht, The Netherlands, 1985; pp. 7–65, ISBN 978-94-010-8752-0.

71. Cressie, N.A.C. Geostatistics. In *Statistics for Spatial Data*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1993; pp. 27–104, ISBN 978-1-119-11515-1.

72. Peterson, E.E.; Theobald, D.M.; Ver Hoef, J.M. Geostatistical modelling on stream networks: Developing valid covariance matrices based on hydrologic distance and stream flow. *Freshw. Biol.* **2007**, *52*, 267–279. [CrossRef]

73. Parr, T.B.; Cronan, C.S.; Ohno, T.; Findlay, S.E.G.; Smith, S.M.C.; Simon, K.S. Urbanization changes the composition and bioavailability of dissolved organic matter in headwater streams. *Limnol. Oceanogr.* **2015**, *60*, 885–900. [CrossRef]

74. Natural Resources Conservation Service Part 618 (Subpart A). Soil Properties and Qualities. In *National Soil Survey Handbook, Title 430-VI*; U.S. Department of Agriculture: Washington, DC, USA, 2017.

75. Soil Science Division Staff; Ditzler, C.; Scheffe, K.; Monger, H.C. Examination and Description of Soil Profiles. In *Soil Survey Manual*; USDA Handbook 18; Government Printing Office: Washington, DC, USA, 2017.

76. Conant, R.T.; Ryan, M.G.; Ågren, G.I.; Birge, H.E.; Davidson, E.A.; Eliasson, P.E.; Evans, S.E.; Frey, S.D.; Giardina, C.P.; Hopkins, F.M.; et al. Temperature and soil organic matter decomposition rates—Synthesis of current knowledge and a way forward. *Glob. Chang. Biol.* **2011**, *17*, 3392–3404. [CrossRef]

77. Simard, M.; Powell, E.N.; Raffa, K.F.; Turner, M.G. What explains landscape patterns of tree mortality caused by bark beetle outbreaks in Greater Yellowstone? *Glob. Ecol. Biogeogr.* **2012**, *21*, 556–567. [CrossRef]

78. Osburn, C.L.; Morris, D.P.; Thorn, K.A.; Moeller, R.E. Chemical and optical changes in freshwater dissolved organic matter exposed to solar radiation. *Biogeochemistry* **2001**, *54*, 251–278. [CrossRef]

79. Bladon, K.D.; Silins, U.; Wagner, M.J.; Stone, M.; Emelko, M.B.; Mendoza, C.A.; Devito, K.J.; Boon, S. Wildfire impacts on nitrogen concentration and production from headwater streams in southern Alberta's Rocky Mountains. *Can. J. For. Res.* **2008**, *38*, 2359–2371. [CrossRef]

80. Noske, P.J.; Lane, P.N.J.; Sheridan, G.J. Stream exports of coarse matter and phosphorus following wildfire in NE Victoria, Australia. *Hydrol. Process.* **2010**, *24*, 1514–1529. [CrossRef]

81. Burke, J.M.; Prepas, E.E.; Pinder, S. Runoff and phosphorus export patterns in large forested watersheds on the western Canadian Boreal Plain before and for 4 years after wildfire. *J. Environ. Eng. Sci.* **2005**, *4*, 319–325. [CrossRef]

82. Earl, S.R.; Blinn, D.W. Effects of wildfire ash on water chemistry and biota in South-Western U.S.A. streams. *Freshw. Biol.* **2003**, *48*, 1015–1030. [CrossRef]

83. Smith, H.G.; Sheridan, G.J.; Lane, P.N.J.; Nyman, P.; Haydon, S. Wildfire effects on water quality in forest catchments: A review with implications for water supply. *J. Hydrol.* **2011**, *396*, 170–192. [CrossRef]

84. Ryan, S.E.; Dwire, K.A.; Dixon, M.K. Impacts of wildfire on runoff and sediment loads at Little Granite Creek, western Wyoming. *Geomorphology* **2011**, *129*, 113–130. [CrossRef]

85. Rhoades, C.C.; Entwistle, D.; Butler, D. The influence of wildfire extent and severity on streamwater chemistry, sediment and temperature following the Hayman Fire, Colorado. *Int. J. Wildfire Sci.* **2011**, *20*, 430–442. [CrossRef]

86. Robichaud, P.R. Measurement of post-fire hillslope erosion to evaluate and model rehabilitation treatment effectiveness and recovery. *Int. J. Wildland Fire* **2005**, *14*, 475–485. [CrossRef]

87. Center for Chemical Process Safety Management Overview. In *Guidelines for Fire Protection in Chemical, Petrochemical, and Hydrocarbon Processing Facilities*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2003; pp. 7–9, ISBN 978-0-470-92504-1.

88. Anderson, H.E. Forest fuel ignitibility. *Fire Technol.* **1970**, *6*, 312–319. [CrossRef]

89. Kinoshita, A.M.; Hogue, T.S. Spatial and temporal controls on post-fire hydrologic recovery in Southern California watersheds. *Catena* **2011**, *87*, 240–252. [CrossRef]

90. Reed, D.E.; Ewers, B.E.; Pendall, E. Impact of mountain pine beetle induced mortality on forest carbon and water fluxes. *Environ. Res. Lett.* **2014**, *9*, 105004. [CrossRef]

91. Berg, B.; Davey, M.P.; Marco, A.D.; Emmett, B.; Faituri, M.; Hobbie, S.E.; Johansson, M.-B.; Liu, C.; McClaugherty, C.; Norell, L.; et al. Factors influencing limit values for pine needle litter decomposition: A synthesis for boreal and temperate pine forest systems. *Biogeochemistry* **2010**, *100*, 57–73. [CrossRef]

92. Gholz, H.L.; Perry, C.S.; Cropper, W.P.; Hendry, L.C. Litterfall, Decomposition, and Nitrogen and Phosphorus Dynamics in a Chronosequence of Slash Pine (*Pinus elliottii*) Plantations. *For. Sci.* **1985**, *31*, 463–478.

93. Delong, M.D.; Brusven, M.A. Storage and decomposition of particulate organic matter along the longitudinal gradient of an agriculturally-impacted stream. *Hydrobiologia* **1993**, *262*, 77–88. [CrossRef]

94. Heinz, M.; Graeber, D.; Zak, D.; Zwirnmann, E.; Gelbrecht, J.; Pusch, M.T. Comparison of organic matter composition in agricultural versus forest affected headwaters with special emphasis on organic nitrogen. *Environ. Sci. Technol.* **2015**, *49*, 2081–2090. [CrossRef] [PubMed]

95. Graeber, D.; Boëchat, I.G.; Encina-Montoya, F.; Esse, C.; Gelbrecht, J.; Goyenola, G.; Gücker, B.; Heinz, M.; Kronvang, B.; Meerhoff, M.; et al. Global effects of agriculture on fluvial dissolved organic matter. *Sci. Rep.* **2015**, *5*, srep16328. [CrossRef] [PubMed]

96. Shang, P.; Lu, Y.; Du, Y.; Jaffé, R.; Findlay, R.H.; Wynn, A. Climatic and watershed controls of dissolved organic matter variation in streams across a gradient of agricultural land use. *Sci. Total Environ.* **2018**, *612*, 1442–1453. [CrossRef] [PubMed]

97. Alvarez, R.; Santanatoglia, O.J.; Garcîa, R. Effect of temperature on soil microbial biomass and its metabolic quotient in situ under different tillage systems. *Biol. Fertil. Soils* **1995**, *19*, 227–230. [CrossRef]

98. Vinolas, L.C.; Vallejo, V.R.; Jones, D.L. Control of amino acid mineralization and microbial metabolism by temperature. *Soil Biol. Biochem.* **2001**, *33*, 1137–1140. [CrossRef]

99. Rumsey, C.A.; Miller, M.P.; Susong, D.D.; Tillman, F.D.; Anning, D.W. Regional scale estimates of baseflow and factors influencing baseflow in the Upper Colorado River Basin. *J. Hydrol. Reg. Stud.* **2015**, *4*, 91–107. [CrossRef]

100. Kaiser, K.; Guggenberger, G. Storm flow flushing in a structured soil changes the composition of dissolved organic matter leached into the subsoil. *Geoderma* **2005**, *127*, 177–187. [CrossRef]

101. Arnold, J.G.; Muttiah, R.S.; Srinivasan, R.; Allen, P.M. Regional estimation of base flow and groundwater recharge in the Upper Mississippi river basin. *J. Hydrol.* **2000**, *227*, 21–40. [CrossRef]

102. Hall, F.R. Base-Flow Recessions—A Review. *Water Resour. Res.* **1968**, *4*, 973–983. [CrossRef]

103. Štursová, M.; Šnajdr, J.; Cajthaml, T.; Bárta, J.; Šantrůčková, H.; Baldrian, P. When the forest dies: The response of forest soil fungi to a bark beetle-induced tree dieback. *ISME J.* **2014**, *8*, 1920–1931. [CrossRef] [PubMed]

104. Kennedy, N.M.; Robertson, S.J.; Green, D.S.; Scholefield, S.R.; Arocena, J.M.; Tackaberry, L.E.; Massicotte, H.B.; Egger, K.N. Site properties have a stronger influence than fire severity on ectomycorrhizal fungi and associated N-cycling bacteria in regenerating post-beetle-killed lodgepole pine forests. *Folia Microbiol.* **2014**. [CrossRef] [PubMed]

105. Walker, J.J.; Pace, N.R. Phylogenetic Composition of Rocky Mountain Endolithic Microbial Ecosystems. *Appl. Environ. Microbiol.* **2007**, *73*, 3497–3504. [CrossRef] [PubMed]

106. Lipson, D.A.; Schmidt, S.K. Seasonal changes in an alpine soil bacterial community in the colorado rocky mountains. *Appl. Environ. Microbiol.* **2004**, *70*, 2867–2879. [CrossRef] [PubMed]

107. Kaňa, J.; Tahovská, K.; Kopáček, J.; Šantrůčková, H. Excess of Organic Carbon in Mountain Spruce Forest Soils after Bark Beetle Outbreak Altered Microbial N Transformations and Mitigated N-Saturation. *PLoS ONE* **2015**, *10*, e0134165. [CrossRef] [PubMed]

108. Condon, L.E.; Hering, A.S.; Maxwell, R.M. Quantitative assessment of groundwater controls across major US river basins using a multi-model regression algorithm. *Adv. Water Resour.* **2015**, *82*, 106–123. [CrossRef]

109. Maguire, D.Y.; James, P.M.A.; Buddle, C.M.; Bennett, E.M. Landscape connectivity and insect herbivory: A framework for understanding tradeoffs among ecosystem services. *Glob. Ecol. Conserv.* **2015**, *4*, 73–84. [CrossRef]

110. Morris, J.L.; Cottrell, S.; Fettig, C.J.; DeRose, J.; Mattor, K.M.; Carter, V.A.; Clear, J.; Clement, J.; Hansen, W.D.; Hicke, J.A.; et al. Bark beetles as agents of change in social–ecological systems. *Front. Ecol. Environ.* **2018**, *16*, S34–S43. [CrossRef]

111. Morris, J.L.; Cottrell, S.; Fettig, C.J.; Hansen, W.D.; Sherriff, R.L.; Carter, V.A.; Clear, J.L.; Clement, J.; DeRose, R.J.; Hicke, J.A.; et al. Marini Lorenzo Managing bark beetle impacts on ecosystems and society: Priority questions to motivate future research. *J. Appl. Ecol.* **2016**, *54*, 750–760. [CrossRef]

112. McGrady, P.; Cottrell, S.; Clement, J.; Cottrell, J.R.; Czaja, M. Local Perceptions of MPB Infestation, Forest Management, and Connection to National Forests in Colorado and Wyoming. *Hum. Ecol.* **2016**, *44*, 185–196. [CrossRef]

113. Arnberger, A.; Ebenberger, M.; Schneider, I.E.; Cottrell, S.; Schlueter, A.C.; von Ruschkowski, E.; Venette, R.C.; Snyder, S.A.; Gobster, P.H. Visitor Preferences for Visual Changes in Bark Beetle-Impacted Forest Recreation Settings in the United States and Germany. *Environ. Manag.* **2018**, *61*, 209–223. [CrossRef] [PubMed]

114. Norden, N.; Angarita, H.A.; Bongers, F.; Martínez-Ramos, M.; la Cerda, I.G.; van Breugel, M.; Lebrija-Trejos, E.; Meave, J.A.; Vandermeer, J.; Williamson, G.B.; et al. Successional dynamics in Neotropical forests are as uncertain as they are predictable. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 8013–8018. [CrossRef] [PubMed]

115. Fites, J.A.; Reiner, A.; Campbell, M.; Taylor, Z. *Fire Behavior and Effects, Suppression, and Fuel Treatments on the Ham Lake and Cavity Lake Fires*; Superior National Forest, Eastern Region, USDA Forest Service: Washington, DC, USA, 2007.

116. Liebsch, D.; Marques, M.C.M.; Goldenberg, R. How long does the Atlantic Rain Forest take to recover after a disturbance? Changes in species composition and ecological features during secondary succession. *Biol. Conserv.* **2008**, *141*, 1717–1725. [CrossRef]

117. Wear, D.N.; Greis, J.G. *Southern Forest Resource Assessment—Technical Report*; US Department of Agriculture, Forest Service, Southern Research Station: Asheville, NC, USA, 2002.

118. Romme, W.H.; Boyce, M.S.; Gresswell, R.; Merrill, E.H.; Minshall, G.W.; Whitlock, C.; Turner, M.G. Twenty Years After the 1988 Yellowstone Fires: Lessons About Disturbance and Ecosystems. *Ecosystems* **2011**, *14*, 20. [CrossRef]

119. Turner, M.G.; Romme, W.H.; Tinker, D.B. Surprises and lessons from the 1988 Yellowstone fires. *Front. Ecol. Environ.* **2003**, *1*, 351–358. [CrossRef]

120. Savage, M.; Mast, J.N. How resilient are southwestern ponderosa pine forests after crown fires? *Can. J. For. Res.* **2005**, *35*, 967–977. [CrossRef]

121. Baker, W.L.; Veblen, T.T.; Sherriff, R.L. Fire, fuels and restoration of ponderosa pine–Douglas fir forests in the Rocky Mountains, USA. *J. Biogeogr.* **2007**, *34*, 251–269. [CrossRef]

122. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**. [CrossRef]

123. NASA Landsat Program. *USGS Landsat 5 TM Raw Scenes (Orthorectified)*; USGS Earth Resources Observation and Science (EROS) Center: Sioux Falls, SD, USA, 2012.

124. Rouse, J.W.; Haas, H.R.; Schell, J.A.; Deering, D.W. *Monitoring Vegetation Systems in the Great Plains with ERTS*; NASA-Scientific and Technical Reports Information Program: Washington, DC, USA, 1974; pp. 301–317.

125. Raffa, K.F.; Aukema, B.H.; Bentz, B.J.; Carroll, A.L.; Hicke, J.A.; Turner, M.G.; Romme, W.H. Cross-scale Drivers of Natural Disturbances Prone to Anthropogenic Amplification: The Dynamics of Bark Beetle Eruptions. *BioScience* **2008**, *58*, 501–517. [CrossRef]

126. Park Williams, A.; Allen, C.D.; Macalady, A.K.; Griffin, D.; Woodhouse, C.A.; Meko, D.M.; Swetnam, T.W.; Rauscher, S.A.; Seager, R.; Grissino-Mayer, H.D.; et al. Temperature as a potent driver of regional forest drought stress and tree mortality. *Nat. Clim. Chang.* **2013**, *3*, 292–297. [CrossRef]

127. Brouillard, B.; Sharp, J.O.; Dickenson, E.R.V.; Hogue, T.S.; Spear, J.R.; Rodriguez, D. *Biogeochemical and Ecological Impacts Resulting from Beetle-Induced Forest Mortality*; Colorado School of Mines; Arthur Lakes Library: Golden, CO, USA, 2017.

128. Johnson, E.W.; Ross, J. *USDA Forest Service Rocky Mountain Region Forest Health Aerial—Survey Accuracy Assessment 2005 Report*; Technical Report R2-06-08; USDA Forest Service: Washington, DC, USA, 2006.

129. Meddens, A.J.H.; Hicke, J.A.; Ferguson, C.A. Spatiotemporal patterns of observed bark beetle-caused tree mortality in British Columbia and the western United States. *Ecol. Appl.* **2012**, *22*, 1876–1891. [CrossRef] [PubMed]

130. Harris, J.L. *2013 Forest Insect and Disease Conditions, Rocky Mountain Region (R2)*; USDA Forest Service: Washington, DC, USA, 2014; p. 77.

131. Penn, C.A.; Bearup, L.A.; Maxwell, R.M.; Clow, D.W. Numerical experiments to explain multiscale hydrological responses to mountain pine beetle tree mortality in a headwater watershed. *Water Resour. Res.* **2016**, *52*, 31433161. [CrossRef]

132. Bearup, L.A.; Maxwell, R.M.; Clow, D.W.; McCray, J.E. Hydrological effects of forest transpiration loss in bark beetle-impacted watersheds. *Nat. Clim. Chang.* **2014**, *4*, 481–486. [CrossRef]

133. Herzog, S.P.; Higgins, C.P.; McCray, J.E. Engineered Streambeds for Induced Hyporheic Flow: Enhanced Removal of Nutrients, Pathogens, and Metals from Urban Streams. *J. Environ. Eng.* **2016**, *142*, 04015053. [CrossRef]

134. Johnston, C.A. Sediment and nutrient retention by freshwater wetlands: Effects on surface water quality. *Crit. Rev. Environ. Control* **1991**, *21*, 491–565. [CrossRef]

135. Karlen, D.L.; Mausbach, M.J.; Doran, J.W.; Cline, R.G.; Harris, R.F.; Schuman, G.E. Soil Quality: A Concept, Definition, and Framework for Evaluation (A Guest Editorial). *Soil Sci. Soc. Am. J.* **1997**, *61*, 4–10. [CrossRef]

136. Quanrud, D.M.; Arnold, R.G.; Gray Wilson, L.; Conklin, M.H. Effect of soil type on water quality improvement during soil aquifer treatment. *Water Sci. Technol.* **1996**, *33*, 419–431. [CrossRef]

137. Tong, S.T.Y.; Chen, W. Modeling the relationship between land use and surface water quality. *J. Environ. Manag.* **2002**, *66*, 377–393. [CrossRef]

138. Soil Survey Staff, Natural Resources Conservation Service. *U.S. General Soil Map (STATSGO2)*; United States Department of Agriculture: Washington, DC, USA, 2017.

139. United States Department of Agriculture. *Natural Resources Conservation Service Soil Data Viewer 6.2—User Guide Version 3*; United States Department of Agriculture: Washington, DC, USA, 2015.

140. O'Connor, D.J. The temporal and spatial distribution of dissolved oxygen in streams. *Water Resour. Res.* **1967**, *3*, 65–79. [CrossRef]

141. Jacobsen, D.; Marín, R. Bolivian Altiplano streams with low richness of macroinvertebrates and large diel fluctuations in temperature and dissolved oxygen. *Aquat. Ecol.* **2008**, *42*, 643–656. [CrossRef]

142. Matthews, K.R.; Berg, N.H. Rainbow trout responses to water temperature and dissolved oxygen stress in two southern California stream pools. *J. Fish Biol.* **1997**, *50*, 50–67. [CrossRef]

143. Beven, K.J. Down to the Basics: Runoff Processes and the Modelling Process—Runoff Generation and Runoff Routing. In *Rainfall-Runoff Modelling: The Primer*; John Wiley & Sons: Hoboken, NJ, USA, 2011; ISBN 978-1-119-95101-8.

144. Eva, H.; Lambin, E.F. Fires and land-cover change in the tropics: A remote sensing analysis at the landscape scale. *J. Biogeogr.* **2000**, *27*, 765–776. [CrossRef]

145. Friedl, M.A.; Sulla-Menashe, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **2010**, *114*, 168–182. [CrossRef]

146. Fry, J.; Xian, G.; Jin, S.; Dewitz, J.; Homer, C.; Yang, L.; Barnes, C.; Herold, N.; Wickham, J. Completion of the 2006 National Land Cover Database for the conterminous United States. *Photogramm. Eng. Remote Sens.* **2011**, *77*, 858–864.

147. Jin, S.; Yang, L.; Danielson, P.; Homer, C.; Fry, J.; Xian, G. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. *Remote Sens. Environ.* **2013**, *132*, 159–175. [CrossRef]

148. Liang, L.; Chen, Y.; Hawbaker, T.J.; Zhu, Z.; Gong, P. Mapping mountain pine beetle mortality through growth trend analysis of time-series Landsat data. *Remote Sens.* **2014**, *6*, 5696–5716. [CrossRef]

149. Clow, D.W.; Nanus, L.; Verdin, K.L.; Schmidt, J. Evaluation of SNODAS snow depth and snow water equivalent estimates for the Colorado Rocky Mountains, USA. *Hydrol. Process.* **2012**, *26*, 2583–2591. [CrossRef]

150. Maurer, E.P.; Wood, A.W.; Adam, J.C.; Lettenmaier, D.P.; Nijssen, B. A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States. *J. Clim.* **2002**, *15*, 3237–3251. [CrossRef]

151. Tang, Q.; Vivoni, E.R.; Muñoz-Arriola, F.; Lettenmaier, D.P. Predictability of Evapotranspiration Patterns Using Remotely Sensed Vegetation Dynamics during the North American Monsoon. *J. Hydrometeorol.* **2011**, *13*, 103–121. [CrossRef]

152. Cayan, D.R.; Dettinger, M.D.; Diaz, H.F.; Graham, N.E. Decadal Variability of Precipitation over Western North America. *J. Clim.* **1998**, *11*, 3148–3166. [CrossRef]

153. Hauer, F.R.; Baron, J.S.; Campbell, D.H.; Fausch, K.D.; Hostetler, S.W.; Leavesley, G.H.; Leavitt, P.R.; Mcknight, D.M.; Stanford, J.A. Assessment of climate change and freshwater ecosystems of the Rocky Mountains, USA and Canada. *Hydrol. Process.* **1997**, *11*, 903–924. [CrossRef]

154. Ross, T.; Lott, N.; McCown, S.; Quinn, D. *The El Niño Winter of '97-'98*; National Climatic Data Center; National Oceanic and Atmospheric Administration: Asheville, NC, USA, 1998.

155. Schoennagel, T.; Veblen, T.T.; Romme, W.H. The Interaction of Fire, Fuels, and Climate across Rocky Mountain Forests. *BioScience* **2004**, *54*, 661–676. [CrossRef]

156. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201. [CrossRef]

157. Gao, Z.; Liu, J.; Cao, M.; Li, K.; Tao, B. Impacts of land-use and climate changes on ecosystem productivity and carbon cycle in the cropping-grazing transitional zone in China. *Sci. China Ser. Earth Sci.* **2005**, *48*, 1479–1491. [CrossRef]

158. Hosen, J.D.; McDonough, O.T.; Febria, C.M.; Palmer, M.A. Dissolved Organic Matter Quality and Bioavailability Changes across an Urbanization Gradient in Headwater Streams. *Environ. Sci. Technol.* **2014**, *48*, 7817–7824. [CrossRef] [PubMed]

159. Daniel, M.H.B.; Montebelo, A.A.; Bernardes, M.C.; Ometto, J.P.H.B.; de Camargo, P.B.; Krusche, A.V.; Ballester, M.V.; Victoria, R.L.; Martinelli, L.A. Effects of Urban Sewage on Dissolved Oxygen, Dissolved Inorganic and Organic Carbon, and Electrical Conductivity of Small Streams along a Gradient of Urbanization in the Piracicaba River Basin. *Water. Air. Soil Pollut.* **2002**, *136*, 189–206. [CrossRef]

160. Westerhoff, P.; Anning, D. Concentrations and characteristics of organic carbon in surface water in Arizona: Influence of urbanization. *J. Hydrol.* **2000**, *236*, 202–222. [CrossRef]