communications earth & environment | Article

A Nature Portfolio journal

https://doi.org/10.1038/s43247-025-02573-x

Factors influencing surface water accumulation in beaver pond complexes across the Western United States

Check for updates

Luwen Wan[1,2] ✉, Emily Fairfax[3] & Kate Maher[1]

North American beavers (*Castor canadensis*) build dams and ponds that alter streamflow, enhance floodplain water storage, and provide refugia during droughts and wildfires. However, drivers of pond area variability remain poorly understood. Here, we quantified the influencing factors that drive pond area and dam length variations using an explanatory modeling approach, after mapping surface water area of beaver ponds and creating beaver pond complexes. Mapped area correlated well with manual delineations ($r^2 = 0.89$), and additive pond area and dam length across 87 complexes followed a significant log-log scaling relationship. Dam length was the strongest covariate of pond area, while woody vegetation height and stream power index were also influential; together, these covariates explained 74% of the variation. Our results provide an empirical foundation to inform site selection and prioritization for beaver restoration, supporting watershed management, climate resilience and ecological conservation strategies in regions with comparable data availability and landscape characteristics.

The North American beaver (*Castor canadensis*) is a keystone species renowned for its remarkable ecosystem engineering capabilities. By building dams and digging canals, beavers create complex wetland ecosystems that enhance biodiversity and improve ecosystem functioning[1]. These modifications increase water storage[2–4], create wildfire refugia[5–7], and improve the resilience of river corridors to hydrological extremes and habitat degradation[8,9]. Beaver activities also influence sediment regimes, biogeochemistry processes, and nutrient cycling, further promoting biodiversity[2,10,11]. Consequently, beaver-related restoration has gained attention as a nature-based solution for restoring ecosystems, and adapting to climate change, particularly in North America[12–15]. However, beaver activities can also present challenges. Increased water storage from beaver damming can inundate private land, leading to financial costs and conflicts with landowners[16]. Additionally, beavers may consume arable crops and fell trees of economic importance, impacting agricultural productivity and forestry[17]. These trade-offs highlight the importance of understanding the environmental conditions that shape beaver pond development and expansion.

Despite both the expansion in beaver populations and growing interest in watershed restoration, the drivers of pond area variation across diverse regional landscape remain poorly understood. Quantifying the environmental and geomorphic factors that influence variation in beaver pond area is essential for linking beaver engineering to hydrological and ecological outcomes, informing targeted restoration efforts, and improving predictions of beaver impacts. While GIS-based models such as the Beaver Restoration Assessment Tool (BRAT) provide broad-scale assessment of dam-building capacity using inputs such as vegetation and stream power, the outputs are typically classification-based and only focus on dam-density[18]. Although recent efforts have improved mapping of beaver pond distributions, relatively few have examined environmental covariates that influence pond surface area directly. For example, a review of 12 studies identified stream gradient as a key determinant of beaver habitat suitability[19]. At the watershed scale, Fitch et al. (2022) delineated beaver-impounded surface water in southeastern Wyoming using the National Agriculture Imagery Program (NAIP) imagery, finding that topography primarily controls pond area in small watersheds, while upstream inputs and riparian space are more influential in larger watersheds[20]. Accurately mapping the spatial footprint of beaver ponds and dams is thus critical for advancing our understanding of pond area variability and for improving the prediction and management of beaver-related hydrological and ecological impacts.

Beaver activities significantly influence surface water dynamics by creating ponds and altering hydrological pathways[2,21,22]. However, existing surface water area datasets inadequately capture these beaver-modified landscapes due to limitations in spatial resolution and classification methods. For example, upstream of the North Fork Platte River in Colorado, 348 beaver dams were identified using high-resolution aerial imagery[23], whereas

[1]Department of Earth System Science, Stanford University, Stanford, CA, USA. [2]Stanford Human-Centered Artificial Intelligence, Stanford University, Stanford, CA, USA. [3]Department of Geography, Environment, and Society, University of Minnesota, Minneapolis, MN, USA. ✉e-mail: luwenwan@stanford.edu

the NHDPlus High Resolution (NHDPlus-HR) dataset captured only 56 water bodies (Fig. 1). This discrepancy underscores the limitations of existing datasets, as even NHDPlus-HR underrepresents beaver-influenced water bodies. Among the 56, only 31 intersected with labeled dams and were visually confirmed as beaver ponds, distinguishing them from lake ponds and swamps (Fig. S1). As a result, beaver dams and ponds are often overlooked in hydrological models and water budgets, leading to gaps in our understanding of water resource distribution and ecosystem dynamics. Previous studies have mapped and digitized beaver ponds at small scales such as headwater streams and reaches, using traditional topographic surveys[3,4,24]. However, these surveys are labor-intensive and time-consuming due to the widespread presence of beaver ponds, their intricate size and shape, and their often remote locations.

Remote sensing imagery provides an efficient and scalable method for identifying beaver wetland complexes. Advanced machine learning models, such as the Earth Engine Automated Geospatial Element(s) Recognition (EEAGER) have been developed to detect beaver dams using high-resolution aerial and satellite imagery[23]. Zhang et al. (2023) mapped beaver ponds across Ontario using multi-temporal and multi-source remote sensing data resampled to 10 m resolution. Their model achieved 72.5% accuracy based on independent validation[25]. Similarly, Fraser et al. (2024) mapped beaver pond changes in the Hudson Bay Lowlands using 30 m Landsat surface water maps and compared a subset of these estimates with pond areas derived from 0.5 m WorldView imagery[26]. They found that WorldView measurements were, on average, 87% larger than the Landsat estimates, and therefore recommended using the Landsat-WorldView regression equation ($R^2 = 0.9$) to adjust Landsat-based pond sizes.

The lack of understanding of the factors influencing beaver pond variation, along with the prerequisite of fine-scale mapping of beaver ponds, particularly in the western United States where beaver restoration efforts are gaining momentum[27], precludes data-driven assessment of reintroduction and management strategies. The objectives of this study are threefold. First, we develop a semi-automated approach to map beaver ponds across diverse landscapes using high-resolution (0.6 m) NAIP imagery. Second, we identify beaver pond complexes using density-based algorithms. These two objectives serve as prerequisites for the third and primary aim: to quantify

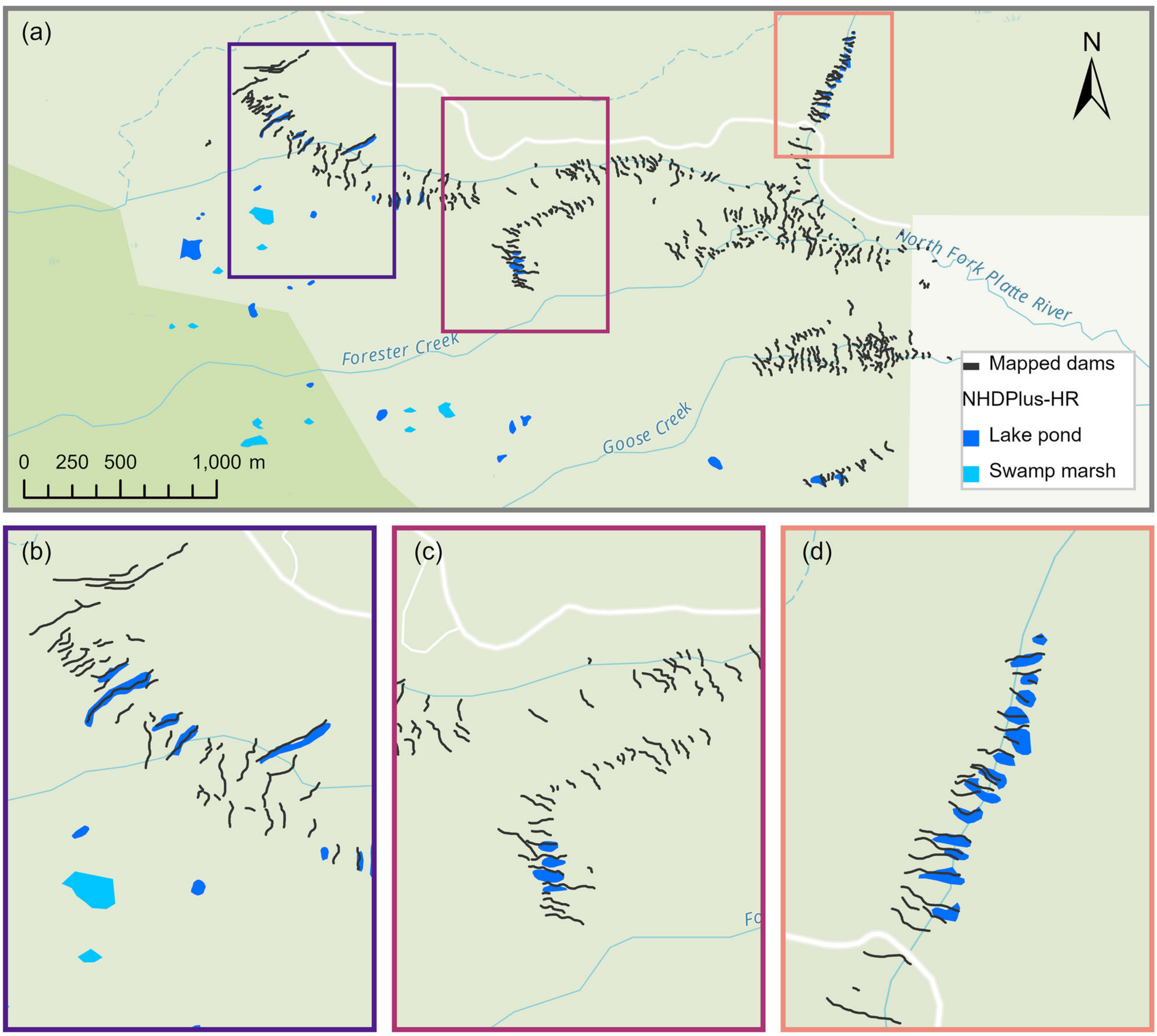


Fig. 1 | Mapped beaver dams ($n = 348$) identified using aerial imagery are shown in black (Source: Fairfax et al. 2023). Water bodies, including lake pond ($n = 45$) and swamp marsh ($n = 11$), are displayed in blue colors, sourced from the NHDPlus High-Resolution dataset. Panel (**a**) shows the upstream region of the North Fork Platte River in Colorado; panels (**b**), (**c**), and (**d**) are zoom-in views corresponding to the colored rectangles in (**a**). Background image source: World Topographic Map. See also Fig. S1 beaver pond visualization with false-color NAIP imagery.

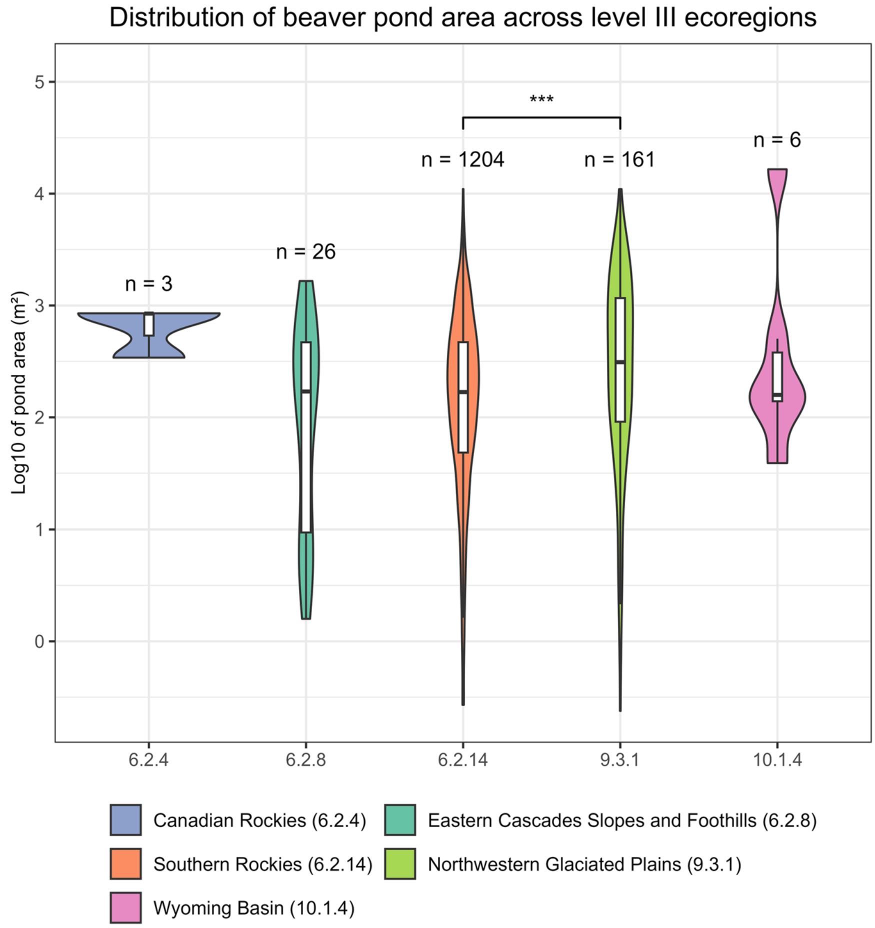


Fig. 2 | Distribution of beaver pond area across Level III ecoregions. The bottom and top of each box represent the first and third quartiles, respectively, with the median shown by the line inside each box. Outliers beyond the whiskers are not displayed. Significant differences (***) between groups were identified ($p \leq 0.005$, Kruskal-Wallis test).

key factors influencing variation in beaver pond area and dam length. The mapped beaver pond locations and sizes from this study can be incorporated into models that evaluate the hydrological and ecological impacts of beaver complexes. By identifying factors driving pond area and dam length variability, this work provides valuable insights with quantitative assessments for land managers and policymakers in prioritizing beaver-related restoration. Additionally, the mapping algorithm developed here can be applied to historical datasets—where NAIP imagery with near-infrared band and dam location data are available—to track changes in ponding areas over time in response to climate variability and anthropogenic activities.

Results

Varying beaver pond size and dam length across diverse landscapes

Of the mapped beaver ponds, 1204 (86%) were located in the Southern Rockies, where the median size was 168 m². The second-most abundant region, the Northwestern Glaciated Plains, had a significantly larger median pond size of 311 m² (Kruskal-Wallis test, $p \leq 0.005$; Fig. 2). Additionally, 26 ponds were mapped in the Eastern Cascades Slopes and Foothills, three in the Canadian Rockies and six in the Wyoming Basin, near the Southern Rockies in Wyoming, with a median size of 158 m².

The 87 beaver pond complexes were unevenly distributed across Level III ecoregions (Fig. 3), so we grouped them into Level I ecoregions for descriptive comparisons, thus no formal statistical comparisons were conducted (Text S1). Total pond area and dam length varied considerably across these broader ecoregions (Fig. S2). In the Northwestern Forested Mountains (Ecoregion 6, $n = 80$)—which includes three Level III ecoregions (6.2.4, 6.2.8 and 6.2.14) as well as one cluster spanning both Ecoregions 6 and 9 near Wyoming (10.1.4)—the median total pond area is 3009 m², and the median dam length is 360 m. The Great Plains (Ecoregion 9, $n = 7$) had a higher median pond area of 8000 m² and a median dam length of 455 m.

At the complex level, dam length and pond area showed a significant log-log scaling relationship ($R^2 = 0.63$, $p < 0.001$), diverging from simple geometric predictions (Fig. 3). A basic geometric model suggests that pond area (A) scales quadratically with dam length (L) (i.e. $A \propto L^2$ or $\pi L^2/8$). However, the observed scaling exponent (0.964) was about half of the geometric prediction, suggesting an elongated elliptical geometry rather than a simple quadratic relationship. On average, additive pond area increases by approximately 10.6 m² per meter of dam length, reflecting that pond area scales jointly with upstream pond width (b) and dam length (L), or $A = \pi b/2 * L$. This model estimates a typical upgradient pond width (b) of ~ 6.8 m, while pond length varies with dam length. Similarly, at the individual pond scale, dam length and pond area also showed a significant positive log-log scaling relationship (Fig. S3), though with lower explanatory power ($R^2 = 0.24$), likely due to greater variability in local factors influencing pond formation at smaller spatial scales.

Factors influencing pond area variation across complexes

Three Generalized Additive Models (GAMs) were developed to analyze covariates influencing pond area across beaver pond complexes. The full model provided a strong fit, explaining 87% of the deviance with an adjusted R^2 value of 0.81 (Model 1a, Table 1). This model included all 16 smooth covariates (Table S1), but only total dam length ($p < 0.001$), woody vegetation height ($p < 0.001$), and median stream power index ($p < 0.05$) were statistically significant. To improve interpretability, we refined the model by retaining only the three significant covariates, resulting in a simplified model that explained 78% of the deviance with an adjusted R^2 of 0.74 (Model 1b).

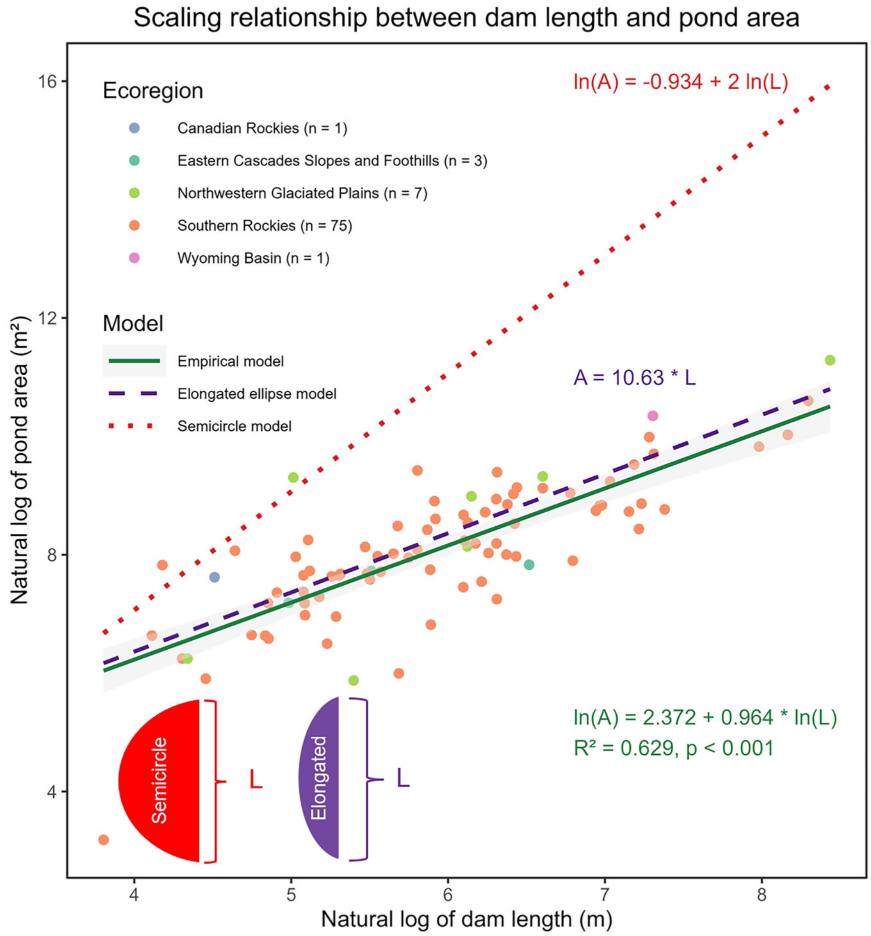


Fig. 3 | Scaling relationship between beaver dam length (L) and pond area (A) at the complex level. The green solid line, with its shaded 95% confidence interval, shows the empirical regression model fit to the log-transformed data. Each point represents an individual dam-pond pair and is color-coded by Level III ecoregion, with sample sizes shown in the legend. The red dotted line represents the predicted relationship under the assumption of semicircle dam geometry, while the purple dashed line represents the predicted relationship for an elongated ellipse dam geometry. The cartoons in the bottom left show idealized semicircle and elongated ellipse dam geometries.

Table 1 | Performance comparison of generalized additive models

Model	Generalized additive model	Log-transformed response variable	Number of covariates	Adjusted R^2	Deviance explained
1a	Full pond GAM	Pond area	16	0.807	87.0%
1b	Simplified pond GAM	Pond area	3	0.741	78.1%
1c	Reduced pond GAM (no dam)	Pond area	15	0.602	71.7%
2a	Full dam GAM	Dam length	15	0.811	87.2%
2b	Simplified dam GAM	Dam length	7	0.755	79.7%
2c	Interactive dam GAM	Dam length	9	0.805	85.6%

In the simplified model, dam length emerged as the strongest covariate ($F = 28.9$, $p < 0.001$; Fig. 4a). Woody height also contributed significantly ($F = 4.4$, $p < 0.001$), exhibiting a more flexible and non-linear relationship with pond area (EDF = 6.3). The median stream power index, calculated as the product of upstream catchment area and the tangent of the terrain slope angle, was also significant but had a weaker effect ($F = 4.4$, $p < 0.05$) compared to dam length and woody height. In addition, we found that these variables influence pond area independently as no significant interaction terms were identified.

Given the strong correlation and significant influence of dam length, we developed an alternative GAM that excluded dam length from the independent variables, referred to as the reduced model (Model 1c). In the absence of dam length, the remaining 15 covariates explained 60.2% of the variability and 72% of the deviance. Ultimately, the simplified model (Model 1b) was selected for analyzing partial effects, as it balances interpretability and predictive power compared to the full and reduced models.

Hydrological and geomorphological variables control beaver dam length

The full dam GAM model (Model 2a, Table 1) explained a substantial proportion of variance, with an adjusted R-squared of 0.81 and 87% deviance explained, indicating that the 15 selected variables (excluding dam length, Table S1) better predict dam length (Model 2a, adjusted $R^2 = 0.81$) than pond area (Model 1c, adjusted $R^2 = 0.6$). These largely static variables directly influence dam length, which represents the maximum ponding potential. In contrast, the mapped pond area reflects actual conditions and is subject to seasonal hydrological variability—covariates not explicitly incorporated here due to data constraints. However, flowlines are included to represent water inflow potential.

To improve interpretability, we simplified Model 2a by retaining only the significant variables, resulting in a refined model (Model 2b) with an adjusted R^2 of 0.755, explaining 80% of the deviance in dam length. Although Model 2b explains slightly less deviance, it retains most of the explanatory power while improving model parsimony and interpretability.

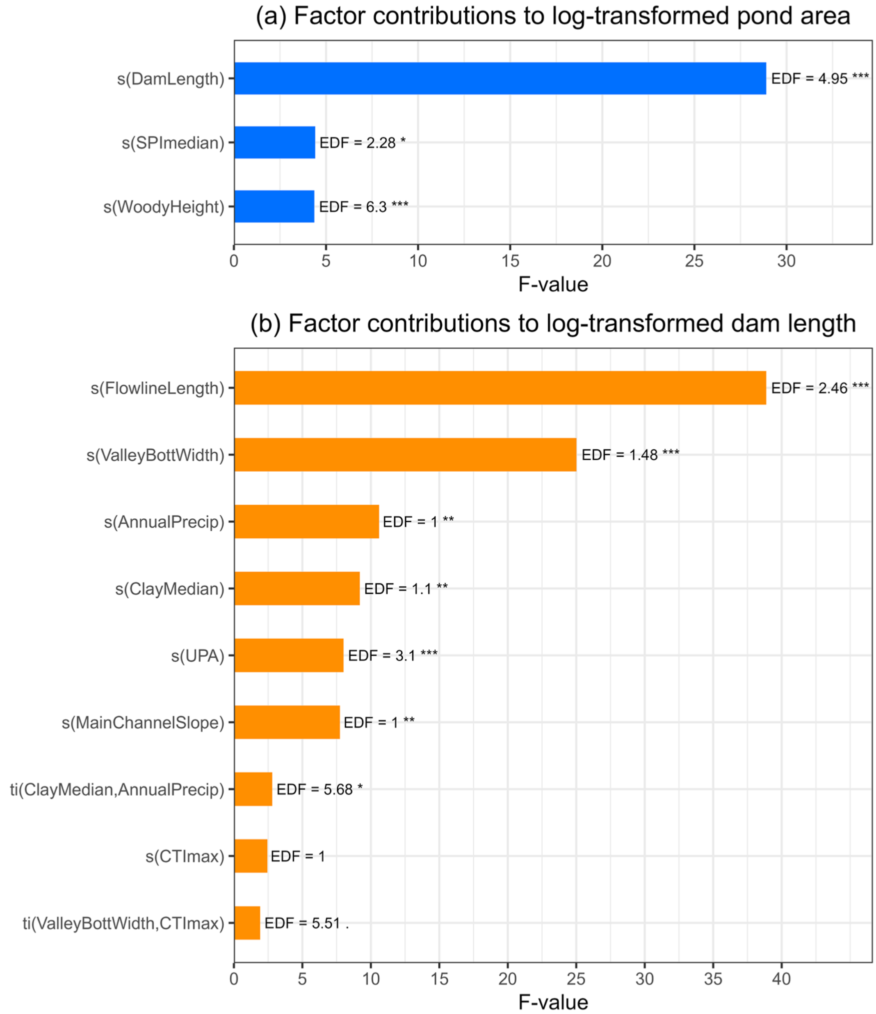


Fig. 4 | Factor contributions from the simplified pond area (model 1b) and interactive dam length (model 2c) models, showing F-values for each covariate, along with Estimated Degrees of Freedom (EDF) and significance levels. Panels (**a**) and (**b**) display the factors contributing to variability in log-transformed pond area and dam length, respectively. F-values represent the contribution of each covariate to explaining the variability. EDF indicates the complexity of the relationship: higher EDF values suggest a more flexible, non-linear relationship. Significance: ***$p < 0.001$, **$p < 0.01$; *$p < 0.05$, and. for $p < 0.1$. Non-significant values ($p \geq 0.1$) are left unmarked.

Unlike the pond area GAMs, the dam length GAM revealed significant interaction terms (Model 2c). To enhance model interpretability while improving model fit, we retained the two strongest interactions (*ti* terms in Fig. 4b), resulting in a higher adjusted R^2 (0.805) and deviance explained (85.6%) compared to the non-interactive model (model 2b). We refer to this final version as the interactive dam GAM, which accounts for joint environmental influences on dam length. This model was used to assess partial effects between log-transformed dam length and significant covariates (Text S2). Total flowline length, valley bottom width and upland contribution area were the strongest covariates ($p < 0.001$, Fig. 4b), while climatic and soil-related variables also played key roles in shaping dam-building potential.

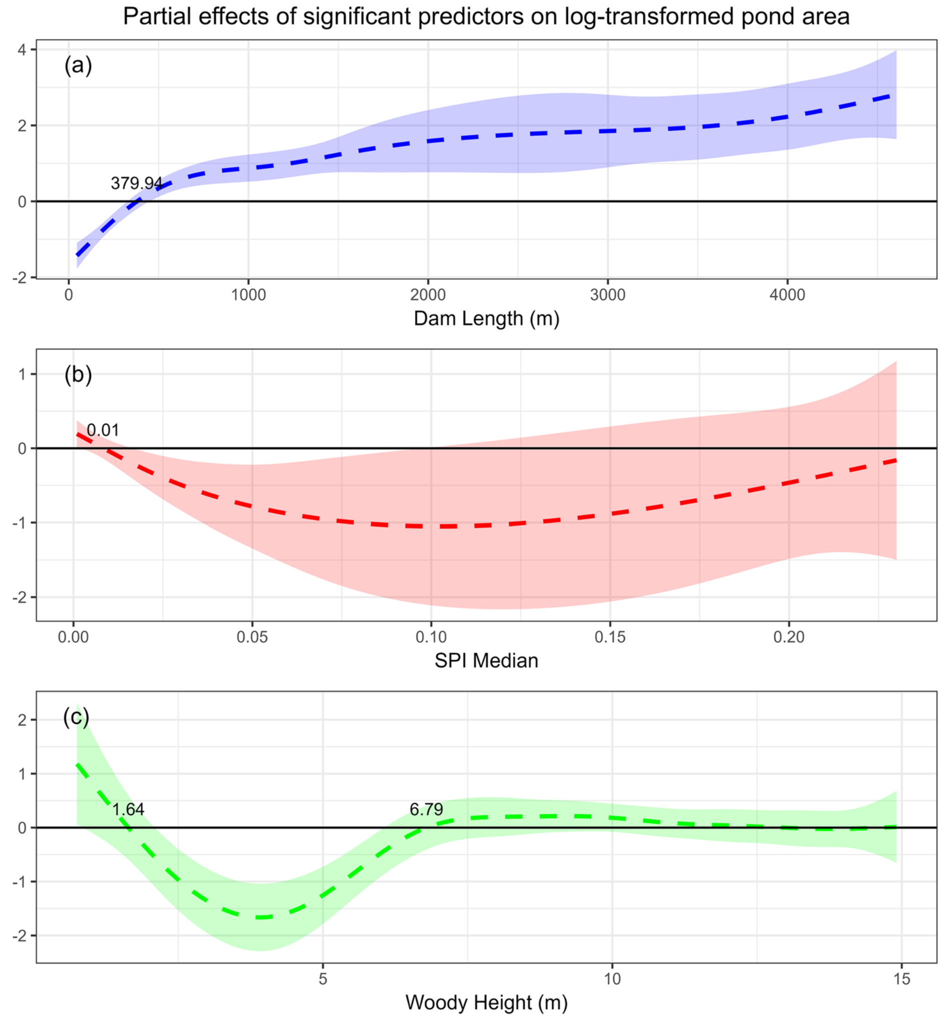
Discussion

Our analysis demonstrates that aerial imagery-based mapping using NAIP imagery is an effective approach for detecting beaver pond areas across diverse regions. At the regional level, smaller ponds in the Southern Rockies (median ~168 m^2) compared to the larger ponds in the NW Glaciated Plains highlight the influence of large-scale terrain differences on beaver complexes. The rugged terrain and the dynamic streamflow of the Intermountain West contrast with the more subdued topography of the glaciated plains, where the latter provides more favorable conditions for larger ponded areas. In a descriptive comparison, beaver pond clusters in Northwestern Forested Mountains (Ecoregion 6, $n = 80$) had a median channel slope of 1.79° and valley bottom width of 121 m, compared to 1.29° and 210 m, respectively in the Great Plains (Ecoregion 6, $n = 7$). Modeling the Northwestern Forested Mountains alone revealed some regional variation in significant covariates (see Text S2), but despite these differences, both models consistently identified dam length and woody vegetation as primary drivers of pond formation. This reinforces their fundamental role in shaping beaver complex dynamics across diverse landscapes.

Nevertheless, looking across geographic regions reveals several important trends. First, dam length is the strongest covariate of pond area, followed by stream power index and woody vegetation height. While these findings align with established paradigms identifying environmental factors that favor extensive beaver pond development, our regression-based approach provides insights into the relative importance of these drivers. Second, existing river restoration tools such as the Beaver Restoration Assessment Tool (BRAT), estimate dam-building capacity—expressed as dam density (dams per kilometer)—for individual stream reaches using inputs such as riparian vegetation and stream power[18]. Our study builds on this by quantifying how landscape-scale covariates—including vegetation height, stream power index, flowline length and valley bottom width—interact to shape dam length variability and surface water accumulation. Unlike the BRAT model, which infers dam-building capacity based on suitability metrics, our approach uses mapped pond area to establish direct, quantitative relationships between beaver engineering and environmental factors. Third, the strong scaling relationship between additive dam length and associated pond area provides a useful framework for estimating pond area when designing Beaver Dam Analogs (BDAs). Below, we discuss how this mechanistic understanding can help prioritize restoration areas that maximize the hydrological and ecological benefits of beaver complexes.

The observed correlation between dam length and pond area is intuitive as longer dams retain a greater volume of water and may increase

Fig. 5 | Partial effects of key covariates on ponding area in beaver pond complexes. Panels show the estimated effects of (**a**) dam length, (**b**) median stream power index, and (**c**) woody vegetation height on log-transformed ponding area, as modeled by Model 1b. These plots show the estimated effect of each covariate while holding the other two variables at the median values, thereby isolating the individual contribution of each factor to variations in the ponding area. In our GAMs, all covariates were normalized to a 0–100 scale to facilitate comparison of their relative effect sizes on the response variables. Here, we back-transformed the values to their original units to improve interpretability.



the hydraulic head leading to upgradient water accumulation. In addition to dam length, additive pond area is influenced by median stream power and woody height. A higher stream power index may limit pond expansion, particularly for dams constructed along the main channel, while higher woody vegetation is likely to enhance the availability of building materials and habitat conditions favorable for beaver dam construction and subsequent pond formation. In some cases, off-channel or side-channel ponds may persist despite high stream power in the mainstem due to reduced flow exposure. While these geomorphic-hydrologic interactions are ecologically important, a detailed mechanistic analysis of flow exposure or dam persistence was beyond the scope of this study and could be an important direction for future research.

The complexity and non-linearity of these variables are demonstrated by using partial effects plots where other covariates are set to median values (Fig. 5 & Figs. S4 & S5). Generally, pond size increases with dam length, but smaller dams yield below-average pond areas. The partial effect shifts from negative to positive at ~380 m, indicating that small dams have a weaker influence on pond area. Beyond this threshold, dam length becomes a stronger covariate of pond size (Fig. 5a). The flattening trend at longer dam lengths (i.e., >1500 m) may indicate diminishing importance, though limited data in this range restricts inference (Fig. S4).

Stream power, in contrast, shows a parabolic relationship, where the negative effect at lower values indicates that low stream power is associated with larger ponds (Fig. 5b). The flattening of the curve at SPI values > 0.05 suggests that stream power becomes a less important covariate of pond area at moderate stream power. The lack of data at high SPI results in large uncertainty, but the upward trend indicates that high SPI may not always limit pond area (Fig. S4). Similarly, woody height is a strong negative covariate at low values, consistent with the formation of smaller ponds in areas with short vegetation, presumably due to forage limitations that limit dam size and colony numbers (Fig. 5c). At moderate values of between 2 and 7 m, woody height has a stronger predictive effect, until leveling off. While these results align with the general expectations of provisioning required to support beaver complexes, they also point to inherent tradeoffs that may enable a wide variety of climates, topographies and ecosystem types to support beavers, pointing to their inherent adaptability as ecological engineers. Viewed collectively, larger pond development tends to occur in areas with longer dam lengths, lower stream power, and moderate woody vegetation height.

The weak dependence of pond area on climatic variables in our model aligns with findings from other studies. For instance, in western Canada, the number of beaver lodges explained over 80% of the variability in open water areas, while precipitation and temperature also contributed but were less important[24]. Similarly, there was no significant correlation between climate metrics and total pond area in southeastern Wyoming[20]. This weaker climate signal could stem from the coarse spatial resolution (1 km) of climate datasets, which may not capture fine-scale variations in water availability that affect beaver pond development. Although climate variables have limited influence in our data set, this does not indicate beaver activity is unaffected by climate. Rather, the integration of hydrologic and geomorphic conditions appears to be a stronger determinant of dam building than direct climatic factors. Environmental predictors of beaver dam length across complexes are provided in Text S5.

Beaver restoration is a promising nature-based solution for enhancing water storage, increasing drought and wildfire resilience, improving water quality, and supporting biodiversity[2,6,10]. Models such as BRAT[18] and the Beaver Dam Capacity (BDC) model[28] are widely used to predict where beavers could establish and how many dams a stream reach could support. While these GIS-based suitability models provide valuable broad-scale assessments, they primarily offer classification-based outputs and evaluate environmental covariates independently, overlooking how multiple covariates interact to shape beaver dam-building potential.

Here, we quantified the direct hydrological and geomorphological influences on dam-building success while also considering variable interactions. For instance, we found that longer dams are primarily located in areas where hydrological and geomorphological conditions are favorable, including well-developed stream networks (e.g., flowline length > 780 m), wide valleys (width > 160 m), and adequate precipitation (> 590 mm per yr) (Fig. S6). Additionally, precipitation interacts with soil permeability (e.g., clay content), demonstrating that both water availability and substrate suitability influence dam-building success (Fig. S7a). In low to moderate clay content areas, increased precipitation enhances dam-building by maintaining sufficient water availability. However, the relationship weakens in high-clay regions where excessive clay limits infiltration and increases surface runoff, reducing the benefits of additional precipitation. As an indicator of drainage class, substrate type was previously identified as a key factor in predicting beaver colony density in Massachusetts[29]. This interaction suggests that precipitation alone is insufficient to predict dam-building potential—soil permeability plays an important role in determining whether additional water can be retained and utilized for dam construction. Similarly, valley morphology strongly influences dam length through its interaction with wetness conditions (Fig. S7b). While wide valleys support long dams even in drier conditions, narrow valleys do not necessarily accommodate long dams even when wetness is high, as geomorphic constraints limit pond expansion.

Additionally, while GIS-based habitat suitability models enable regional assessments, their application at local scales can be misleading without ground-truthed data. For instance, a local-scale assessment in Park County, Colorado found beaver capacity to be 6–7 times lower than predicted by BRAT[30]. In contrast, in the Happy Jack Recreation Area, Wyoming, the BRAT model underestimated dam capacity because beavers used materials not classified as preferred vegetation[31]. Such discrepancies highlight the importance of integrating site-specific hydrologic and geomorphic conditions when prioritizing beaver restoration sites. Our findings suggest that landscape-scale assessments that incorporate environmental factor interactions provide a more reliable framework for determining where dam-building is most likely to succeed. This approach enables restoration projects to maximize hydrological and ecological benefits and improving restoration effectiveness, though it requires greater investment in data collection and analysis. We note, however, that biotic and anthropogenic stressors, including hunting, trapping, predation, disease and land-use conflicts, can also influence beaver occupancy and long-term restoration success. These social and ecological constraints will need to be integrated with physical site characteristics.

The broader ecological and hydrological benefits of beaver restoration extend beyond dam construction to include pond and wetland formation, which enhance water retention, habitat creation, groundwater recharge, and carbon sequestration[2,10,22,32]. While models like BRAT estimate dam densities, they do not explicitly predict pond area—a key determinant of water storage and ecosystem benefits. Our aerialimagery-based mapping and established scaling relationships between dam length and pond area provide a more direct method for assessing these hydrological benefits, improving the integration of beaver restoration into water management strategies. Additionally, predicting pond expansion is relevant for designing Beaver Dam Analogs (BDAs)—where pond expansion could be estimated using the scaling relationship between dam length and pond area. By incorporating additional covariates such as woody vegetation height and stream power, restoration projects can better identify locations where BDAs are most likely to replicate natural beaver ecosystem functions.

While beaver ponds and their associated water storage provide a range of benefits, they are not without tradeoffs. Pond expansion can lead to nuisance flooding and tree mortality, raising concerns from landowners, infrastructure managers and water users. Thus, predicting pond size and its implications for stakeholders is often a key consideration. Integrating our findings with existing climate resilience and watershed management frameworks could facilitate a delicate balance between ecological benefits and potential risks, ensuring that beaver restoration remains both effective and sustainable.

One limitation of our beaver pond area mapping is the reliance on NAIP imagery to detect surface water accumulation, which is typically acquired for the same location every 2 to 3 years during the growing season. This constraint may result in missing visible ponded areas, even when beaver dams are present. There is also a mismatch in imagery acquisition dates between the high-resolution imagery used to map the dams and the NAIP imagery used to delineate area, which may introduce temporal inconsistencies in the representation of beaver-modified hydrology. Additionally, our approach depends on accurately identifying beaver dams so that ponds that are not associated with labeled dams may be excluded, particularly if NAIP imagery is captured after dam labeling, as new ponds may have formed. Furthermore, the varying spatial resolutions of the covariates used in the GAMs (ranging from 10 to 1,000 meters) could introduce bias, although these datasets represent the best available for the study area.

Given the strong relationship between dam length and pond area, future work could address how changing hydrologic regimes, such as altered precipitation patterns or increased drought frequency, impact long-term pond persistence by integrating high-resolution, time-series imagery. Datasets with finer spatial and temporal resolutions (e.g., Planet Labs and Google's aerial imagery) would enable more precise tracking of pond area changes over time. In addition, advancing surface water detection methods beyond NDWI could help capture inundated areas obscured by flooded vegetation, which may be missed in our current estimates. Incorporating this temporal dimension would be especially valuable for understanding beaver impacts on riparian and watershed hydrology and biogeochemistry in the context of climate change.

Conclusions

This study identified dam length as the most significant covariate of pond area, explaining 74% of pond area variability, along with stream power and woody vegetation height, using generalized additive models. When considered together, conditions supporting more extensive pond development are associated with longer dam lengths, reduced stream power and moderate woody vegetation height. Additionally, seven variables and their interactions explained 81% of the variability in beaver dam length, with hydrological and geomorphological covariates and their interactions with soil conditions emerging as key determinants. Thus, valley geometry alone does not determine dam length, but rather a balance between valley shape, hydrological inputs, and soil conditions. We also developed an aerial imagery-based technique for mapping beaver pond areas across four western states, a semi-automated approach that can be adapted to other regions where high-resolution imagery and beaver dam data are available. Our mapped beaver pond areas are generally smaller in the Northwestern Forested Mountains compared to the Great Plains, likely due to differences in valley morphology and hydrologic conditions. These findings enhance our understanding of where and why beaver dams and ponds form, providing empirical constraints on the relationships between dam length, pond expansion and landscape suitability.

Our results emphasize the importance of site-specific hydrologic and geomorphic conditions for setting meaningful restoration targets. In addition, by mapping the pond area using high resolution imagery and quantifying the key factors influencing pond size and dam length, this work offers critical data for hydrological and ecological modeling. Prioritizing beaver

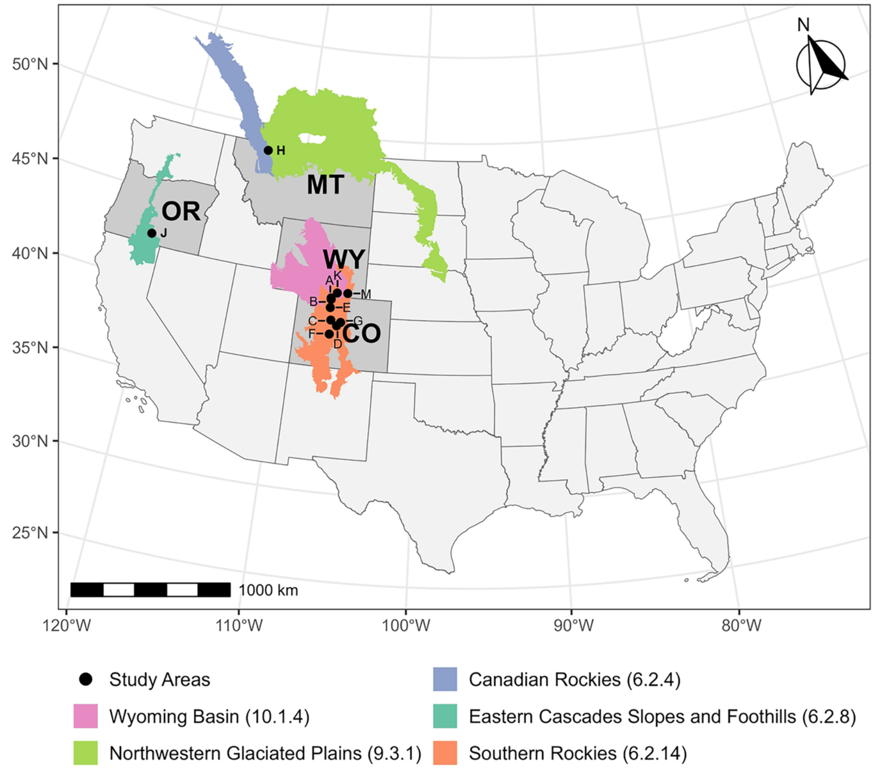


Fig. 6 | Study areas. Eleven study areas (black filled circles, enlarged for visibility and labeled A-H, J-K, M) across four western U.S. states (Colorado, Wyoming, Montana and Oregon) and are overlaid with five level III ecoregions. Note: A and B are located very close together and may appear as one circle at this scale.

activity in regions with suitable conditions maximizes their hydrological and ecological benefits—such as floodplain reconnection, groundwater recharge, and habitat enhancement—and provides a foundation for policymakers and land managers to design effective, targeted conservation strategies.

Methods

Study area

We mapped beaver ponds across Colorado, Wyoming, Montana and Oregon (Fig. 6), focusing on 11 study areas with known beaver dams[23]. These areas (black filled circles) encompass three level I ecoregions–broad ecological and geographical zones defined by similar climate, geology and soils—including the Northwestern Forested Mountains, Great Plains and North American Desserts[33]. A hierarchical framework further refines them into five Level III ecoregions, including Canadian Rockies, Eastern Cascades Slopes and Foothills, Southern Rockies, Northwestern Glaciated Plains and the Wyoming Basin[34].

Mapping beaver ponds using high-resolution aerial imagery

We mapped surface water area (hereafter referred to as "area") associated with known beaver dams using 0.6 m resolution aerial imagery from the National Agriculture Imagery Program (NAIP), processed in Google Earth Engine (GEE) (Fig. 7). Beaver dam locations (latitude and longitude) and lengths were manually labeled using high-resolution imagery (0.2 m or better) from a proprietary image database, with data collected between 2014 and 2021[23]. Across the eleven study areas, we identified a total of 1917 beaver dams. Of these, 1198 (62%) of the labeled beaver dams were sourced from the EEAGER model evaluation polygons, while the remaining 719 (38%) dams were manually identified around Laramie, WY (study area M) using high resolution imagery following the same criteria applied in the EAGER model and other prior studies that delineate beaver dams[5,6,13,23]. These dams range in length from a few meters to 300 meters (median = 30 m). They represent diverse landscapes but are limited to the model's coverage[23]. Thus, the spatial distribution of labeled beaver dams in our dataset may not fully represent the broader ecological and geographical variability of beaver activity across the western United States.

In total, we accessed 53 NAIP tiles collected between June and September of 2015, 2021, and 2022 via the GEE platform (Table S2; Text S3). Specifically, we used 2021 NAIP imagery for Polygons A–H, as 2022 imagery was not available, and 2022 imagery for Polygons J and K to match the evaluation period used by the EEAGER model and ensure consistency—although the original dam labeling dates span from 2014 to 2020 (Table S2). For Polygon M, we used 2015 NAIP imagery to align with the timeframe during which the dams were recorded. We acknowledge that this introduces a temporal mismatch for some sites, as the dams may have been constructed prior to the imagery date. However, given that beaver ponds often persist for multiple years and our analysis is conducted at the complex scale, we expect the majority of mapped pond areas still reflect the hydrological footprint of those dams. To improve computational efficiency and reduce shadow effects, we restricted mapping to riparian zones and generated cloud-free NAIP mosaic scenes for each study area.

We calculated the Normalized Difference Water Index (NDWI) using green and near-infrared (NIR) reflectance following[35], Eq. (1). This formulation enhances surface water detection and differs from other NDWI variations, such as[36], which uses shortwave infrared (SWIR) to assess vegetation water content and stress. NDWI is positively correlated with water content, as water typically exhibits higher reflectance in the green band and lower reflectance in the NIR band, resulting in positive NDWI values. In riparian zones, beaver ponds are distinct from surrounding land features and have higher NDWI values[37,38]. Based on this, we applied a threshold of zero to classify images into water and non-water (Fig. 7b). The resulting water-class rasters were converted into vector polygons using the *reduceToVectors* function in GEE. By intersecting these polygons with labeled beaver dam locations, we delineated areas within the riparian zones.

$$\mathrm{NDWI} = (\mathrm{Green} - \mathrm{NIR})/(\mathrm{Green} + \mathrm{NIR}) \tag{1}$$

Some partially drained ponds were not initially captured due to lower water levels during the summer mapping period compared to when the dams were originally labeled. To account for this, we applied a 5 m buffer on both sides of the beaver dam lines and re-intersected them with the pond polygons, ensuring that adjacent ponds missed due to seasonal water levels

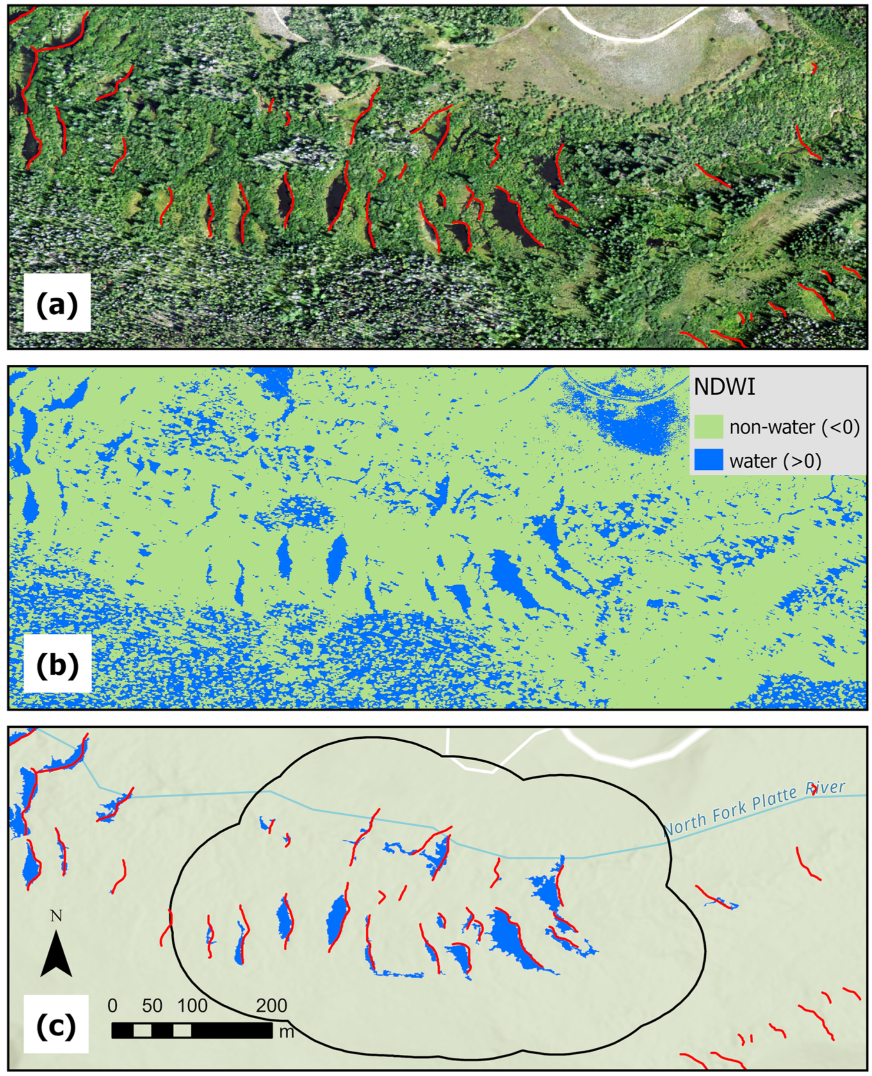


Fig. 7 | Workflow for mapping individual beaver ponds and creating beaver pond complexes. **a** Manually identified beaver dams (red) overlaid on the NAIP mosaic. **b** Water and non-water masks created using an NDWI threshold of zero. **c** Beaver ponds (blue) generated by intersecting water polygons with mapped dams. The black outline represents a 100-meter buffer applied to a beaver pond cluster identified using the DBSCAN clustering algorithm, referred to as a beaver pond complex. Background image source: World Topographic Map.

fluctuation were included. Disconnected areas associated with each beaver dam were merged into a single pond for analysis. However, some water areas merged due to closely spaced beaver dams. These connected surface water bodies and their associated dams were excluded from the pond-level distribution analysis, reducing the dataset to 1272 dams, but were retained for complex-scale distribution and factor analyses.

Using this workflow, we successfully mapped water areas for 73% (1400) of the labeled beaver dams. Mapping success was lower in some regions, for instance, in the North American Desert where only 65.5% of dams had detectable areas (Fig. S8a). One limitation of the NDWI-based approach is its reduced sensitivity to flooded vegetation, which may cause underestimation in ponds that inundate vegetated floodplains. To assess the accuracy of our aerial mapping, we randomly selected 3% of the beaver ponds and compared their mapped areas with manually delineated areas. In total, 45 ponds were manually drawn, with a mean area 10.5% smaller than the mapped pond area (slope = 0.895). The coefficient of determination (R^2) was 0.89, with a mean absolute error (MAE) of 89 m^2, indicating a good fit (Fig. S9). Based on this accuracy assessment, we analyzed pond size distributions across ecoregions using the Shapiro-Wilk test for normality and compared median pond sizes among ecoregions using the Kruskal-Wallis test (Text S1).

Creating beaver pond complexes using a density-based clustering algorithm

Beavers construct dams in spatial clusters, commonly known as beaver dam complexes[39]. In this study, we define these clusters as beaver pond complexes (Fig. 7c) to better reflect their spatial structure, as beavers modify landscapes by creating interconnected ponds and channels that influence hydrology, vegetation and wildlife habitat. To delineate these complexes, we grouped individual beaver ponds to beaver pond clusters using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, implemented with the R package *dbscan*[40,41]. To account for beaver movement ranges and hydrological connectivity among nearby ponds, we created beaver pond complexes by applying a 100 m buffer to each cluster, representing the maximum distance beavers typically travel from their ponds in search of food[18,42]. More details on DBSCAN are provided in Text S2.

Using this approach, we delineated 87 beaver pond complexes (Fig. S10). A preliminary Generalized Additive Model (GAM) analysis showed that total pond area at the complex level is strongly correlated with dam length (adjusted R^2 = 0.88, Fig. S2c). This result suggests that dam length is a strong covariate of complex-scale pond area, prompting further investigation into the factors influencing both total pond area and dam length across beaver pond complexes.

Investigating factors influencing pond area and dam length variation across complexes

To identify factors influencing additive ponding areas and dam lengths across 87 beaver pond complexes, we implemented a systematic three-step approach:

(1) Variable selection - A set of 25 variables was selected based on prior research on beaver pond and dam habitat drivers[19,43–45].

(2) Variable exclusion - Pearson's correlation coefficients (r) were calculated to assess multicollinearity among independent variables, followed by variance inflation factor (VIF) analysis to refine the variable set.
(3) Generalized Additive Models (GAMs) - Relationships between ponding area and independent variables, as well as dam length and independent variables (excluding dam length as a covariate) were examined.

These variables (Table S1) were grouped into five categories: (1) soil characteristics (e.g., clay percentage, saturated soil water content); (2) vegetation metrics (e.g., vegetation cover and height); (3) climate variables (e.g., snow water equivalent, precipitation, temperature); (4) topographic indices (e.g., compound topographic index, stream power index, stream transport index); and (5) hydrological indices and stream network metrics.

Data were compiled from seven publicly available datasets since ground-based measurements were infeasible at the spatial scale of this study. Among the 25 variables selected, climate variables were dynamic, corresponding to the year when NAIP imagery was acquired, while the remaining variables were static. Static datasets ranged in spatial resolution from 10 to 90 m, and values were extracted for each beaver pond complex using default summary statistics (e.g., median, maximum). Climate variables were available at a coarser resolution (1 km), such that each beaver pond complex (ranging from 46,000 to 1,500,000 m²) typically intersected fewer than two pixels. To address this, we first calculated relevant climate statistics (e.g., max, sum, mean) over the appropriate temporal window and then applied the built-in weighted reducer function in GEE to derive a representative mean value for each complex. Additional details on variable derivation are provided in Text S4.

For Step 2, Pearson's correlation analysis identified variable pairs with $r > 0.7$, where the variable more strongly correlated with pond area was retained. Additional response variables (e.g., ponding area relative to cluster area, ponding area per stream length) were tested but exhibited weaker relationships with independent variables, thus pond area was determined as the target variable. For dam length analysis, the total dam length per complex was used. This process eliminated five variables (saturated hydraulic conductivity, summer precipitation, mean temperature, valley bottom area percent and beaver dam counts) due to high intercorrelation, as well as four vegetation variables (tree and shrub height and cover) due to data availability constraints (Table S1). The final 16 variables (indicated by * in Table S1) exhibited pairwise correlation coefficient (r) below 0.7, indicating low to moderate relationships (Fig. S11). VIF analysis showed that most variables had values below or slightly above 5, except for spring precipitation (PrecipSpring, VIF = 8.09) which was retained due to its hydrological relevance in sustaining ponding during the early growing season.

GAMs were selected to examine the relationships between pond area and independent variables at the complex level (Table 1). As an explanatory modeling approach here, GAMs offer a flexible and interpretable framework for assessing nonlinear relationships between covariates and response variables. They also provide partial effects natively, allowing us to examine the marginal effect of each covariate on the response while holding other covariates constant, which facilitate clearer interpretation of individual variable contributions. This contrasts with other models such as XGBoost, which require additional interpretation tools. Additionally, simple linear regression was tested but explained substantially less variability in pond area.

Separate GAMs were also developed to analyze dam length, excluding pond area due to its strong correlation with dam length (adjusted $R^2 = 0.88$, Fig. S2c). To ensure comparability across variables, the response variables (pond area and dam length) were transformed using the natural logarithm, while the independent variables were normalized to a 0–100 scale. The distributions of independent variables prior to transformation are shown in Fig. S12. Model performance was evaluated using adjusted R^2 and deviance explained. Factor importance was assessed using F-values and corresponding p-values (statistical significance), and Effective Degrees of Freedom (EDF) where higher EDF indicates greater nonlinearity. Partial effect plots were used to visualize each covariate's influence while controlling other variables as their median values. Additionally, pairwise interactions were examined using tensor product smooths ($ti()$) to determine whether environmental factors jointly influence dam construction and pond area. Further details on GAM methodology are provided in Text S2.

Data availability

Polaris 30 m Probabilistic Soil Properties US, Landfire Mosaics LF, Snow Data Assimilation System (SNODAS) and Hydrography 90 m Layers used in this paper can be accessed from the awesome-gee-community-catalog (https://gee-community-catalog.org/)[46], Daymet V4: Daily Surface Weather and Climatological Summaries, USGS 3DEP 10 m National Map and NAIP: National Agriculture Imagery Program imagery can be obtained from the Earth Engine's public data archive (developers.google.com/earth-engine/datasets). The NHDPlus HR data are available from USGS. We have made the beaver dam location data publicly available on HydroShare. http://www.hydroshare.org/resource/f078d94b1f99460f8f374dbe018a01bc. We have also made the mapped beaver pond and cluster datasets, along with associated predictor variables, publicly available on HydroShare. https://www.hydroshare.org/resource/f55b079d07ac4a3f9e1f05f997763da4.

Code availability

GEE code for mapping beaver ponds and extracting influencing factors, and R notebooks for data processing, analysis, and visualization, are available on GitHub. https://github.com/LuwenWan/BeaverPond_Cluster_Factor.

Received: 12 March 2025; Accepted: 10 July 2025;
Published online: 11 August 2025

References

1. Wohl, E. Resilience in River Corridors: How Much Do We Need?. *Perspect. Earth Space Sci.* **5**, e2023CN000226 (2024).
2. Larsen, A., Larsen, J. R. & Lane, S. N. Dam builders and their works: Beaver influences on the structure and function of river corridor hydrology, geomorphology, biogeochemistry and ecosystems. *Earth-Sci. Rev.* **218**, 103623 (2021).
3. Puttock, A., Graham, H. A., Cunliffe, A. M., Elliott, M. & Brazier, R. E. Eurasian beaver activity increases water storage, attenuates flow and mitigates diffuse pollution from intensively-managed grasslands. *Sci. Total Environ.* **576**, 430–443 (2017).
4. Westbrook, C. J., Cooper, D. J. & Baker, B. W. Beaver dams and overbank floods influence groundwater–surface water interactions of a Rocky Mountain riparian area. *Water Resour. Res*. **42**, W06404 (2006).
5. Fairfax, E. et al. Impacts of beaver dams on riverscape burn severity during megafires in the Rocky Mountain region, western United States. in (2024).
6. Fairfax, E. & Whittle, A. Smokey the Beaver: beaver-dammed riparian corridors stay green during wildfire throughout the western United States. *Ecol. Appl.* **30**, e02225 (2020).
7. Wohl, E., Marshall, A. E., Scamardo, J., White, D. & Morrison, R. R. Biogeomorphic influences on river corridor resilience to wildfire disturbances in a mountain stream of the Southern Rockies. *USA. Sci. Total Environ.* **820**, 153321 (2022).
8. Brazier, R. E. et al. Beaver: Nature's ecosystem engineers. *Wiley Interdiscip. Rev. Water* **8**, e1494 (2021).
9. Thompson, S., Vehkaoja, M., Pellikka, J. & Nummi, P. Ecosystem services provided by beavers Castor spp. *Mammal Rev* **51**, 25–39 (2021).
10. Grudzinski, B. P. et al. A global review of beaver dam impacts: Stream conservation implications across biomes. *Glob. Ecol. Conserv.* **37**, e02163 (2022).

11. Martin, S. L., Jasinski, B. L., Kendall, A. D., Dahl, T. A. & Hyndman, D. W. Quantifying beaver dam dynamics and sediment retention using aerial imagery, habitat characteristics, and economic drivers. *Landsc. Ecol.* **30**, 1129–1144 (2015).
12. Adamchak, C., Lininger, K. B. & Hinckley, E.-L. S. Animating the Critical Zone: Beavers as Critical Zone Engineers. *Front. Water* **7**, 1547094 (2025).
13. Fairfax, E. & Small, E. E. Using remote sensing to assess the impact of beaver damming on riparian evapotranspiration in an arid landscape. *Ecohydrology* **11**, e1993 (2018).
14. Jordan, C. E. & Fairfax, E. Beaver: The North American freshwater climate action plan. *Wiley Interdiscip. Rev. Water* **9**, e1592 (2022).
15. Pollock, M. M. et al. Using Beaver Dams to Restore Incised Stream Ecosystems. *BioScience* **64**, 279–290 (2014).
16. Campbell-Palmer, R., Gow, D., Rosell, F., Dickinson, H. & Parker, H. *The Eurasian Beaver Handbook: Ecology and Management of Castor Fiber*. (Pelagic Publishing, Exeter, 2016).
17. Yarmey, N. T. & Hood, G. A. Resident Perceptions of Human–Beaver Conflict in a Rural Landscape in Alberta, Canada. *Human–Wildlife Interact*. **14**, 476–486 (2020).
18. Macfarlane, W. W. et al. Modeling the capacity of riverscapes to support beaver dams. *Geomorphology* **277**, 72–99 (2017).
19. Touihri, M., Labbé, J., Imbeau, L. & Darveau, M. North American Beaver (*Castor canadensis* Kuhl) key habitat characteristics: review of the relative effects of geomorphology, food availability and anthropogenic infrastructure. *Écoscience* **25**, 9–23 (2018).
20. Fitch, K., Nippgen, F., Albeke, S. E. & Paige, G. B. Where the wild beavers are: Climate and landscape controls on beaver pond area in snow-dominated rangeland headwaters. *Ecohydrology* **15**, e2418 (2022).
21. Dewey, C. et al. Beaver dams overshadow climate extremes in controlling riparian hydrology and water quality. *Nat. Commun.* **13**, 6509 (2022).
22. Dittbrenner, B. J., Schilling, J. W., Torgersen, C. E. & Lawler, J. J. Relocated beaver can increase water storage and decrease stream temperature in headwater streams. *Ecosphere* **13**, e4168 (2022).
23. Fairfax, E. et al. EEAGER: A Neural Network Model for Finding Beaver Complexes in Satellite and Aerial Imagery. *J. Geophys. Res. Biogeosci.* **128**, e2022JG007196 (2023).
24. Hood, G. A. & Bayley, S. E. Beaver (Castor canadensis) mitigate the effects of climate on the area of open water in boreal wetlands in western Canada. *Biol. Conserv.* **141**, 556–567 (2008).
25. Zhang, W., Hu, B., Brown, G. & Meyer, S. Beaver pond identification from multi-temporal and multi- sourced remote sensing data. *Geo-Spat. Inf. Sci.* **27**, 953–967 (2023).
26. Fraser, R. H., Olthof, I. & Berezanski, D. Large multi-decade beaver ponding changes in the subarctic Hudson Bay Lowlands, Canada observed using satellite remote sensing. *Environ. Res. Lett.* https://doi.org/10.1088/1748-9326/ad36d6 (2024).
27. Fairfax, E. & Westbrook, C. The Ecology and Evolution of Beavers: Ecosystem Engineers that Ameliorate Climate Change. https://doi.org/10.1146/annurev-ecolsys-102722-122317 (2024).
28. Graham, H. A. et al. Modelling Eurasian beaver foraging habitat and dam suitability, for predicting the location and number of dams throughout catchments in Great Britain. *Eur. J. Wildl. Res.* **66**, 42 (2020).
29. Howard, R. J. & Larson, J. S. A Stream Habitat Classification System for Beaver. *J. Wildl. Manag.* **49**, 19 (1985).
30. Beardsley, M. Park County Beaver Restoration Assessment. (2022).
31. Moravek, J. A. et al. Using beaver capacity models: the importance of local knowledge. *Restor. Ecol.* **n/a**, e70135.
32. Wohl, E., Dwire, K., Sutfin, N., Polvi, L. & Bazan, R. Mechanisms of carbon storage in mountainous headwater rivers. *Nat. Commun.* **3**, 1263 (2012).
33. Omernik, J. M. Ecoregions of the Conterminous United States. *Ann. Assoc. Am. Geogr.* **77**, 118–125 (1987).
34. Omernik, J. M. & Griffith, G. E. Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework. *Environ. Manage.* **54**, 1249–1266 (2014).
35. McFeeters, S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **17**, 1425–1432 (1996).
36. Gao, B. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **58**, 257–266 (1996).
37. Kaha, M. & Lang, M. Spectral response of forest disturbances induced by beaver caused floods: A case study in Estonia using Sentinel-2 MSI image time series. *Balt. For.* **30**, id647–id647 (2024).
38. Smufer, F., Casas-Ruiz, J. P., St-Pierre, A. & del Giorgio, P. A. Integrating Beaver Ponds into the Carbon Emission Budget of Boreal Aquatic Networks: A Case Study at the Watershed Scale. *Ecosystems* **26**, 1309–1325 (2023).
39. Hafen, K. C., Wheaton, J. M., Roper, B. B., Bailey, P. & Bouwes, N. Influence of topographic, geomorphic, and hydrologic variables on beaver dam height and persistence in the intermountain western United States. *Earth Surf. Process. Landf.* **45**, 2664–2674 (2020).
40. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* 226–231 (AAAI Press, Portland, Oregon, 1996).
41. Hahsler, M., Piekenbrock, M. & Doran, D. dbscan: Fast Density-Based Clustering with R. *J. Stat. Softw.* **91**, 1–30 (2019).
42. Jenkins, S. H. A Size-Distance Relation in Food Selection by Beavers. *Ecology* **61**, 740–746 (1980).
43. Dittbrenner, B. J. et al. Modeling intrinsic potential for beaver (Castor canadensis) habitat to inform restoration and climate change adaptation. *PLoS ONE* **13**, e0192538 (2018).
44. Scrafford, M. A., Tyers, D. B., Patten, D. T. & Sowell, B. F. Beaver Habitat Selection for 24 Yr Since Reintroduction North of Yellowstone National Park. *Rangel. Ecol. Manag.* **71**, 266–273 (2018).
45. St-Pierre, M. L., Labbé, J., Darveau, M., Imbeau, L. & Mazerolle, M. J. Factors Affecting Abundance of Beaver Dams in Forested Landscapes. *Wetlands* **37**, 941–949 (2017).
46. Roy, S., Swetnam, T., Trochim, E., Schwehr, K. & Pasquarella, V. samapriya/awesome-gee-community-datasets: Community Catalog (1.0.4). Zenodo. (2023).

Acknowledgements

The authors gratefully acknowledge support from the Stanford HAI Postdoctoral Fellowship Program and Woods Institute for the Environmental Ventures Project at Stanford University. We are grateful to Kuai Fang (Stanford University), George Hilley (Stanford University) and Kamini Singha (Colorado School of Mines) for their valuable suggestions for improving the quality of this work.

Author contributions

L.W., E.F., and K.M. conceptualized the study. L.W. performed the analyses and wrote the manuscript with input from all co-authors. All authors contributed to manuscript editing. K.M. acquired funding.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43247-025-02573-x.

Correspondence and requests for materials should be addressed to Luwen Wan.

Peer review information *Communications Earth & Environment* thanks Johnston John M. and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Haihan Zhang and Alireza Bahadori. A peer review file is available.