ORNL/TM-2022/2459

An Assessment of Non-Powered Dam Hydropower Development Opportunities in the United States

Scott DeNeale
Carly Hansen
Jenberu Feyyisa
Gbadebo Oladosu

October 2022

Approved for public release.
Distribution is unlimited.

ORNL IS MANAGED BY UT-BATTELLE LLC FOR THE US DEPARTMENT OF ENERGY

Cover credits: Images courtesy of Scott DeNeale.

(top-left) Byrd Creek Dam; located in Crossville, Tennessee; photograph dated November 10, 2019.

(top-right) Lake Sequoyah Dam; located in Highlands, North Carolina; photograph dated August 20, 2021.

(bottom) A decommissioned turbine runner; located at Watts Bar Dam in Spring City, Tennessee; photograph dated November 20, 2018.

ORNL/TM-2022/2459

Environmental Sciences Division

AN ASSESSMENT OF NON-POWERED DAM HYDROPOWER DEVELOPMENT OPPORTUNITIES IN THE UNITED STATES

Scott DeNeale
Carly Hansen
Jenberu Feyyisa
Gbadebo Oladosu

October 2022

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, TN 37831
managed by
UT-BATTELLE LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

CapEx	capital expenditures per kilowatt
DOE	US Department of Energy
EHA	Existing Hydropower Assets
FERC	Federal Energy Regulatory Commission
LCOE	levelized cost of energy
MOU	memorandum of understanding
NPD	non-powered dam
ORNL	Oak Ridge National Laboratory
Q30	30% exceedance flow
t-SNE	t-distributed stochastic neighbor embedding
USACE	US Army Corps of Engineers

ACKNOWLEDGMENTS

US Department of Energy, Water Power Technologies Office

The authors would like to acknowledge and express their appreciation to the US Department of Energy (DOE) Water Power Technologies Office (WPTO) for overseeing and funding this study to perform a critical review of hydropower geotechnical foundation practice and innovation opportunities. The following DOE WPTO staff were heavily involved in reviewing this report and supporting this study:

- Kathryn Jackson
- Kyle DeSomber
- Colin Sasthav
- Timothy Welch
- Jennifer Garson

Oak Ridge National Laboratory

The following individuals from Oak Ridge National Laboratory provided technical review and support for this report:

- Olivia Shafer

EXECUTIVE SUMMARY

Retrofitting non-powered dams (NPDs) to add hydropower offers multiple benefits over traditional, new hydropower development. Since much of the civil works infrastructure already exists, many retrofits require minimal new construction and can leverage existing water conveyances and discharge for generation. Although NPDs vary considerably in terms of their defining characteristics, identifying similarities helps support targeted investment and find opportunities to develop solutions for common challenges. With the publication of NPD-related data in 2022 (Hansen et al. 2022b), the US Department of Energy and relevant stakeholders are equipped with key information across roughly 89,000 US NPDs. These data represent the best-available information to date and extend beyond previous assessments of potential power capacity by describing design, operational, socioeconomic, environmental aspects of NPDs (Hadjerioua et al. 2012).

To capitalize on these efforts to improve breadth and depth of NPD data access, this study uses a data-driven approach to assess NPD hydropower development opportunities in the United States. To help describe project feasibility drivers for NPDs, recent NPD retrofits were reviewed to examine variability with respect to a variety of characteristics. Based on the assessment of recent retrofits and data availability, several attributes were selected to characterize the remaining NPD population: (1) owner type, (2) 30% exceedance flow (a common design flow that describes the flow that is exceeded by 30% of the flow in the record), (3) hydraulic head, and (4) maximum reservoir storage. These four characteristics were used as inputs to a statistical clustering analysis (a common technique for grouping individuals of a population based on similarities or how closely associated individuals are to one another) of a set of 2,709 NPDs with at least 100 kW potential capacity and recently retrofit dams with available data. With the large dataset of NPDs, the clusters help describe how the population breaks down into different types (i.e., how many types of dams there are and what portion of the population belongs to each cluster).

Key takeaways from NPD retrofit and population clustering analysis include the following:

- Recent NPD retrofits have most often used an approach that goes through existing intake/conveyance systems rather than adding generation via a bypass or siphon.
- Dam ownership plays a major role in differentiating NPDs—nearly all NPD clusters consist of a single dam owner type.
- Federally owned NPDs are the most common type of retrofit and have higher distributions of Q30 flow, head, and reservoir storage compared with other clusters.
- Each cluster contains a mix of dam purposes, as well as dams with locks and without.

Comparisons of cost estimates are also provided for this subset of NPDs. Results demonstrate that **NPDs with higher estimated capacity are associated with lower costs.** Median capital expenditures per kilowatt (CapEx) and levelized cost of energy (LCOE) for the 100 dams with the greatest capacity are 30%–-40% lower than other NPDs. The top 100 NPDs by CapEx and LCOE make up between 15% and 33% of the total capacity among the subset of NPDs (n = 2,668). **The majority of these low-cost, high-feasibility NPDs are large (with respect to estimated flow, head, and storage) and owned by federal or local government entities.** These observations are consistent with other hydropower and energy cost analyses, which demonstrate the impact of economies of scale (i.e., large projects are generally more economically feasible than small projects).

In summary, the statistical analysis of US NPDs describes patterns that are relevant to economic feasibility of powering NPDs. Owner type and project size (in terms of head, flow, and storage) were effective in forming distinct, descriptive clusters that represent the opportunity space for NPD development and clearly reflect the dominance of particular types of dams which have been successfully retrofit. Additional data improvement efforts will be needed in future work to address limitations of data quality. The results of this study may prove useful to project developers considering what kind of projects to target for development, as well as to the US Department of Energy for informing follow-on investment strategies.

The defining characteristics of clusters and the estimated project cost metrics should not be considered a promise for project feasibility or success. Rather, the identification of certain characteristics that are common among successful projects and patterns of feasibility observed across clusters can aid developers by supporting a targeted approach to further data refinement and more detailed evaluations of potential hydropower development.

1. INTRODUCTION

Non-powered dams (NPDs; defined as "dams that do not have any electricity generation equipment installed"[1]) represent both critical infrastructure (DeNeale et al. 2022; DeNeale et al. 2019; Hansen et al. 2021) and energy development opportunities (Hadjerioua et al. 2012). Totaling more than 89,000 dams across the United States,[2] NPDs come in a variety of shapes and sizes, serve numerous purposes, and exhibit a wide range of characteristics that make each dam unique (Hansen et al. 2022b); however, similarities do exist. To improve the understanding of energy development potential, an improved understanding of similarities among NPDs is needed. To this end, this report aims to provide a detailed review and analysis of NPD development opportunities in the United States.

1.1 Motivation

Increasing renewable energy has been a long-standing goal of global governments and communities as the world addresses the increasing, negative impacts of climate change. For example, the Biden Administration has a goal of achieving 100% clean electricity by 2035.[3] As one of the leading sources of renewable power, hydropower represents 6.3% of the US electricity generation and 31.3% of total renewable generation.[4] Despite its slow growth over the past few decades, there remains sizable undeveloped hydropower potential in the United States, including NPDs.

Over the past decade, multiple resource assessments have been conducted to estimate the potential from adding hydropower at NPDs. The 2012 NPD resource assessment led by the US Department of Energy's (DOE's) Oak Ridge National Laboratory (ORNL) identified 12.1 GW of NPD technical development potential (Hadjerioua et al. 2012). A subsequent 2016 economic assessment funded by DOE identified the potential to add 3.6 GW at NPDs through 2030 (US Department of Energy 2016). These resource assessments provide programmatic support for hydropower development at NPDs and offer valuable information for decision makers.

Various organizations have recently been increasing efforts toward low-impact NPD development. Since 2018, Stanford University has been spearheading the Uncommon Dialogue[5], an effort that includes commonly underrepresented groups, such as Native American tribes, for more community-centered projects and encourages increased stakeholder engagement around NPDs. Included in this initiative is the identification of three key opportunities in hydropower, river restoration, and public safety. These opportunities are referred to as the three R's: rehabilitation, retrofit, and removal. The Uncommon Dialogue continues to push for legislative action and investments toward addressing the three R's, including a recent legislative package geared toward expediting the NPD licensing process to a maximum of 2 years for qualifying NPDs.[6]

Retrofitting is of most relevance for NPD development. DOE Water Power Technologies Office's 2022 Multi-Year Program Plan identified a number of NPD-related goals, including developing data sets and

---

[1] Definition from DOE: https://www.energy.gov/eere/water/glossary-hydropower-terms (Accessed September 28, 2022).
[2] According to the 2021 US Army Corps of Engineers' National Inventory of Dams (Accessed September 28, 2022).
[3] https://www.energy.gov/articles/biden-administration-launches-bipartisan-infrastructure-law-initiative (Accessed September 28, 2022).
[4] According to the US Energy Information Administration, values as of 2021. https://www.eia.gov/tools/faqs/faq.php?id=427&t=3 (Accessed September 28, 2022).
[5] Uncommon Dialogue on Hydropower, River Restoration, and Public Safety | Stanford Woods Institute for the Environment (Accessed September 28, 2022)
[6] https://www.environmentallawandpolicy.com (Accessed September 28, 2022).

interactive geospatial tools to identify NPD development potential and site characteristics; developing, testing, and validating new technologies for low-impact hydropower growth at NPDs; and supporting standardized and modular approaches to NPD hydropower project design. Much of these initiatives stem from R&D efforts funded by national laboratories (US Department of Energy 2022).

Beyond R&D, DOE also remains in active collaboration with other federal agencies through partnerships and collaborations. A 2010 memorandum of understanding (MOU)[7] among DOE, the US Army Corps of Engineers (USACE), and the US Bureau of Reclamation resulted in a subsequent MOU between USACE and the Federal Energy Regulatory Commission (FERC) for streamlining hydropower development at USACE NPDs. A 2020 MOU[8] among DOE, USBR, and USACE further aims to enhance collaboration between agencies to meet the need for reliable and affordable hydropower.

1.2 Objectives and Scope

Altogether, DOE's various initiatives surrounding hydropower retrofits at NPDs have furthered the understanding and capabilities among decision makers for investigating development opportunities. This study builds upon previous DOE-funded efforts, including resource assessments, classification research, and cost analysis. The objective of this report is to help identify the most promising development opportunities for NPDs in the United States, and to support NPD decision makers, primarily hydropower project developers considering development at one or more NPDs.

To accomplish this objective, this study assesses characteristics that are believed to influence NPD development feasibility, develops statistics-based groupings of NPDs based on multivariate analysis, and provides preliminary determinations of feasibility based on available data. Available data stem largely from data collected by ORNL in its recent research on NPD data needs and research gaps (Hansen et al. 2021) and the NPD Characteristics Inventory data set (Hansen et al. 2022a).

Rather than conducting analyses for the roughly 90,000 US NPDs, this study focuses on those most likely to be economically feasible. Given economies of scale, larger-scale projects are often more feasible. Thus, this study's scope focuses on NPDs with capacities previously estimated as having 100 kW or more of hydropower development potential (Hadjerioua et al. 2012) at approximately 3,300 US NPDs.

Beyond the background research presented in Section 1.3, this report provides a review of recent NPD retrofits in the United States; clustering and pre-feasibility cost analyses of NPD development opportunities, and summary and discussion related to challenges, limitations, opportunities, and next steps.

1.3 Background Research

Ultimately, DOE's goal through this study and related research is to increase renewable hydropower generation at NPDs by leveraging foundational research and informing hydropower development and deployment at NPDs through low-impact solutions. Along these lines, DOE recently funded ORNL to conduct research related to NPD classification, data access, and retrofitting. These research efforts,

[7] https://www.energy.gov/articles/doe-doi-and-army-corps-engineers-sign-memorandum-understanding-hydropower (Accessed September 28, 2022).
[8] https://www.energy.gov/eere/water/articles/hydropower-memorandum-understanding-2020 (Accessed September 28, 2022).

funded as a part of the Standard Modular Hydropower Technology Acceleration project,[9] have two web-based applications:[10] the NPD Explorer and NPDamCAT Apps (Carter et al. 2022).

These research, data, and tool development efforts that support understanding of opportunities build on decades of research at national laboratories and federal agencies. Initial estimates of potential for undeveloped sites in the United States were produced by Idaho National Laboratory (Conner et al. 1998). More recently, potential capacity was assessed specifically for NPDs based on the best-available public dam locations and flow information (Hadjerioua et al. 2012). More detailed estimates have also been developed based on agency-specific information (US Department of the Interior Bureau of Reclamation 2011; US Army Corps of Engineers 2013).

Additionally, several key data sets have been developed that are closely related to NPDs or are relevant to NPD development because they describe existing hydropower and water/energy infrastructure. These data are used widely within the hydropower industry and research community and provide authoritative information on the status of the US hydropower fleet and development pipeline. These include ORNL's Existing Hydropower Assets (EHA) data set (Johnson et al. 2022) and US Hydropower Development Pipeline data set (Johnson and Uría-Martínez 2022). Section 2 leverages information from the EHA to determine which hydropower projects have recently become operational by retrofitting NPDs. Figure 1 maps NPDs in the United States, as well as operational NPD retrofit projects that came online from 2000 to 2021.

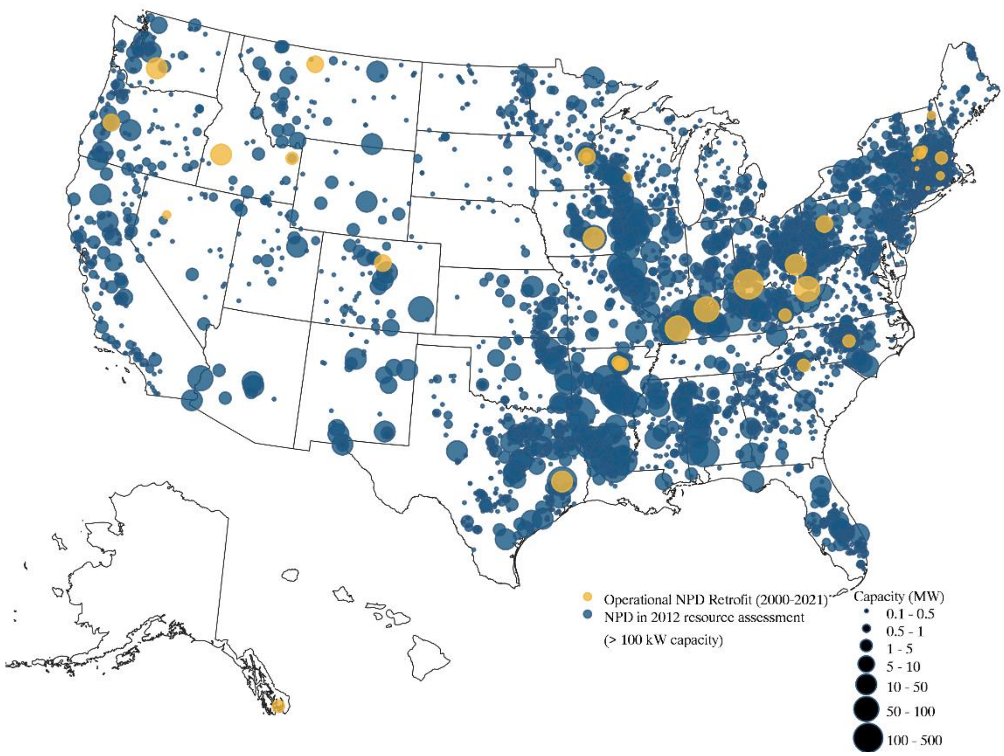


Figure 1. US NPDs and operational NPD retrofits. Blue dots represent NPDs estimated as having at least 100 kW of power potential in the 2012 NPD Resource Assessment and are scaled by capacity. Gold dots represent NPDs that were retrofitted from 2000 to 2021, totaling 588 MW in installed capacity.

[9] https://smh.ornl.gov/ (Accessed September 28, 2022).
[10] https://hydrosource.ornl.gov/tool/npd_tools (Accessed September 28, 2022).

DOE also recently funded ORNL to conduct a cost analysis of hydropower options at NPDs (Oladosu et al. 2021). That analysis followed a similar approach to the approach used herein; however, the objectives and scope of the two efforts vary. Rather than using groups or classification to describe types or the makeup of dams, the team used clustering to identify representative or reference dams that could be used for detailed cost modeling. The results of the 2020 cost analysis have helped inform DOE’s programmatic goals related to NPDs.

2. REVIEW OF RECENT NPD RETROFITS

To better assess hydropower development opportunities across the US NPD population, the various NPD retrofit projects that have recently been developed should be considered. Since these NPDs were successfully commissioned, one may assume that they represented attractive opportunities from both a technical and economic perspective and were capable of achieving hydropower development while meeting environmental and other regulatory requirements. However, improved understanding of the characteristics and factors that contributed to the success of these retrofits will help refine this assumption and inform future NPD development opportunities and R&D initiatives. This section presents a summary of the data collection and analyses conducted on recent NPD retrofits.

2.1 Data Collection

As stated, NPDs come in a variety of shapes and sizes, and the setting of each dam carries certain implications that may affect design. Hansen et al. (2022b) discussed the various characteristics related to NPDs and presented a framework for custom classification of dams. At the highest level, NPD characteristics fall into five main themes: design-related, environmental, hydropower opportunity–related, operational, or socioeconomic. Using this framework, ORNL conducted a data collection effort and published the NPD Characteristics Inventory (Hansen et al. 2022a), which provides data across 178 characteristics. A summary of subcategories of characteristics, organized by theme, is provided in Table 1. Although others exist, these 178 characteristics represent those for which data are currently available across most NPDs. Among these characteristics, a variety may be assumed to influence cost. To analyze features of recent NPD retrofits, reasonably reliable data are needed.

Table 1. Summary of subcategories of characteristics included in the NPD Characteristics Inventory.
Source: Hansen et al. (2022a)

Theme	Subcategory
Design	Age
Design	Architectural/structural type
Design	Connection to the grid
Design	Material
Design	Size/dimension
Design	Water conveyance
Environmental	Climate
Environmental	Ecology
Environmental	Geology
Environmental	Hydrology
Environmental	Land
Environmental	Land protections
Environmental	Soil type in the catchment
Environmental	Soil type in the watershed
Environmental	Water quality
Hydropower opportunity	Energy regulatory agency
Hydropower opportunity	Generation potential
Hydropower opportunity	Relationship to other renewable energies
Operational	Operating entity
Operational	Purpose
Operational	Relationship to other dams
Operational	Reservoir operation
Socioeconomic	Attitudes toward renewable energy and climate issues
Socioeconomic	Cultural significance

Theme	Subcategory
Socioeconomic	Dam safety
Socioeconomic	Demographics
Socioeconomic	Location
Socioeconomic	Ownership
Socioeconomic	Political context
Socioeconomic	Recreation
Socioeconomic	Regulatory authority
Socioeconomic	Water use

To support recent NPD retrofit analysis, data were collected from a variety of sources. The National Inventory of Dams (NID; US Army Corps of Engineers (2021)) was used to obtain information about location, dam ownership, and physical characteristics of the dam (i.e., dam height). Project descriptions in FERC licenses (obtained from the FERC eLibrary[11]) were also used to fill in details about dam height and verify locations. Other data sources included the NPD Characteristics Inventory (Hansen et al. 2022a), which uses exceedance flow (the flow that is expected to be exceeded by some percentage of the given record) to describe the flow at each NPD in the conterminous United States. For this study, the 30% exceedance flow (Q30; the amount of flow which is equalled or exceeded 30% of the time)—a common measure of flow that is used in initial design or estimates of power potential—was examined. This Q30 value represents the flow that is exceeded by 30% of the flow in the record, which comes from historical (1980–2015) reanalysis of flow data (Ghimire et al. 2022). Finally, locations of existing substations were obtained from the Homeland Infrastructure Foundation-Level Data repository (Homeland Infrastructure Foundation-Level Data 2022).

Major steps taken to assemble the recent retrofit data include the following:

- Assembling data for projects that have been completed and listed as operational in the most recent EHA data set since 2000 (Johnson et al. 2022): This includes compiling details from 36 projects reviewed by Hansen et al. (2021), as well as identifying 6 additional projects that involve powering previously powered or mechanically powered (e.g., mills) facilities. Of the 42 projects identified as NPD retrofits, 18 have become operational since the 2012 NPD Resource Assessment was completed (i.e., the power plant reports the operational year from 2012 to 2021).

- Aggregating project data across the various sources into a single data set, resulting in 42 unique retrofits: **In total, 42 recent retrofit projects have been identified for the period of 2000 to 2021.**

- Augmenting the data set of recent retrofits to reflect information from recent updates to the USACE NID (including identifiers that may have changed between the 2010 NID and the 2021 NID) and the NPD Characteristics Inventory, which is based on the 2019 USACE NID data and other recent data products data

Finally, estimated project costs (including construction and equipment, design, preparation of license application, and other costs) were obtained for 25 of the 42 recent retrofit projects. In some cases, estimates were based on project costs described in FERC license documents; other cases used reported costs that were published in articles from industry journals or project websites. Although many projects had estimates of partial costs (e.g., estimates for design or construction), only estimates that were described as "total" or "overall" were used. There is some uncertainty in these cost estimates because they

[11] elibrary.ferc.gov

may have been based on projected costs (not reflecting any changes in equipment or plans) or they may not include all costs associated with the hydropower retrofit, even though they are described as overall costs. Project costs were not available at 17 of the recent NPD retrofits.

2.2 Summary Statistics

Of the 42 unique retrofits (totaling 588 MW in added installed capacity from 2000-2021), 6 are not represented in the NID, and most of those projects are previously powered dams or mills. Nevertheless, the projects are representative of powering NPDs. **Thus, data for a total of 42 projects were used for *general* retrofit assessment.**

Given that the data used for prior analysis (Hansen et al. 2021) included categorization of 36 NPD retrofits based on a retrofit method, this subset of projects is useful for further analysis. The retrofit method categories (around, over, or through) refer to the general manner in which the water conveyance to the generating units was modified, as described in detail by DeNeale et al. (2022). To briefly summarize, an *around* method would require bypassing water or constructing conveyance infrastructure from the impoundment, around the structure, and directing it to the generating units downstream. In contrast, *over* configuration refers to siphon or use of an existing spillway, and *through* refers to using existing outlet works or adding new outlet works through the main structure. **Thus, data for a total of 36 projects were used for more *detailed* retrofit assessment.**

The following plots show relevant summaries of NPD retrofit characteristics based on available data. Data which were missing are represented by "NA" in the plots. Given the limited sample size, it is not appropriate to conduct a clustering analysis on the NPD retrofits alone. However, by summarizing this information, some of the characteristics of recent NPD retrofits can be generalized.

The general retrofit assessment plots (black and gray bars) include data for the 42 total NPD retrofits from 2000–2021, and the detailed retrofit assessment plots (blue and orange bars) include data for a subset (i.e., the 36 total NPD retrofits for which retrofit approach data were collected). Additional efforts to characterize the remaining 6 retrofits may occur in subsequent research activities beyond this report. Based on the plots, the following observations can be made:

- General retrofit assessment (*n* = 42):
  - 67% of the 42 recent retrofits became operational from 2010 to 2019 (Figure 2).
  - No predominant spillway type (evaluated in Oladosu et al. (2021) was observed (Figure 3).
  - 50% of the 42 recent retrofits were classified as high-hazard dams[12] (Figure 4).
  - 20% of the 42 recent retrofits have a lock or are a lock and dam facility.
- Detailed retrofit assessment (n = 36):
  - 81% of the 36 recent retrofits followed a through approach, and 19% followed an around approach. None followed an over approach. Most through retrofits went through an existing intake or outlet work (Figure 5).
  - 50% of the 36 retrofits were federally owned,[13] though they were all developed by a nonfederal entity. An additional 28% were privately owned. Of the 7 through retrofit approaches, 4 were at federally owned dams.

[12] "Dams assigned the high hazard potential classification are those where failure or mis-operation will probably cause loss of human life." https://www.ferc.gov/sites/default/files/2020-04/fema-333.pdf (Accessed September 28, 2022). This percentage compares with only 15% of the NPD population which is classified as high-hazard, according to the 2021 USACE NID.

[13] This percentage compares with only 3.5% of the NPD population which is federally owned, according to the 2021 USACE NID.

- The timeline from application to operation varied across retrofits; through approaches had generally slightly shorter timelines than around approaches.
- Although a positive correlation exists between dam storage volume and dam height, it is not linear, and some distinct patterns were observed in the retrofit approach. Around retrofits were concentrated toward the lower end of the range of dam height but spanned the full range of dam storage compared with through dams; this lower height range for around retrofits may be attributable to the reduced civil works cost to accomplish the water conveyance at shorter dams.

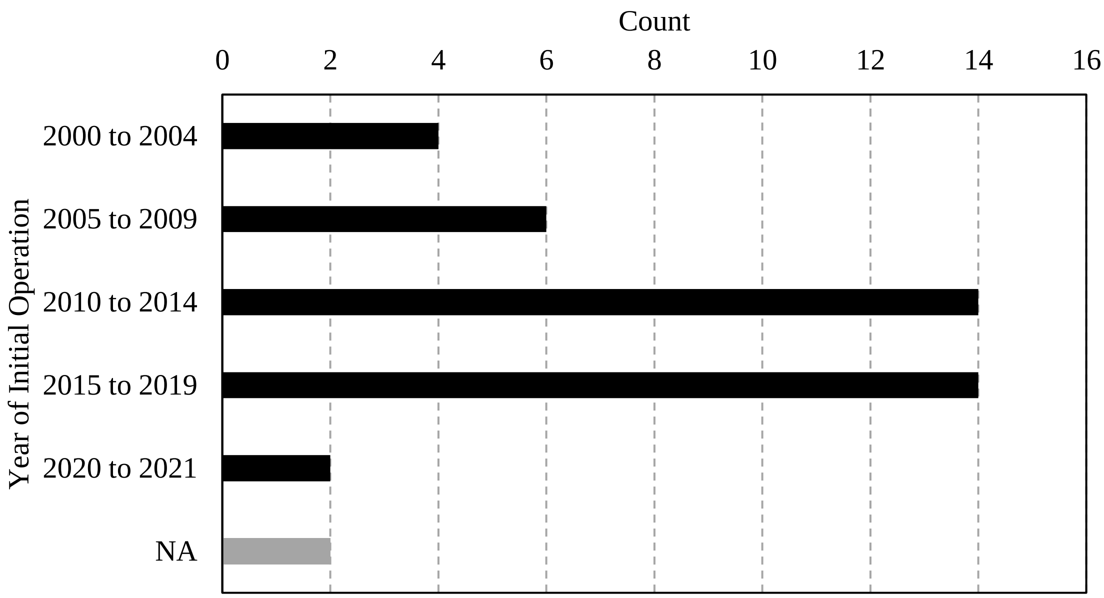


Figure 2. NPD retrofits by year of operation. Data source: Martinez, Johnson, and Shan (2021).

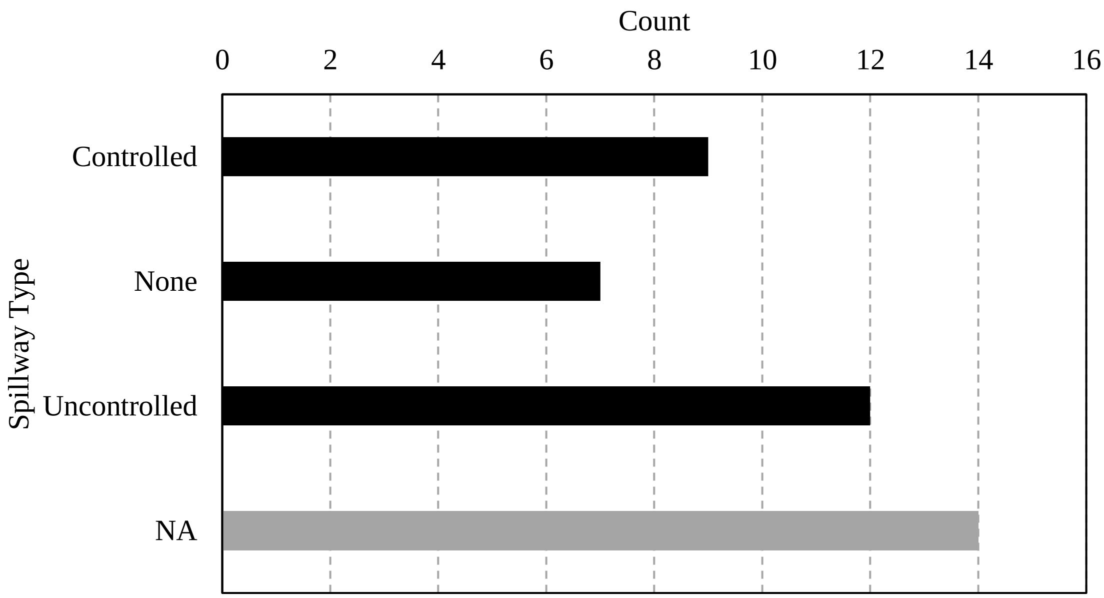


Figure 3. NPD retrofits by spillway type. Data source: US Army Corps of Engineers (2021).

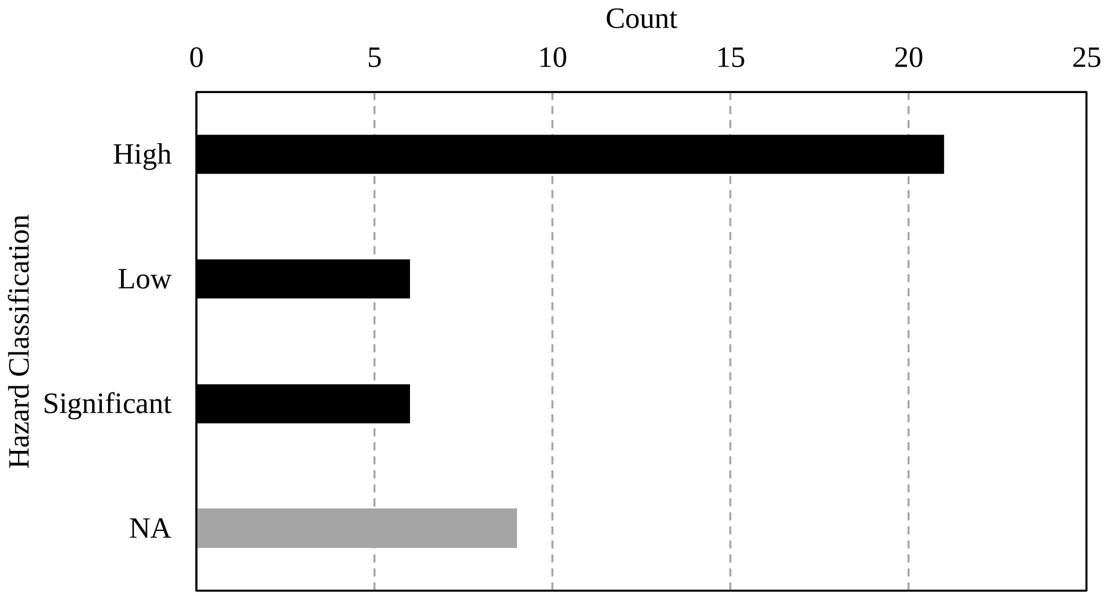


Figure 4. NPD retrofits by hazard classification. Data source: US Army Corps of Engineers (2021).

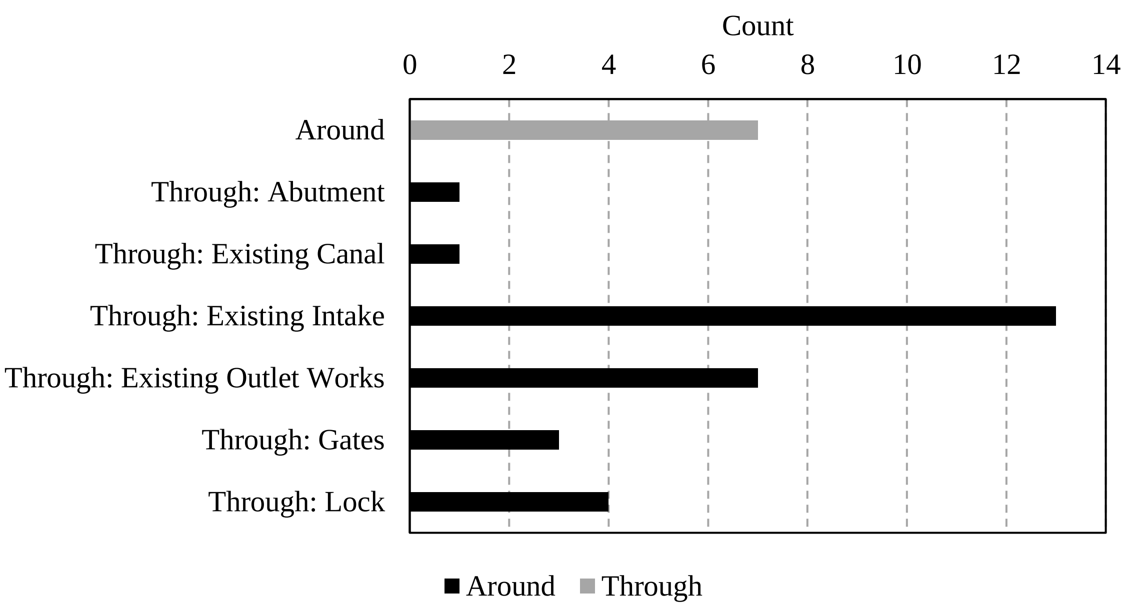


Figure 5. NPD retrofits by retrofit method. Data source: ORNL.

In addition to the general and detailed retrofit assessments, a high-level review of retrofit costs was conducted for 25 retrofits for which cost data were collected; 19 are federally owned, and 6 are nonfederally owned. As shown in Figure 6, retrofit cost is highly positively correlated ($R^2 = 0.80$) with project size (capacity in megawatts). This trend is consistent with other literature on hydropower costs (O’Connor et al. 2015a; O’Connor et al. 2015b). A distinction of federal versus nonfederal costs is included in the plot, which shows that 13 of the 14 retrofits above 5 MW were federally owned.

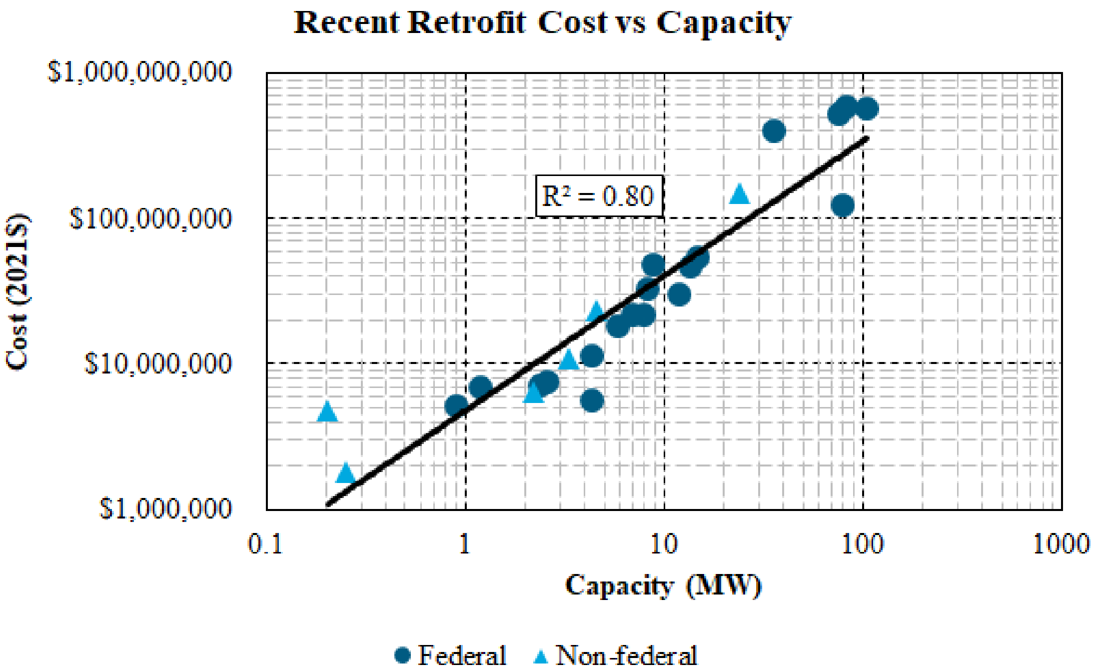


Figure 6. Recent retrofit cost versus capacity by owner type. Data source: ORNL.

3. ANALYSIS OF NPD DEVELOPMENT OPPORTUNITIES

Based on the review of recent NPD retrofits and data availability, development opportunities were evaluated for remaining NPDs through the lens of pre-feasibility development decisions. A multitude of factors have been found to influence the success and timeline of an NPD retrofit (Hansen et al. 2021). Additionally, data describing NPDs have been periodically updated. The following content describes the steps conducted to collect and summarize available data, perform a clustering analysis that characterizes dams and describes differences between dam clusters, and estimate pre-feasibility costs for these different types of dams.

3.1 Data Collection

The NID catalogs locations and key characteristics of dams in the United States and was used to obtain information about retrofits, as described in Section 2.1. Based on information provided by state and federal agencies, the NID lists purposes served by each facility, including whether the dam supports hydropower. According to the NID Data Dictionary, these are the "current purpose(s) for which the reservoir is used" (US Army Corps of Engineers 2022). The 2012 NPD Resource Assessment used the NID from 2010 and evaluated potential power capacity at dams that did not list hydropower as a purpose. The subset of NPDs identified through that assessment with estimated potential capacity greater than 100 kW (n = 3,299) formed the starting point for this analysis. However, since the publication of this resource assessment, several updates were needed to address the following:

- **Dams that have been retrofit for hydropower generation:** The EHA data set describes currently operational hydropower plants. The EHA data set published in 2021 was joined to the NPD subset using both unique identifiers (the National Inventory of Dams ID, NIDID) and spatial joins. In total, 27 NPDs from the 2012 Resource Assessment with power potential greater than 100 kW have been linked to operational power plants.

- **Dams that have been removed:** American Rivers maintains an annually updated register of dams that have been removed (American Rivers 2021). Most removal sites include the NIDID, allowing the dam removal database to be linked to the NPD subset, which identified 88 NPDs with potential capacity greater than 100 kW that are reported as having been removed.

- **Updated characteristics and purposes of inventoried dams:** The NID itself has been updated several times since the 2012 NPD Resource Assessment. In some cases, the identifiers have changed and some structures are no longer listed, resulting in some structures from the 2012 NPD Resource Assessment being excluded in the NPD Characteristics Inventory (Hansen et al. 2022a). This NPD Characteristics Inventory is based on the NID published in 2019 and describes parameters of interest to the clustering analysis, including detailed flow statistics (i.e., percent exceedance flow) and hydraulic head based on up-to-date reported descriptions from the NID. Additionally, there are cases in which an NPD reports hydropower as a purpose, despite not being linked to a power plant. This may reflect an NPD that is an upstream diversion or storage dam or reregulation structure downstream; alternatively, it could be an error from the reporting agency that provided details to USACE. Of the 3,299 NPDs with potential capacity greater than 100 kW, at least 26 are identified as supporting hydropower by the most 2021 version of the NID, despite not being linked to power plants. These updates help provide a more accurate baseline for NPD locations and status of current NPD facilities.

3.2 Summary Statistics

Key characteristics directly used in calculating the potential power capacity of NPDs are flow, hydraulic head, and capacity factor. As mentioned in Section 3.1, some characteristics have been re-calculated as underlying data sets are updated (i.e., new versions of the NID), and new data sets support more detailed analysis (i.e., flow information based on the Dayflow dataset rather than monthly runoff). In total, 2,883 dams remain after eliminating dams that have already been retrofitted, have been removed, or cannot be joined to the inventory of characteristics. Roughly one-third are owned by local government agencies (county/city/service districts), followed by 30% that are privately owned. The remaining are owned by state (15.5%) or federal (13.5%) agencies, public utilities (5.7%), or multiple entities (cooperatively owned). Recreation is the most common primary purpose (33%), followed by flood control (21.4%) and water supply (15.1%). However, multiple purposes are reported for many of these dams (27%).

Distributions of characteristics used directly in calculating potential capacity (flow and hydraulic head) along with other NPD attributes that relate to hydropower development through operations or the relationship of NPDs and grid infrastructure (i.e., distance to the nearest substation) and are widely available are shown in Figure 7. Reservoir storage (in million cubic meters [MCM]) is related to the size of the dam and hydraulic head but provides a different perspective; storage also reflects operations on a coarse level. NPDs with no storage may be indicative of run-of-river releases. Distance to the nearest substation could be important because connections to existing grid infrastructure may lower costs and be indicative of an existing energy market/demand that could be supported by hydropower. Values for most variables are highly skewed toward low values, so they appear relatively evenly distributed around the mean values in the box plots of Figure 7 owing to the log scale. Because of the clear pattern observed in NPD retrofits where more than half are federally owned dams, retrofits are shown by owner type (federal versus nonfederal).

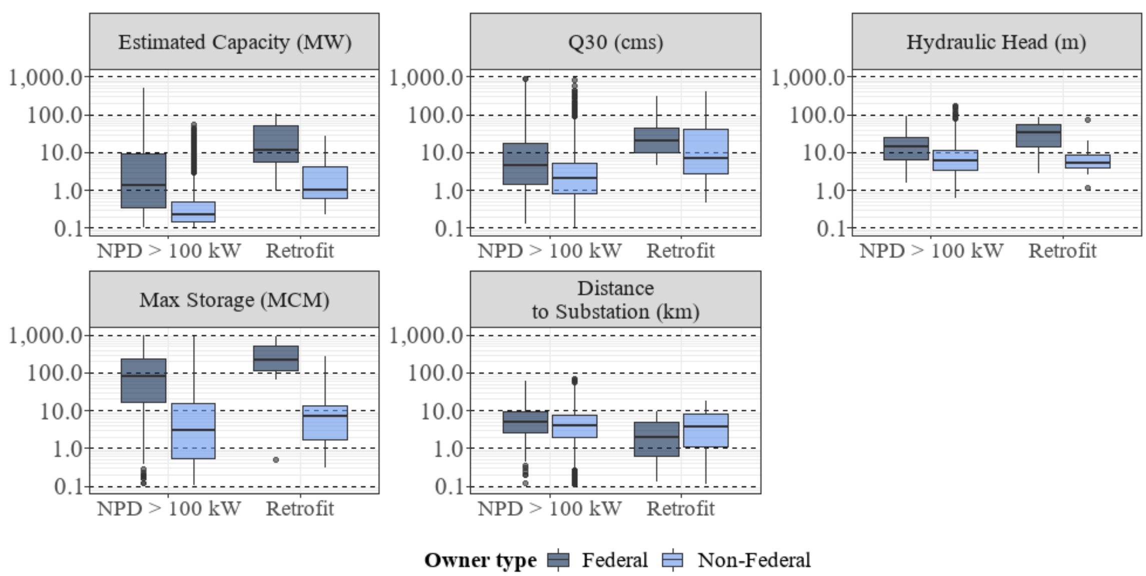
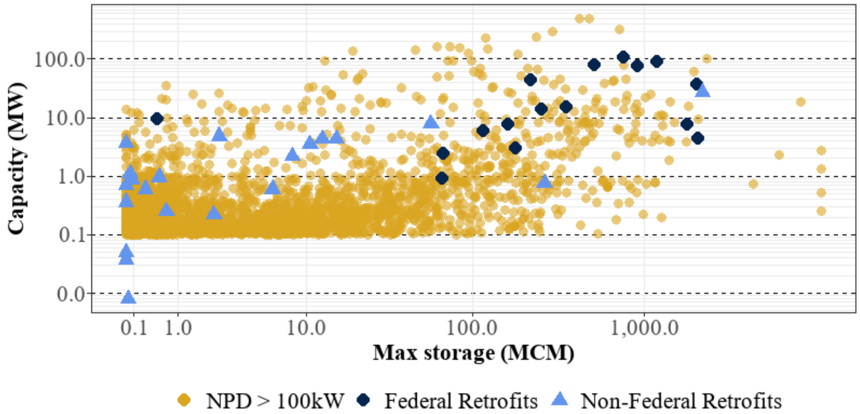


Figure 7. Distribution of key attributes related to hydropower potential, operations, and relationship to existing grid infrastructure. Note there are 172 NPDs that report a maximum storage value of 0 but are not shown in the plot. Data sources: NID and ORNL.

Comparisons of flow, head, storage, and distance distributions between federally and nonfederally owned NPDs are as follows:

- Federal retrofits have higher distributions of flow, head, and storage than federal NPDs. Thus, **federal retrofits generally have more attractive flow, head, and storage parameters compared with the remaining federal NPD population.**
- Federal retrofits have a lower distribution of substation distance (i.e., distance to the nearest substation) than federal NPDs. Thus, **federal retrofits generally have more attractive substation distances compared with the remaining federal NPD population.**
- Nonfederal retrofits have a higher distribution of flow and storage than nonfederal NPDs. Thus, **nonfederal retrofits generally have more attractive flow and storage parameters compared with the remaining nonfederal NPD population.**
- Nonfederal retrofits have similar distributions of head and substation distance compared to nonfederal NPDs.
- Federal NPDs have more high-flow NPDs than nonfederal NPDs. Thus, **federal NPDs generally convey higher flow rates, which correlate with higher power potential.**
- Federal NPDs have a higher distribution of hydraulic head than nonfederal NPDs. Thus, **federal NPDs generally have higher head available, which correlates with higher power potential.**

Additional patterns are observed at the intersect between capacity and descriptions of storage, and substation distance at NPD retrofits and the NPD population (Figure 8). This helps determine whether there is a connection between the size (generation) of the project and these other variables. Flow (Q30) and hydraulic were not compared because they are directly used in determining capacity. Again, because of the large share of recent NPD retrofits that are federally owned, projects are distinguished as federal versus nonfederal.



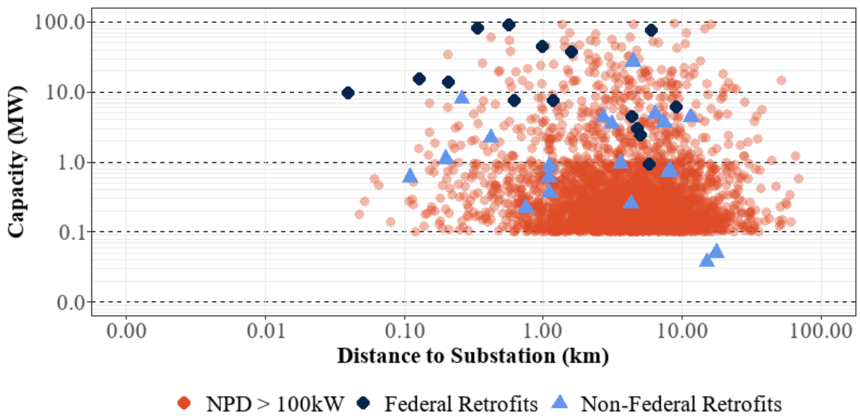


Figure 8. Estimated capacity versus maximum storage and substation distance. Data source: ORNL.

Several observations can be made from these comparisons:

- **NPD retrofit capacity and reservoir storage are positively correlated.** This is intuitive because larger reservoir storage will increase with the size of dam, which directly influences the capacity calculation.
- Reservoir storage varies widely for NPDs, spanning from 0 (this does not mean storage was not reported, but rather that there is no storage at the dam) to 10,508 MCM. Although there is no linear correlation between storage and capacity for NPDs, **NPDs with the largest estimated capacity are concentrated toward the larger end of the spectrum of reservoir storage**.
- No correlation is visible for NPDs or retrofits and substation distance. However, it is helpful to note that **substation distance is generally lower for NPD retrofits than the rest of the NPD population.**[14]
- **Federally owned retrofits were generally of higher capacity** than nonfederal retrofits. **Federally owned retrofits generally had higher substation distance** than nonfederal retrofits.
- **The remaining NPD population generally has lower capacity and longer substation distance than recent retrofits**. Most undeveloped NPDs had a potential power capacity of 100 kW to 1 MW and a substation distance of 1 to 10 km. These observations may help

[14] Given data limitations, it is not known whether substation distances for retrofits represent predevelopment or post-development information.

support the hypothesis that **power potential and substation distance may affect project feasibility**, given the distinction between the retrofits and the NPD population. However, the team could not distinguish whether the substation distance data available represents pre- or post-development distances.

3.3 Clustering Analysis

This section describes the methodology used to group NPDs into clusters. Section 3.3.1 provides an overview of the general objectives of clustering and reviews previous applications of clustering of hydropower and hydrologic resources. Section 3.3.2 details the methods of analysis. All analyses were completed using R statistical software version 4.0.3. Specifically, the package "cluster" was integral to the analyses (Maechler et al. 2022).

3.3.1 Clustering Overview

Clustering is a common technique for grouping individuals of a population based on similarities or how closely associated individuals are to one another. It is a form of unsupervised learning, which means inputs are not labelled and the algorithm does not know beforehand what the outcome should be. In other words, there is no assumption of the relationships between individuals or which individuals should be assigned to certain clusters. Many types of clustering applications exist, even within water resources, infrastructure, and hydropower research. Objectives of clustering include understanding the diversity of a population and being able to apply specific strategies or technologies depending on the needs or characteristics of a category. Examples include grouping streams based on characteristics relevant to small hydropower development considerations (Bevelhimer et al. 2018), identifying streams with similar profiles (Clubb et al. 2019), and determining which fish species have similar stressors (Pracheil et al. 2016). In each case, clustering analysis is used to determine similarities within groups in a population and dissimilarities between different groups. For this study, clustering was used to create groups of similar dams to facilitate comparisons between those NPDs that have some theoretical potential for development (at least 100 kW of potential capacity) and the NPDs that have recently been retrofit and are now generating hydroelectricity. The clusters will help describe how the population breaks down into different types (i.e., how many types of dams there are and what portion of the population belongs to each cluster).

3.3.2 Methodology

Preliminary steps for clustering analysis included data collection and preparation (i.e., formatting and filtering data), which are summarized in Section 2.1 and Section 3.1. The input data set consisted of NPDs and recent NPD retrofits and their attributes describing various physical and socioeconomic characteristics. Variables used in the final clustering analysis were selected based on several criteria:

- **Data availability:** Variables describing the dams need to be widely available, or else there is a significant loss in information. Only complete cases, or those dams that have values for all variables, can be used in the clustering analysis.

- **Relevance to hydropower development:** One of the key factors in pre-feasibility analysis is the potential power capacity, which is a direct function of the design flow and hydraulic head. There is also interest in incorporating the entities that own (and in many cases, operate) the dam and who would therefore have a major influence on the licensing process and what modifications can be made to the dam or its operations. Other variables that reflect the operations of the dam are the storage capacity, reported purpose, and presence of a lock. Additionally, proximity to existing infrastructure was hypothesized as a possible metric that could reflect development feasibility.

- **Resulting algorithm performance:** Different combinations of variables can lead to unique sets of clusters. The best clustering results are parsimonious; they can maximize how well similarities within groups are represented using the fewest meaningful groups possible.

A methodology was designed to accommodate different types of input variables (numerical and categorical) and optimize the number of clusters. It can be summarized in the following steps: calculating a dissimilarity matrix, applying a PAM (Partitioning Around Medoids) model, and evaluating the silhouette information of resulting clusters. Figure 9 outlines the specific methodology used in this study, following the general approach to unsupervised clustering outlined by Xu and Tian (2015). Table A.2 in the appendix summarizes results from various combinations of variables that were evaluated using this approach. Commentary is provided to describe the considerations that went into the final selection. The variables used in the final clustering analysis were Q30, hydraulic head, maximum storage, and dam owner type. In total, 2,709 NPDs and retrofits were complete cases or had values for all variables.

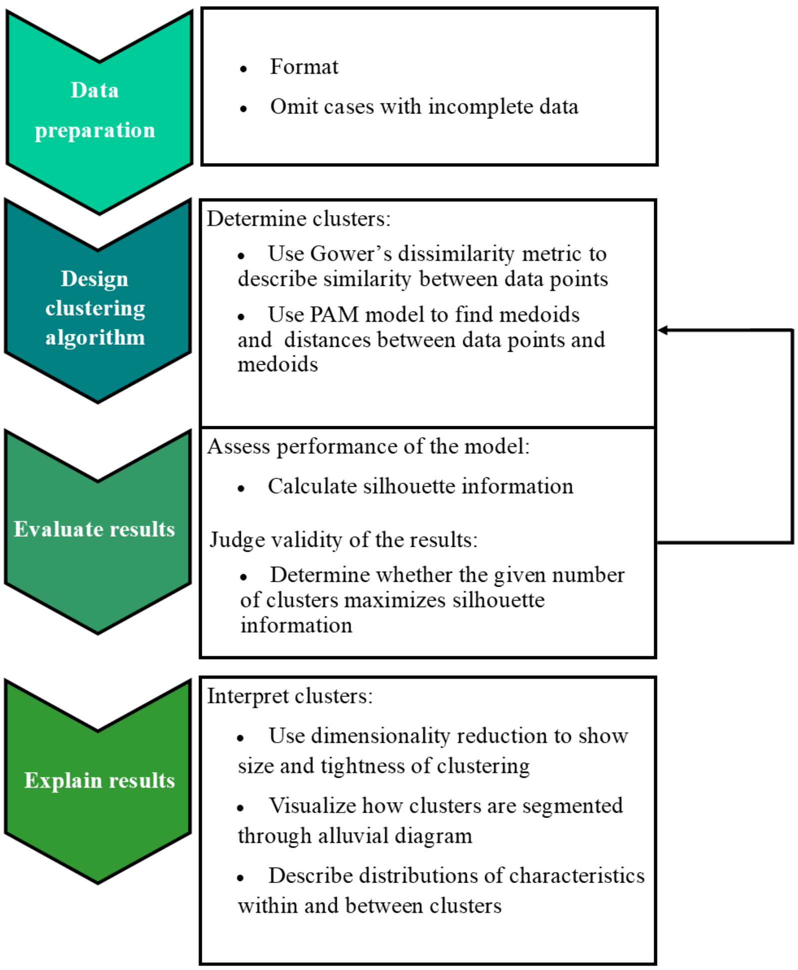


Figure 9. Clustering workflow applied to the NPDs and NPD retrofits.

3.3.2.1 Dissimilarity matrix

First, the Gower's distance was calculated, which describes the dissimilarity (or distance) between pairs of points for data that includes both numerical and categorical variables (Gower 1971). When calculated over the entire data set of NPDs, Gower's dissimilarity function produces a dissimilarity matrix of distances scaled between 0 and 1 for numerical variables and assigns either a value of 0 or 1 for categorical variables. A value of 0 indicates identical records in a pair of NPDs. A value of 1 indicates that the pair is as dissimilar as can be. The overall distance between one NPD and another is the average of the distances calculated for each variable.

3.3.2.2 PAM model

One of the most widely applied clustering algorithms is *k*-means, which groups data into k clusters by minimizing the distance between points and a mean value (i.e., the sum of squares) within clusters. Similarly, the PAM model calculates the distance to a central point but uses median values rather than mean of a cluster (Kaufman and Rousseeuw 1990). Similar to k-means clustering, PAM groups together elements that show a high degree of similarity to each other and dissimilarity to elements in other groups. However, the PAM modelling approach is generally considered more robust than k-means clustering because it is less sensitive to outliers. The input data set of NPDs and NPD retrofits has a high variability and several outliers for the numerical variables: Q30 (cms), hydraulic head (m), and maximum reservoir storage volume (MCM; Figure 10). The highly skewed distributions of numerical data illustrate why median-based clustering is important (as opposed to clustering based on means).

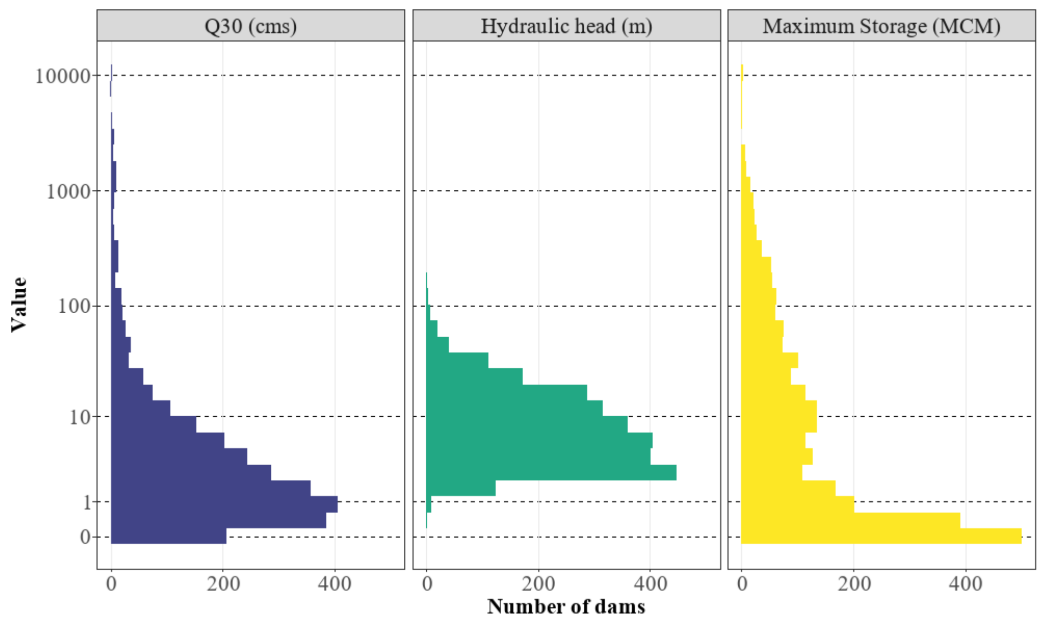


Figure 10. Histograms of numerical data for NPDs and NPD retrofits used in the clustering analysis. Each of the numerical variables are positively skewed, with tails in the distributions extending orders of magnitude greater than the median values.

First, *k* elements from the input data set were selected as medoids (corresponding to the number of clusters specified, *k*). Then, the remaining input data points were assigned to the nearest medoid and the sum of distances of all the data points to the medoids was calculated. The process was repeated with a new set of medoids until the algorithm finds a minimum sum of distances between data points and their assigned medoids.

3.3.2.3 Silhouette information

The optimal number of clusters is determined by maximizing the silhouette information, which refers to the difference between the average distance within the cluster and the mean of the next nearest cluster normalized by the maximum distance. The silhouette information values range from −1 to 1, where 1 indicates that a data point is very similar to the others in its assigned cluster, whereas −1 indicates the data point is more similar to members of another cluster. The average silhouette score for a given cluster is sometimes referred to as the *silhouette width* and is calculated from the silhouette scores of all data points assigned to that cluster. This is used to measure how "cohesive" data points are within a given cluster, whereas the overall average silhouette score across different clusters can be used to describe how well data points have been sorted (Rousseeuw 1987).

3.3.3 Results

To determine the optimal number of clusters based on owner type, Q30, hydraulic head, and maximum storage, the average silhouette width was calculated for between 2 and 10 clusters. Resulting silhouette widths for a given number of clusters are shown in Figure 11. The average silhouette width increased (indicating better grouping of the elements within each cluster) from 2 to 5 clusters and then decreased from 6 to 10 clusters.

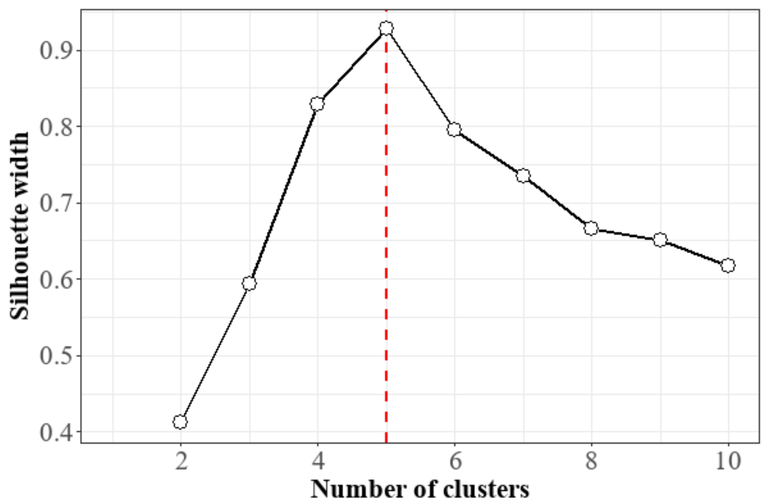


Figure 11. Average silhouette width resulting from different numbers of clusters. The red dashed line indicates that the optimal number of clusters is 5 (where silhouette width is at a maximum).

The optimal number of medoids or clusters is 5, with the greatest average silhouette width across clusters of 0.93. For the remaining analysis, NPDs were assigned to one of five clusters. Figure 12 shows the resulting silhouette information for each data point (individual NPD) in the five clusters. Average silhouette width for each cluster varied between 0.83 and 0.95, indicating that elements were appropriately clustered. Only 3 data points (out of 2,709) in all 5 clusters had a negative silhouette width, which indicates the clustering algorithm performed well and the dams are generally assigned to clusters with the other dams that are most similar to them.

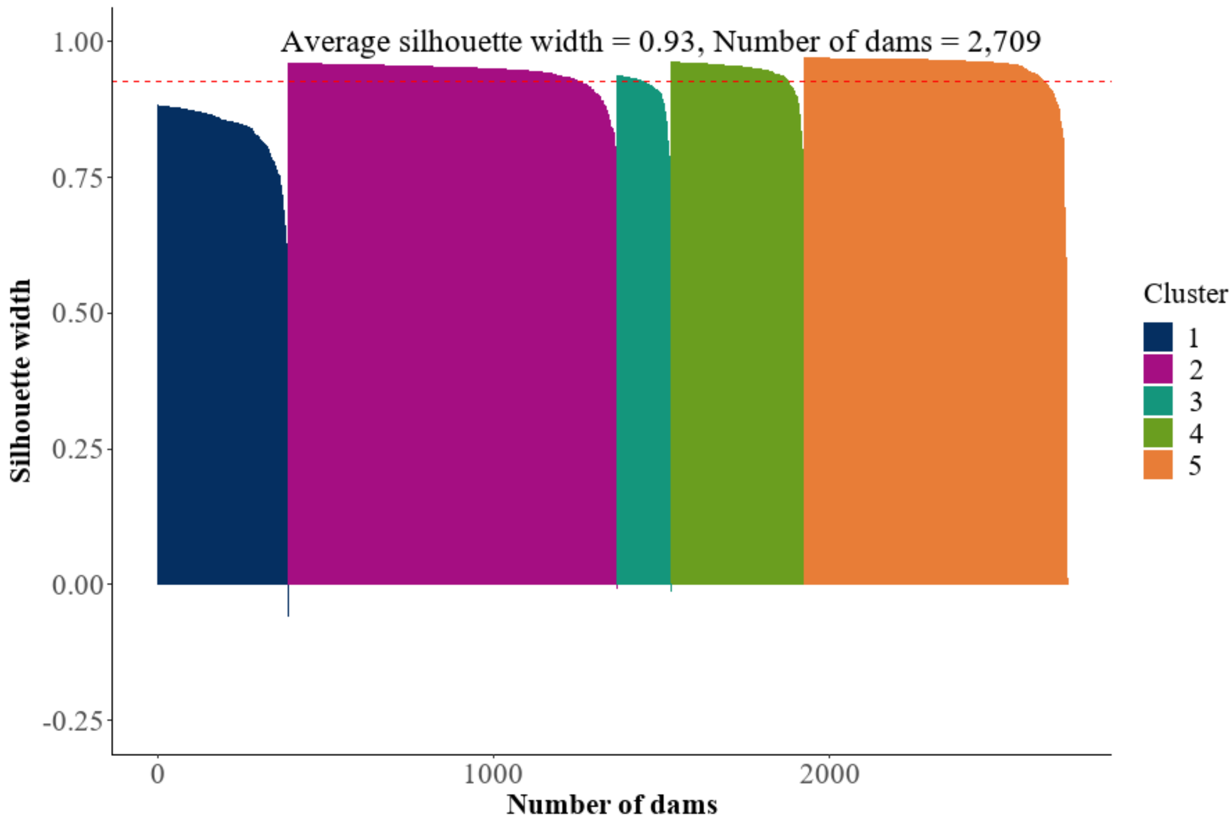


Figure 12. Silhouette information by cluster. Silhouette width for each element is shown in descending order and grouped by cluster. A silhouette width near 1 indicates the element is well clustered, whereas a smaller or negative value indicates that the element may be more similar to other clusters. The red dashed line indicates the average silhouette width across all clusters. The number of dams on the horizontal axis shows the relative size of cluster.

Sizes of clusters are unevenly distributed. Cluster 2 and 5 have the largest share of dams (978 and 783 dams, respectively), whereas Cluster 3 has 158 dams. An alluvial diagram illustrates these differences in resulting clusters by segmenting data and ranking the clusters for each variable (Figure 13). This provides a fairly comprehensive visualization of clustering results, including size of clusters and comparison between clusters across the each of the input variables. The width of each band corresponds to cluster size. Values of the dam that forms the medoid for each cluster are shown for each of the variables.

As shown in Figure 13, Cluster 1 is characterized by federally owned dams that have relatively high Q30, high hydraulic head, and large storage capacity. Conversely, Cluster 5 contains dams that are privately owned, are characterized by low Q30, hydraulic head, and storage. Although these two clusters are highly dissimilar, they contain the largest percentage of recent NPD retrofits (80.5%), even though they do not contain the largest number of dams. The largest group of dams is contained in Cluster 2, which is evident by the widest band in the alluvial diagram. Differences between the proportions of NPDs and recent retrofits are illustrated in Figure 14.

	Description	% retrofits	% NPDs
1	High flow, head, and storage federal dams, most often flood control	51.2	13.9
5	Low flow, head, and storage, privately owned, used for water supply/recreation	29.3	28.9
2	Low flow, medium head and storage, local gov, mixed purposes	14.6	36.4
3	Low flow, high head, low storage, public utility, mixed purposes	4.9	5.9
4	Low flow, medium head, large storage, largely unknown or recreation	0	14.8

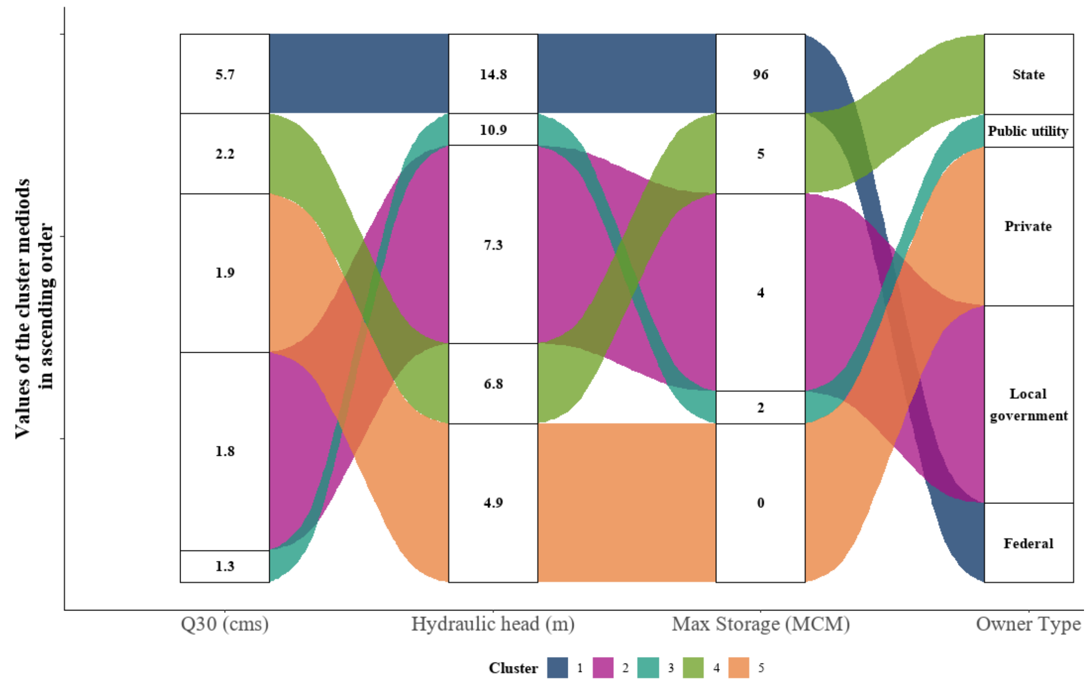


Figure 13. Alluvial diagram of clusters with the median value of the cluster for each variable. Numerical categories are shown in ascending order (increasing from bottom to top). Note that the median of the maximum storage for Cluster 5 is 0, which is commonly reported for dams with no practical storage. The table lists clusters in order of the percentage of retrofits included in each cluster to highlight which clusters have had the greatest number of successful retrofit projects. Note the differences in the relative share of retrofits and NPDs > 100 kW potential capacity in each cluster.

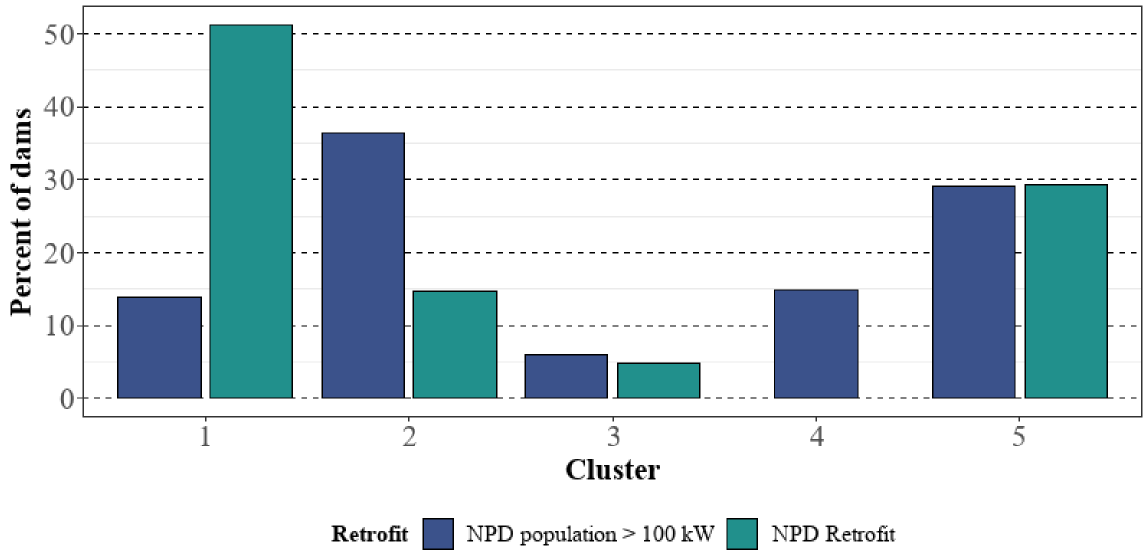


Figure 14. Distribution of NPD population and retrofits across clusters.

Figure 15 focuses on the variable that has the biggest impact on clustering results: owner type. The clusters are fairly homogenous, which reflects the outsized influence that owner type has on the clustering results. The only exception is dams that are co-owned by local governments and private entities; these dams are distributed across the five clusters, but no more than two are in any cluster (<1% of any one group). The largest share of dams are owned by local governments. This may include counties, municipalities, and/or special service districts. The next largest clusters are privately owned, and then nearly the same share are owned by federal agencies and state governments. When considering the distribution of retrofits, there are clear differences in patterns of ownership between the general NPD population and dams that have been successfully retrofit. More than half of the NPDs that were retrofit in the past 20 years are federally owned, even though this is a relatively small portion of NPDs. In contrast, despite state-owned dams being roughly the same proportion as federal dams, no state-owned NPDs have been recently retrofitted.

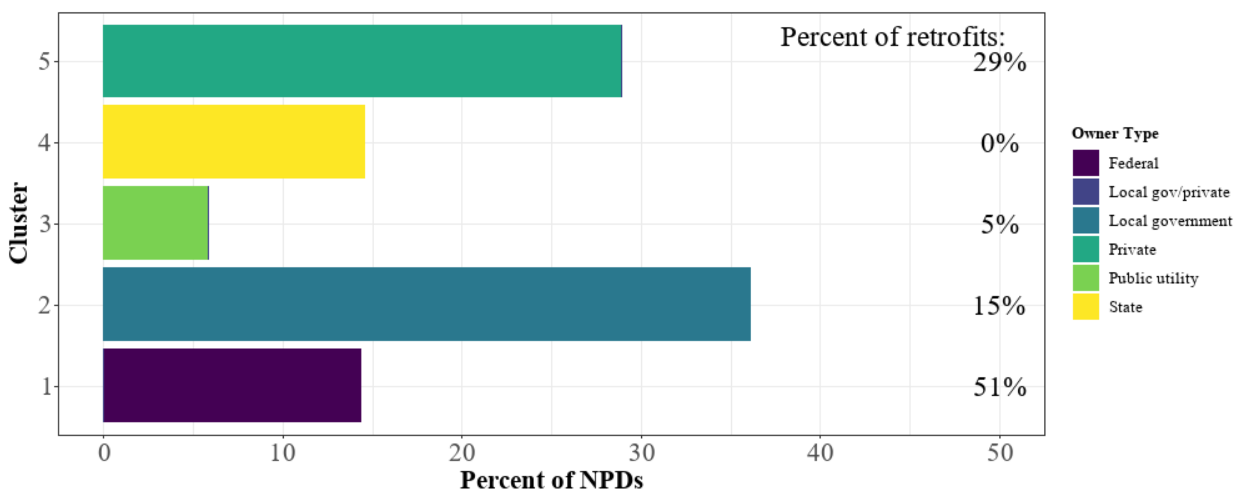


Figure 15. Owner type by cluster. Clusters are generally homogenous with respect to owner type. Note that less than 1% in any category are co-owned by a local government and private entity.

Additional comparisons of the distribution of clustered dams between the NPD population and the recent NPD retrofits provide insight into patterns of development. Results are shown from a variety of perspectives, with the following observations:

- Based on Figure 14, the majority of NPD retrofits fall into Clusters 1 and 5, whereas the majority of the entire NPD population falls into Clusters 2 and 5. Based on Figure 15, Cluster 1 is predominantly federally owned dams, Cluster 5 is predominantly privately owned dams, and Cluster 2 is predominantly dams owned by the local government. Thus, **most recent NPD retrofits are federally or privately owned, whereas most of the NPD population is privately owned or owned by the local government**.
- Figure 16 and Figure 17 illustrate that **clusters are not homogenous when it comes to primary purpose or whether there is a lock.** In most clusters, there is a higher percentage of dams used primarily for navigation or that have locks in the NPD retrofit subset compared with the subset of remaining NPDs.
- Figure 18, Figure 19, and Figure 20 show the distribution of NPD population across three variables: Q30, hydraulic head, and maximum storage, respectively. Main observations include the following:
    - **Cluster 1 NPDs are associated with higher flow estimates, and all other clusters have roughly the same distribution of flow;**
    - **Cluster 1 NPDs are associated with higher hydraulic head estimates; and**
    - **The distribution of maximum storage is highly dependent on the cluster.**
- The **general takeaways** from this analysis include the following:
    - Owner type appears to play a major role in differentiating NPDs, with nearly all NPD clusters consisting of a single dam owner type.
    - Cluster 1 NPDs have higher flow and head distributions compared with other clusters. This observation is supported by the summary statistics presented in Section 3.2.
    - Differences in substation distance between clusters are statistically significant, but in practical terms, the range in median values for the clusters is less than 2 km. Because of this small variation, ownership (which is very cluster-specific) does not appear to be a differentiating factor for substation distance among the NPD population.

These represent the main takeaways from the clustering analysis. Additional observations may be made from the plots provided. For example, the comparison of distributions by cluster show which clusters are most similar to each other. Additionally, the distributions by percentile highlight what the extreme values are for each cluster (i.e., the values at low and high percentiles). This information is also insightful when considering the pre-feasibility cost estimation, as presented in Section 3.4.

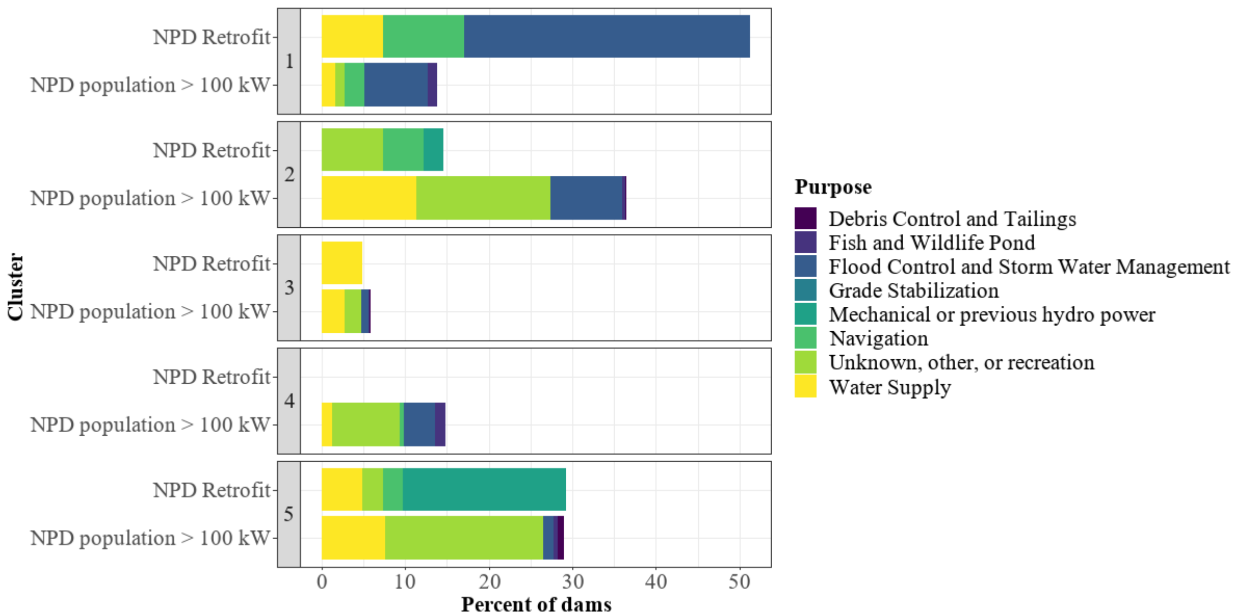


Figure 16. Distribution of NPD population and retrofits across owner types.

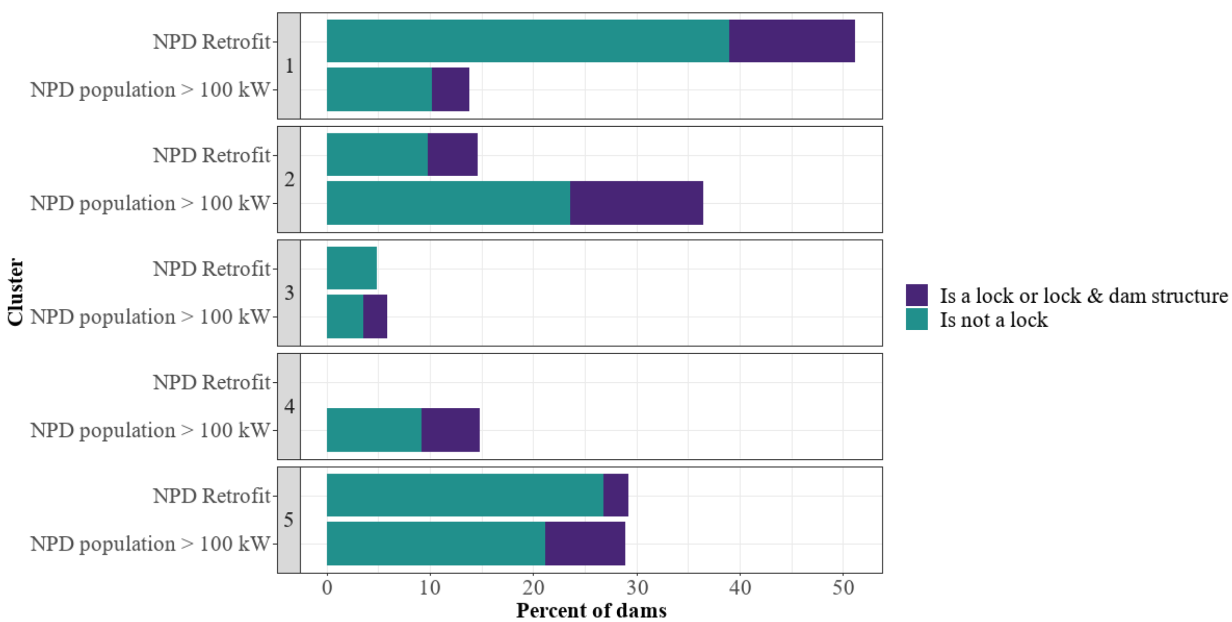


Figure 17. Distribution of NPD population and retrofits by presence of a lock.

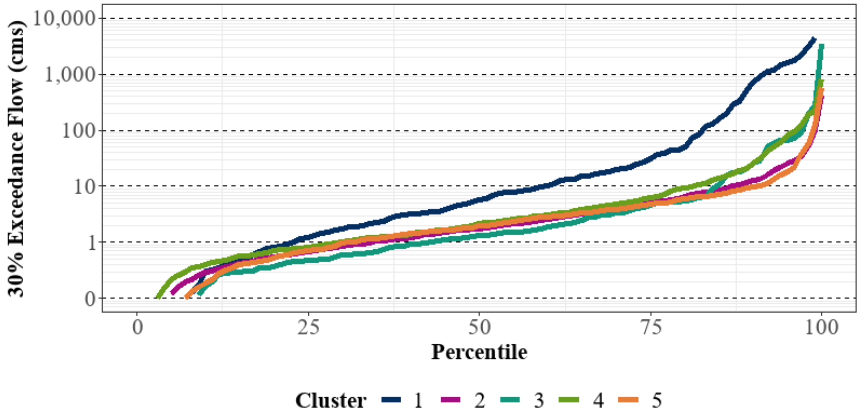


Figure 18. Q30 distribution across clusters for the NPD population.

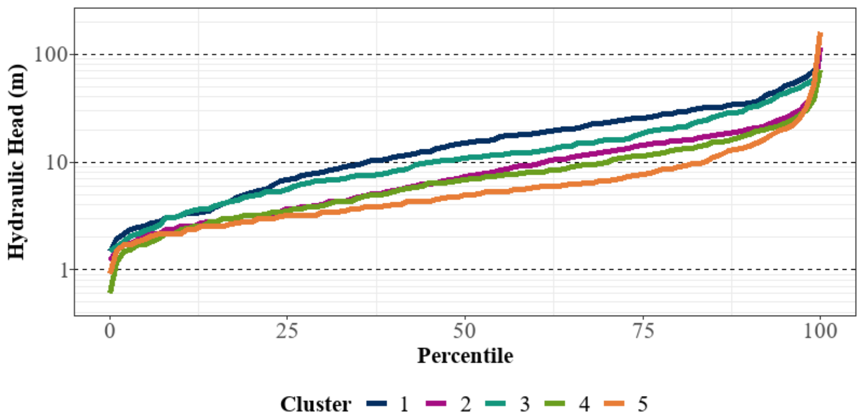


Figure 19. Hydraulic head distribution across clusters for the NPD population.

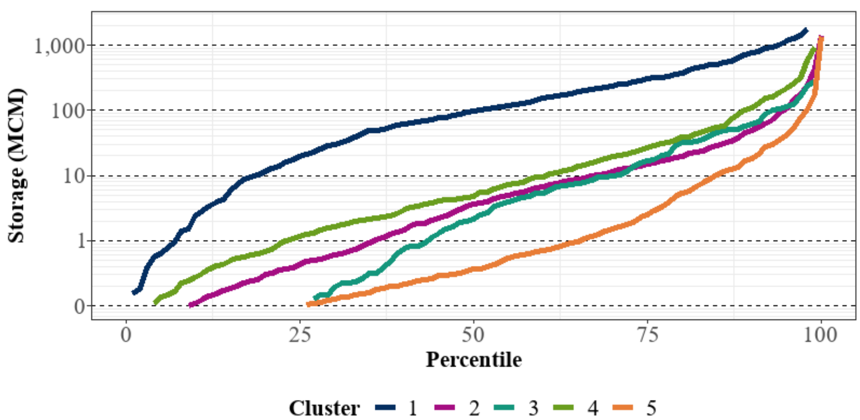


Figure 20. Storage distribution across clusters for the NPD population.

3.4 Pre-Feasibility Cost Estimation

3.4.1 Methodology

Pre-feasibility cost estimates for the sites evaluated in this study were extracted from the results of a recent analysis of project costs for the 50,000+ sites included in the 2012 NPD Resource Assessment (Hadjerioua et al. 2012). The cost analysis was performed using a reduced-form model of NPD hydropower costs (Oladosu 2022), which consisted of a set of parametric equations developed from detailed cost simulations for 20 reference sites by Oladosu et al. (2021). The equations determine the design flow, head, plant capacity, and plant cost components for a given site. These equations can be combined with available data to evaluate NPD designs and costs for many sites at a time, reducing the significant resources required to perform detailed cost simulations for hundreds of sites. Application of the model to a given site begins with identifying the reference sites that best match its infrastructure and water resource data. Thus, the reference sites serve as templates or archetypes that embed key information on dam infrastructure, water resource, and design data for NPD hydropower.

Estimates from the reduced-form NPD cost model depended on both the equations and the available site data. In addition, the model equations were based on specific baseline technology choices for the reference sites. The resulting pre-feasibility costs estimates, while providing greater insights than previously possible, do not capture NPD project costs that depend on actual construction requirements, technology choices, site conditions, and other site-specific data, but do provide a starting point for further analysis with such information. The estimates presented in this study were based on infrastructure information from the NID database and monthly flow/head estimates included in the 2012 NPD Resource Assessment (Hadjerioua et al. 2012). Future efforts will incorporate improved data on the input variables. A public version of this model is under development that would allow users to provide exogenous inputs for design variables, such as design flow and head.

3.4.2 Cost Estimates and Comparison Across Clusters

Cost estimates are provided here for the NPD population analyzed in this study. As shown in Figure 21 and Figure 22, NPDs with higher capacity are associated with lower costs, both in terms of capital expenditures per kilowatt (CapEx, $/kW) and levelized cost of energy (LCOE, $/kWh). There is a strongly positive correlation between CapEx and LCOE, meaning that projects with lower CapEx also have lower LCOE, and vice versa. This relationship is consistent with other hydropower and energy cost analyses (O’Connor et al. 2015a; O’Connor et al. 2015b), which demonstrates the impact of economies of scale (i.e., larger projects are generally more economically feasible than small projects).

CapEx for NPDs ranges from $1,840/kW to $52,852/kW, with a median value of $13,794/kW. LCOE for NPDs ranges from $0.04/kWh to $0.73/kWh, with a median value of $0.27/kWh. Figure 21 and Figure 22 show the distribution of CapEx and LCOE, respectively, among the six clusters identified in Section 3.3.3. As expected—given the distribution of project size—Cluster 1 (large, federally owned NPDs) had the lowest distribution of CapEx and LCOE. For Cluster 1, the median values of CapEx ($9,505/kW) and LCOE ($0.16/kWh) were 30%–40% lower than the general NPD median. Distributions of CapEx and LCOE for other clusters (nonfederal) are higher and relatively similar to each other.

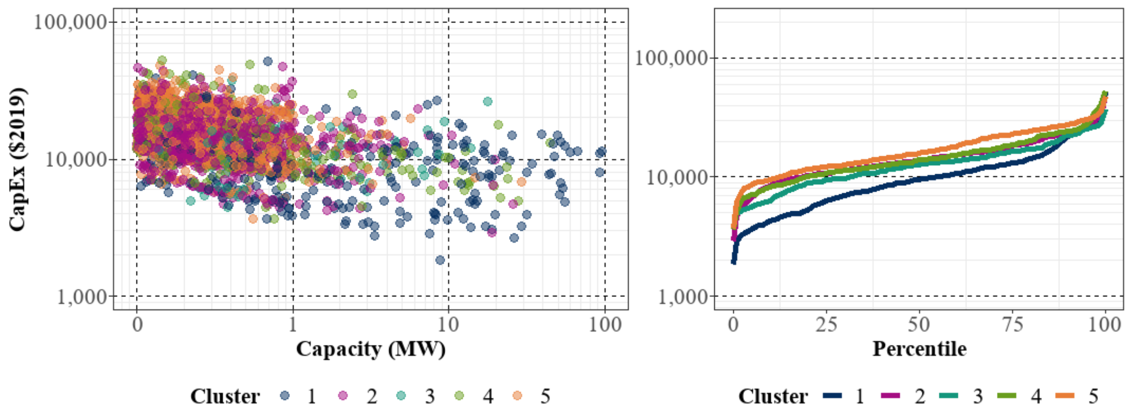


Figure 21. Estimated CapEx versus capacity for the NPD population and distribution of CapEx by cluster.

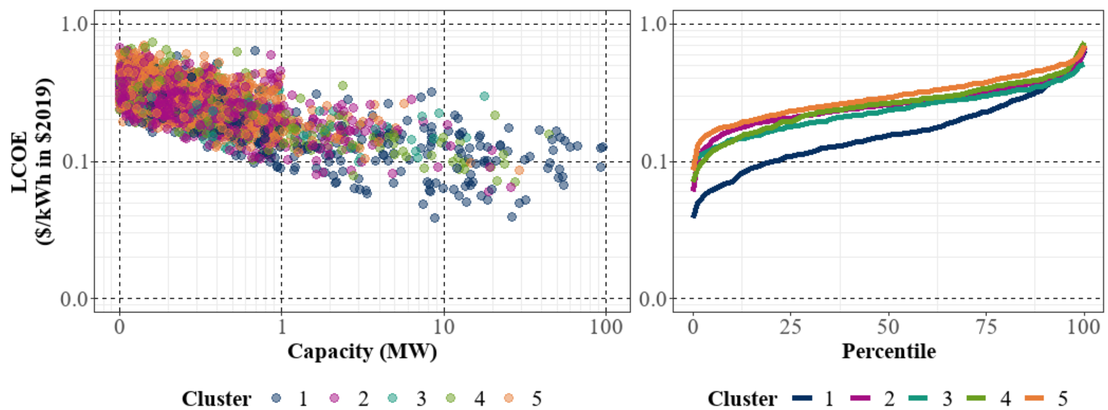


Figure 22. Estimated LCOE versus capacity for the NPD population and distribution of LCOE by cluster.

Figure 23 and Figure 24Figure 24. Cluster distribution for LCOE and CapEx among top 100 NPDs ranked by capacity. NPDs are limited to the 100 dams with the greatest capacity (MW). Note that CapEx and LCOE were not available for 27 of the top 100 dams ranked by capacity.

show the distribution and spread of the top 100 NPDs, ranked by CapEx, LCOE, and capacity, across clusters. Consistent with the other comparisons, Cluster 1 dominated the subset of lowest-cost (i.e., most economically feasible) NPDs, indicating that the majority of high-feasibility NPDs are federally owned. Although the largest portion of these high-capacity and lowest-cost NPDs are within Cluster 1 (64%–78%, depending on which metric used to rank dams), 5%–13% of the 100 dams with the greatest CapEx, LCOE, or capacity are in Cluster 4, which has no recent NPD retrofits.

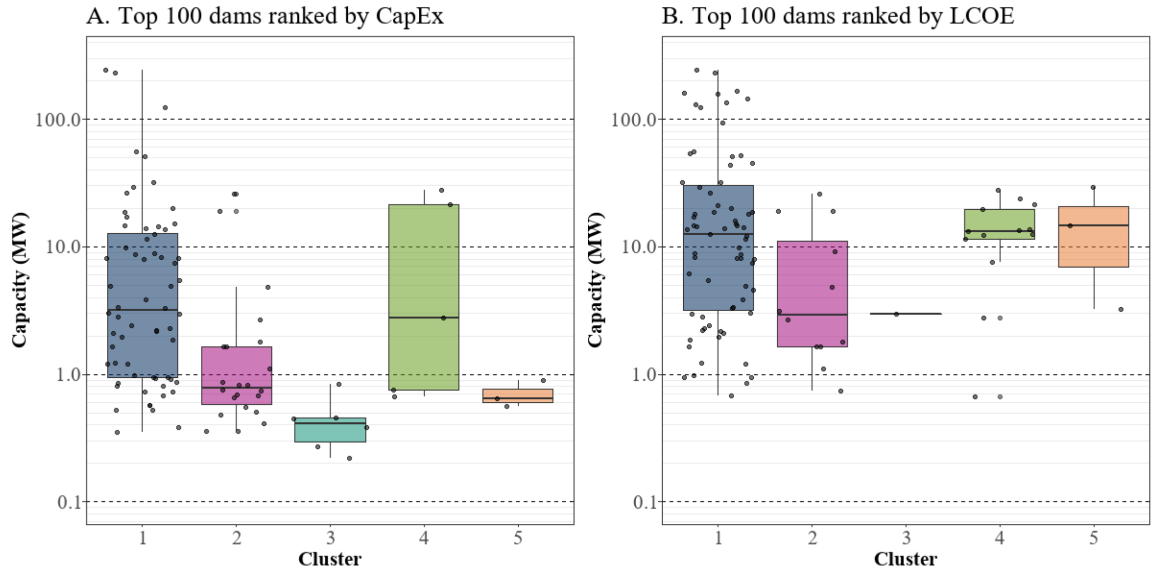


Figure 23. Cluster distribution for capacity among the 100 NPDs ranked by (A) CapEx and (B) LCOE. NPDs are limited to the 100 dams with the lowest CapEx or lowest LCOE ($/kWh, $2019). Capacity ranges from 0.22 to 241.7 MW for dams in (A), and from 0.67 to 241.7 MW for dams in (B).

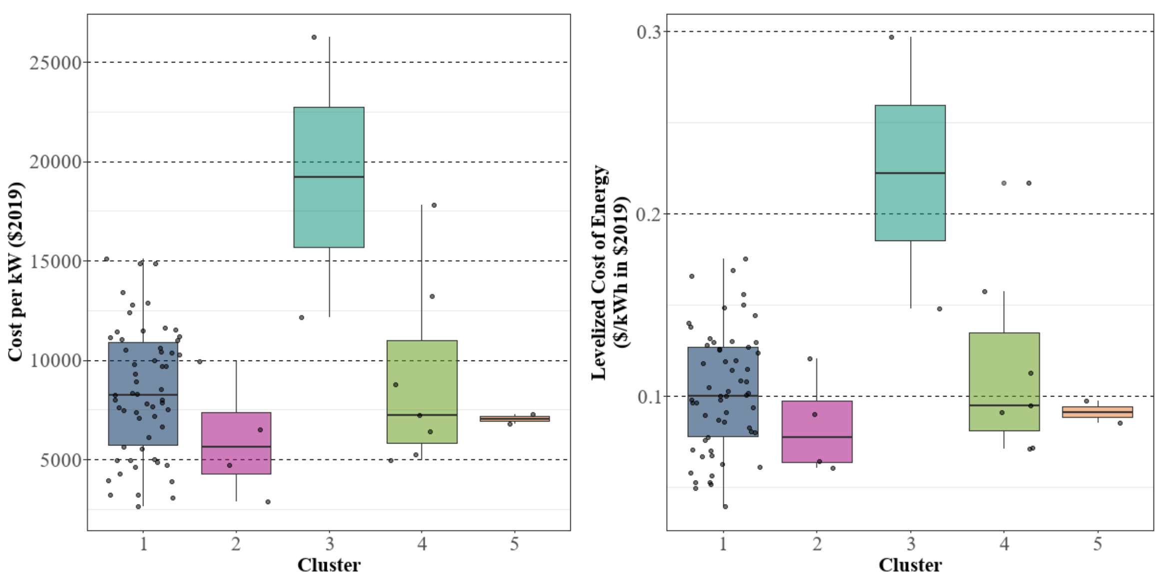


Figure 24. Cluster distribution for LCOE and CapEx among top 100 NPDs ranked by capacity. NPDs are limited to the 100 dams with the greatest capacity (MW). Note that CapEx and LCOE were not available for 27 of the top 100 dams ranked by capacity.

4. SUMMARY AND DISCUSSION

The results of the clustering analysis provide a valuable demonstration of the types of opportunities for hydropower development at existing NPDs. The methodologies used in the clustering analysis follow a typical workflow (Xu and Tian 2015) to create data-driven groupings of NPDs. Section 4.1 reviews challenges encountered during the analysis and limitations that need to be considered when interpreting results. Section 4.2 provides a brief summary of the analysis of NPD retrofits and clustering of NPDs.

4.1 Challenges and Limitations

Although clustering is a useful method for examining characteristics and types of dams, there are several significant limitations to the clustering results and their application to development decisions. First, dam clusters were determined by only four characteristics. These four characteristics were chosen because of data availability and the amount of information they contain. Flow and hydraulic head are closely related to theoretical potential capacity; owner type may serve as a proxy for the support or funding mechanisms that could promote hydropower development; and substation distance may reflect existing infrastructure and ease of connecting to the grid. Despite the justification for focusing on these characteristics, many other factors would likely be useful in reflecting the type(s) of dams that have high development feasibility. These factors are often difficult to incorporate in a data-driven approach, such as public perception or local support for dam retrofit, which has not yet been measured or described in a consistent way that would apply to a national data set. However, numerous other characteristics that could result in favorable conditions for certain dams (e.g., a site-specific need for improving fish passage combined with other structural or environmental rehabilitation needs), but increasing dimensions is not without penalty. Unless the variable does not share information with other variables already used (e.g., storage capacity, which is correlated with hydraulic head), higher dimensionality can artificially result in closer distances (less dissimilarity) between data points.

Still, each of these numerical characteristics is imperfect and carries some amount of uncertainty because of the source data or assumptions used to produce estimates. First, Q30 is based on simulated flow records from limited stream gauges, which may not represent all the current release constraints that are in place for an existing dam. Second, hydraulic head estimates used in this clustering analysis are rough approximations of general conditions based on physical characteristics of the dam. In reality, hydraulic head can vary seasonally or annually. The simple relationship used to estimate hydraulic head from NID-based dam height has been used in previous studies (Hadjerioua et al. 2012) but has not been validated across the different types of dams in the NPD subset. Finally, the Q30 and substation distance both rely on spatial joins to other data sets to determine which river reach and substation are nearest. This means that errors in the georeferencing of the NPDs or errors in locations of the river network and substations could lead to inaccurate calculated distances or assigned flow values.

There are also uncertainties related to substation distance. These distances were calculated based on Euclidean distance to the nearest substation locations provided by the Homeland Infrastructure Foundation-Level Data. First, this does not reflect geographic/topographic obstacles that may complicate the connection between an NPD and the substation. Furthermore, this may reflect substations that were built as part of the NPD retrofit project. Additional information about the grid is needed to know if there is appropriate capacity at the substation and within the transmission.

Despite the relative availability of data for the selected characteristics, some dams previously identified in the 2012 NPD resource assessment do not have current information (i.e., attributes are not reported or identifying information has changed so the dams could not be found in the most recent characteristics inventory). In total, 214 dams that had been identified as having >100 kW potential capacity fall into this category and were not included in the clustering analysis. These are mostly privately owned or state-owned dams, and they have a total estimated potential capacity of 197 MW (roughly 2% of the total capacity of dams that did contain all the necessary data to be included in the clustering analysis). Improvements in data coverage may help in characterizing these remaining dams and identify other NPDs that either have large capacity or share other characteristics (but were not previously included in the 2012 NPD resource assessment).

4.2 Comparison with the Population of NPDs

The majority of successful retrofits in the past two decades have been dominated by large federal dams with high flow, hydraulic head, and reservoir storage. The makeup of retrofits is disproportionately concentrated in this type of dam relative to the remaining population of NPDs. Although federal dams are 51% of the retrofits, they are only 13% of the NPDs with >100 kW potential capacity and less than 4% of all NPDs. However, these are not the only significant group of retrofits. Nearly one-third of the retrofits are privately owned dams with low flow, low head, and little-to-no storage. In fact, 30% of the NPDs with >100 kW potential capacity and 63% of all NPDs are privately owned, and most have very low head (50% of all NPDs have a reported head <5.1 m).

Ranking dams by cost and capacity emphasized the dominance of larger, federal dams among dams that may be suitable for more detailed feasibility studies. However, the fact that other types of dams were included in the top 100 subsets indicates that opportunities for NPD development are diverse. In particular, there may be opportunities that have little precedence (i.e., no retrofits fit into the group of state-owned dams with low flow and high storage). Because this group of dams lacks examples of successful retrofits but makes up a significant portion of the NPD subset by percentage (15%) and share of the total capacity of NPDs with >100 kW potential capacity (8%), it may be worth exploring what obstacles are preventing development.

5. OPPORTUNITIES AND NEXT STEPS

The clustering analysis described in this report highlights clear patterns in NPD data – for both successfully completed projects and projects that have been identified as having >100 kW potential capacity. This information adds to the general understanding of the types of facilities that exist and which conditions or characteristics tend to lead to more successful development outcomes, but it remains far from a final solution or exhaustive assessment of which existing facilities can or should be developed. Further work is needed to:

- explore other factors that contribute to project feasibility such as funding or regulatory policies,
- consider how projects might be prioritized with respect to metrics other than project costs,
- improve representation of key characteristics by increasing levels of detail and communicating variability or ranges of circumstances where development would be feasible, and
- increase accuracy and validating the input information to ensure physical, operational, environmental, and socio-economic characteristics of retrofits and NPDs reflect reality.

As described in Section 4.1, a variety of other factors may play significant roles in the feasibility or success of NPD development, and many of these factors are not yet described at the national level. For example, to address the influence that public perception has on NPD project success, a dedicated study in the social sciences is needed to properly evaluate this factor. Additionally, the major focus thus far has been to examine physical and management-related characteristics that would hypothetically result in fewer barriers to development. However, energy and water security in areas that may be underserved by water and energy resource management may be an equally, if not, more important motivating factor for NPD development. Further research is needed to connect existing NPD data to information related to climate, energy, and environmental justice, which could identify facilities that have increased motivation and justification for development beyond maximizing profits.

Validation of assumptions (e.g., using a single value for hydraulic head based on the reported dam height and assuming this is representative of the capacity at the dam) is also necessary. This particular characteristic may be further enhanced as new data sets describing surface water level variability can be tapped for national-scale descriptions of variable hydraulic head. More nuanced hydraulic head information could be summarized and used in a similar manner as variable flow; for example, a hydraulic head duration curve could be generated to provide percent head exceedance and model potential generation throughout the year. More detailed description of this fundamental characteristic could shed additional light on the types of operations or physical characteristics that factor into hydropower development feasibility.

Further analysis of the accuracy of NPD data, including reported purposes, hydropower status, and characteristics derived from spatial proximity to other features or data sets, is needed to judge how well the NPD opportunities are being described. A formal assessment and comparison against facility records or data provided with greater detail (e.g., local or state-based inventories rather than the NID) would be a practical way to improve confidence in the larger data set of NPDs and associated analyses. Reported reservoir storage was helpful in improving clustering performance and distinguishing groups of dams. However, inventoried storage (and the large number of facilities that report no storage) may not reflect realistic operations. Detailed storage/release records could help verify storage behavior and better describe dam operations, which could have a major impact on how NPDs are viewed as candidates for hydropower retrofit. Although the clusters or types of dams may not drastically change, more detailed information would ensure that the clusters reflect the true range of physical and operational characteristics at NPDs. Greater confidence in the descriptions of characteristics across different types of dams will ultimately help approach complex design aspects of NPD retrofitting, including turbine sizing, efficiencies of operation, and foundation requirements that factor into project feasibility.

Over the past two decades, hydropower development at NPDs has been successful under a variety of circumstances and in diverse facilities/settings. This is clearly reflected in the operational NPD retrofits that were distributed across different clusters. However, the dominance of clusters 1 and 2 among NPD retrofits and NPDs with high estimated LCOE, CapEx, and capacity indicates that there are certain types of infrastructure or facilities that are more favorable than others for development. The defining characteristics of clusters and the estimated project cost metrics should not be considered promising for project feasibility or success. Rather, the identification of certain characteristics that are common among successful projects and patterns of feasibility observed across clusters can aid developers by supporting a targeted approach to further data refinement and more detailed evaluations of potential hydropower development. This may lead to a more efficient approach to development as certain projects are prioritized and detailed studies of similar types of dams can be justified by a record of project success. Follow-on investment strategies may also target specific subsets of dams and supporting data collection or analysis at facilities with limited information.

6. REFERENCES

American Rivers. 2021. "American Rivers Dam Removal Database." 2021. https://doi.org/https://doi.org/10.6084/m9.figshare.5234068.

Bevelhimer, M., C. DeRolph, and A. Witt. 2018. "Site Classification for Standard Modular Hydropower Development: Characterizing Stream Reaches by Module Need." Oak Ridge, Tennessee: Oak Ridge National Laboratory. https://doi.org/10.2172/1463991.

Carter, Forest, Scott Deneale, Chris Derolph, and Carly Hansen. 2022. "NPD Classification Tools - User Guide NPD Explorer and NPDamCAT Apps." Oak Ridge National Laboratory. https://doi.org/10.2172/1855640.

Clubb, F. J., B. Bookhagen, and A. Rheinwalt. 2019. "Clustering River Profiles to Classify Geomorphic Domains." *Journal of Geophysical Research: Earth Surface* 124 (6): 1417–39. https://doi.org/10.1029/2019JF005025.

Conner, A. M., J. E. Francfort, and B. N. Rinehart. 1998. "US Hydropower Resource Assessment Final Report." US Department of Energy Office of Energy Efficiency and Renewable Energy. https://doi.org/10.2172/1218171.

DeNeale, S., C. Sasthav, M. Musa, C. Hansen, K. Stewart, and P. Matson. 2022. "Non-Powered Dam Retrofit Exemplary Design for Hydropower Applications." Oak Ridge, Tennessee: Oak Ridge National Laboratory.

DeNeale, S. T., G. B. Baecher, K. M. Stewart, E. D. Smith, and D. B. Watson. 2019. "Current State-of-Practice in Dam Safety Risk Assessment." *Oak Ridge National Laboratory*. Oak Ridge, TN, USA.

Ghimire, G., C. Hansen, S. Gangrade, S.-C. Kao, P. Thornton, and D. Singh. 2022. "Dayflow: CONUS Daily Streamflow Reanalysis, Version 1." Oak Ridge National Laboratory. https://doi.org/10.21951/Dayflow/1847639.

Gower, J. C. 1971. "A General Coefficient of Similarity and Some of Its Properties." *Biometrics* 27 (4): 857–71.

Hadjerioua, Boualem, Yaxing Wei, and Shih-Chieh Chieh Kao. 2012. "An Assessment of Energy Potential at Non-Powered Dams in the United States." *Hydropower and Energy Potential at Non-Powered Dams*. Oak Ridge, TN, USA, USA.

Hansen, C., C. DeRolph, F. Carter, and S. DeNeale. 2022a. "US Non-Powered Dam Characteristics Inventory." Oak Ridge, Tennessee: Oak Ridge National Laboratory. https://doi.org/10.21951/US_NPD_Characteristics/1860464.

Hansen, C., M. Musa, C. Sasthav, and S. DeNeale. 2021. "Hydropower Development Potential at Non-Powered Dams: Data Needs and Research Gaps." *Renewable and Sustainable Energy Reviews* 145: 111058. https://doi.org/10.1016/j.rser.2021.111058.

Hansen, C., C. Sasthav, M. Musa, and S. DeNeale. 2022b. "Non-Powered Dam Custom Analysis and Taxonomy Framework." Oak Ridge, Tennessee: Oak Ridge National Laboratory.

Homeland Infrastructure Foundation-Level Data. 2022. "Electric Substations."

Johnson, M. M., S.-C. Kao, and R. Uria-Martinez. 2022. "Existing Hydropower Assets." Oak Ridge, Tennessee: Oak Ridge National Laboratory. https://doi.org/10.21951/EHA_FY2022/1865282.

Johnson, M., and R. Uría-Martínez. 2022. "US Hydropower Development Pipeline Data." Oak Ridge, Tennessee: Oak Ridge National Laboratory. https://doi.org/10.21951/HMR_PipelineFY22/1860473.

Kaufman, L., and P. Rousseeuw. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley & Sons.

Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, K. Hornik, M. Studer, P. Roudier, et al. 2022. "Cluster R Package Version 2.1.3."

Martinez, R., M. Johnson, and R. Shan. 2021. "U.S. Hydropower Market Report (January 2021 Edition)." Oak Ridge, Tennessee: Oak Ridge National Laboratory. https://doi.org/10.2172/1763453.

O'Connor, P. W., Q. Fen, K. Zhang, and S. T. DeNeale. 2015. "Hydropower Baseline Cost Modeling." Oak Ridge, Tennessee: Oak Ridge National Laboratory.

O'Connor, P. W., K. Zhang, S. T. DeNeale, D. R. Chalise, and E. Centurion. 2015. "Hydropower Baseline Cost Modeling, Version 2." Oak Ridge, Tennessee: Oak Ridge National Laboratory. https://doi.org/10.2172/1185882.

Oladosu, G. A. 2022. "A Reduced-Form Model for Hydropower Cost Assessment at U.S. Non-Powered Dams."

Oladosu, G., L. George, and J. Wells. 2021. "2020 Cost Analysis of Hydropower Options at Non-Powered Dams." Oak Ridge, Tennessee: Oak Ridge National Laboratory. https://doi.org/10.2172/1770649.

Pracheil, B. M., R. A. McManamay, M. S. Bevelhimer, C. R. DeRolph, and G. F. Čada. 2016. "A Traits-Based Approach for Prioritizing Species for Monitoring and Surrogacy Selection." *Endangered Species Research* 31 (1): 243–58. https://doi.org/10.3354/esr00766.

Rousseeuw, P. J. 1987. "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis." *Journal of Computational and Applied Mathematics* 20: 53–65.

US Army Corps of Engineers. 2013. "Hydropower Resource Assessment at Non-Powered USACE Sites."

———. 2021. "National Inventory of Dams." 2021.

———. 2022. "National Inventory of Dams Data Dictionary." 2022.

US Department of Energy. 2016. "Hydropower Vision A New Chapter for America's 1st Renewable Electricity Source." Washington, DC. https://doi.org/10.2172/1502612.

———. 2022. "Water Power Technologies Office Multi-Year Program Plan." Water Power Technologies Office.

US Department of the Interior Bureau of Reclamation. 2011. "Hydropower Resource Assessment at Existing Reclamation Facilities." Resources Office.

Xu, D., and Y. Tian. 2015. "A Comprehensive Survey of Clustering Algorithms." *Annals of Data Science* 2 (2): 165–93. https://doi.org/10.1007/s40745-015-0040-1.

Xu, Dongkuan, and Yingjie Tian. 2015. "A Comprehensive Survey of Clustering Algorithms." *Annals of Data Science* 2 (2): 165–93. https://doi.org/10.1007/s40745-015-0040-1.

APPENDIX A. COMPARISON OF ALTERNATIVE CLUSTERING VARIABLES

Table A.1. Summary of model performance and considerations for selection of variables in clustering analysis

Variables used in clustering	Number of dams with data	Optimal number of clusters	Average silhouette width	Comments
Owner type, Q30, head, max reservoir storage	2,709	5	0.93	Relatively minimal loss of information owing to lack of data. Reported values are still subject to the accuracy of the reporting agency, and estimated values (Q30, hydraulic head) are still limited by simplifying assumptions about accuracy and that single values can adequately represent the head (which may be variable). NPD retrofits are distributed across owner type, and differences in distributions of Q30, head, and storage are more substantial than when clustered according to other categorical variables.
Owner type, Q30, head, distance to substation	2,781	5	0.87	In theory, close proximity to grid infrastructure may support or encourage hydropower retrofits at NPDs. However, the uncertainty of information (when the substation was built) make it an unreliable variable in describing retrofits. Additionally, there was not a lot of practical difference in the variability between groups (e.g., 1.5 km difference between median values of different groups).
Presence of lock, Q30, head	2,803	2	0.94	There was clearly a disproportionate number of retrofits that are locks (greater percentage of lock retrofits than of lock NPDs). However, sorting dams into two groups that each contain a sizeable portion of retrofits does not further the understanding of where feasible opportunities exist. There is very little difference between the distribution of flow and head (which has implications for design) between these two groups, so they are not very useful clusters.
Purpose, Q30, head, storage	2,803	6	0.94	Half of the clusters are much smaller than the others. Many dams are multipurpose, so the dominance of purpose places a lot of importance on information that may not be entirely accurate or able to represent the type of dam (e.g., a dam may serve purposes with equal importance, but there can—by definition—be only one purpose). This necessarily oversimplifies the type of dam.