Trade-Off Analysis Methodology to Evaluate Alternatives for Water Resource Management

Technical Memorandum M&S-2021-G1364
Manuals and Standards (M&S) Program

U.S. Department of the Interior August 2021

Mission Statements

The U.S. Department of the Interior protects and manages the Nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors its trust responsibilities or special commitments to American Indians, Alaska Natives, and affiliated Island Communities.

The mission of the Bureau of Reclamation is to manage, develop, and protect water and related resources in an environmentally and economically sound manner in the interest of the American public.

Cover Photo: Swiftcurrent Creek below Many Glacier, Montana. (Reclamation, Chuck Heinje)

Trade-Off Analysis Methodology to Evaluate Alternatives for Water Resource Management

Technical Memorandum M&S-2021-G1364
Manuals and Standards (M&S) Program

prepared by

Technical Service Center
Economic Analysis Group
Matthew Elmer, Ph.D.

reviewed by

Technical Service Center
Economic Analysis Group
Randy Christopherson

and

Mustapha Alhassan, Ph.D.

Acknowledgments

This technical memorandum was conducted under fiscal year 2021 U.S. Bureau of Reclamation Manuals and Standards (M&S) Program funding (project G1364).

Contents

Abbreviations

AHP – Analytical Hierarchy Process
ASP – Agency Specific Procedures
BRI – Benefit Relevant Indicators
DOI – U.S. Department of the Interior
D&S – Directives and Standards
EPA – U.S. Environmental Protection Agency
EQ – Environmental Quality
IDC – Interest During Construction
MADA – Multi-Attribute Decision Analysis
MCA – Multi-Criteria Analysis
MCDA – Multi-Criteria Decision Analysis
MAUA – Multi-Attribute Utility Analysis
M&I – Municipal and Industrial
M&S – Manuals and Standards
NED – National Economic Development
OM&R – Operation, Maintenance, and Replacement
OSE – Other Social Effects
P&G – Economic and Environmental Principles and Guidelines for Water and Related Land Resources Implementation Studies
P&R – Principles and Requirements for Federal Investments in Water Resources
PR&G – Principles, Requirements and Guidelines for Water and Land Related Resources Implementation Studies
PV – Present Value
QCA – Qualitative Comparative Analysis
Reclamation – U.S. Bureau of Reclamation
RED – Regional Economic Development
SDM – Structured Decision Making
SMART – Simple Multi-Attribute Rating Technique
TBL – Triple Bottom Line
UKCLG – U.K. Department of Communities and Local Government
USACE – U.S. Army Corps of Engineers
USFS – U.S. Forest Service
WSRV – West Salt River Valley

1.0 Introduction

Public benefits encompass environmental, economic, and social considerations. When evaluating the public benefits of alternatives for water resource management, the U.S. Department of the Interior (DOI) Agency Specific Procedures (ASP) for Implementing the Council on Environmental Quality's *Principles, Requirements and Guidelines for Water and Land Related Resources Implementation Studies* (PR&G) require that the U.S. Bureau of Reclamation (Reclamation) include monetary and non-monetary effects and allow for the inclusion of quantified and unquantified measures. Past assessments done by Reclamation have struggled to include both quantified and unquantified considerations together in a single analysis. That which can be measured is typically quantified and sometimes monetized, while that which is not easily measured is evaluated qualitatively and independent of quantified effects. A trade-off analysis allows for both quantified and unquantified considerations to be evaluated together, identifying the strengths and weaknesses of each alternative and indicating how alternatives perform across several considerations at once. This technical memorandum provides a framework and methodology for conducting a trade-off analysis in a systematic and consistent manner.

Reclamation is regularly tasked with evaluating alternatives pertaining to water resource management, for example, alternative reservoir operations, alternative infrastructure projects, alternative policies, and alternative adaptation strategies to address future water supply and demand imbalances. Each alternative comes with positive and negative effects, and a trade-off analysis recognizes that there are effects that are not easily quantified that are still important to consider when making decisions. The inclusion of quantified and unquantified considerations means that a trade-off analysis includes a broader range of effects than a typical economic analysis. As such, a trade-off analysis accommodates settings with multiple objectives where there is no unique solution that optimizes all objectives simultaneously, sometimes referred to as "wicked problems." A trade-off analysis is a mandatory component of Reclamation's WaterSMART (Sustain and Manage America's Resources for Tomorrow) Basin Studies. However, past trade-off analyses have varied substantially since existing guidance is general and purposefully avoids being prescriptive on methodology. This technical memorandum offers a systematic and flexible framework for conducting a trade-off analysis that is consistent and applicable to a wide range of settings and Reclamation work.

2.0 Background

2.1 WaterSMART Basin Studies

In 2009, Congress passed the SECURE Water Act directing DOI to develop a sustainable water management policy. Subsequently, in 2010 DOI established WaterSMART, which combined existing programs with new initiatives to create a broad framework for managing the country's water supplies. The WaterSMART program focuses on improving water conservation and helping water resource managers ensure sufficient supplies to meet competing water demands for municipal, industrial, irrigation, hydropower, recreational, fish and wildlife, and ecosystem use. Through the WaterSMART Basin Study program, Reclamation works cooperatively with States, Tribes, other Federal agencies, non-governmental organizations, and local partners to identify adaptation strategies to address current and future regional water shortages associated with drought, climate change, population growth, and other stressors on water resources across the West.

Under Reclamation's Directives and Standards (D&S) WTR 13-01, each basin study must include four key elements; (1) projections of future water supply and demand, (2) an analysis of how existing water operations and infrastructure will perform under changing water realities, (3) development of strategies to meet current and future water demands, and (4) a trade-off analysis of identified strategies. The final element, a trade-off analysis comparing adaptation strategies, has varied significantly in methodology across basin studies that have been completed to date. This stems from a lack of guidance on conducting a trade-off analysis and applying a consistent methodology across studies. Existing guidance for conducting a trade-off analysis for basin studies is limited to Reclamation's D&S WTR 13-01 Sec. 9 Part D, which states that each basin study must include:

> *A quantitative or qualitative trade-off analysis of the adaptation strategies identified. Such analysis will examine all proposed strategies in terms of their ability to meet the study objectives, the extent to which they minimize imbalances between water supply and demand and address possible impacts of climate changes, the level of stakeholder support, the relative cost (when available), the potential environmental impacts, and other attributes common to the strategies.*

This language is general and without details on how one might conduct a trade-off analysis that is technically sound and consistent across basin studies. The ASP offers some additional information on the evaluation of alternatives and trade-offs (covered in the next section) but stops short of providing guidance on formally carrying out a trade-off analysis. This lack of guidance can reduce the quality and usefulness of trade-off analyses done by Reclamation. This technical memorandum provides detailed, yet flexible guidance on conducting a trade-off analysis in a systematic manner to provide useful information and support decision making.

2.2 Existing Guidance

When evaluating public benefits of alternatives for water resource management, DOI's ASP for implementing the PR&G require that Reclamation include monetary and non-monetary effects and allow for the inclusion of quantified and unquantified measures. The ASP further state that the level of detail required to support alternative analyses may vary but should be sufficient to inform the decision-making process efficiently and effectively. It goes on to say that alternative plans should clearly identify and evaluate the trade-offs among stakeholders and resources, and that the viability of an alternative should be determined through an evaluation of its

(1) acceptability, (2) efficiency, (3) effectiveness, and (4) completeness. These conditions are defined in the *Principles and Requirements for Federal Investments in Water Resources* (P&R) and adopted for the PR&G.

Completeness is the extent to which an alternative provides and accounts for all features, investments, and/or other actions necessary to realize the planned effects, including any necessary actions by others. Effectiveness is the extent to which an alternative alleviates the specified problems and achieves the specified opportunities. Efficiency is the extent to which an alternative alleviates the specified problems and realizes the specified opportunities at the least cost. Acceptability is the viability and appropriateness of an alternative from the perspective of the Nation's general public and consistency with existing Federal laws, authorities, and public policies. The ASP require that analysis must explicitly address the extent to which an alternative achieves each of these conditions and states that the evaluation must be systematic and include both quantitative and qualitative components. A trade-off analysis offers a framework for systematically evaluating both quantitative and qualitative components. That said, the ASP do not require a trade-off analysis and therefore do not provide any formal guidance on the process of conducting a trade-off analysis.

Under the guidance on ecosystem services, the ASP state that the ecosystem service concept provides an analytical framework which fully articulates the trade-offs inherent in a decision and provides additional information to the decision maker. This framework is well suited for trade-offs that involve many competing values, and it starts from the assumption that all ecosystem services should be evaluated. The framework is intended to equally consider services that are market commodities and those that are not, as well as services that provide use and non-use values. The ASP note that this approach can be used to address the full range of benefits and costs associated with a proposed alternative, and that there are many different metrics and methods that might be used, including qualitative or quantitative, and monetary or non-monetary approaches. While a trade-off analysis often goes beyond ecosystem services, this language offers a starting point for understanding the role and justification for a trade-off analysis when evaluating alternatives for water resource management.

A key component of the ASP is documenting and displaying trade-offs in a manner that informs decision making. The guidance states that such displays should be understandable, transparent, and constructed in a generally consistent fashion. This includes a combination of both tables and explanatory materials to help inform a decision. Displays of trade-offs should facilitate the evaluation and comparison of alternative plans necessary to make a determination and reflect; (1) the effectiveness of given plans in solving the problems and taking advantage of the opportunities identified in the planning process; (2) what must be given up in monetary and nonmonetary terms to enjoy the benefits of the various alternative plans, relative to the baseline; and (3) the differences among alternative plans. This is often accomplished through a table summarizing the trade-offs, relative to the baseline, resource-by-resource. These tables often include both quantitative and qualitative information. That said, it is not obvious how decision makers might make use of all this information, and a lack of guidance on aggregating the information in a systematic and consistent manner can limit the usefulness of the information for decision making.

There are many different approaches to describing and measuring the changes associated with an alternative. In the context of ecosystem services, the ASP require, at a minimum, a qualitative discussion of the relative value of each alternative. The guidance goes on to say that a qualitative assessment can provide additional information and more fully describe the trade-offs among alternatives, though the analysis should quantify as many effects as possible and monetize as many of the quantified effects as possible given data and resource constraints. While the ASP repeatedly emphasizes the need for including both quantitative and qualitative information, the guidance stops short of discussing ways to meaningfully aggregate and utilize all of this information to help inform decision making. In general, the evaluation of qualitative measures is considered independent of quantified measures. This is where a trade-off analysis comes in, offering a systematic framework for evaluating all relevant considerations at once in a consistent and transparent fashion.

2.3 Trade-Off Analysis Overview

The term "trade-off analysis" has been used in several settings and by several professions with varying definitions. D&S WTR 13-01 does not provide a concrete definition for a trade-off analysis, so here a trade-off analysis is defined as an evaluation of alternatives to identify strengths and weaknesses using a systematic analysis of all relevant considerations, both quantified and unquantified. Past assessments have struggled to include both quantitative and qualitative considerations in a single analysis. In many cases, that which can be measured is quantified, while that which is not easily measured is evaluated qualitatively and independent of quantified considerations. A trade-off analysis allows for both quantified and unquantified considerations to be evaluated simultaneously, identifying the strengths and weaknesses of each alternative and indicating how each alternative performs across several considerations at once. The inclusion of quantified and unquantified considerations means the analysis includes a broader range of effects than a typical economic analysis which generally aims to evaluate impacts purely in monetary terms. A trade-off analysis also applies to settings with multiple objectives where there is no unique solution that optimizes all objectives simultaneously.

A framework that is consistent with ASP guidance and useful to adopt for a trade-off analysis is triple bottom line (TBL). TBL is an accounting framework that evaluates impacts along three dimensions; social, environmental, and economic. Economic effects include market and non-market benefits and impacts on regional economic activity in terms of output, income, and employment. This also encompasses financial effects, such as the capital cost of implementing an alternative as well as ongoing costs for operation, maintenance, and replacement (OM&R). Environmental effects reflect impacts on natural resources and the environment, such as changes in water quality, riparian vegetation, pollution levels, habitat condition, fish and wildlife health, and overall ecosystem function. Social considerations include impacts on the community, for example, in terms of health, safety, equity, and quality of life.

Although superseded by the PR&G in 2015, the 1983 *Economic and Environmental Principles and Guidelines for Water and Related Land Resources Implementation Studies* (P&G) previously defined four accounts to consider when evaluating alternatives. These accounts are consistent with a TBL framework and capture the potential range of effects across economic, environmental, and social considerations. As such, these accounts provide a useful starting point for determining which effects ought to be included in a trade-off analysis. The National Economic Development (NED) account measures economic costs and benefits to the nation in monetary terms; the Regional Economic Development (RED) account measures changes in the distribution of regional economic activity; the Environmental Quality (EQ) account represents impacts on ecological, cultural, and aesthetic attributes of significant natural and cultural resources that cannot be measured in monetary terms; and the Other Social Effects (OSE) account reflects impacts on the community in terms of life, health, and safety as well as any other considerations not captured elsewhere. In general, NED and RED are the only accounts quantified, but they are not additive since RED reflects impacts resulting from NED effects. Meanwhile, EQ and OSE accounts are typically examined independently and qualitatively.

Given that impacts are not easily quantifiable along every dimension, several important considerations can get left out of the evaluation of alternatives. At best, unquantified considerations often get assessed descriptively and independent from quantified effects, which means that impacts across criteria cannot be aggregated in any systematic way. Sometimes a scoring process is used, but there is often little transparency and consistency with this process, limiting the reproducibility and insights of the scoring results. A trade-off analysis addresses these limitations, systematically incorporating both quantitative and qualitative effects into a single assessment that aggregates all effects in a transparent and meaningful way to help inform decision making. While the goal of a trade-off analysis is fairly straightforward, the process of combining quantitative and qualitative considerations into a single assessment comes with several challenges. It is therefore important that the analyst understands the nuances of this process so that the results are meaningful and helpful to inform decision making.

A trade-off analysis is an application of multi-criteria decision analysis (MCDA), which pertains to evaluating complex decision making under multiple and often conflicting objectives. MCDA addresses trade-offs that occur when a decision leads to a desirable change in one or more objectives while simultaneously resulting in an undesirable change in another objective. Most multi-criteria problems have conflicting criteria and as a result there is no unique solution that can optimize all criteria simultaneously. In the literature, MCDA is sometimes used interchangeably with "multi-attribute decision analysis" (MADA). This is similar, but unique, from the "analytical hierarchy process" (AHP), from "multi-attribute utility analysis" (MAUA), and from "fuzzy set analysis," which is sometimes called "qualitative comparative analysis" (QCA). There is an extensive literature on each of these (see the *Journal of Multi-Criteria Decision Analysis*), but they are all generally considered a form of "multi-criteria analysis" (MCA) and "structured decision making" (SDM). These approaches are intended to deal with decision making in a structured and consistent way while handling large amounts of complex information, differing primarily in how they combine the data and information. The U.K. Department of Communities and Local Government (UKCLG) provides an overview of various MCA techniques, but ultimately MCDA is recommended for government decision making (UKCLG, 2009).

For this technical memorandum, the focus is on applying MCDA for a trade-off analysis. The goal of MCDA is to provide an overall ordering of options, from the most preferred to the least preferred option. In general, there are conflicts and trade-offs between objectives, meaning no alternative is optimal across all considerations at once. MCDA is a way of looking at complex problems that are characterized by any mixture of monetary and non-monetary objectives, of breaking the problem into more manageable pieces to allow data and judgements to be brought to bear on the pieces, and then of reassembling the pieces to present a coherent overall picture to decision makers (UKCLG, 2009). MCDA simply provides a way to disaggregate a complex problem to determine the extent to which alternatives achieve different objectives. The purpose is to aid decision making, but not necessarily make a decision, so the process is sometimes referred to as a "decision support tool."

In order to compare quantified and unquantified criteria, a trade-off analysis involves putting all considerations on an equivalent scoring scale. For quantified effects, converting impacts to a scoring scale requires careful attention to ensure that the magnitude of differences (proportionality) is retained in the process. For unquantified effects, alternatives must be measured qualitatively (e.g., low-high or increases/decreases) for each criteria being considered. When done properly, the scoring process allows one to compare and rank alternatives on a relative basis across all considerations at once, both quantitative and qualitative. Given that some considerations may be viewed as more important than others, different weighting schemes can be employed to explore how each alternative performs under different sets of preferences.

Assessments that evaluate multiple criteria and use "scoring and weighting" to compare alternatives are not new to Reclamation or government reports more broadly. However, the analysis is not always referred to as MCA nor carried out in a systematic and consistent manner. Historically, MCA has been most commonly used by federal agencies to evaluate ecosystem services, which is consistent with the guidance found in the ASP. In this setting, it is useful to distinguish between "means" versus "end" objectives to ensure there is a direct link to benefits for society. Measurement is often easier for means objectives, but end objectives are what people actually care about. Such measures are often referred to as "benefit relevant indicators" (BRIs). The use of BRIs ensures that assessments measure outcomes that are relevant to human welfare, rather than measures that might not be directly or even indirectly relevant. This is an important point and should guide the selection, definition, and measurement of criteria to be used in a trade-off analysis. BRIs go beyond a narrative description and are intended to capture whether there is a demand for the change that might result from an alternative, rather than a change that may have no meaningful effect on social welfare (Olander et al. 2015). BRIs can also be measures of a disservice that result in lower rather than higher benefits.

According to Olander et al. (2015), BRIs fulfill two important needs. The first is the need for indicators that are socially comprehensible, in that lay audiences can relate them clearly to their own wellbeing. The second need is for indicators that enhance the accuracy of social evaluations. Just because lay audiences see a causal connection between a change and their welfare doesn't mean they will see it accurately (Olander et al. 2015). The authors go on to say that an indicator becomes "benefit relevant" when it is defined and cast in units that resonate with stakeholders as something that affects their welfare. For example, "numbers of catchable fish" is more relevant to fishers than other measures such as dissolved oxygen content in the water or an index of biotic integrity, even though water quality might directly influence fish populations (Olander et al. 2015). If those that benefit from a change cannot easily understand why an indicator is relevant to their welfare, the indicator is unlikely to be effective. This consideration is important to keep in mind when selecting, defining, and measuring evaluation criteria for a trade-off analysis to ensure that the analysis and results are meaningful.

2.4 Trade-Off Analysis and Planning

A trade-off analysis is often considered a non-monetary approach to evaluating alternatives, but it can also be used in combination with monetary values, such as those found in an economic or financial analysis. A trade-off analysis is thus not a substitute for such analysis, but rather, can serve as a complementary/secondary analysis that includes non-monetary considerations. A trade-off analysis is a mandatory component of Reclamation's WaterSMART Basin Studies and used to evaluate adaptation strategies that reduce future water shortages in particular areas. That said, a trade-off analysis conforms with existing guidance under the PR&G and DOI's ASP and is widely applicable to other Reclamation work. For example, a trade-off analysis can be used to examine alternative reservoir operations, alternative infrastructure projects, or alternative policies. A trade-off analysis is flexible and broadly applicable to any planning process, and the analysis can accommodate simple or complex details, ultimately depending on the availability of information and scope of the analysis. In general, the planning process involves six distinct steps:

1. Identify Problems, Needs, and Opportunities
2. Inventory Existing Resources and Conditions
3. Formulate Alternatives
4. Evaluate Effects of Alternatives
5. Compare Alternatives
6. Select the Recommended Alternative

This process is typically iterative, meaning it is common to bounce between the various steps as a study evolves and progresses. In general, WaterSMART Basin Studies cover steps 1-5 of the planning process at an appraisal level, stopping short of recommending a preferred alternative, and the trade-off analysis addresses steps 4 and 5. Step 1 pertains to identifying the problem and need, which then defines the scope and objective of the study. For basin studies, the problem is an expectation of future water shortages, or imbalances between water supply and demand in a particular area. The need is then for alternatives that help reduce future shortages using a supply-side or demand-side approach. Step 2 entails taking inventory of existing conditions and establishing the baseline for taking no action in the area of interest, which generally includes forecasting future conditions without any alternative in place. Step 3 consists of formulating alternatives to address the problem and need.

Next, step 4 entails evaluating the effects of each alternative, which may be assessed quantitively and/or qualitatively across a range of considerations tailored to the area of interest. This is where a trade-off analysis enters the planning process. Step 5 is to compare alternatives across those effects of interest. This is ultimately the goal of a trade-off analysis. The final step in the planning process is to select a recommended alternative, but basin studies are generally at an appraisal level and only intended to screen alternatives. The trade-off analysis therefore identifies strengths and weaknesses of each alternative and helps inform which alternatives deserve further attention and a feasibility level planning analysis with a narrower scope and higher level of detail.

Under the 1983 P&G, a trade-off analysis can serve as a secondary analytical approach, assuming that the primary selection criterion (NED benefits outweigh NED costs) is met. A positive net NED benefit means that the alternative is economically justifiable and monetized benefits exceed monetized costs. If net NED benefit is not positive, the alternative is typically eliminated from consideration, unless a "Secretarial Exception" is granted in order to implement an alternative that achieves a non-economic goal. Under the ASP and current guidance (PR&G, 2015), economic justification is no longer the primary selection criterion and assessments are required to include monetary and non-monetary effects and allow for the inclusion of quantified and unquantified measures. This lends itself to a trade-off analysis as a means to evaluate quantified and unquantified effects together. This helps document and justify the selection of the preferred alternative, which is especially important if an alternative is not economically justified or does not provide the highest monetized net NED benefit of the alternatives being considered.

In 2002, the U.S. Army Corps of Engineers (USACE) developed a *Trade-Off Analysis Guidebook for Planning* which provides an extensive discussion on conducting a trade-off analysis for water resource planning. Another useful resource, UKCLG (2009), is a guidance manual for MCA aimed at government officials and non-specialists evaluating options for policy and other decisions, including those related to the environment. In 2013, the U.S. Forest Service (USFS) produced a technical report on *The Science of Decision Making* that covers SDM and trade-off analysis looking at forest and grassland management. These guidance documents provide a great deal of information, some of which is beyond the scope of this technical memorandum. That said, many of the key insights from these resources are referenced throughout this document. Readers interested in more information and other applications for trade-off analysis are encouraged to refer to these resources, as well as others referenced.

3.0 Conducting a Trade-Off Analysis

A trade-off analysis can be separated into six distinct steps (or "sub-steps" that address steps 4 and 5 of the planning process). The first step is to determine the potential impacts arising from the alternatives being considered, which should include all dimensions in the TBL framework. This is necessary to inform the second step, which is the selection of the evaluation criteria to be used for comparing the alternatives. The third step is to measure the criteria of interest, both those quantifiable and unquantifiable. Next, impacts for each criteria must be scored on an equivalent scale. The fifth step is to weight criteria scores to address the fact that some criteria might be viewed as more important than others. The final step is to compare alternatives, realizing that the analysis does not necessarily identify a "best" choice, but rather, identifies the key trade-offs encountered in selecting between alternatives. Each of these six steps are described in detail in the following sections.

3.1 Impacts from Alternatives

The first step in conducting a trade-off analysis is to determine which impacts might result from the alternatives being considered. These effects should reflect a TBL framework that considers impacts along social, environmental, and economic dimensions. In some cases, such as small-scale projects, there may not be impacts along all dimensions (e.g., social effects). Conversely, large-scale alternatives, such as those commonly evaluated for basin studies, are likely to have impacts along several dimensions. A criterion is broadly defined as a principle or standard by which something may be judged. It generally represents a mechanism by which a larger objective can be achieved, for example NED benefit and NED Cost are criteria used to judge if an alternative provides a positive net NED benefit to society. Candidate criteria for comparing alternatives should be based on those effects most likely to result from the alternatives being considered, as well as those effects most relevant to the scope and objective of the study. As previously discussed on BRIs, criteria are also more effective when defined and selected to ensure there is a meaningful link to human welfare. That said, a trade-off analysis is flexible and can accommodate broadly or narrowly defined criteria.

Table 3.1 provides a list of some common criteria to consider for the purpose of evaluating alternatives for water resource management. This list is not exhaustive, and criteria should be defined and tailored to analysis at hand. There is often some overlap between criteria, which may warrant grouping, and in some cases criteria can be divided into more refined categories (this is discussed in the next section). As previously mentioned, the four accounts in the P&G offer a good starting point for selecting criteria, and as such, the criteria listed in Table 3.1 could conceivably be grouped into one of the four accounts. That said, USACE (2002) notes that the P&G accounts include a wide variety of effects, some of which may influence the decision at hand, and others which may not weigh directly in the decision process. As such, the list provided below simply groups potential criteria along social, environmental, and economic dimensions consistent with a TBL framework. Just as there is overlap between some criteria, there is also some overlap across these dimensions, meaning one should not assume that these criteria and categories are rigid.

Table 3.1 – Potential Criteria for Consideration

Economic	Environmental	Social
Residential Benefit	Instream Flow	Safety
Commercial Benefit	Water Quality	Health
Institutional Benefit	Air Quality	Education
Industrial Benefit	Pollution	Local Public Services
Irrigation Benefit	Fish and Wildlife	Aesthetic Amenities
Hydropower Benefit	Endangered Species	Environmental Justice
Recreation Benefit	Erosion and Sedimentation	Equity
Fish and Wildlife Benefit	Soils and Vegetation	Risk
Flood Risk Benefit	Aquatic Habitat	Cultural Resources
Transportation Benefit	Riparian Habitat	Historic Significance
Power Ancillary Services	Land Habitat	Tourism
Ecosystem Services	Sustainability	Lodging Availability
Regional Output	Conservation	Public Perception
Regional Income	Biodiversity	Political Perception
Regional Employment	Ecosystem Function	Regulatory Barriers
Capital Cost	Land Use	Administrative Complexity
OM&R Cost	Connectivity	Policy Compliance
Household Cost	Land Subsidence	Energy Use and Mix
Business Cost	Aquifer Recharge	Adaptation
Ability to Pay	Depth to Groundwater	Resilience
Property Value	Heat Island Effect	Mitigation

3.2 Criteria Selection

Selecting the evaluation criteria is a crucial part of a trade-off analysis and comparing alternatives. NED benefits and NED costs are examples of criteria and the primary focus of an economic analysis, which generally entails an objective to maximize net benefit to society. While NED costs and benefits are an important part of any analysis, the reliance on only monetized criteria unavoidably excludes several important considerations, and including RED, EQ, and OSE considerations independently limits the usefulness of the information. A trade-off analysis therefore supports a richer and more complex analysis that incorporates the full range of effects that might result from alternatives, both quantified and unquantified. That said, the usefulness of the analysis depends on the criteria selected for evaluation and how those criteria are defined and measured.

The criteria used to compare alternatives should be based on those effects most likely to result from the alternatives being considered and most relevant to the scope and objective of the study. Some criteria, although possibly import to consider, are not useful as evaluation criteria because they do not distinguish between alternatives. If the effects for a criterion are identical across alternatives (e.g., all have "no impact" or all have a "high impact") then the criterion does not help distinguish between alternatives and is of little use as an evaluation criterion. However, including such criteria may be important to show that they were considered, and to highlight that there is no variation across alternatives, which itself may be informative and important for decision making. For example, if each alternative improves water quality to a similar extent, it may be important to leave this as a criterion, even though it will not help to distinguish between the alternatives in the trade-off analysis.

Roy (1985) defines a set of criteria as "coherent" if the following three properties are satisfied: (1) exhaustiveness, (2) consistency, and (3) non-redundancy. Exhaustiveness is satisfied when no important criterion has been forgotten. Consistency means that when comparing two alternatives, if one alternative scores better for a criterion, then it must be true that that alternative is preferred to the other for this criterion. If a set of criteria is exhaustive and consistent, then the set is considered non-redundant if removing any single criterion leads the remaining criteria to no longer be exhaustive or consistent. Put differently, if one criteria is wholly a subset of another criteria, or if there is interdependence or overlap between two criteria, then including both may be redundant and give additional preference to the overlapping components. An example might be to have a broad criterion for "environmental effects" as well as a specific criterion for "water quality," which is one of the considerations already captured under "environmental effects." This would lead to water quality unintentionally receiving additional importance in the final results.

USACE (2002) further defines criteria as being "effective" when they are: (1) directional, (2) concise, (3) complete, and (4) clear. Directional is defined as there being a clear preference for the direction of the criteria score, for example, a preference to minimize or maximize the criteria. Concise pertains to providing the smallest number of criteria that allows all significant impacts to be assessed. Complete means that no significant impact is excluded. Clear means that criteria are appropriately separated between quantitative and qualitative measures and measurement is well defined. Determining a "coherent" (Roy, 1985) and "effective" (USACE, 2002) set of criteria therefore requires special attention when selecting and defining criteria.

Another key consideration when selecting criteria is to decide when criteria should be grouped rather than included individually. If alternatives score similarly across certain criteria, then including each one individually may add little to the analysis, and this may warrant combining criteria into a single criterion. For example, alternatives may score similarly in terms of effects on water quality and fish habitat, in which case these considerations should be captured with a single criterion. In general, highly correlated criteria should be combined or removed to avoid redundancy. For example, the Los Angeles Basin Study (2016) used a correlation analysis to remove highly correlated criteria and narrow down the set of evaluation criteria. Making sure that criteria do not significantly overlap ensures "linear additivity" and is a key difference between MCDA and MAUA, where MAUA formally accounts for interdependencies between criteria as well as uncertainty (UKCLG, 2009). These features generally make MAUA more complex and less transparent than MCDA, requiring more expertise and time.

Combining criteria and narrowing the set used for evaluation not only ensures independence and linear additivity, but also helps to simplify the analysis and makes the results easier to interpret. Pomerol and Barba-Romero (2000) suggest using no more than 20 criteria to allow for meaningful distinctions. That said, a trade-off analysis is flexible and can be as simple or complex as desired. When combining criteria one can focus solely on broad categories or use a hierarchical breakdown with subcategories to accommodate additional detail. For example, one might include NED benefits as a broad category, or further separate this between municipal, industrial, irrigation, hydropower, etc. subcategories of benefits for additional resolution. Categories such as municipal benefit can be separated even further, for example, between residential, commercial, and institutional benefits. Such a breakdown is useful when subcategories can be meaningfully distinguished and measured separately, or when there is a desire to give different preference to subcategories. Otherwise, such refinement may not be warranted and only adds to the complexity. This is discussed further under criteria weighting and demonstrated in the example provided later on.

3.3 Criteria Measurement

Before measuring criteria, one must separate them between qualitative and quantitative criteria based on data availability and the practicality of measurement. In general, quantitative criteria can be measured numerically, while qualitative criteria cannot. The distinction between cardinal and ordinal data is used in this technical memorandum to differentiate quantitative versus qualitative criteria. While there exists several definitions of quantitative versus qualitative data, cardinal versus ordinal offers a straightforward and generalizable distinction for the purpose of conducting a trade-off analysis. It should also be noted that throughout this document the distinction between "quantitative" and "qualitative" is used synonymously with "quantified" versus "unquantified" measures, which is the language found in the ASP.

Qualitative (ordinal) data (e.g., low/medium/high or increases/decreases) reflects the relative position of an alternative for that criterion, but the distance between categories is not necessarily meaningful. Qualitative data is sometimes referred to as categorical data, as it is typically not measured by real numbers, but rather, categories defined by properties and attributes. This sort of data provides a meaningful ranking but lacks information on the magnitude of differences. Meanwhile, cardinal data has a quantitative interpretation and reflects magnitude (proportionality), such as NED benefits measured in monetary terms, or water quality measured by turbidity. This sort of data is numerical, where differences indicate proportionality across alternatives, meaning that ratios have a meaningful interpretation. For example, if NED benefits are estimated at $10,000 for one alternative and $20,000 for another, then the latter offers twice the benefit as the former. In general, quantitative data is objective (i.e., not influenced by one's personal values) while qualitative data tends to be more subjective (i.e., potentially influenced by one's personal values). Nonetheless, both serve an important role in decision making and evaluating alternatives, and a trade-off analysis is unique from other analyses due to the ability to incorporate both quantitative and qualitative information.

There are some criteria that can be measured quantitatively but are most reasonably measured qualitatively. For example, Reclamation cost estimators often provide estimates for capital and OM&R costs, and there are six levels of cost estimates defined in Reclamation's D&S FAC 09-01 Sec. 2 Part A. In particular, these estimates are defined as preliminary, appraisal, feasibility, post-authorization, pre-validation, and independent government cost estimates. That said, in some cases it is not feasible to get quantitative cost estimates. In such an instance, costs could still be considered qualitatively and incorporated into a trade-off analysis. An example of this is the West Salt River Valley (WSRV) Basin Study (Reclamation, 2021). In this study, both supply-side and demand-side alternatives were evaluated, and each alternative consisted of several underlying components, some structural and others non-structural. In this setting, it was deemed impractical to obtain cost estimates that were comparable across alternatives, so capital and OM&R costs were measured qualitatively using a low/medium/high categorical scale with alternatives evaluated on a purely relative basis.

Any criterion that can be measured quantitatively can also be measured qualitatively, but not necessarily vice-versa. As such, qualitative measurements lend themself to settings where it is impractical or even impossible to use quantitative measurement. For example, with criteria such as sustainability or equity, there is no universal way to measure these quantitatively, which means that qualitative measurement is most reasonable. That said, when given the choice, quantitative measurement is generally preferable to qualitative measurement as it is more objective and informative.

Whenever qualitative data is used, it is important to understand what the data represents. Ideally, the data should mean the same thing to everyone, but in practice, this is not always possible due to potential points of disagreement (USACE, 2002). First, individuals may differ in their interpretation of the criterion itself. Criteria such as sustainability, biodiversity, equity, and risk are but a few examples that lack a clear universal definition and understanding. Second, even when there is common agreement on the definition of the criterion, there may be disagreement about the meaning of the categories used for measurement. For example, individuals may not agree on what constitutes a "low," "medium," and "high" classification. Carefully defining and communicating different categories is the only practical solution to these problems (USACE, 2002). For example, low water quality could be described in terms of specific thresholds such as those defined by the U.S. Environmental Protection Agency (EPA). If a precise definition is not feasible, a scenario can be used to describe it. For example, low water quality could be described as requiring treatment before going towards a particular use.

When relying on qualitative data, it is important to get input from those with expertise related to the criterion. For example, if fish habitat is a criterion being considered, it would be important to work with a fish biologist or someone with expertise in this area to determine how to define and measure the criterion. How the criterion gets defined will determine whether it might be considered a BRI or not. For example, "change in habitat acreage" is not as concretely linked with social welfare as "change in the population of species X." When defining and measuring qualitative criteria, it is also important to get input and consensus among all relevant stakeholders. For example, basin studies involve partnerships across States, Tribes, other Federal agencies, non-governmental organizations, and local entities, so it is crucial that all parties have an opportunity to weigh in and a consensus is reached for criteria selection, definition, and measurement.

Gregory et al. (2012) discusses SDM for environmental decision making involving multiple and diverse stakeholders. The book goes into detail on how to have a rigorous, inclusive, defensible, and transparent process with insights from cognitive psychology, facilitation, and negotiation (Gregory et al., 2012). The authors review key methods and discuss examples based on their experiences in communities, boardrooms, and stakeholder meetings. For the WSRV Basin Study, several meetings were held to determine the set of criteria and how they should be defined and measured, which involved an iterative process and input from all entities involved in the project. Though reaching a consensus can be time-consuming when several parties are involved, this is a vital step to ensure that the process is transparent, and the results are meaningful. It is also important to clearly state all assumptions so that the analysis is reproducible and understandable to those outside the process.

In the WSRV Basin Study, quantitative criteria were assessed based on the water volume provided by each alternative. Water volume was used to quantify a monetary benefit as well as assess regional economic impacts on income, employment, and output. The remaining criteria were measured qualitatively using a survey to evaluate impacts on a low/medium/high relative basis. This was done using a 1 to 3 (low to high) scale, and respondents were able to use decimal places if they felt that the impact was somewhere between these categories. Respondents were also able to enter zero if they thought there was no impact. The WSRV Basin Study involved numerous stakeholders, from cities and municipal systems to irrigation districts and regulatory bodies. The survey was sent to all entities involved in the study, and when an entity provided more than one survey, the responses were first averaged to ensure that each entity was given equal preference in the process. This resulted in qualitative criteria being measured by the full range of experts and stakeholders involved in the study. In general, the responses were rather consistent across criteria and alternatives. That said, response variation was reported to highlight the areas of uncertainty and potential disagreement. These steps are crucial to highlight the limitations of qualitative measurement and ensure that the analysis is transparent and meaningful.

3.4 Criteria Scoring

After the criteria of interest have been measured, quantitively or qualitatively, the next step involves scoring all criteria on an equivalent scale. For economic assessments, the main rationale for monetization is to ensure that everything gets evaluated on an equivalent scale. However, it is often difficult, and sometimes even perceived as inappropriate, to derive monetary values for all relevant considerations. This is why the inclusion of qualitative considerations is important, in general, capturing those effects that cannot be monetized or even quantified in some manner. That said, the downside of trying to evaluate quantitative and qualitative considerations together is that there is a wide range of scales used for measurement, which then makes it difficult to make comparisons. For a trade-off analysis, scoring is used to put all criteria on an equivalent scale. The consequences of different criteria scales and the importance of scoring criteria on an equivalent scale is sometimes overlooked in MCA (see Monat (2009) for further discussion on measurement scales).

To put criteria on an equivalent scale, it is generally necessary to normalize (rescale) criteria to ensure that some criteria are not implicitly given additional influence in the analysis. This is required to aggregate effects across the criteria and is necessary since a trade-off analysis compares alternatives purely on a relative basis. Without normalization, the analysis will implicitly give additional weight to those criteria with a larger scoring scale and less weight to those with a smaller scale. For example, if the scores for a qualitative criterion range from 0 to 10, and for another criterion the scores range from 0 to 5, the former has twice the potential to impact final scores, resulting in the former criterion having more influence than the latter criterion. Explicitly weighting criteria to account for some criteria being more important than others is discussed in the next section, but prior to explicit weighting, it is necessary to normalize and score criteria on the same scale to ensure no implicit weighting.

Before normalization, best practice is to set up a decision matrix (sometimes called a "performance matrix" or "consequence table") and convert all negative impacts to positive impacts so that the goal is a straightforward maximization problem. A decision matrix summarizes the performance of each alternative for each criterion. Converting negative impact criteria (i.e., those where a larger value is less desirable, such as costs) allows them to be maximized while maintaining the appropriate directionality of the criteria. A simple method for maximizing negative impacts is to change the sign of measurement. For example, if one alternative has a capital cost of $1,000 and another has a capital cost of $2,000, changing the sign and maximizing would identify -$1,000 as performing twice as well for capital cost. An alternative is to take the reciprocal. Using the previous example, a cost of $1,000 would become 0.001 and $2,000 would become 0.0005, and maximization would again identify the former alternative as performing twice as well. Both of these approaches maintain proportionality (the magnitude of differences) while converting negative impacts into values that can be maximized.

Recall that directionality is a key condition that USACE (2002) defines for an "effective" criterion. That said, not all criteria can be defined with a single directionality. USACE (2002) provides the examples of temperature and pH, both of which have a range, but the optimum is not necessarily at either end of the scale. In these instances, one can take the absolute value of the difference between the optimum value and the actual measurement. For example, suppose the desired pH level is 7 and one alterative is expected to result in a pH of 6 and another in a pH of 9. The transformation would result in |7-6|=1 and |7-9|=2. Small numbers now reflect values closer to the optimum, so these transformed values reflect a single directionality with values to be minimized. As before, one can now change the sign or take the reciprocal for maximization. It is therefore always possible to prepare a decision matrix of criteria to be maximized, and using the reciprocal approach ensures that all values are positive, which helps to simplify the normalization process.

There are a variety of normalization techniques one can use to ensure that all criteria are evaluated on an equivalent scoring scale, generally from 0 to 1. USACE (2002) provides a detailed discussion on different normalization methods, while here a brief overview is provided for four of the most common approaches. These methods are referred to as, "percentage of maximum," "percentage of range," "percentage of total," and "unit vector" (USACE, 2002). The percentage of maximum approach involves dividing each value by the maximum value for that criterion. This is the most commonly used technique, but the values will not span the entire 0 to 1 interval. However, this approach preserves proportionality and ensures that at least one alternative scores a one. Meanwhile, the percentage of range approach forces the scale to fully span the 0 to 1 interval, but it does not preserve the proportionality of the original values. The unit vector technique maintains proportionality while transforming the scale to 0 to 1, but the values will not necessarily include the extreme values of zero and one. Lastly, the percentage of total maintains proportionality while forcing the values to sum to one. Table 3.2 shows each of these methods.

Table 3.2 – Normalization Methods

The normalized value (*v*) for the measurement of a criterion (*x*) can be calculated for each alternative (*i*) according to one of the following methods:	
Percentage of Maximum	$v_i = \frac{x_i}{\max x_i}$
Percentage of Range	$v_i = \frac{x_i - \min x_i}{\max x_i - \min x_i}$
Percentage of Total	$v_i = \frac{x_i}{\sum_i x_i}$
Unit Vector	$v_i = \frac{x_i}{(\sum_i x_i^2)^{1/2}}$

If there is no basis for favoring one normalization method over another, the percentage of maximum technique is generally suggested (USACE, 2002). This approach is straightforward and ensures that at least one alternative scores the maximum possible score, which then ensures that all criteria are treated equivalently before explicitly determining criteria weights (discussed in the next section). This approach is also flexible and allows one to easily convert the 0 to 1 normalized values to a different scale, such as 0 to 10 or 0 to 100. The percentage of maximum technique is therefore used in the example provided in this technical memorandum. The WSRV Basin Study also provides an example of this technique, where normalized scores are ultimately multiplied by 3 to generate a 0 to 3 scale that signifies "no impact," "low impact," "medium impact," and "high impact" for a straightforward interpretation.

3.5 Criteria Weighting

To address the fact that some criteria might be deemed more important than others, a trade-off analysis typically involves criteria weighting to explicitly assign different preferences to each criterion in the set of evaluation criteria. Assigning weights to the criteria is often the most contentious task because it is, by definition, the most subjective task (USACE, 2002). To ensure transparency and highlight the sensitivity to criteria weighting, best practice is to generally report the unweighted results as well as the weighted results. In some cases, the results may be quite similar, in which case criteria weighting becomes less controversial. This was the case for the WSRV Basin Study where the unweighted rankings were identical to the weighted rankings. Nonetheless, criteria weighting is often an important step in a trade-off analysis and can drastically influence the results.

A weight is a measure of the relative importance of a criterion. Assigning weights is a method for incorporating human judgments about the relative value of a criterion (USACE, 2002). Put differently, weighting relies on subjectively expressed expert preferences and judgment. The subjective nature of this step is a contributing factor for why a trade-off analysis cannot identify a unique solution. Conversely, incorporating subjective preferences into the analysis is also a key strength of a trade-off analysis, as it supports a richer and more complex assessment of alternatives. Clemen and Reilly (2001) offers a comprehensive overview on the technical aspects of MCA and the elicitation of expert opinion and assessment of weights. Meanwhile, Hajkowicz et al. (2000) provides a detailed discussion on different weighting techniques used in natural resource management. In particular, the authors evaluate "fixed-point," "rating," "ordinal ranking," "graphical," and "paired comparison" weighting approaches. These techniques are those most commonly used, but several others also exist.

For this technical memorandum, attention is placed on the fixed-point and rating methods, as they are both straightforward and provide cardinal weights which are easily incorporated into a trade-off analysis. Techniques such as ordinal ranking and graphical weighting are also straightforward, but assumptions and additional steps are needed to generate cardinal weights and incorporate the weights into the analysis. Meanwhile, the paired comparison approach (which is an application of AHP) is often not straightforward, and the complexity increases with the number of criteria. Another alternative is the "simple multi-attribute rating technique" (SMART), which first identifies the most important criterion and then weights others on a relative basis (USFS, 2013). Fortunately, Hajkowicz et al. (2000) finds that different weighting methods tend to result in similar weights, meaning the approach used should have minimal consequence on the final results.

The fixed-point technique involves a fixed number of points (e.g., 10 or 100), with points then distributed across the set of criteria with more points reflecting greater importance. With this approach, percentages are often used, where the weights sum to 100 percent. Equivalently, one can use decimals such that the weights sum to one, which makes the weights normalized so that no further steps are needed. If a different approach is used, then normalization using the percentage of total method is suggested to ensure that the weights sum to one. The fixed-point method is simple and transparent, making it ideal for a trade-off analysis. However, since points are fixed, this approach means that the only way to give greater importance to one criterion is to give less to another. This is a potential weakness of this method, particularly if making such a choice is difficult (USACE, 2002). An alternative approach is the rating method, which avoids making trade-offs between weights.

The rating method uses a numerical scale, for example 1 to 5, where the low end reflects low importance, and the high end reflects high importance. The number of integers used is arbitrary, but the interval distance between the values is typically assumed to be identical (i.e. linear). This is similar to the rating process used for rating one's satisfaction with a purchase or experience, which often involves a 1 to 5-star rating scale. This approach uses a common scale for each criterion and there is no restriction on the number of points assigned to each criterion. This means that there is no trade-off when determining weights, as is the case with the fixed-point method. Furthermore, the ratings are easily converted to a decimal or percentage using the percentage of total normalization method. The fixed-point and rating methods are therefore quite similar, with the key difference being that the former involves trade-offs when assigning importance while the latter is unconstrained.

For the WSRV Basin Study, a survey was sent to all stakeholders on the study to determine the relative importance of each criterion. Criteria importance was determined using the rating method, which involved a 0 to 3 scale representing "no importance," "low importance," "medium importance," and "high importance." Respondents were also permitted to use decimals to accommodate responses between these categories. To compare criteria importance, responses were normalized using the percentage of maximum method, which indicated that the criterion of lowest importance was weighted at 35 percent of the criterion considered most important, and the remaining criteria were weighted anywhere from 50-80 percent of the most important criterion (Reclamation, 2021). However, for the actual trade-off analysis, the percentage of total normalization method was used so that criteria weights summed to one and the original 0 to 3 (low to high) scoring scale was maintained. Normalizing criteria weights is not mandatory but preserving the proportionality of the weights is essential (USACE, 2002).

To calculate the weighted criteria scores, one simply multiplies criteria weights (ideally normalized as a percentage of total) by the normalized criteria scores. The resulting values for each alternative reflect both the magnitude and relative importance of the effect captured by each criterion. It is standard practice to normalize the values one last time before conducting the trade-off analysis, but this is not required. The main reason for doing so is to convert the weighted criteria scores to the desired scale, and this is typically necessary when comparing the unweighted results with the weighted results to ensure that both are on an equivalent scale. When including subcategories for the set of criteria, weighting can be used to assign different importance to the subcategories. That said, special attention is needed to ensure that proportionality and the desired scale is maintained throughout the weighting process. This is covered in more detail in the example trade-off analysis provided below.

3.6 Interpreting Results

As previously emphasized, there is no unique solution in a trade-off analysis and the analysis does not necessarily identify the “best” alternative. Instead, a trade-off analysis highlights the strengths and weaknesses of each alternative and indicates how each alternative performs across several considerations at once. The analysis cannot identify a unique solution since the impacts of alternatives are multidimensional, the criteria are often conflicting, and the analysis involves a degree of uncertainty and subjectivity. As such, the results of the analysis are typically used to rank alternatives based on the merits of the assumptions and framework used. Beyond ranking the alternatives, the results can also be used to indicate the magnitude of differences across alternatives. This is typically done by reporting the weighted criteria scores in relation to the top-ranking alternative using the percentage of maximum normalization method. This provides additional information that shows how alternatives compare proportionally to one another considering several criteria at once, which can help to determine when the difference between rankings is rather small or perhaps quite large.

4.0 Example Trade-Off Analysis

To demonstrate how to conduct a trade-off analysis, this section provides a hypothetical example. This entails going through the six distinct steps from the previous section to illustrate the process and highlight some of the key aspects of a trade-off analysis.

4.1 Example – Impacts from Alternatives

Suppose there are five alternatives being considered. The first step is to determine the potential impacts that might result from the alternatives as well as the effects most relevant to the objective of the study. As previously emphasized, there may be a large number of criteria worth considering, but the most useful evaluation criteria are those which are intuitively linked with human welfare and help distinguish between alternatives. Additional criteria can remain in the analysis to show that they were considered or discussed independently and excluded from the trade-off analysis. For this example, suppose the alternatives are all relatively large-scale, as is often the case for basin studies. In this setting, there is likely to be effects to consider along social, environmental, and economic dimensions.

4.2 Example – Criteria Selection

Once potential criteria have been identified, it is time to determine the set of criteria used to evaluate alternatives. In general, it is best to narrow the set of criteria as much as possible. This involves grouping criteria that overlap and subdividing criteria where additional detail is desired. Recall that Roy (1985) defines a set of criteria as "coherent" if it is exhaustive, consistent, and non-redundant, and USACE (2002) defines the set as "effective" when criteria are directional, concise, complete, and clear. Given these considerations, suppose the following nine criteria in Table 4.1 are used as the set of evaluation criteria, many with subcategories. The criteria are defined such that the first three reflect economic considerations, the next three reflect environmental considerations, and the last three reflect social considerations. A brief definition of each criterion and subcategory is provided below.

The first criterion, *NED Benefits*, reflects monetary benefits from a national perspective. This generally comes from changes in water supply volume or reliability and accrues to residential, commercial, institutional, industrial, irrigation, hydropower, and recreational users of water. This may also include flood control benefits, navigation benefits, fish and wildlife benefits, and ecosystem services (which generally includes recreation, fish and wildlife, and other benefits linked with the environment, some of which are included elsewhere in the criteria list). To simplify things, the municipal and industrial (M&I) subcategory is used to capture municipal (residential, commercial, and institutional) and industrial uses of water. These categories are difficult to untangle and the benefit estimation process is often similar across these uses, warranting a single subcategory. Meanwhile, irrigation use is often quite distinct, and the benefit estimation is unique from M&I benefits. Subcategories for hydropower and recreation benefits are also included, and both involve a unique benefit estimation process.

Regardless of how monetary benefits get estimated, one could avoid the use of subcategories and simply include the criterion *NED Benefits* as the sum of monetary benefits across all relevant subcategories. However, suppose the study team has a desire to include subcategories and assign different preferences across them. For example, perhaps M&I and irrigation use are the main authorized and reimbursable purposes of a project, while hydropower and recreation are secondary uses and constrained by M&I and irrigation delivery obligations. In this setting, there could be a desire to give additional preference to M&I and irrigation benefits over hydropower and recreation benefits. This can be accomplished through weighting, as shown later in the example.

Table 4.1 – Example Evaluation Criteria Definitions

Criterion/Subcategory		Definition
(1) NED Benefits		Monetary benefits from a national accounting stance.
	M&I Benefit	Benefit to municipal and industrial (M&I) water users.
	Irrigation Benefit	Benefit to irrigated agriculture.
	Hydropower Benefit	Benefit to hydropower production.
	Recreation Benefit	Benefit to recreation directly or indirectly linked with water.
(2) NED Costs		Monetary costs from a national accounting stance.
	Capital Cost	Cost of initial planning, design, land acquisition, and construction.
	OM&R Cost	Ongoing costs required to assure continued benefits over the life of the project.
(3) RED Effects		Changes in the distribution of regional economic activity.
	Employment	Change in regional employment.
	Income	Change in regional income (employee compensation and proprietor income).
	Output	Change in regional output of goods and services.
(4) Physical Environment		Effects on the abiotic environment.
	Water Quality	Chemical, physical, and biological characteristic of water.
	Erosion and Sedimentation	Deposition of eroded material.
	Aquifer Recharge	Recharge of groundwater basins.
(5) Biological Environment		Effects on the biotic environment.
	Aquatic Habitat	Habitat for fish and aquatic species.
	Riparian Habitat	Habitat for species along rivers and streams.
	Endangered Species	Impact on species at risk of extinction.
(6) Sustainability		Ability to maintain conditions without depletion of natural resources.
(7) Social Effects		Effects on people and communities.
	Environmental Justice	Disparate effects based on race, color, national origin, or income.
	Public Perception	Public view of the alternative, locally and nationally.
(8) Administrative Barriers		Obstacles from regulatory barriers, policy compliance, administrative complexity, or political perception.
(9) Adaptation and Resilience		Ability to recover from or adapt to adverse future changes.

The second criterion, *NED Costs*, reflects monetary costs from a national perspective. This generally captures the cost of capital and OM&R. Capital cost includes initial planning, design, land acquisition, and construction costs. These are financial costs and must be paid before an alternative can start providing benefits. It is also common to include interest during construction (IDC) to capture the opportunity cost of expenses that are incurred before each alternative begins to produce benefits. Meanwhile, OM&R cost reflects the ongoing cost for continued benefits, which stems from operation, maintenance, and replacement expenses over the lifecycle of a project. For an economic or financial analysis, capital cost and OM&R cost are typically calculated separately and either a present value (PV) or annualized value is calculated for comparison purposes. Keeping these costs separate can be important since some alternatives might be formulated to minimize either capital cost or OM&R cost, and project beneficiaries may care more about OM&R cost since they are responsible to pay for OM&R while they may be granted relief from repayment of capital cost.

Capital cost and OM&R cost are included as subcategories here which also allows them to be measured differently, in this case, in total versus annually. Using a PV or annualized value does not influence the results of the trade-off analysis due to the normalization process, provided that the measurement scales are consistent across alternatives. Capital and OM&R costs are most applicable to a supply-side alternative, but it is common for basin studies and other studies to include demand-side alternatives as well. For example, conservation approaches based on incentives, regulations, or price changes. In these settings, one might define similar categories as capital and OM&R, but perhaps broaden the definition to "up-front cost" and "ongoing cost" or "fixed cost" and "variable cost" to encompass both supply- and demand-side alternatives. If relevant, one could further separate cost categories for more detail, such as costs to households, costs to businesses, costs to farmers, costs to treatment facilities, or costs to different levels of government/taxpayers.

The choice of including subcategories for *NED Benefits* and *NED Costs* highlights the flexibility of a trade-off analysis. In general, one should only include subcategories when there is a reasonable justification for doing so. Otherwise, adding subcategories only increases the complexity of the analysis, and doing so might provide little to no insight when comparing alternatives. When there is overlap between subcategories or difficulty in trying to distinguish between them, this generally warrants a single broad criterion without the use of subcategories. The ability to measure subcategories quantitively versus qualitatively is also important to consider. The choice of criteria and subcategories should therefore always be tailored to the setting at hand and the scope and objective of the study.

The third criterion, *RED Effects*, encompasses impacts on the regional economy in terms of income, employment, and output. These impacts result from NED costs and benefits and affect the region where the alternative takes place. In general, both costs (spending) and benefits generate positive effects on the regional economy. However, these effects are not additive with NED effects as they typically reflect shifts from other parts of the country, region, or economy, in which case it may be considered zero-sum from a national accounting stance. These impacts are sometimes interpreted as economic "ripple effects" from the direct costs and benefits, which generally affect the location of the alternative, but can also have spill-over effects on other areas. As shown, RED effects are separated between subcategories for income, employment, and output for the example trade-off analysis. This is done, not to accommodate different preferences (weights) for these subcategories, but rather, because it is not possible to aggregate them due to overlap and different measurement scales. Income is a component of output, both measured in monetary terms, while employment is measured in the number of jobs and linked to income and output. This is covered in more detail later on and each of these subcategories are weighted equally since there is no obvious reason to give preference to any.

The next criterion, *Physical Environment*, includes subcategories for water quality, erosion and sedimentation, and aquifer recharge. Note that erosion and sedimentation overlap some with water quality. That said, it is assumed that there is a meaningful difference between these subcategories, enough to warrant including them separately. For example, water quality is a broad category and can encompass numerous considerations beyond sediment levels, such as nutrient levels, pH, and temperature. Water quality also overlaps some with subcategories included under the criterion *Biological Environment*, which are aquatic habitat, riparian habitat, and endangered species. It is again assumed that there is a meaningful distinction between these subcategories, as well as an ability to measure them separately. If this was not the case, one could simply include a broad criterion that encompasses numerous environmental considerations.

Such a broad environmental criterion might also include sustainability, which is included here separately as the sixth criterion. *Sustainability* is a broad criterion that generally includes environmental considerations as well as others, so it is again assumed that there is a meaningful difference and ability for measurement. For example, an alternative aiming to harness more surface water or pump groundwater might put strain on existing resources, while an alternative to reuse effluent or encourage conservation might be considered more sustainable. This criterion is therefore considered separate and distinctly measurable from the subcategories reflected under the criteria for physical environment and biological environment. Sustainability does not have a universal definition, so as with other qualitative criteria, a careful and clear definition is required.

The next criterion, *Social Effects*, captures environmental justice and public perception. Environmental justice reflects whether an alternative has a disparate impact on individuals based on factors such as race, color, national origin, or income. This may be relevant depending on the beneficiaries and location of an alternative, which could for example, disproportionately impact tribal, urban, or rural communities. Public perception pertains to how the general public views the alternative. Some alternatives are contentious with the local population, or the rest of the public outside of the project area. Certain types of alternatives can also face public perception hurdles, such as direct potable reuse of effluent, which is often stigmatized, or increased prices and regulations, which is generally not favored. Other alternatives might be viewed in a relatively positive light from all perspectives.

The eighth criterion, *Administrative Barriers*, is left as a broad criterion which includes considerations such as regulatory barriers, policy compliance, administrative complexity, and political perception. This category is kept broad since it is difficult to separate and measure these underlying considerations. For example, regulatory barriers and policy compliance might include state and local rules on land and water use. These could also overlap with administrative complexity, which pertains to hurdles associated with implementing and carrying out an alternative. For example, this might include necessary agreements between multiple parties, be it different levels of government, or different local interests, such as tribal, business, or agricultural entities. Political perception is also linked with these considerations, which relates to whether an alternative is likely to come with political impediments. *Administrative Barriers* is therefore left as a broad criterion, which means that it is important to clearly define what the criterion encompasses when it comes to measurement (discussed in the next section).

The last criterion, *Adaptation and Resilience*, is also kept as a broad criterion. In general, adaptation represents the ability for a community to adapt to adverse future changes, such as drought, climate change, or natural disturbances. Resilience is similar but pertains to the ability to recover from adverse changes. The key distinction is that adaptation refers to changing as things change, while resilience refers to returning to how things were after things change. In practice it is difficult to distinguish between these two, and the underlying factors that improve one tend to also improve the other. For example, using groundwater recharge facilities to firm water supply improves both adaptation and resilience. That said, both of these are also similar to mitigation, which generally pertains to preventing adverse changes from occurring in the first place, but can also include avoiding, reducing, rectifying, or compensating for impacts. Mitigation could therefore warrant inclusion as an independent criterion, for example, if an alternative is linked with safety considerations and the likelihood of a dam or canal failure event and subsequent flood impacts.

4.3 Example – Criteria Measurement

Now that the evaluation criteria and subcategories are decided and defined, it is time to determine how to measure each. Note that this process is often iterative, meaning that it is common to go back and refine the list of evaluation criteria as one begins measurement. This may arise due to difficulty during measurement or due to a realization that criteria ought to be grouped or further divided. The first step in measuring criteria is to determine those which can be measured quantitively versus those that must be measured qualitatively. Recall that the distinction between cardinal and ordinal data is used in this technical memorandum to differentiate quantitative versus qualitative criteria. In general, quantitative measurement is preferable since it is more informative and objective. However, it is often difficult to measure everything quantitatively, so one must rely on qualitative measurement to ensure that all relevant criteria are included in the analysis. Accommodating both quantitative and qualitative considerations is the key strength of a trade-off analysis over other forms of assessment.

Recall that it is important that each criterion has directionality. This is particularly crucial when it comes to measurement. Criteria such as *NED Benefits* and *NED Costs* have clear directionality. Specifically, higher benefits and less costs are more desirable. Meanwhile, the directionality of other criteria, such as *Social Effects*, is not immediately obvious. This means that it is important to clearly define and assign directionality for the purpose of measurement. When a criterion is divided into subcategories, the subcategories are what gets measured, which are then scored and ultimately aggregated to capture the main criterion. The criteria without subcategories are instead measured directly. Table 4.2 provides a brief description of how the criteria and subcategories are measured.

For the quantified effects, notice that some are measured in total (e.g., NED Benefits) and others are measured on an annual basis (e.g., OM&R and RED Effects), which has no bearing on the trade-off analysis. For those effects measured qualitatively, the measurement scale also varies. Water quality could be measured quantitatively in several ways, but in this example, a simple qualitative scale is used to indicate whether there is a decrease, increase, or no impact. This might encompass changes in turbidity, total dissolved solids, nutrient levels, pH, temperature, etc. considerations. The downside of using such a broad qualitative measurement scale is that there is no distinction between alternatives which may impact these different aspects of water quality. The upside of using a broad qualitative approach is that there is no need to decide between these different measures, and an improvement along any of these dimensions is captured by the criterion.

Table 4.2 – Example Evaluation Criteria Measurement

Criterion/Subcategory		Measurement
NED Benefits		
	M&I Benefit	Monetary (total)
	Irrigation Benefit	Monetary (total)
	Hydropower Benefit	Monetary (total)
	Recreation	Monetary (total)
NED Costs		
	Capital	Monetary (total)
	OM&R	Monetary (annual)
RED Effects		
	Employment	Number of Jobs (annual)
	Income	Monetary (annual)
	Output	Monetary (annual)
Physical Environment		
	Water Quality	Qualitative (1=decreases, 2=no impact, 3=increases)
	Erosion and Sedimentation	Qualitative (1=increases, 2=no impact, 3=decreases)
	Aquifer Recharge	Groundwater volume (annual)
Biological Environment		
	Aquatic Habitat	Reservoir/lake surface area
	Riparian Habitat	River/stream surface area
	Endangered Species	Qualitative (0=no impact, 1=supports)
Sustainability		Qualitative (1=decreases, 2=no impact, 3=increases)
Social Effects		
	Environmental Justice	Qualitative (0=concerns exist, 1=no impact)
	Public Perception	Qualitative (1=negative, 2=no impact, 3=positive)
Administrative Barriers		Qualitative (1=high, 2=medium, 3=low, 4=none).
Adaptation and Resilience		Qualitative (1=no impact, 2=low, 3=medium, 4=high).

For the criteria and subcategories measured qualitatively, notice the scale and categories used, which are defined to impose directionality. For these, the categories are represented numerically, and defined such that a higher value is preferable. For example, water quality is defined such that 1=decreases, 2=no impact, and 3=increases. The same scale is used for the criterion for sustainability. Meanwhile, erosion and sedimentation represents a negative impact and is thus defined in the opposite direction, where 1=increases, 2=no impact, and 3=decreases. For public perception, 1=negative, 2=no impact, 3=positive. This makes it easier to set up a decision matrix with criteria to be maximized, avoiding the need to convert negative impacts by changing the sign or taking the reciprocal.

A similar approach is used for administrative barriers, but instead the scale goes from 1 to 4 where 1=high, 2=medium, 3=low, and 4=none. For adaptation and resilience, the scale again goes from 1 to 4, but instead 1=no impact, 2=low, 3=medium, and 4=high. These definitions ensure that a higher number is preferable. For these two criteria, decimals are permitted, which allows one to measure effects between the defined categories, for example medium-low or medium-high impacts. For the other categorical scales, permitting the use of decimals does not make as much sense. That said, when using a survey for qualitative measurement, it is common to wind up with decimals when relying on an average across respondents. In this instance, one can use the median, round the average, or use the average and retain the decimals to capture variation across responses.

Some criteria and subcategories are measured qualitatively using a binary (0 or 1) scale. The implied assumption is that the alternatives being evaluated only increase/decrease or have no impact along these considerations. A binary scale is used for endangered species where 0=no impact and 1=supports. For environmental justice, a binary scale is used where 0=concerns exist and 1=no impact. The variation in measurement across criteria and subcategories highlights the flexibility of a trade-off analysis. That said, keep in mind that it is important that the alternatives rely on the same scale for the criteria and subcategories used. This is crucial for the normalization and scoring process covered in the next section. Keep in mind that when relying on qualitative measurement, there is always a degree of subjectivity, so it is ideal to get input from those with expertise related to the criterion. It is also important to get input and consensus among all study members and key stakeholders when defining and measuring criteria. This is necessary to ensure that the process is transparent, and the results are meaningful.

4.4 Example – Criteria Scoring

Once all criteria and subcategories have been measured, quantitively or qualitatively, the next step is to score alternatives along each consideration using an equivalent scale. To score everything along the same scale, normalization is needed. This is required to aggregate effects across criteria and necessary since a trade-off analysis compares alternatives purely on a relative basis. Without normalization, the analysis will implicitly give additional weight to those criteria with a larger scoring scale and less weight to those with a smaller scale. Before normalization, best practice is to set up a decision matrix and convert all negative impacts to positive impacts so that the goal is a straightforward maximization problem. Table 4.3 shows the decision matrix for the five alternatives in this hypothetical example, indicating the performance of each alternative across each consideration. Note that negative effects are not yet converted in the table below. This is done in the following table.

Table 4.3 – Example Decision Matrix

Criterion/Subcategory	Alternative 1	Alternative 2	Alternative 3	Alternative 4	Alternative 5
NED Benefits					
M&I	$2.5b	$1.5b	$3.5b	$3.0b	$1.5b
Irrigation	$2.0b	$3.5b	$0.7b	$1.5b	$0.6b
Hydropower	$1.6b	$0.8b	$0.3b	$1.0b	$0.3b
Recreation	$0.6b	$0.1b	$0.2b	$1.5b	$0.1b
NED Costs					
Capital Cost	$1.6b	$1.5b	$2.5b	$2.2b	$2.5b
OM&R Cost	$10m	$15m	$5m	$30m	$30m
RED Effects					
Income	$50m	$65m	$60m	$55m	$30m
Employment	100 jobs	75 jobs	150 jobs	120 jobs	75 jobs
Output	$0.8b	$0.6b	$1.0b	$0.7b	$0.5b
Physical Environment					
Water Quality	2	2	1	3	1
Eros. and Sed.	2	3	1	2	1
Aquifer Recharge	100 AF	35 AF	40 AF	120 AF	20 AF
Biological Environment					
Aquatic Habitat	25 acres	20 acres	10 acres	30 acres	10 acres
Riparian Habitat	20 acres	10 acres	25 acres	15 acres	5 acres
Endang. Species	0	1	0	1	0
Sustainability	2	3	1	2	1
Social Effects					
Enviro. Justice	1	0	1	1	0
Pub. Perception	1	2	3	2	1
Administrative Barriers	1.5	3	2	3.5	1
Adapt. and Resilience	3.5	1.5	4	3	1

At this point, it is worthwhile to see if any alternative dominates, or is dominated by, the other alternatives. If any one alternative dominates all others (i.e., performs better along at least one criterion or subcategory, and not worse along the remaining criteria and subcategories), it should be chosen, and no further analysis is needed. If any alternative is dominated by the others (i.e., performs worse along at least one criterion or subcategory, and not better along the remaining criteria and subcategories), it can be eliminated. In these instances there are no trade-offs between alternatives and the number of alternatives can be narrowed down.

This is the case with Alternative 5, which is dominated by the other alternatives. Looking at Table 4.3, this alternative performs the worst along several criteria and subcategories, and not better along any criterion or subcategory. This alternative can therefore be removed from the analysis, and the remaining alternatives carried forward to evaluate in the trade-off analysis. Table 4.4 shows the remaining four alternatives and takes the reciprocal of capital cost and OM&R cost so that everything is now to be maximized. Taking the reciprocal of negative effects is generally preferable to changing the sign as it simplifies the normalization process. To avoid large decimals, one can change the unit of measurement before taking the reciprocal (e.g., from dollars to millions of dollars).

Table 4.4 – Example Modified Decision Matrix

Criterion/Subcategory	Alternative 1	Alternative 2	Alternative 3	Alternative 4
NED Benefits				
M&I	$2.5b	$1.5b	$3.5b	$3.0b
Irrigation	$2.0b	$3.5b	$0.7b	$1.5b
Hydropower	$1.6b	$0.8b	$0.3b	$1.0b
Recreation	$0.6b	$0.1b	$0.2b	$1.5b
NED Costs (reciprocal)				
Capital Cost	0.63	0.67	0.40	0.45
OM&R Cost	0.10	0.07	0.20	0.03
RED Effects				
Income	$50m	$65m	$60m	$55m
Employment	100 jobs	75 jobs	150 jobs	120 jobs
Output	$0.8b	$0.6b	$1.0b	$0.7b
Physical Environment				
Water Quality	2	2	1	3
Eros. and Sed.	2	3	1	2
Aquifer Recharge	100 AF	35 AF	40 AF	120 AF
Biological Environment				
Aquatic Habitat	25 acres	20 acres	10 acres	30 acres
Riparian Habitat	20 acres	10 acres	25 acres	15 acres
Endang. Species	0	1	0	1
Sustainability	2	3	1	2
Social Effects				
Enviro. Justice	1	0	1	1
Pub. Perception	1	2	3	2
Administrative Barriers	1.5	3	2	3.5
Adapt. and Resilience	3.5	1.5	4	3

Now that the decision matrix represented by Table 4.4 excludes dominant/dominated alternatives and represents all effects to be maximized, the next step is normalization. For this example, the percentage of maximum technique is used, which scores everything from 0 to 1 and forces at least one alternative to score a one for every criterion and subcategory. This ensures that there is no implicit weighting when aggregating criteria scores, prior to determining explicit weights to capture different importance for criteria and subcategories. Table 4.5 provides a decision matrix with normalized scores. For those criterion with subcategories, the normalized scores for subcategories are aggregated and then normalized again using the percentage of maximum method to derive a 0 to 1 score for each criterion. This is shown in parentheses above the subcategories. Doing so is important so that each criterion is treated equivalently. Without this, if one simply aggregated all subcategory scores to derive a total score, those criterion with more subcategories would have more influence on the overall score.

Table 4.5 – Example Normalized Decision Matrix

Criterion/Subcategory	Alternative 1	Alternative 2	Alternative 3	Alternative 4
NED Benefits	**(0.92)**	**(0.69)**	**(0.52)**	**(1.00)**
M&I	0.71	0.43	1.00	0.86
Irrigation	0.57	1.00	0.20	0.43
Hydropower	1.00	0.50	0.19	0.63
Recreation	0.40	0.07	0.13	1.00
NED Costs	**(0.90)**	**(0.83)**	**(1.00)**	**(0.53)**
Capital Cost	0.94	1.00	0.60	0.68
OM&R Cost	0.50	0.33	1.00	0.17
RED Effects	**(0.76)**	**(0.72)**	**(1.00)**	**(0.80)**
Income	0.77	1.00	0.92	0.85
Employment	0.67	0.50	1.00	0.80
Output	0.80	0.60	1.00	0.70
Physical Environment	**(0.81)**	**(0.73)**	**(0.38)**	**(1.00)**
Water Quality	0.67	0.67	0.33	1.00
Eros. and Sed.	0.67	1.00	0.33	0.67
Aquifer Recharge	0.83	0.29	0.33	1.00
Biological Environment	**(0.63)**	**(0.79)**	**(0.51)**	**(1.00)**
Aquatic Habitat	0.83	0.67	0.33	1.00
Riparian Habitat	0.80	0.40	1.00	0.60
Endang. Species	0.00	1.00	0.00	1.00
Sustainability	**0.67**	**1.00**	**0.33**	**0.67**
Social Effects	**(0.67)**	**(0.33)**	**(1.00)**	**(0.83)**
Enviro. Justice	1.00	0.00	1.00	1.00
Pub. Perception	0.33	0.67	1.00	0.67
Administrative Barriers	**0.43**	**0.86**	**0.57**	**1.00**
Adapt. and Resilience	**0.88**	**0.38**	**1.00**	**0.75**

Looking at Table 4.5, notice that at least one alternative receives a one for each criterion and subcategory, which indicates the best-performing alternative. Alternative 1 does not perform the best along any criterion and Alternative 2 performs the best only along Sustainability. Meanwhile, Alternative 3 performs the best along NED Costs, RED Effects, Social Effects, and Adaptation and Resilience, while Alternative 4 performs the best along NED Benefits, Physical Environment, Biological Environment, and Administrative Barriers. This highlights some of the trade-offs when deciding between the alternatives. Aggregating the normalized criteria scores would provide the unweighted performance of each alternative, before assigning different levels of importance to criteria and subcategories, which is done in the next section. This aggregation is shown later on, and best practice is to compare the overall unweighted performance with the overall weighted performance.

At this stage, it is desirable to visually examine the unweighted performance of each alternative along each consideration. With several criteria and subcategories, either a bar chart or radar chart (sometimes called a "spider chart") is suggested. Figures 4.1 through 4.4 below provide a radar chart for each alternative. Radar charts are shown across the nine criteria, as well as broken out for the subcategories, which are grouped along economic, environmental, and social considerations. This information is the same as that reported in Table 4.5, but instead presented visually. This makes it easy to quickly see how each alternative performs along each dimension, prior to weighting them for importance.

Looking at Figure 4.1 we see that, although Alternative 1 does not perform the best for any criterion, it performs relatively well along all of them, other than administrative barriers. For economic considerations, this alternative performs the best for the hydropower benefit subcategory and does not do well in terms of recreation and irrigation benefits. This alternative also comes with a relatively low capital cost but a high OM&R cost. Looking at environmental considerations, Alterative 1 performs moderately, but offers nothing for endangered species. For social considerations, this alternative does well along environmental justice and adaption and resilience, but poorly along public perception and administrative barriers.

From Figure 4.2 we see that Alternative 2 comes with several trade-offs. In particular, this alternative performs well in terms of irrigation benefit, capital cost, and regional income, but poorly in terms of recreation benefit, M&I benefit, and OM&R cost. For environmental considerations, this alternative performs well in terms of sustainability, endangered species, and erosion and sedimentation, but poorly in terms of aquifer recharge and riparian habitat. Along social considerations, Alternative 2 offers nothing for environmental justice and little for adaptation and resilience but does perform well for public perception and administrative barriers.

Figure 4.3 shows the unweighted performance for Alternative 3. As previously indicated by Table 4.3, this alternative is the top-performer in terms of NED costs, regional economic effects, social effects, and adaptation and resilience. That said, Alternative 3 primarily benefits M&I and the regional economy, but does little in terms of irrigation, hydropower, and recreation benefits. Looking at environmental considerations, we see that this alternative does very poorly along nearly every dimension, other than riparian habitat. For social considerations, this alterative performs very well, but it is not the top-performer for administrative barriers.

Looking at Figure 4.4 we see that Alternative 4 is relatively well-rounded. This alternative is the top-performing alternative along several dimensions and performs relatively well along the remaining considerations. That said, this alternative comes with the highest OM&R cost and a relatively high capital cost, as well as relatively low irrigation and hydropower benefits. For environmental considerations, Alternative 4 appears to outperform the others overall, but it does not necessarily perform the best along every dimension. The same applies for social considerations. This highlights that even when an alternative performs well overall, there still exists several trade-offs that are important to consider. These radar charts highlight how each alternative performs along each consideration before weighting, making it easy to quickly visualize the trade-offs. After weighting for importance, some of these trade-offs may become even more important, and the relative performance between alternatives might change. This is covered in detail in the next section.

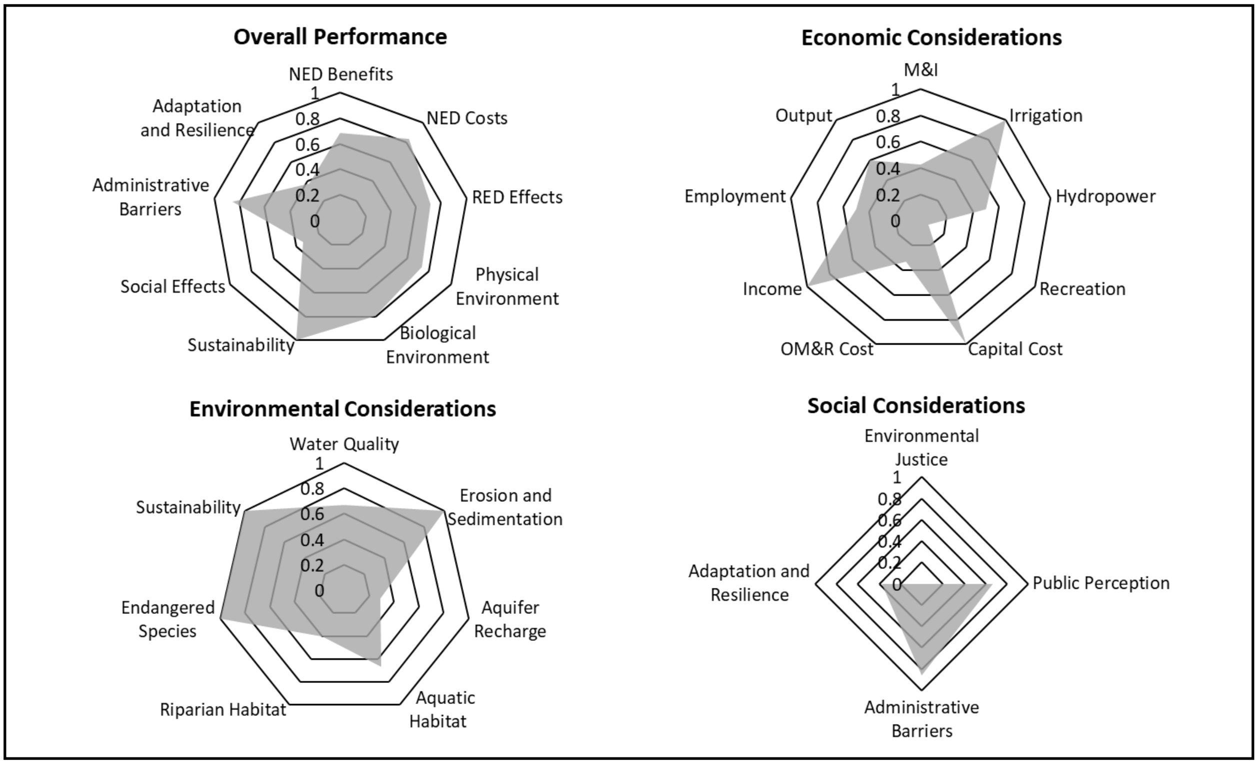

Figure 4.1 – Example Alternative 1 Performance.

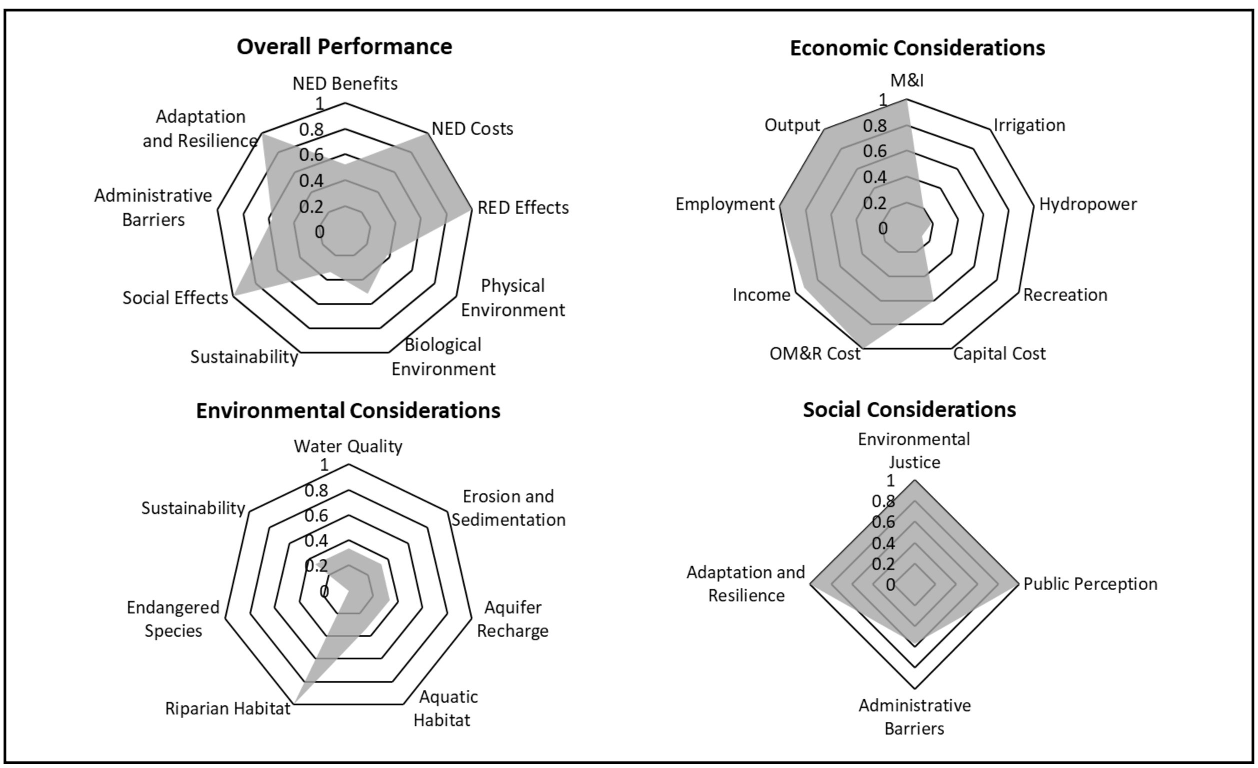

Figure 4.2 – Example Alternative 2 Performance.

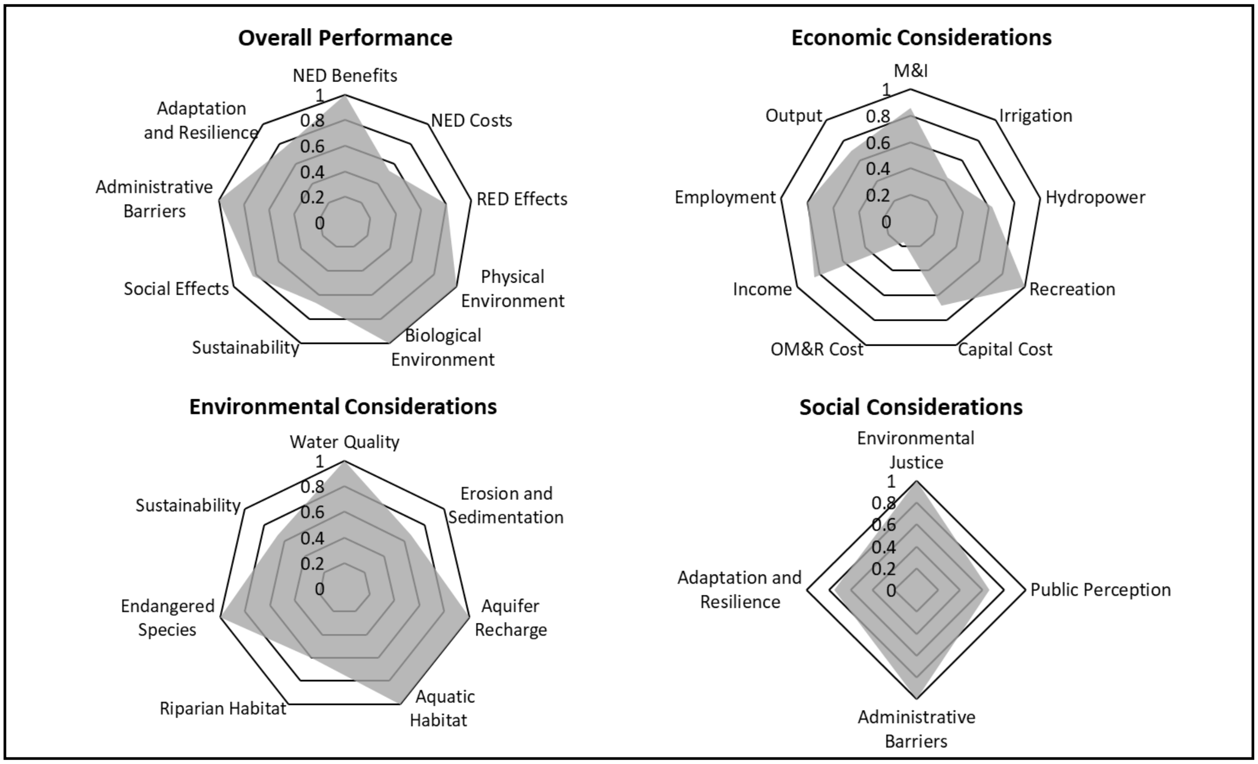

Figure 4.3 – Example Alternative 3 Performance.

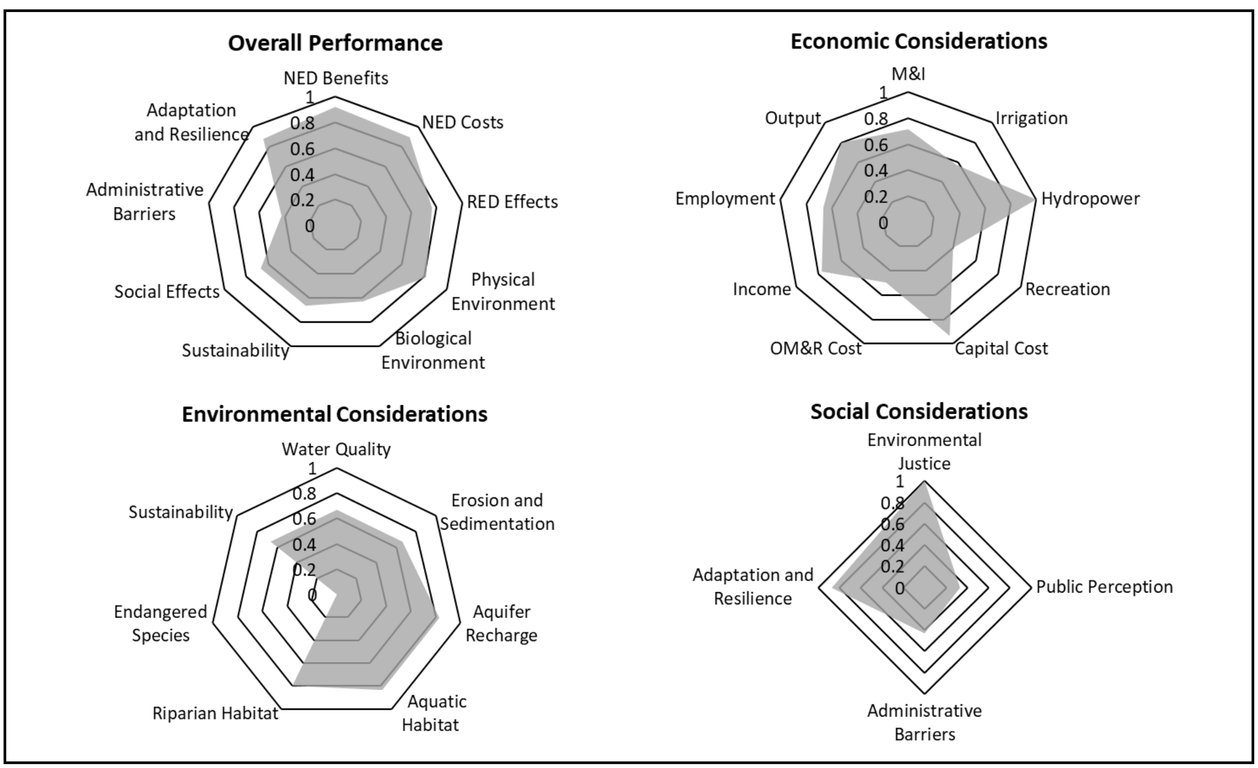

Figure 4.4 – Example Alternative 4 Performance.

4.5 Example – Criteria Weighting

In many cases, some evaluation criteria are deemed more important than others. Criteria weighting is used to assign different levels of importance to criteria. Weighting is the most subjective part of a trade-off analysis, so the process and assumptions should always be transparent. It is also best practice to compare the unweighted performance of alternatives with the weighted performance to highlight the influence of weighting on the final results. Earlier in this technical memorandum, attention was placed on the fixed-point and rating methods of weighting. The fixed-point and rating methods are similar, but the key difference is that the former involves trade-offs when assigning importance while the latter is unconstrained. Specifically, the fixed-point technique is constrained since greater importance can only be given to one criterion by giving less to another. Nonetheless, the fixed-point method is straightforward and therefore used for the example provided here.

To implement fixed-point weighting, percentages are assigned to each criterion and subcategory such that final weights add to 100 percent. Table 4.6 below shows the weights assumed for each criterion and subcategory in our example. Ideally, these weights are based on the preferences and expert judgment of those involved in the study and the weights align with the overall objectives of the study. Different individuals will always have different views on which weights ought to be used, so in general, getting input from all perspectives and reaching a consensus on weighting will provide the most meaningful set of weights to use. When consensus is not easily obtained and there exists a wide range of views, multiple sets of weights can be tested, highlighting how alternatives perform under different weighting schemes. This then leaves it up to the reader and decision makers as to which set they view as the most appropriate. This might also indicate that the choice of weights has minimal bearing on the final results, in which case weighting becomes less controversial.

Looking at Table 4.6, a 15 percent weight is assigned to *NED Benefits*, *Physical Environment*, *Biological Environment*, and *Adaptation and Resilience*, a 10 percent weight is assigned to *NED Costs*, *Sustainability*, and *Social Effects*, and a 5 percent weight is assigned to *RED Effects* and *Administrative Barriers*, for a total of 100 percent. In effect, those criteria with a 10 percent weight are given twice the importance as those with a 5 percent weight, and those criteria with a 15 percent weight are given thrice the importance as those with a 5 percent weight. For criteria without subcategories, these are the final weights applied. For those criteria with subcategories, weights are assigned to the subcategories to sum to 100 percent for each criterion. The subcategory weights are then multiplied by the respective criterion weight to derive the final weight used for each subcategory. To calculate the weighted performance of each alternative, these weights are applied as a decimal so that the weights sum to one (i.e., no normalization is needed), and the decimals are multiplied by the normalized values from Table 4.5. The resulting decision matrix of weighted values is shown in Table 4.7.

Table 4.6 – Example Criteria and Subcategory Weights

Criterion	Criterion Weight	Subcategory	Subcategory Weight	Final Weight
NED Benefits	15%	M&I Benefit Irrigation Benefit Hydropower Benefit Recreation Benefit	30% 30% 20% 20%	4.5% 4.5% 3% 3%
NED Costs	10%	Capital Cost OM&R Cost	60% 40%	6% 4%
RED Effects	5%	Income Employment Output	33.3% 33.3% 33.3%	1.67% 1.67% 1.67%
Physical Environment	15%	Water Quality Erosion and Sedimentation Aquifer Recharge	60% 20% 20%	9% 3% 3%
Biological Environment	15%	Aquatic Habitat Riparian Habitat Endangered Species	33.3% 33.3% 33.3%	5% 5% 5%
Sustainability	10%	N/A	N/A	10%
Social Effects	10%	Environmental Justice Public Perception	80% 20%	8% 2%
Administrative Barriers	5%	N/A	N/A	5%
Adaptation and Resilience	15%	N/A	N/A	15%
			Total	**100%**

After weighting, the values become smaller and it is no longer the case that at least one alternative scores a one for each criterion and subcategory, as was the case with the unweighted scores in Table 4.5. The proportionality of these values now reflects the performance of each alternative as well as the importance of each consideration. At this point, one might be inclined to try and normalize the weighted values in Table 4.7 so that they are comparable with the unweighted values in Table 4.5. However, doing so would remove the proportionality generated by the weighting process. Instead, one must aggregate the weighted values across all subcategories and criteria, which ensures that the weights are maintained. After aggregation, the weighted total scores can be normalized and compared with the normalized unweighted total scores, as shown in the next section.

Table 4.7 – Example Weighted Decision Matrix

Criterion/Subcategory	Alternative 1	Alternative 2	Alternative 3	Alternative 4
NED Benefits				
M&I	0.032	0.019	0.045	0.039
Irrigation	0.026	0.045	0.009	0.019
Hydropower	0.030	0.015	0.006	0.019
Recreation	0.012	0.002	0.004	0.030
NED Costs				
Capital Cost	0.056	0.060	0.036	0.041
OM&R Cost	0.020	0.013	0.040	0.007
RED Effects				
Income	0.013	0.017	0.015	0.014
Employment	0.011	0.008	0.017	0.013
Output	0.013	0.010	0.017	0.012
Physical Environment				
Water Quality	0.060	0.060	0.030	0.090
Eros. and Sed.	0.020	0.030	0.010	0.020
Aquifer Recharge	0.025	0.009	0.010	0.030
Biological Environment				
Aquatic Habitat	0.042	0.033	0.017	0.050
Riparian Habitat	0.040	0.020	0.050	0.030
Endang. Species	0.000	0.050	0.000	0.050
Sustainability	0.067	0.100	0.033	0.067
Social Effects				
Enviro. Justice	0.080	0.000	0.080	0.000
Pub. Perception	0.007	0.013	0.020	0.007
Administrative Barriers	0.021	0.043	0.029	0.050
Adapt. and Resilience	0.131	0.056	0.150	0.113

4.6 Example – Interpreting Results

To derive the total unweighted score for each alternative, the normalized unweighted values from Table 4.5 are summed across the nine criteria. This total score treats all nine criteria equivalently. Since at least one alternative scores a one for each criterion, the maximum possible overall score is nine and the total scores range from 0 to 9. That said, note that a score of nine would only be possible if an alternative dominates all others, in which case a trade-off analysis is not actually warranted. To derive the total weighted score for each alternative, the weighted values from Table 4.7 are summed across all criteria and subcategories. This total score accounts for the importance of each criterion as well as the importance of underlying subcategories. After weighting, the maximum possible overall score is one with total scores ranging from 0 to 1. Again, the maximum score would only be possible if an alternative dominates all others.

In order to compare the unweighted performance of each alternative with the weighted performance, both the unweighted and weighted total scores are normalized using the percentage of maximum method. Table 4.8 shows the total unweighted and weighted score for each alternative, both before and after normalization. Although Alternative 5 was previously removed from the analysis since it is dominated by all other alternatives, it is included here for discussion.

Table 4.8 – Example Unweighted and Weighted Results

Alternative	Unweighted Performance			Weighted Performance		
	Total Score	Normalized Total Score	Rank	Total Score	Normalized Total Score	Rank
Alternative 1	6.66	0.88	2nd	0.71	0.90	2nd
Alternative 2	6.33	0.84	3rd	0.60	0.77	4th
Alternative 3	6.32	0.83	4th	0.62	0.79	3rd
Alternative 4	7.58	1.00	1st	0.79	1.00	1st
Alternative 5	2.83	0.37	5th	0.27	0.34	5th

Looking back at the unweighted normalized values in Table 4.5, recall that Alternative 3 performs the best along four criteria, Alternative 2 performs the best along one criterion, and Alternative 1 does not perform the best along any criterion. One might look at this fact and naively conclude that Alternative 3 must outperform Alternative 2 and both must outperform Alternative 1. However, both the unweighted and weighted scores shown in Table 4.8 actually indicate that Alternative 1 outperforms both Alternative 2 and Alternative 3. This occurs because even if an alternative doesn't perform the best for any one criterion, it may perform well across numerous criteria together, leading it to outperform other alternatives that only perform well along certain criterion. This highlights the strength of a trade-off analysis, which identifies how alternatives perform across several considerations at once, accounting for a myriad of effects as well as the importance of those effects.

Another important insight shown in Table 4.8 is that Alternative 2 and Alternative 3 perform very similar overall, both before and after weighting. Prior to weighting, Alternative 2 slightly outperforms Alternative 3, but after weighting for importance, Alternative 3 slightly outperforms Alternative 2. This highlights the role of weighting, which can impact the relative performance of each alternative, and potentially change which alternative is deemed preferable. When alternatives score similarly prior to weighting, it is not uncommon for weighting to switch the relative ranking of those alternatives. Before weighting, the total score for Alternative 2 is 84 percent of the top-performing alternative and the total score for Alternative 3 is 83 percent of the top-performing alternative. This is indicated by the normalized total scores in Table 4.8. After weighting, the total score for Alternative 2 is 77 percent of the top-performing alternative and the total score for Alternative 3 is 79 percent of the top-performing alternative. This shows that the overall performance is similar between these alternatives, and that the relative performance switches after weighting. Meanwhile, the total score for Alternative 1 is 88 percent of the top-performing alternative before weighting and 90 percent after weighting, indicating that it outperforms both Alternative 2 and Alternative 3, before and after weighting.

In this example, Alternative 4 is the top-performing alternative, both before and after weighting. This suggests that Alternative 4 is a clear winner. However, it is not always the case that the top-performing alternative is the same with and without weighting, in which case the validity of the weighting process is crucial. Even when an alternative outperforms the others before and after weighting, it is still not appropriate to simply consider that alterative the "best." This is because weighting is subjective, and a set of weights could conceivably be devised such that the top-performing alternative differs with versus without weighting. This holds true as long as there exists no dominant alternative, meaning that there exists trade-offs between alternatives. This is why a trade-off analysis cannot ever identify a "best" alternative, but rather, identifies the top-performing alternative under a given framework and set of assumptions. That said, Alternative 5 is definitively the worst-performing alternative. This is not just because it performs the worst before and after weighting, but because the alternative is dominated by all other alternatives, as previously discussed in Section 4.4. Put differently, there is no conceivable set of weights that could possibly lead Alternative 5 to perform better than any other alternative. This is why this alternative can be removed from the analysis early on.

It is important to realize that the information presented in Table 4.8 is purely relative and only pertains to the alternatives included in the analysis. Put differently, this information shows how alternatives perform relative to one another. This information cannot be used to describe the performance of an alternative in isolation or compared with other alternatives not included in the analysis. The information presented in Table 4.3 showed how alternatives perform in absolute terms for each criterion and subcategory. After the analysis proceeds and normalization is done, the information is no longer absolute and instead becomes relative, reflecting how the alternatives perform relative to one another. This feature is sometimes overlooked, and this highlights why it is important to ensure that the list of alternatives and evaluation criteria are exhaustive, at least for the scope and objective of the study at hand. If a new alterative, criterion, or subcategory is added to the framework, this necessitates going back and redoing the analysis from the normalization and scoring step. The one exception to making changes is weighting, which comes after the normalization step. After the results are presented, there may be a desire to revisit the weighting process, in some cases to vary the weights used to see how this might influence the final results. This can serve as a sensitivity analysis for the weighting assumptions, which are again the most subjective part of a trade-off analysis.

Another way to present the results is to break them down into groupings, such as economic, environmental, and social considerations. Recall that the example provided here includes three economic criteria, three environmental criteria, and three social criteria. The values presented in Table 4.5 showed how each alternative performs along each criterion and subcategory, but it can also be useful to see the performance grouped along each of these dimensions. This helps to further highlight where the trade-offs are across the alternatives and identify strengths and weaknesses. Table 4.9 shows the performance of each alternative along economic considerations, both before and after weighting for the importance of criterion and subcategories. The unweighted score comes from summing the normalized values from Table 4.5 for the three economic criteria. The weighted score comes from summing the weighted values from Table 4.7 for the economic subcategories, capturing the importance of each subcategory as well as the economic criterion that it falls under.

Table 4.9 – Example Economic Performance

Alternative	Unweighted Performance			Weighted Performance		
	Score	Normalized Score	Rank	Score	Normalized Score	Rank
Alternative 1	2.59	1.00	1st	0.213	1.00	1st
Alternative 2	2.24	0.87	4th	0.190	0.89	3rd
Alternative 3	2.52	0.98	2nd	0.188	0.88	4th
Alternative 4	2.33	0.90	3rd	0.193	0.91	2nd
Alternative 5	1.27	0.49	5th	0.102	0.48	5th

Table 4.10 shows the performance of each alternative along environmental considerations, both before and after weighting. The unweighted score comes from summing the normalized values from Table 4.5 for the three environmental criteria. The weighted score comes from summing the weighted values from Table 4.7 for the environmental criteria and subcategories. Note that one of the three environmental criteria does not have subcategories, in which case the criterion score is used.

Table 4.10 – Example Environmental Performance

Alternative	Unweighted Performance			Weighted Performance		
	Score	Normalized Score	Rank	Score	Normalized Score	Rank
Alternative 1	2.11	0.79	3rd	0.253	0.75	3rd
Alternative 2	2.53	0.95	2nd	0.302	0.90	2nd
Alternative 3	1.22	0.46	4th	0.150	0.45	4th
Alternative 4	2.67	1.00	1st	0.337	1.00	1st
Alternative 5	0.85	0.32	5th	0.105	0.31	5th

Table 4.11 shows the performance of each alternative along social considerations, both before and after weighting. The unweighted score comes from summing the normalized values from Table 4.5 for the three social criteria. The weighted score comes from summing the weighted values from Table 4.7 for the social criteria and subcategories. Note that two of the three social criteria do not have subcategories, in which case the criterion score is used.

Table 4.11 – Example Social Performance

Alternative	Unweighted Performance			Weighted Performance		
	Score	Normalized Score	Rank	Score	Normalized Score	Rank
Alternative 1	1.97	0.76	3rd	0.239	0.86	3rd
Alternative 2	1.57	0.61	4th	0.112	0.40	4th
Alternative 3	2.57	1.00	2nd	0.279	1.00	1st
Alternative 4	2.58	1.00	1st	0.256	0.92	2nd
Alternative 5	0.70	0.27	5th	0.058	0.21	5th

Looking at Tables 4.9 through 4.11, we see how the alternatives perform along economic, environmental, and social considerations. While Alternative 4 is the top-performing alternative overall, it does not necessarily outperform the other alternatives along each of these dimensions. In fact, the results show that Alternative 4 is only the top-performer for environmental considerations, both before and after weighting. Meanwhile, Alternative 1 performs the best along economic considerations, before and after weighting. If a decisionmaker decides that they only truly care about economic considerations, they would advocate for Alternative 1, followed by Alternative 3, and then finally Alternative 4. For social considerations, Alternative 3 is essentially tied with Alternative 4 for performing the best before weighting, but Alternative 3 outperforms Alternative 4 after weighting. This again highlights that a trade-off analysis cannot ever identify the "best" alternative, as there are important trade-offs when selecting between them, and there is always an element of subjectivity. If this wasn't so, this would imply that there exists a dominant strategy, in which case a trade-off analysis is not warranted, and the choice becomes obvious.

As shown in Table 4.10, the reason Alternative 3 performs so poorly overall is because it does not do well along environmental considerations. However, this alternative does perform well along both economic and social dimensions. Meanwhile, Alternative 2 performs poorly along social and economic considerations, which is why it does not rank well overall. Looking at Alterative 5, it does not outperform any alternative along any dimension, before or after weighting. As previously emphasized, this is always to be expected whenever an alternative is dominated. Even though Alternative 4 does not perform the best along every dimension, it does perform relatively well along all of them, which explains why it is the top-performing alternative overall. Separating the performance of each alternative along each of these dimensions helps to highlight where the overall ranking and performance in Table 4.8 comes from. This helps to inform decision making and accomplish the main purpose of a trade-off analysis; to identify the strengths and weaknesses of each alternative and determine the overall performance across both quantifiable and unquantifiable considerations.

5.0 Conclusion

This technical memorandum provides guidance on conducting a systematic trade-off analysis that conforms with Reclamation's D&S WTR 13-01 and DOI's ASP for implementing the PR&G. A trade-off analysis helps to evaluate so-called "wicked problems," which are those that are highly complex with no unique solution due to conflicting objectives and preferences. A trade-off analysis seeks to evaluate such problems using a systematic and consistent approach, not necessarily to identify the "best" alternative, but rather, to highlight the strengths and weaknesses of each alternative and indicate how each alternative performs across numerous considerations at once. This entails an evaluation of the effects of each alternative, both quantitative and qualitative. The methodology presented here is applicable to WaterSMART Basin Studies as well as a broad range of Reclamation studies evaluating alternatives. The framework allows one to conduct a simple or complex analysis, depending on the scope and objective of the study and availability of information. This helps improve decision making and the selection process between alternatives.

5.1 Limitations

By accommodating just about any range of measurement scales and including both quantitative and qualitative measures, a trade-off analysis is quite flexible and powerful to assist decision making. However, the analysis does not actually identify whether any alternative is economically justifiable (NED benefits outweigh NED costs) and thus welfare enhancing from a national stance. For example, since the WSRV Basin Study measured capital and OM&R costs on a qualitative scale from low-high, costs are measured purely on a relative basis and therefore not comparable with the monetized benefits. For the example analysis provided in this memorandum, capital cost was measured as a total across the construction period while OM&R was kept at the annual level, excluding the need to assume a project lifecycle and calculate a PV. This highlights the flexibility of a trade-off analysis, but also emphasizes that a trade-off analysis is not a substitute for a formal economic analysis, which aims to evaluate whether alternatives actually provide a positive net economic benefit to the nation.

An economic assessment aims to evaluate costs and benefits on a monetary basis since it provides the most objective and accurate universal measurement scale for comparing the effects of each alternative. The normalization process used for a trade-off analysis results in alternatives being compared purely on a relative basis with scores that only have meaning in relation to one another. This is not as powerful or informative as monetary valuations, which are more universally generalizable and allow for comparisons across alternative as well as categories. That said, monetary values can be included in a trade-off analysis, and the key advantage is the flexibility to go beyond that which can be monetized, or even quantified, and measure impacts using qualitative scales. A trade-off analysis is therefore a useful tool for decision making, particularly for screening alternatives, but it should never be considered the only tool for making decisions. If one conducts a detailed and complex trade-off analysis gathering original information, it may be more resource intensive than other forms of analysis, while a simple and highly qualitative analysis with only existing quantitative data can be a quick and effective way to screen alternatives. The rationale for including a trade-off analysis therefore depends on the particular planning process and the scope and objective of the study at hand.

References

Army Corps of Engineers (2002). Trade-Off Analysis Planning and Procedures Guidebook. U.S. Army Corps of Engineers, Institute for Water Resources (IWR 02-R-2).

Bureau of Reclamation (2007). Reclamation Manual, Directives and Standards FAC 09-01 – Cost Estimating. U.S. Bureau of Reclamation.

Bureau of Reclamation (2016). Reclamation Manual, Directives and Standards WTR 13-01 – Basin Studies. U.S. Bureau of Reclamation.

Bureau of Reclamation (2016). Los Angeles Basin Study – The Future of Stormwater Conservation, Task 6 – Trade-Off Analysis & Opportunities. U.S. Bureau of Reclamation, Technical Service Center, Economic Analysis Group, Steven Piper. California-Great Basin Report.

Bureau of Reclamation (2021). West Salt River Valley Basin Study – Economic and Trade-Off Analysis. U.S. Bureau of Reclamation, Technical Service Center, Economic Analysis Group, Matthew Elmer. Lower Colorado Basin Report.

Clemen, Robert and Terence Reilly (2001). Making Hard Decisions. 2nd ed. revised. Pacific Grove, CA: Duxbury Press.

Council on Environmental Quality (2013). Principles and Requirements for Federal Investments in Water Resources. U.S. Council on Environmental Quality (P&R).

Department of Communities and Local Government. (UKCLG, 2009). Multi-Criteria Analysis: A Manual. U.K. Department of Communities and Local Government.

Department of the Interior (2015). Agency Specific Procedures for Implementing the Council on Environmental Quality's Principles, Requirements, and Guidelines for Water and Land Related Resources Implementation Studies. U.S. Department of Interior (707 DM 1 HB).

Forest Service (2013). The Science of Decisionmaking: Applications for Sustainable Forest and Grassland Management in the National Forest System. U.S. Department of Agriculture Forest Service (GTR WO-88).

Hajkowicz, Stefan, Geoff McDonald, and Phil Smith (2000). An Evaluation of Multiple Objective Decision Support Weighting Techniques in Natural Resource Management. *Journal of Environmental Planning and Management* 43(4): 505-518.

Monat, Jamie (2009). The Benefits of Global Scaling in Multi-Criteria Decision Analysis. *Judgment and Decision Making* 4(6): 492-508.

Olander, Lydia, Robert J. Johnston, Heather Tallis, Jimmy Kagan, Lynn Maguire, Steve Polasky, Dean Urban, James Boyd, Lisa Wainger, and Margaret Palmer (2015). Best Practices for Integrating Ecosystem Services into Federal Decision Making. Durham: National Ecosystem Services Partnership, Duke University.

Pomerol, Jean-Charles and Sergio Barba-Romero (2000). Multicriterion Decision in Management: Principles and Practice. International Series in Operations Research & Management Science, Kluwer Academic Publishers; Boston, Massachusetts.

Roy, Bernard (1985, translation 1996). Multicriteria Methodology for Decision Aiding. Nonconvex Optimization and Its Applications, Kluwer Academic Publishers; Boston, Massachusetts.

Water Resources Council (1983). Economic and Environmental Principles and Guidelines for Water and Related Land Resources Implementation Studies. U.S. Water Resources Council (P&G). Gregory, Robin, Lee Failing, Michael Harstone, Graham Long, Tim McDaniels, and Dan Ohlson (2012). Structured Decision Making: A Practical Guide to Environmental Management Choices. Chichester, West Sussex, UK: Wiley-Blackwell.