# Scaling Up Pareto Optimization for Tree Structures with Affine Transformations: Evaluating Hybrid Floating Solar-Hydropower Systems in the Amazon

**Marc Grimson[1], Rafael Almeida[5], Qinru Shi[2], Yiwei Bai[1], Héctor Angarita[3], Felipe Siqueira Pacheco[4], Rafael Schmitt[3], Alexander Flecker[4], Carla P. Gomes[1]**

[1]Department of Computer Science, Cornell University, USA
[2]Center for Applied Mathematics, Cornell University, USA
[3]Natural Capital Project, Stanford University, USA
[4]Department of Ecology and Evolutionary Biology, Cornell University, USA
[5]School of Earth, Environmental, and Earth Sciences, University of Texas Rio Grande Valley, USA
mg2425@cornell.edu, rafael.almeida@utrgv.edu, qs63@cornell.edu, yb263@cornell.edu, haac@stanford.edu, fsp9@cornell.edu, asf3@cornell.edu, gomes@cs.cornell.edu

## Abstract

Sustainability challenges inherently involve the consideration of multiple competing objectives. The Pareto frontier – the set of all optimal solutions that cannot be improved with respect to one objective without negatively affecting another – is a crucial decision-making tool for navigating sustainability challenges as it highlights the inherent trade-offs among conflicting objectives. Our research is motivated by the strategic planning of hydropower in the Amazon basin, one of the earth's largest and most biodiverse river systems, where the need to increase energy production coincides with the pressing requirement of minimizing detrimental environmental impacts. We investigate an innovative strategy that pairs hydropower with Floating Photovoltaic Solar Panels (FPV). We provide a new extended multi-tree network formulation, which enables the consideration of multiple dam configurations. To address the computational challenge of scaling up the Pareto optimization framework to tackle multiple objectives across the entire Amazon basin, we further enhance the state-of-the-art algorithm for Pareto frontiers in tree-structured networks with two improvements. We introduce affine transformations induced by the sub-frontiers to compute Pareto dominance and provide strategies for merging sub-trees, significantly increasing the pruning of dominated solutions. Our experiments demonstrate considerable speedups, in some cases by more than an order of magnitude, while maintaining optimality guarantees, thus allowing us to more effectively approximate the Pareto frontiers. Moreover, our findings suggest significant shifts towards higher energy values in the Pareto frontier when pairing hybrid hydropower with FPV solutions, potentially amplifying energy production while mitigating adverse impacts.

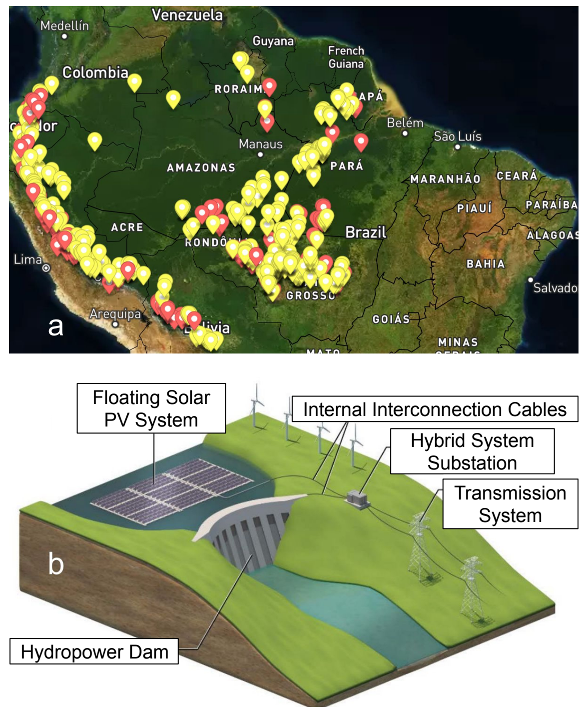


Figure 1: (a) Existing (red) and proposed (yellow) hydropower dams in the Amazon basin and (b) Hybrid hydropower dam with floating photovoltaic solar panels (FPV) system (Lee et al. 2020).

## Introduction

Computational Sustainability (Gomes et al. 2019) is a field within Computer Science that aims to use Artificial Intelligence to work towards a sustainable future. Problems in Computational Sustainability often require finding a balance between conflicting concerns, as captured for example, in the Sustainable Development Goals that aim to ensure economic and social equity across all strata of society while maintaining a healthy environment (United Nations General Assembly 2015). Such trade-offs between conflicting objectives naturally translate to Multi-Objective Combinatorial Optimization Problems (MOCOP) (Ehrgott and Gandibleux 2000; Ehrgott, Gandibleux, and Przybylski 2016; Wiecek et al. 2008). Our goal is to understand these trade-offs and enable stakeholders to make informed decisions based on their preferences and constraints.

Our work is motivated by the need to identify portfolios of hydropower dams that maximize economic, social, and environmental needs (Almeida et al. 2022b). Curbing climate change requires an extraordinary increase in energy generation from renewable sources in the coming decades. While wind and solar power are growing rapidly, hydropower remains the largest source of renewable electricity globally, and construction of new dams is expected to continue. Most new hydropower projects are projected for the Global South, where latent hydropower potential is large. Hotspot areas for future hydropower development include pristine, biologically diverse regions such as the Amazon basin (Flecker et al. 2022). As of 2014, there were over 3,700 large dams ($>$1MW installed capacity) proposed, planned, or under construction across the world (Zarfl et al. 2014), and over 350 in the Amazon basin alone. In addition to energy generation, dams may negatively impact surrounding ecosystems, blocking sediment and river connectivity, and harming biodiversity. Thus, to minimize these adverse effects, it is important to carefully select which dams to build, considering trade-offs among competing objectives (Almeida et al. 2019; Winemiller et al. 2016; Ziv et al. 2012).

Floating Photovoltaic Solar Panels (FPV) have emerged as an attractive application of solar PV that allows for systems to be floated on water bodies. When integrated into hybrid systems with hydropower, FPV can enhance the overall efficiency and value of hydropower systems (Lee et al. 2020). See Figure 1. Placing solar arrays on reservoirs offers several advantages. These floating panels, which can be simply anchored through mooring lines, stay cool near the water, making them about 5% more efficient than land-based panels. Additionally, they shield the water surface, potentially reducing evaporation and preserving water for hydropower, drinking, and irrigation. Moreover, integrating these panels with the pre-existing grid infrastructure of hydropower reservoirs can decrease transmission costs. This "Floatovoltaics" approach can also lower the carbon intensity of some hydropower, particularly in cases where methane emissions from submerged plant matter make them as carbon-intensive as fossil-fuel power plants. Nevertheless, additional environmental and social impacts must be assessed (Almeida et al. 2022a).

Traditionally, the benefits and impacts of dams have been addressed on a site-by-site basis. However, there is growing recognition in research (Opperman et al. 2023), policy (Castaño et al. 2019), and practice domains to move towards decision processes that assess the complex cumulative interactions of projects at river-basin scales to identify optimal development portfolios, with the objective of minimizing the consequences on the integrity of ecological processes of river systems. We have partnered with governmental agencies, academics, non-governmental organizations, and the private sector in Amazonian countries, for data acquisition, ground-truthing of data, and the translation of results to policy- and decision-makers. Given the interdisciplinary nature of this project, our team involves researchers across multiple disciplines (e.g., ecology, hydrology, policy) in addition to computer science. As part of these efforts, we have organized numerous working group meetings to collectively make the developed AI tools and methodologies more accessible to partners in addressing sustainability solutions, and to ensure the developed models are grounded in reality, with tools provided to partners for analysis (see the website https://www.cs.cornell.edu/gomes/udiscoverit/amazon-ecovistas/).

Previous work framed the problem of selecting which dams to build in the Amazon basin as a tree-structured MOCOP arising from converting the underlying river network to a rooted, directed tree network (Gomes-Selman et al. 2018; Wu et al. 2018). The goal is to find the Pareto frontier: the set of all solutions such that no solution is dominated by any other feasible solution. In other words, for a given solution, there is no way to simultaneously improve upon some objective without compromising another. Due to the tree structure, Dynamic Programming (DP) may be used to compute the exact Pareto frontier (Wu et al. 2018). In addition, a rounding technique applied to the exact DP algorithm provides a fully polynomial-time approximation scheme (FPTAS) (Wu et al. 2018; Wu, Sheldon, and Zilberstein 2014a). The FPTAS finds a solution set of polynomial size, which approximates the Pareto frontier within an arbitrary small $\epsilon$ factor and runs in time that is polynomial in the size of the instance and $1/\epsilon$. The DP approach finds the Pareto frontier of a node by combining the frontiers of each child to form the new frontier. This greatly reduces the search space of the problem as we need only consider the Cartesian product of optimal solutions and decisions from each child. However, even with the FPTAS, the number of solutions that need to be considered may be too large to process, especially given the additional challenge of considering two configurations for the dams, with and without FPV. Therefore it is critical to scale up Pareto optimization for tree structures.

**Our contributions.** In this paper, we investigate the innovative strategy of pairing hydropower with Floating Photovoltaic Solar Panels (FPV). To address the scalability challenge, we propose three enhancements to the state-of-the-art algorithm for Pareto frontiers in tree-structured networks (Gomes-Selman et al. 2018) while maintaining the same optimality guarantees of the exact Pareto frontier algorithm and the FPTAS. We then validate our approaches with two use cases in the Amazon River basin. More specifically: **(1)** We provide a new extended ***multi-tree network*** formulation, which enables the consideration of multiple dam configurations. **(2)** We formulate the problem of joining the subfrontiers of a node as an ***affine transformation that preserves Pareto optimality***, allowing for efficient dominance checks. **(3)** We ***dynamically select the children*** to join pairwise based on a ranking heuristic, unlike the previous approach following a static order. **(4)** We validate our approach using real data from the Amazon, the world's largest and most biodiverse river basin, considering six energy and environmental objectives. Our experiments demonstrate a remarkable *reduction in the number of solutions considered by over an order of magnitude or more in most cases*, without losing any guarantees on finding non-dominated solutions, as well as ***better approximate Pareto frontiers*** in practice, especially for extremely large problem instances. **(5)** Our findings also show ***positive shifts towards higher energy values in the Pareto frontier when pairing hybrid hydropower with FPV solutions***, amplifying energy production while mitigating adverse impacts.

## Related Work

For unstructured general multi-objective optimization problems, genetic algorithms have been used for approximating the Pareto frontiers, with Non-dominated Sorting Genetic Algorithm(s) (NSGA (Srinivas and Deb 1994), NSGA-II (Deb et al. 2002), and NSGA-III (Deb and Jain 2013)) and Multi-objective Evolutionary Algorithm Based on Decomposition (MOEA/D) (Zhang and Li 2007) being among the most popular methods. However, because these algorithms do not consider the underlying structure of the problem, they are unable to provide the same theoretical guarantees.

Other methods for identifying the Pareto frontier use ray-based techniques (Lin et al. 2019; Ma, Du, and Matusik 2020; Mahapatra and Rajan 2021; Nowak and Küfer 2020) to identify Pareto optimal solutions by finding solutions that match a given preference vector or that move along the Pareto surface in the direction of rays to identify other solutions. These methods are gradient-based, requiring a defined gradient, but our problem domain has a discrete input domain and lacks any well-defined gradients, making ray-based techniques a poor fit for this problem.

Our methods are closely related to Binary Decision Diagrams (BDDs) (Bergman and Cire 2016) which use a compact decision diagram representation of the problem and identify shortest paths through the BDD as non-dominated solutions, which are closely related to DP problems(Hooker 2013). However, these methods assume linear separability in the objective functions with respect to the parameters (Bergman et al. 2016), whereas, for our problem domain, the objective functions are non-separable and are dependent upon all decisions made along paths from the root to a node.

Other approaches, such as the Compression and Expansion (CE) technique (Bai et al. 2023), involve the parallel generation of partial Pareto frontiers, optimized with respect to fewer criteria, employing an existing Pareto solver. These partial Pareto frontiers, evaluated on all the criteria are then merged, with dominated solutions removed. In our experimental section, we use our approach as the underlying Pareto solver for the CE method, which synergistically leads to improved results.

Finally, to the best of our knowledge, this is the first quantitative work computing the Pareto frontier with hybrid floating solar-hydropower systems.

## Preliminaries

In the hydropower dam portfolio selection problem, our goal is to determine the subset of the proposed dams to build to jointly optimize a set of criteria. Note, we cannot optimize each criterion independently, as maximizing one criterion may require sacrificing other criteria. Given our goal, we set forth some definitions for the underlying optimization problem.

**Pareto Dominance.** For a given solution $\pi$, $z(\pi) = (z^1(\pi), \ldots, z^d(\pi))$ is the values of the $d$ objectives. A solution $\pi$ dominates solution $\pi'$ - written as $z(\pi) \succ z(\pi')$ - if and only if the following two properties hold: (1) for all $1 \leq i \leq d, z^i(\pi) \geq z^i(\pi')$; and (2) there exists at least one strict inequality, $1 \leq j \leq d$ such that $z^j(\pi) > z^j(\pi')$.

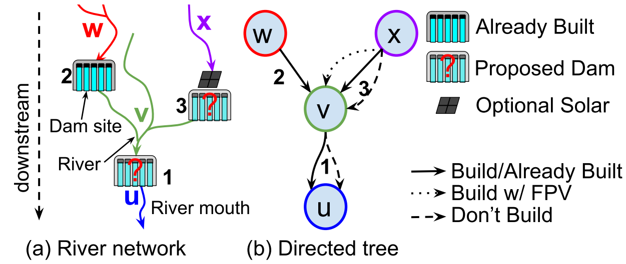


Figure 2: Converting a river network (a) into a directed multi-tree (b). Dam sites, represented by numbers, become directed edges, with one edge per decision. Dams that are already built may have a single edge (dam 2), whereas proposed dams may contain two edges (build/not build) (dam 1), but can be extended to more decisions (dam 3) where the options include building hydro only, hydro + FPV, or not building. Each edge has associated river ecosystem services rewards ($s$) and passage probabilities ($p$) depending on the decision. The contiguous river segments, undisturbed by dam sites, represented by letters, become nodes with associated river ecosystem services values ($r$). The river section starting from the mouth of the river, $u$, is the root of the tree.

**Pareto Frontier.** Given multiple competing objectives, we aim to find the Pareto frontier or Pareto set. Let $\mathcal{P}$ be the set of all feasible solutions, we define the Pareto set as $\{\pi \in \mathcal{P} | z(\pi) \not\prec z(\pi'), \forall \pi' \in \mathcal{P}\}$. For example, consider three solutions $(\pi_1, \pi_2, \pi_3)$, with objective values $z(\pi_1) = (10, 4, 3)$, $z(\pi_2) = (9, 4, 2)$, and $z(\pi_3) = (8, 5, 3)$. Solution $\pi_1$ dominates $\pi_2$ as it has strictly greater values for its first and third objectives and equal values for the second. However, $\pi_1$ does not dominate $\pi_3$, and vice versa, as $\pi_1$ has a greater value in its first objective than $\pi_3$, but $\pi_3$ has a greater value in its second objective. Additionally, $\pi_2$ and $\pi_3$ do not dominate each other. Since $\pi_1$ and $\pi_3$ are not dominated by any other solution, the Pareto set is $\{\pi_1, \pi_3\}$.

**$\epsilon$-approximate Pareto Frontier.** Given a Pareto frontier $P$, an approximate Pareto frontier $P'$ is said to $\epsilon$ approximate $P$ if and only if for every $\pi \in P$, there exists a solution $\pi' \in P'$ such that $z^i(\pi') \geq (1 - \epsilon) z^i(\pi)$ for all criteria $i$.

## Problem Formulation

Here we introduce our new problem formulation that extends the original problem's tree-structured network layout. In the original problem formulation(Wu et al. 2018), the underlying river network is converted into a rooted, directed tree, motivated by the work in (Wu, Sheldon, and Zilberstein 2014a,b). Here, a node in the tree is a contiguous portion of the river network undisturbed by potential dams, which typically consists of multiple river segments, with the dam locations becoming edges in the tree. In this original formulation, each node $u$ has associated ecosystem service values $r_u = (r_u^1, \ldots, r_u^d)$, and each edge $(u, v)$ had associated with it a decision variable $\pi_{uv}$ that represents the status of that dam, associated ecosystem service values $s_{uv} = (s_{uv}^1, \ldots, s_{uv}^d)$, and passage probabilities $p_{uv}$ and $q_{uv}$ which represent the percentage of a given service that passes through a dam location when built or not built respectively. In this formulation, the associated ecosystem service values and passage probabilities associated with the dam differ depending on the status of the decision variable, which both restricts the formulation to two decisions and requires multiple variables representing similar constructs. We, therefore, propose a new formulation that takes into account many possible decisions and results in a cleaner representation even for just two decisions.

Instead of formulating the problem as a simple tree with an edge representing multiple potential decisions simultaneously, we now convert the underlying river network to a rooted, directed multi-tree. Here we define a directed multi-tree as a directed acyclic multigraph where each connected pair of nodes $u$ and $v$ may contain multiple parallel edges, in other words, any path between two nodes $u$ and $v$ must always pass through the same nodes, but potentially along different distinct parallel edges, which is how it differs from a standard tree. Each edge, which we will now refer to as $(u, v)_j$ where $j$ is an index that differentiates the edges, now represents a different choice, and a solution $\pi$ in the network, defined as a decision made for every dam in the river, is a spanning tree of the graph. In addition, each edge $(u, v)_j$ now has a single $s_{uvj}$ vector for the ecosystem service values associated with a dam and the representative decision and a singular $p_{uvj}$ that represents the passage probability for the given decision. This new formulation is shown in Figure 2, where already-built dams only have a single edge, but unbuilt dams have two or three if FPV is an option.

*Input:* We are given a directed connected multi-tree $T = (V, E)$, where each node $u \in V$ has a set of rewards $\{r_u^1, \ldots, r_u^d\}$, and each edge $(u, v)_j \in E$ is associated with a set of objective values $\{s_{uvj}^{(1)}, \ldots, s_{uvj}^{(d)}\}$ and passage probabilities $\{p_{uvj}^{(1)}, \ldots, p_{uvj}^{(d)}\}$. Each edge $(u, v)_j \in E$ is also associated with a decision on the dam. A solution $\pi$ is a spanning tree of the multi-tree, where exactly one edge from each set of parallel edges is selected. For a given solution $\pi$, and tree $T_u$, a tree rooted at node $u$, we can compute the value of $d$ objectives $z(\pi, T_u) = (z^1(\pi, T_u), \ldots, z^d(\pi, T_u))$. Let $\mathrm{P}(u, v) \subset \pi$ be the set of selected edges in the path from node $u$ to any node $v$ in the tree $T_u$.

$$\begin{aligned} z^i(\pi, T_u) = r_u^i + \sum_{(v,w,k)\in(\pi\cap E_u)} r_w^i \prod_{(x,y,j)\in \mathrm{P}(u,w)} p_{xyi} \\ + s_{vwk}^i \prod_{(x,y,j)\in \mathrm{P}(u,v)} p_{xyi} \end{aligned} \tag{1}$$

*Output:* The Pareto frontier with respect to the $d$ objectives within the set of all feasible solutions $\mathcal{P}$.

## Main Algorithm

We can take advantage of the tree structure of the river network and compute the Pareto frontier at each node recursively since the criteria we care about share the characteristic that an optimal solution on $T$ is also optimal on a subtree $T_u$ rooted at node $u \in V$, i.e. the subtree induced on the set of all nodes above $u$ including $u$.

From the problem formulation, we recursively define the $z^i(\pi, T_u)$ values for tree $T$ rooted at node $u$ and solution $\pi$:

$$z^i(\pi, T_u) = r_u^i + \sum_{(u,v,j)\in\pi} s_{uvj}^i + p_{uvj} z^i(\pi, T_v) \tag{2}$$

In (Gomes-Selman et al. 2018; Wu et al. 2018), the authors propose a DP algorithm that takes, as input, a directed tree, and outputs the set of non-dominated solutions at the root of the tree. The algorithm, which is shown in more detail in the supplementary materials (SI), recursively generates the Pareto frontier of all the children of a node and joins the sub-frontiers of each child together to produce the Pareto frontier of the parent by finding the set of non-dominated solutions from the combination of solutions and decisions at each child. In practice, the size of the frontier grows exponentially in the number of objectives, making it infeasible to calculate the exact Pareto frontier for any substantial number of objectives. To alleviate this issue, (Wu et al. 2018) proposes a rounding scheme that $\epsilon$-approximates the Pareto frontier in polynomial time. The steps performed on a node $u$ can be broken into four cases: leaf nodes, single child nodes, two child nodes, and greater than two child nodes, with each case given a more detailed inspection in SI.

## Sub-frontier Transformations

When joining two children together, any combination of left child solution and decisions made on the left and right dams defines an affine transformation that is applied to the entirety of the right child solution set. More concretely, for a given node $u$, children $v$ and $w$, let $\pi_v$ and $\pi_w$ be some specific solutions[1] from the Pareto frontiers associated with the trees $T_v$ and $T_w$ respectively. Let $\pi_u$ be a new solution containing the solutions $\pi_v$ and $\pi_w$ and let $(u, v, j)$ and $(u, w, k)$ be the edges selected from all parallel edges of $(u, v)$ and $(u, w)$ respectively. Consider Equation 2, expanding out the sum:

$$\begin{aligned} z^i(\pi_u, T_u) = & r_u^i + s_{uwk}^i + p_{uwk}^i z^i(\pi_w, T_w) \\ & + s_{uvj}^i + p_{uvj}^i z^i(\pi_v, T_v) \end{aligned} \tag{3}$$

By fixing $\pi_w$, $(u, v, j)$ and $(u, w, k)$, we see that $r_u^i + s_{uwk}^i + p_{uwk}^i z^i(\pi_w, T_w) + s_{uvj}^i$ forms a constant term, and $p_{uvj}^i z^i(\pi_v, T_v)$ is a scalar multiplied by the value $z^i(\pi_v, T_v)$ for all solutions $\pi_v$ associated with $T_v$. Thus we obtain an affine transformation $a^i z^i(\pi_v) + b^i$ for each $j$, where

$$\begin{aligned} b^i((u, w, k), (u, v, j), \pi_w) = r_u^i + s_{uwk}^i + s_{uvj}^i \\ + p_{uwk}^i z^i(\pi_w, T_w) \end{aligned} \tag{4}$$

$$a^i((u, w, j)) = p_{uvj}^i \tag{5}$$

[1] We are slightly overloading the usage of $\pi$. When used with a single subscript, such as $\pi_u$ it refers to a specific node solution which is a vector of decision variables, one for each edge in the tree rooted at node $u$. When used with two subscripts $\pi_{uv}$, it refers to the specific decision variable associated with edge $(u, v)$.

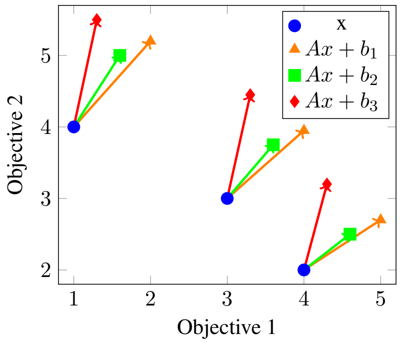


Figure 3: Example of transform dominance. For the blue circles, we have the initial points of a sub-frontier along with a series of affine transformations $Ax + b$ for different values of $b$. The transformation $b_1 \succ b_2$. All points in $Ax + b_2$ are dominated by at least one point in $Ax+b_1$, but neither points in $Ax + b_1$ nor $Ax + b_3$ dominate each other.

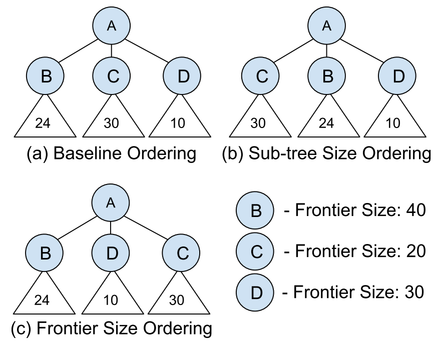


Figure 4: Example ranking strategies. The tree has root node A with three children B, C, and D. The algorithm selects two of the children to join first and produces an intermediate node to be joined with the last child. Assume we join the left two nodes first. Each of the nodes B, C, and D have a sub-tree containing some number of nodes, as well as the size of their Pareto frontier. In (a) we see the baseline ordering which sorts by listed order. In (b) we see the ordering that ranks by the size of the sub-trees. Finally, in (c) we see the ordering that ranks by the size of the Pareto frontier.

Because the scalar $a^i$ is dependent only on the edge $(u, v, j)$, for a fixed $(u, v, j)$, for all possible decisions $(u, w, k)$ and solutions $\pi_w$ associated with tree $T_w$, the objective values from $z^i(\pi_v, T_v)$ are scaled by the same amount, and the difference in the resulting position of the transformed solution is dependent only on the shift caused by $b^i$. Therefore, for a given $(u, w, k)$, if a shift $b((u, w, k),(u, v, j), \pi_w)$ induced by edge $(u, w, k)$ and solution $\pi_w$ is dominated by another shift $b'((u, w, k'),(u, v, j), \pi'_w)$ induced by decision $(u, w, k')$ and solution $\pi'_w$, then every point shifted by $b$ must be dominated by at least one point shifted by $b'$.

**Lemma 1.** *Let $A \in \mathbb{R}^{d\times d}$ be a diagonal matrix, $B$ be a set of vectors in $\mathbb{R}^d$, and $X$ be a set of non-dominated points in $\mathbb{R}^d$. If a vector $b \in B \succ b' \in B$, that is $b$ dominates $b'$, then every point in $\{Ax+b'|x \in X\}$ is dominated by at least one point in $\{Ax + b|x \in X\}$.*

The proof is provided in the SI. This result allows us to check for dominance in the set of transformations first, and remove any transformations that are dominated. If we have two children $w$ and $v$ with $m$ and $n$ solutions respectively, and the decisions to build or not build each dam, without considering transform domination we must consider $4mn$ solutions. If we instead first consider the set of transformations that are non-dominated, of which there are $m' \leq 4m$, we first must consider $4m$ transformations and $m'n$ solutions. As long as the resulting number of transformations $m' \leq 4m\frac{n-1}{n}$, where $n$ may be quite large, then we will consider fewer solutions and transformations combined. Figure 3 shows an example of how a dominated transform results in the entire frontier being dominated.

## Child Ranking

Another contribution concerns the ordering in which children are joined when there are more than two children to choose from. In the original algorithm (Gomes-Selman et al. 2018), children were joined using the input order. Our work looks at two different child orderings: one sorting by the subtree size of the nodes and the other sorting by the size of the Pareto frontier of the nodes. See Figure 4. For the basins we examined, we find it best to rank the children in descending order, choosing to join the children with either the largest sub-trees or the largest frontiers first. Consider an example with a node with three children: $u$, $v$, and $w$, with 4, 8, and 10 solutions respectively. First, consider the case where we keep all solutions on each step. If we join the smallest two first, $u$ and $v$, we produce and keep all $4 \cdot 8 = 32$ solutions. Next, we join these 32 solutions with the 10 solutions in $w$, producing $10 \cdot 32 = 320$. In all, we have to search through $320 + 32 = 352$ solutions. Instead, if we join the two largest first, $v$ and $w$, we must search through 400 solutions. However, if we were to keep a different percentage of values from each of the joins, it is possible that joining the largest first is better. For example, if joining $u$ and $v$ keeps 50%, but joining $v$ and $w$ only keeps 10%, then if we join the smallest first, we generate $4 \cdot 8 = 32$ solutions as before but keep 16. Next, we join $16 \cdot 10$ solutions to produce 160 solutions, thus needing to consider $32 + 160 = 192$ solutions. If we instead join the largest two first, we only consider 112 solutions. Thus, under similar circumstances, this may result in fewer solutions considered by joining the largest nodes first, which is what we see in general for our datasets.

## Experiments

Our goal is to scale up the exact or approximate Pareto frontier. In the following experiments, we consider real data based on the Amazon and Marañón river basins (Flecker et al. 2022). Each experiment always includes energy as an objective, as it is the only objective that is optimized by building, thus any run without energy will have a single trivial solution of building no dams. Other criteria considered

| Basin | Criteria | $\epsilon$ | Baseline Order | | Node Size Order | | Frontier Size Order | |
|---|---|---|---|---|---|---|---|---|
| | | | No Transforms | Transforms | No Transforms | Transforms | No Transforms | Transforms |
| A | ECB | 0.01 | >86400 | >86400 | 3317 | **2599** | 4002 | 2813 |
| | ECD | 0.01 | >86400 | >86400 | 6157 | **4490** | 7000 | 4871 |
| | ECG | 0.025 | 25567 | 12199 | 449 | **369** | 561 | 421 |
| | ECS | 0.01 | 18787 | 9708 | 992 | 762 | 926 | **653** |
| | EBD | 0.1 | >86400 | >86400 | 18689 | **14807** | 21605 | 18213 |
| | EBG | 0.05 | >86400 | >86400 | 14362 | **12406** | 15558 | 13663 |
| | EBS | 0.1 | 21510 | 17246 | 7374 | 6288 | 6207 | **5826** |
| | EDG | 0.05 | >86400 | >86400 | 5323 | **4782** | 8350 | 7497 |
| | EDS | 0.05 | 27229 | 22834 | 3181 | 2775 | 2615 | **2260** |
| | EGS | 0.05 | 37276 | 33858 | 3729 | **3173** | 7265 | 6348 |
| M | EBCDG | 0.025 | 36556 | 29270 | 5059 | 4766 | 4410 | **4123** |
| | EBCDS | 0.025 | >86400 | >86400 | 8942 | **8609** | 9887 | 8959 |
| | EBCGS | 0.025 | 36046 | 26730 | 3344 | **3130** | 3569 | 3321 |
| | EBDGS | 0.025 | >86400 | >86400 | 7903 | 7738 | 8168 | **7464** |
| | ECDGS | 0.025 | 13841 | 11714 | 2254 | **2015** | 2253 | 2076 |
| | All | 0.05 | >86400 | >86400 | 53372 | 50387 | 46016 | **43236** |

Table 1: Running time in seconds for different combinations of criteria and epsilon values, considering different child orderings and the inclusion of the affine transforms. Three criteria were used against the full Amazon River basin (A) and five or six criteria were used against the Marañón (M), a sub-basin of the Amazon. Each run was given 86,400 seconds (or 24 hours) to complete, jobs that did not complete within that time are listed as taking over 86,400 seconds. The criteria are Energy (E), Connectivity (C), Sediment (S), Degree of Regulation (D), Biodiversity (B), and Greenhouse Gases (G). The fastest running time for each set of parameters is written in bold. The baseline consists of no ordering and no transforms.

are greenhouse gas emissions, river connectivity, biodiversity loss, degree of regulation of the water flow in the dams, and sediment transport. The data are from the SI in (Flecker et al. 2022). All experiments are run using 12 threads on Intel® Xeon® 3.47 GHz CPUs with 100 GB of memory.

To determine the impact of our methods, we ran multiple experiments on the Amazon River basin and Marañón River basin (a sub-basin of the Amazon) with and without different combinations of using the affine transformation dominance and child ranking methods, the results of which are in Table 1. We ran each combination with $\epsilon$ values between $0.01$ and $0.1$ and a maximum running time of 24 hours. Here, we see that considering the dominance in the affine transformation always outperforms not doing so. In the case of child ranking, both node size and frontier size ordering resulted in much faster run times than without considering the order in which the nodes are joined. Typically the frontier size and node size ordering have similar performance, but in some cases the node size ordering performed twice as fast, thus we would recommend using the node size ordering. In many cases, the algorithm did not complete within a 24-hour allotted period when running with no order, and the inclusion of the different order heuristics results in an order of magnitude increase in performance in most cases. Finally, we see that combining both affine transformation dominance and child ordering heuristics results in further improvements over using any of the methods individually.

Next, we look at how the introduction of floating photovoltaics may impact the energy system and needs of the Amazon through an analysis of the Pareto frontiers with and without floating photovoltaics as an option for increased energy production. First, we consider the current system as-is with hydropower dam planning at each site. Second, we consider a system where each hydropower dam reservoir can be fitted with floating photovoltaic energy production, allowing for each dam site to potentially produce more energy for a similar ecosystem service investment. For floating photovoltaics, we assume up to 5% of the reservoir or 30 km$^2$ of solar panels may be built at each dam site, whichever value is smaller, using data from (Almeida et al. 2022a). To get a good candidate set of solutions without requiring large $\epsilon$ values for the approximation across all six available objectives, we use the Compress and Expansion methods introduced in (Bai et al. 2023) using our enhancements as the underlying method used by Expansion and Compression. We follow the recommendations in the original paper and use the combination of Compression-3,4,5 and Expansion-3,4, with each compression method using equal weights. We next discuss the results of these analyses and comparisons.

## Discussion

Floating solar panels allow for the dual use of hydropower dam reservoirs both for energy production from the dam as well as the solar panels taking up space that was otherwise already flooded by the reservoir. As such, by adding solar arrays to the reservoirs, we expect dams to be able to produce more energy for a similar ecological impact, allowing energy planners to be able to reach desired energy targets more sustainably. Here, we examine the impacts on both the overall energy system when including floating solar as well as the impacts on optimal dam portfolios and the individual dams within those portfolios to determine if and when cer-

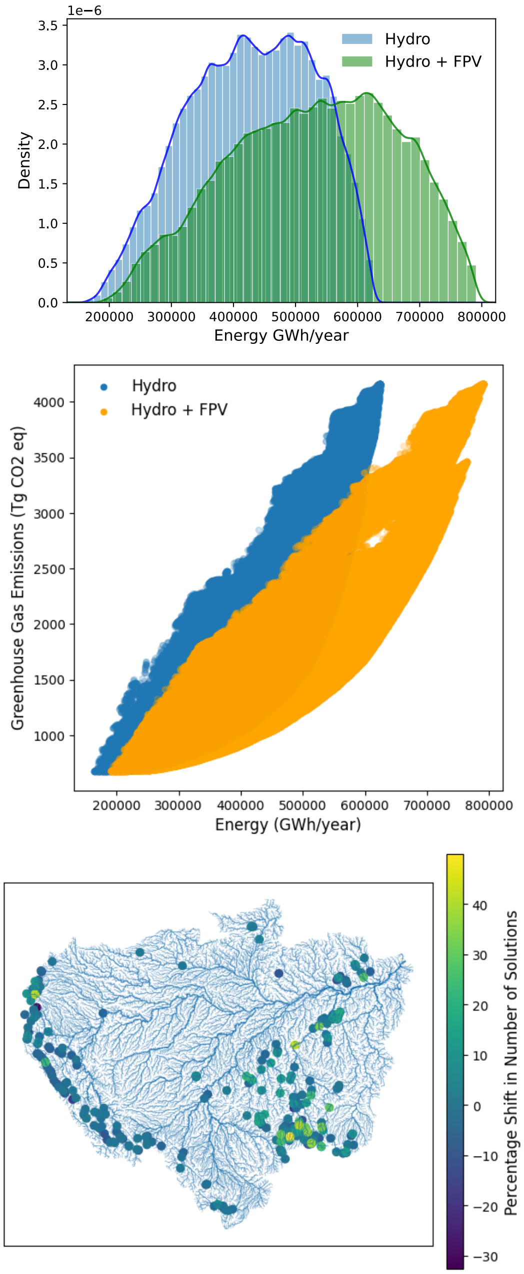


Figure 5: *Top*: Distribution of energy values among Pareto frontier solutions for hydro-power energy only versus hydropower plus floating photovoltaic energy. There is a shift towards higher energy when including FPV, and the distribution is wider as well, with more solutions at the higher end. *Middle*: visualization of the Pareto frontier, optimized w.r.t. six criteria, projected onto energy vs greenhouse gas. Other Pareto frontier plots are provided in SI. *Bottom*: Shift in the percentage of solutions that a proposed dam is Pareto optimal when introducing FPV. A positive shift means that a dam location is more appealing (i.e., more often Pareto optimal) when FPV is included in the energy system. We observe FPV increases the frequency of lowland dams in the east.

tain types of dams may become more attractive when considering the addition of floating solar.

When considering the energy system as a whole, we look to see how the Pareto frontier shifts and changes due to the introduction of FPV. In Figure 5, we see that when including FPV, as expected, the energy output by the system as a whole increases, but perhaps more interestingly the distribution widens, with many more solutions in the higher ranges of energy values. In the SI we show a similar widening pattern for different projected 2D Pareto frontiers.

We now shift our attention towards visualizing and analyzing dams within the Amazon River basin. For each of the two Pareto frontiers we calculate, for each dam, the percentage of solutions a given dam appears in and then compare these percentages on a dam-by-dam basis between the two frontiers. In the map in Figure 5, we map out each dam that is not already built (since dams already built are trivially in all solutions) and see how the percentage shifts as FPV is included. As all other objectives remain the same, any shift in the solution space for a given dam is due to the relation between including FPV versus not, as some dams that have poor ecological impacts may gain large improvements to their total energy output when including FPV that would have otherwise made the dam a less attractive option. For example, we see a general increase in the percentage of solutions dams that appear in the optimal solutions in the lowlands in the eastern part of the Amazon. The eastern region of the Amazon tends to have larger reservoirs on average, which often results in negative impacts on many ecosystem services. However, in the case of FPV, larger reservoirs mean more space for solar arrays to be placed, resulting in FPV causing these dam locations to become more attractive.

## Conclusion

Our work significantly improves the computation of exact and approximate Pareto frontiers for tree-structured networks. By leveraging underlying affine transformations and employing an intelligent selection of children, we have achieved a remarkable reduction in algorithm runtime, sometimes exceeding an order of magnitude. Consequently, we are now able to empirically approximate frontiers with greater accuracy. Our analysis of adding floating photovoltaic energy (FPV) to hydropower dams revealed a shift in the spatial distribution of the dams included in the Pareto frontier. Notably, dam sites in the eastern lowland regions of the Amazon become more attractive when incorporating FPV. This observation underscores the importance of considering various aspects of the energy system beyond just ecological impacts. Strategically pairing hydropower dams with FPV can yield more environmentally amenable outcomes, with dams producing more energy and lesser additional ecological impact. By identifying better energy portfolios in significantly less time, compared to previous algorithms, we provide interested parties with valuable insights for decision-making. We hope our research will catalyze further exploration and studies in this area, encouraging researchers and interested parties to delve deeper into the potential of combining hydropower dams with floating photovoltaic energy for sustainable and efficient energy solutions.

## Acknowledgements

This project is partially supported by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Futures program; the National Science Foundation (NSF); the Air Force Office of Scientific Research (AFOSR); and the Cornell Atkinson Center for a Sustainable Future (ACSF).

## References

Almeida, R. M.; Schmitt, R.; Grodsky, S. M.; Flecker, A. S.; Gomes, C. P.; Zhao, L.; Liu, H.; Barros, N.; Kelman, R.; and McIntyre, P. B. 2022a. Floating solar power: evaluate trade-offs. *Nature*, 606: 246–249.

Almeida, R. M.; Schmitt, R. J.; Castelletti, A.; Flecker, A. S.; Harou, J. J.; Heilpern, S. A.; Kittner, N.; Kondolf, G. M.; Opperman, J. J.; Shi, Q.; Gomes, C. P.; and McIntyre, P. B. 2022b. Strategic planning of hydropower development: balancing benefits and socioenvironmental costs. *Current Opinion in Environmental Sustainability*, 56: 101175.

Almeida, R. M.; Shi, Q.; Gomes-Selman, J. M.; Wu, X.; Xue, Y.; Angarita, H.; Barros, N.; Forsberg, B. R.; García-Villacorta, R.; Hamilton, S. K.; et al. 2019. Reducing greenhouse gas emissions of Amazon hydropower with strategic dam planning. *Nature Communications*, 10(1): 1–9.

Bai, Y.; Shi, Q.; Grimson, M.; Flecker, A.; and Gomes, C. P. 2023. Efficiently Approximating High-Dimensional Pareto Frontiers for Tree-Structured Networks Using Expansion and Compression. In Cire, A. A., ed., *Integration of Constraint Programming, Artificial Intelligence, and Operations Research*, 1–17. Cham: Springer Nature Switzerland. ISBN 978-3-031-33271-5.

Bergman, D.; and Cire, A. A. 2016. Multiobjective optimization by decision diagrams. In *International Conference on Principles and Practice of Constraint Programming*, 86–95. Springer.

Bergman, D.; Cire, A. A.; van Hoeve, W.-J.; and Hooker, J. 2016. *Introduction*, 1–9. Cham: Springer International Publishing. ISBN 978-3-319-42849-9.

Castaño, R. S.; Rodríguez, C. A.; Huertas, M. L. R.; Botello, W. A. P.; Pinilla, H. S. M.; Fajardo, S. L. Z.; Forero, C. A. B.; Triana, D. F. F.; Velosa, I. F. A.; Vasquez, C. Y. E.; and Cifuentes, A. V. 2019. Reporte de Alertas de Cuencas Hidrográficas. Instrumento de regionalización, Porce, Nechí y Bajo Cauca. Technical report, ANLA, Colombia.

Deb, K.; and Jain, H. 2013. An evolutionary many-objective optimization algorithm using reference-point-based non-dominated sorting approach, part I: solving problems with box constraints. *IEEE transactions on evolutionary computation*, 18(4): 577–601.

Deb, K.; Pratap, A.; Agarwal, S.; and Meyarivan, T. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE transactions on evolutionary computation*, 6(2): 182–197.

Ehrgott, M.; and Gandibleux, X. 2000. A survey and annotated bibliography of multiobjective combinatorial optimization. *OR-spektrum*, 22(4): 425–460.

Ehrgott, M.; Gandibleux, X.; and Przybylski, A. 2016. Exact Methods for Multi-Objective Combinatorial Optimisation. In Greco, S.; Ehrgott, M.; and Figueira, J. R., eds., *Multiple Criteria Decision Analysis: State of the Art Surveys*, 817–850. New York, NY: Springer New York. ISBN 978-1-4939-3094-4.

Flecker, A. S.; Shi, Q.; Almeida, R. M.; Angarita, H.; Gomes-Selman, J. M.; García-Villacorta, R.; Sethi, S. A.; Thomas, S. A.; Poff, N. L.; Forsberg, B. R.; Heilpern, S. A.; Hamilton, S. K.; Abad, J. D.; Anderson, E. P.; Barros, N.; Bernal, I. C.; Bernstein, R.; Cañas, C. M.; Dangles, O.; Encalada, A. C.; Fleischmann, A. S.; Goulding, M.; Higgins, J.; Jézéquel, C.; Larson, E. I.; McIntyre, P. B.; Melack, J. M.; Montoya, M.; Oberdorff, T.; Paiva, R.; Perez, G.; Rappazzo, B. H.; Steinschneider, S.; Torres, S.; Varese, M.; Walter, M. T.; Wu, X.; Xue, Y.; Zapata-Ríos, X. E.; and Gomes, C. P. 2022. Reducing adverse impacts of Amazon hydropower expansion. *Science*, 375(6582): 753–760.

Gomes, C.; Dietterich, T.; Barrett, C.; Conrad, J.; Dilkina, B.; Ermon, S.; Fang, F.; Farnsworth, A.; Fern, A.; Fern, X.; Fink, D.; Fisher, D.; Flecker, A.; Freund, D.; Fuller, A.; Gregoire, J.; Hopcroft, J.; Kelling, S.; Kolter, Z.; Powell, W.; Sintov, N.; Selker, J.; Selman, B.; Sheldon, D.; Shmoys, D.; Tambe, M.; Wong, W.-K.; Wood, C.; Wu, X.; Xue, Y.; Yadav, A.; Yakubu, A.-A.; and Zeeman, M. L. 2019. Computational Sustainability: Computing for a Better World and a Sustainable Future. *Commun. ACM*, 62(9): 56–65.

Gomes-Selman, J. M.; Shi, Q.; Xue, Y.; Garcia-Villacorta, R.; Flecker, A. S.; and Gomes, C. P. 2018. Boosting efficiency for computing the Pareto frontier on tree structured networks. In *International Conference on the Integration of Constraint Programming, Artificial Intelligence, and Operations Research*, 263–279. Springer.

Hooker, J. N. 2013. Decision Diagrams and Dynamic Programming. In Gomes, C.; and Sellmann, M., eds., *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*, 94–110. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-642-38171-3.

Lee, N.; Grunwald, U.; Rosenlieb, E.; Mirletz, H.; Aznar, A.; Spencer, R.; and Cox, S. 2020. Hybrid floating solar photovoltaics-hydropower systems: Benefits and global assessment of technical potential. *Renewable Energy*, 162: 1415–1427.

Lin, X.; Zhen, H.-L.; Li, Z.; Zhang, Q.; and Kwong, S. 2019. Pareto Multi-Task Learning. *Advances in Neural Information Processing Systems*, 32.

Ma, P.; Du, T.; and Matusik, W. 2020. Efficient Continuous Pareto Exploration in Multi-Task Learning. In *International Conference on Machine Learning*, 6522–6531. PMLR.

Mahapatra, D.; and Rajan, V. 2021. Exact Pareto Optimal Search for Multi-Task Learning: Touring the Pareto Front. *arXiv preprint arXiv:2108.00597*.

Nowak, D.; and Küfer, K.-H. 2020. A Ray Tracing Technique for the Navigation on a Non-convex Pareto Front. *arXiv preprint arXiv:2001.03634*.

Opperman, J. J.; Carvallo, J. P.; Kelman, R.; Schmitt, R. J. P.; Almeida, R.; Chapin, E.; Flecker, A.; Goichot, M.; Grill, G.; Harou, J. J.; Hartmann, J.; Higgins, J.; Kammen, D.; Martin, E.; Martins, T.; Newsock, A.; Rogéliz, C.; Raepple, J.; Sada, R.; Thieme, M. L.; and Harrison, D. 2023. Balancing renewable energy and river resources by moving from individual assessments of hydropower projects to energy system planning. *Frontiers in Environmental Science*, 10: 1036653.

Srinivas, N.; and Deb, K. 1994. Muiltiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary computation*, 2(3): 221–248.

United Nations General Assembly. 2015. Transforming our world: the 2030 Agenda for Sustainable Development. https://sdgs.un.org/2030agenda. Accessed: 2022-11-23.

Wiecek, M. M.; Ehrgott, M.; Fadel, G.; and Figueira, J. R. 2008. Multiple criteria decision making for engineering. *Omega*, 36(3): 337–340.

Winemiller, K. O.; McIntyre, P. B.; Castello, L.; Fluet-Chouinard, E.; Giarrizzo, T.; Nam, S.; Baird, I. G.; Darwall, W.; Lujan, N. K.; Harrison, I.; Stiassny, M. L. J.; Silvano, R. A. M.; Fitzgerald, D. B.; Pelicice, F. M.; Agostinho, A. A.; Gomes, L. C.; Albert, J. S.; Baran, E.; Petrere, M.; Zarfl, C.; Mulligan, M.; Sullivan, J. P.; Arantes, C. C.; Sousa, L. M.; Koning, A. A.; Hoeinghaus, D. J.; Sabaj, M.; Lundberg, J. G.; Armbruster, J.; Thieme, M. L.; Petry, P.; Zuanon, J.; Vilara, G. T.; Snoeks, J.; Ou, C.; Rainboth, W.; Pavanelli, C. S.; Akama, A.; van Soesbergen, A.; and Sáenz, L. 2016. Balancing hydropower and biodiversity in the Amazon, Congo, and Mekong. *Science*, 351(6269): 128–129.

Wu, X.; Gomes-Selman, J.; Shi, Q.; Xue, Y.; Garcia-Villacorta, R.; Anderson, E.; Sethi, S.; Steinschneider, S.; Flecker, A.; and Gomes, C. 2018. Efficiently Approximating the Pareto Frontier: Hydropower Dam Placement in the Amazon Basin. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).

Wu, X.; Sheldon, D.; and Zilberstein, S. 2014a. Rounded dynamic programming for tree-structured stochastic network design. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 479–485.

Wu, X.; Sheldon, D.; and Zilberstein, S. 2014b. Stochastic network design in bidirected trees. In *Proceedings of the 27th International Conference on Neural Information Processing Systems-Volume 1*, 882–890.

Zarfl, C.; Lumsdon, A.; Berlekamp, J.; Tydecks, L.; and Tockner, K. 2014. A Global Boom in Hydropower dam Construction. *Aquatic Sciences*, 77: 161–170.

Zhang, Q.; and Li, H. 2007. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on evolutionary computation*, 11(6): 712–731.

Ziv, G.; Baran, E.; Nam, S.; Rodríguez-Iturbe, I.; and Levin, S. A. 2012. Trading-off fish biodiversity, food security, and hydropower in the Mekong River Basin. *Proceedings of the National Academy of Sciences*, 109(15): 5609–5614.