Constraint-aware Pareto Optimization for Tree-Structured Networks: Addressing Decarbonization Targets with Hydropower Expansion

Marc Grimson[1], Zhongdi Qu[1], Yue Mao[1], Aaron M Ferber[1], Felipe Siqueira Pacheco[2], Sebastian Heilpern[3], Héctor Angarita[4], Alexander Flecker[2], Carla P Gomes[1]
[1]Department of Computer Science, Cornell University
[2]Department of Ecology and Evolutionary Biology, Cornell University
[3]Department of Public and Ecosystem Health, Cornell University
[4]Natural Capital Project, Stanford University
mg2425@cornell.edu, zq84@cornell.edu, ym277@cornell.edu, amf272@cornell.edu, fsp9@cornell.edu, s.heilpern@cornell.edu, haac@stanford.edu, asf3@cornell.edu, gomes@cs.cornell.edu

Abstract

Addressing global sustainability challenges as outlined by the United Nations (UN) Sustainable Development Goals (SDGs) often requires navigating many potentially conflicting societal objectives simultaneously. For instance, increasing hydropower production enhances renewable energy supply but may adversely impact people and nature. Understanding these trade-offs is crucial, and the Pareto frontier - the set of solutions that cannot be improved with respect to one objective without negatively affecting another - is a valuable framework. Strategic hydropower planning concerns finding energy portfolios that achieve decarbonization targets, while balancing energy production with socioeconomic and environmental impacts. Previous work has considered exact and approximate algorithms for Pareto optimization for tree-structured networks, such as rivers, for hydropower planning. However, such approaches do not account for bounding constraints, such as realistic energy production targets, critical in real-world applications. Herein, we propose a novel approach for constraint-aware Pareto optimization for tree-structured networks, incorporating objective bounds to ensure more realistic and robust solution outcomes. We apply our constraint-aware Pareto approach to the strategic planning of hydropower expansion, *considering energy bounds to adhere to the UN's net zero by 2050 decarbonization targets*, in the Magdalena River basin, home to more than 80% of Colombia's population. Our analysis demonstrates how lower and upper bounds can significantly modify the unconstrained Pareto frontier, revealing that feasible Pareto solutions can be dominated by infeasible solutions, and thus may be ignored by constraint-agnostic solvers. Our results highlight the importance of considering real-world constraints in multi-objective problems such as optimizing hydropower expansion to meet both energy and sustainability goals.

Introduction

Computational Sustainability harnesses Artificial Intelligence (AI) to balance environmental, economic, and societal needs for a sustainable future, as captured by the United Nations (UN) Sustainable Development Goals (SDGs) (Gomes et al. 2019; United Nations 2015). Real-world decision-making is needed to achieve such goals, which often requires considering multiple competing objectives. Such decision-making problems translate to Multi-Objective Combinatorial Optimization Problems (MOCOP), and have received growing attention in recent years from both computer science and applied scientific communities (Ehrgott and Gandibleux 2000; Ehrgott, Gandibleux, and Przybylski 2016; Wiecek et al. 2008).

Our work herein is motivated by the problem of hydropower expansion, a key source of renewable energy prominently featured in global efforts to decarbonize energy systems and curb climate change. Nevertheless, damming rivers to generate hydropower can also lead to adverse environmental and socioeconomic impacts, from trapping sediment and changing river hydrology to fragmenting rivers, altering biodiversity, and displacing populations. Hydropower is slated to rapidly expand in biodiverse river systems with large untapped potential, many of which are located in the Global South. These systems include the Magdalena River basin in Colombia, which is a globally important biodiversity hotspot and Colombia's economic powerhouse. Of the more than 3,700 dams >1MW proposed globally, 143 are located in the Magdalena River basin.

To date, strategic basin planning for hydropower often considers the full possible range of energy buildout for a region. In reality, all proposed hydropower dams are unlikely to be built; instead, most countries have estimated a range of future hydropower buildouts needed to meet different decarbonization targets. Rather than considering the full range of energy buildout scenarios when evaluating multi-objective trade-offs, including objective bounds can more realistically identify potential portfolios that pave a path towards decarbonization while also minimizing the adverse socioeconomic and environmental impacts of hydropower expansion.

In this work, in collaboration with domain scientists across a number of disciplines in ecology, energy, and policy to tackle this interdisciplinary challenge, we address the constrained hydropower dam portfolio selection problem, which consists of selecting a subset of proposed dam sites, considering multiple competing objectives, where several objectives have upper and lower bound constraints. We aim to identify the Pareto frontier of feasible solutions, where no solution can be improved in one objective without compro-

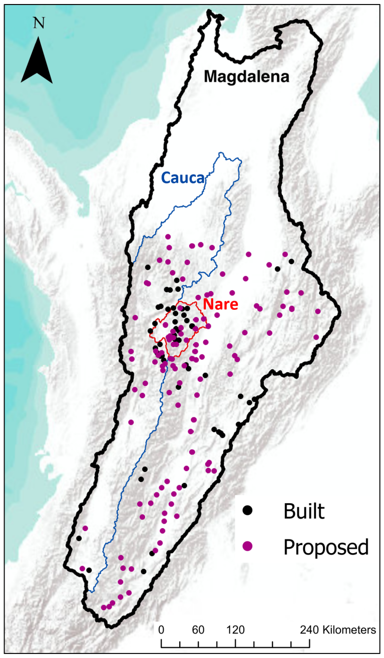


Figure 1: Magdalena River basin with outlines for the Cauca (blue) and Nare (red) sub-basins

mising others. Previous work on this problem converted the river network into a directed tree-structured network, giving rise to a dynamic programming (DP) formulation of the problem with guarantees on finding the exact Pareto set with an extension that allows for a fully polynomial-time approximation scheme (FPTAS) of the problem that finds an approximate Pareto set (Wu et al. 2018; Gomes-Selman et al. 2018; Grimson et al. 2024). However, such approaches do not consider constraints on the objectives, only ever constructing the unconstrained Pareto set. Herein we consider constraints on the objectives, for example, setting minimum or maximum demands on the amount of energy generated, or for maintaining greenhouse gas emissions within set targets. Our work shows that incorporating objective lower and upper bounds into Pareto optimization can significantly deviate from the unconstrained Pareto frontier; solutions that are optimal in the constrained case can become dominated and thus eliminated in the unconstrained case. Our findings underscore the importance of accounting for real-world constraints in multi-objective problems, such as those concerning balancing energy and sustainability goals.

Our contributions: **(1)** We propose a **constraint-aware Pareto optimization algorithm for tree-structured networks**, considering both lower and upper bound objective values. **(2) We prove that: a)** our constraint-aware Pareto algorithm always finds a Pareto frontier at least as good as removing solutions that violate the constraints from the unconstrained frontier, measured by the hypervolume covered by the frontiers and **b)** our constraint-aware Pareto algorithm always finds the true constrained Pareto frontier, with lower bounds. **(3)** We developed a **dataset of synthetic problem instances** to further understand the performance of our constraint-aware Pareto approach against the unconstrained Pareto with post-hoc removal of infeasible solutions. **(4)** We empirically demonstrate **good performance of our constraint-aware Pareto approach** namely: **a)** in general it **finds more of the true Pareto optimal solutions** than removing solutions that violate the constraints from the unconstrained frontier (in some cases 50% more), **b)** consistently **finds closer points to the true optimal solutions** for cases where both methods miss the true Pareto solutions and is **c) between 1.5 to 9 times faster** on large scale instances as we increase the number of objectives and tighten the bounds. **(5)** We show that our ***constraint-aware Pareto algorithm performs well and provides valuable insights*** on a ***real-world case study on the Magdalena River basin in Colombia, with 143 dams in consideration, based on the energy needs of Colombia for the business-as-usual (UPME 2018) and the desired decarbonization targets (Gobierno de Colombia 2021) by 2050***. **(6)** We additionally **provide our framework as a decision support tool at** https://www.cs.cornell.edu/gomes/udiscoverit/magdalena-ecovistas, to enable efficient policy and trade-off exploration of the proposals.

Related Work

For general constrained and unconstrained multi-objective optimization problems, many approaches use genetic algorithms, such as the Non-dominated Sorting Genetic Algorithm(s) (Srinivas and Deb 1994; Deb et al. 2002; Deb and Jain 2013) and MOAE/D (Zhang and Li 2007). These methods may get stuck in local neighborhoods or fail to obtain particularly large solution spaces since they lack strong formal guarantees on the Pareto frontier. So far, theoretical analysis of these algorithms only applies to relatively few and simplistic objectives (Zheng, Liu, and Doerr 2022; Doerr and Qu 2023a,c,b).

Other methods revolve around identifying a few subsets of solutions at a time based on preference vectors or rays (Lin et al. 2019; Ma, Du, and Matusik 2020; Mahapatra and Rajan 2021; Nowak and Küfer 2020), by linearly scalarizing the objectives to solve smaller problems and merge the resulting frontiers (Bai et al. 2023). These methods however cannot guarantee global Pareto optimality and scaling to get good coverage of the entire frontier can be difficult.

Another general structured discrete multi-objective optimization method is the Binary Decision Diagram (BDD) (Bergman and Cire 2016), which constructs decision diagram trees on the decision space and identifies shortest paths through the decision diagram as non-dominated solutions. BDDs, however, assume linear separability of the objective functions (Bergman et al. 2016).

Recent work employed machine learning to find heuristic Pareto frontiers. For example, (Sierra-Altamiranda et al. 2024) reduced runtime while preserving Pareto optimality of generated solutions by learning to project the objective space to a smaller dimensionality. In (Jiang et al. 2024), the authors approach multi-objective games, generating Pareto solutions with reinforcement learning. In (Mısır and Cai 2023), the authors propose learning to schedule multi-objective solvers.

Finally, multi-objective sequential decision-making problems arise in reinforcement learning and leverage techniques inspired by multi-objective optimization (Hayes et al. 2022). However, without taking advantage of the underlying problem structure, these methods fail to scale to large numbers of objectives (> 5) and decision variables.

Previous work (Wu et al. 2018; Gomes-Selman et al. 2018; Grimson et al. 2024; Qu et al. 2024) leverages tree structure to efficiently generate approximate Pareto frontiers in solely unconstrained settings. These methods were shown to outperform other Pareto methods on specifically the tree-structured problem by taking advantage of that structure.

We propose a **constraint-aware Pareto optimization algorithm for tree-structured networks**, considering lower and upper bound objective values, which outperforms previous approaches both in scalability, Pareto solutions found, and runtime, improving directly on the previous state-of-the-art approach for unconstrained tree-structured problems.

Preliminaries

In the constrained hydropower dam portfolio selection problem, we are given a set of potential dam locations and a set of d objectives. Policymakers are interested in identifying the optimal set of solutions given competing objectives, where a solution is a set of dams to be built, and some of the objectives may be constrained to be between upper and lower bounded values. The goal is to identify the Pareto optimal set of solutions for $d >= 2$ objectives where at least one objective o is constrained to be between $l_o <= o <= u_o$. Without loss of generality, we assume that all the objectives are to be maximized. We have the following definitions:

Pareto Dominance For a given solution π, let $z(\pi) \in \mathbb{R}^d$ be the set of d objective values for π. A solution π dominates another solution π' ($\pi \succ \pi'$) if for all objectives $0 \leq i < d, z^i(\pi) \geq z^i(\pi')$ and there exists at least one objective $0 \leq j < d$ such that $z^j(\pi) > z^j(\pi')$.

Pareto Optimal Set Let $\mathcal{S}$ be the set of all feasible solutions, the Pareto optimal set is then defined as $\{\pi \in \mathcal{S} | z(\pi) \not\prec z(\pi'), \forall \pi' \in \mathcal{S}\}$. For example, if we have three solutions $z(\pi_1) = (1,2,3), z(\pi_2) = (2,1,1), z(\pi_3) = (1,1,2)$, we see that $\pi_1 \not\succ \pi_2$ and $\pi_2 \not\succ \pi_1$, however $\pi_1 \succ \pi_3$. Thus the Pareto set of these solutions is $\{\pi_1, \pi_2\}$.

Hypervolume Hypervolume (Zitzler et al. 2003) is a metric to measure the quality of a Pareto optimal set. Specifically, given a Pareto optimal set, we normalize each objective value $z_i(\pi)$ to $[0,1]$ by scaling it to $z_i(\pi)^* = \frac{|z_i(\pi) - z_i(\pi)_{\min}|}{|z_i(\pi)_{\max} - z_i(\pi)_{\min}|}$, where $z_i(\pi)_{\max}$ and $z_i(\pi)_{\min}$ are the objective values from the reference points. We compute the hypervolume of the objective space dominated by the solutions, with higher hypervolume indicating better coverage of the objective space.

Problem Formulation

We are given a set of locations L along a river network R where each location is either an already built hydropower dam or a proposed location, along with a set of d objectives,

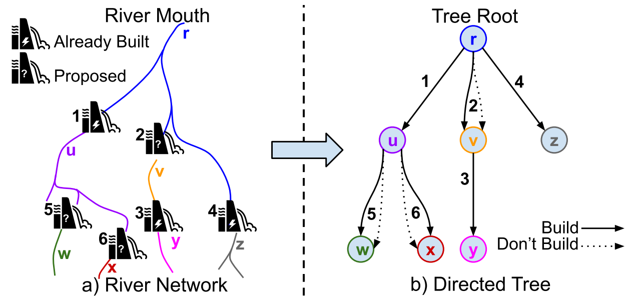


Figure 2: Converting a river network (a) to a multi-edged tree (b). Contiguous sections of the river between dams (letters) become nodes. Dams (numbers) become multi-edges connecting river sections, with an edge per decision.

where each objective has an associated value at each dam location $l \in L$ or along each river segment $r \in R$. We then convert the river network R and locations L to a multi-edged tree $T = (V, E)$ (Grimson et al. 2024). A node u is defined as the contiguous segments of the river R undisturbed by dam locations and each dam location $l \in L$ becomes a multi-edge $\{(u,v,0), ..., (u,v,m-1)\}$ connecting downstream to upstream river sections and containing $m \geq 1$ edges (see Figure 2). Each edge (u,v,i) as part of a multi-edge set represents each of the possible decisions at that location. For example, for already built dams, there may only be a single edge, but for proposed locations, there may be two edges representing the decisions to build and not build. More than two edges are possible when considering different dam configurations such as floating photovoltaics (Grimson et al. 2024). Each node u has objective values $r_u \in \mathbb{R}^d_{\geq 0}$. Each edge $(u,v,i) \in E$ also has for each objective a reward $s_{uvi} \in \mathbb{R}^d_{\geq 0}$, in addition to a set of passage probabilities $p_{uvi} \in [0,1]^d$ representing the proportion of an objective that passes through the edge. We also have lower and upper bounds $l \in \mathbb{R}^d$ and $u \in \mathbb{R}^d$ for the objectives, where a solution π is feasible if and only if $l^i \leq z(\pi)^i \leq u^i, \forall i \in [d]$.

Here, a solution π is defined as a spanning tree of the multi-edge tree, selecting exactly one edge from each set. We define a solution's objectives for tree T with root r as:

$$z^i(\pi, T) = r_r^i + \sum_{(v,w,k)\in\pi} \left(r_w^i \prod_{(x,y,j)\in \mathrm{P}(r,w,\pi)} p_{xyj}^i + s_{vwk}^i \prod_{(x,y,j)\in \mathrm{P}(r,v,\pi)} p_{xyj}^i \right)$$

where $P(u,v,\pi)$ is the path from u to v through the edges in π. A recursive formulation, using T_u as the sub-tree rooted at node u, is $z^i(\pi, T_u) = r_u^i + \sum_{(u,v,j)\in\pi} \left(s_{uvj}^i + p_{uvj}^i z^i(\pi, T_v) \right)$

In the original unconstrained algorithm (Wu et al. 2018), the authors employed dynamic programming (DP) to efficiently calculate the Pareto optimal set by recursively generating a Pareto optimal set of partial solutions at each node from the Pareto optimal set of partial solutions at the children. Specifically, the algorithm generates all possible solutions by taking the Cartesian product of each solution in a child's frontier with each possible edge decision between the parent and the child. It then determines the Pareto optimal set of partial solutions at a node as the non-dominated solutions. However, this algorithm does not consider constraints and may result in infeasible solutions.

Constraint-aware Pareto Optimization

Previous work on multi-objective optimization for tree-structured problems considered an unconstrained setting (Wu et al. 2018; Gomes-Selman et al. 2018; Grimson et al. 2024). In this work, we introduce box constraints on the objective space of the form $l \leq z(\pi) \leq u \in \mathbb{R}^d$.

Lower Bounds

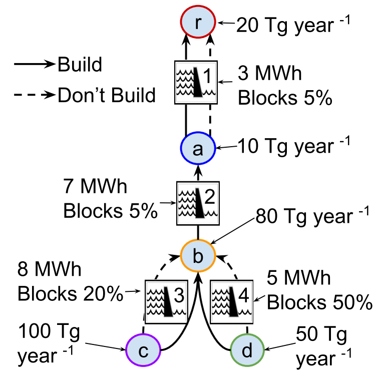
We first consider lower bounds $z(\pi) \geq l$. Lower bounds result in a Pareto optimal set that is a subset of the unconstrained Pareto optimal set.

Lemma 1. *Let $\mathcal{S} = \{\pi \subseteq E\}$ be the set of all feasible solutions where π is a spanning tree of E, let $z(\pi) \in \mathbb{R}^d$ be the d objective values for a solution π, and let $l \in \mathbb{R}^d$ be a set of lower bounds on each objective. Let $\mathcal{P} \subseteq \mathcal{S}$ be the Pareto optimal set to the unconstrained problem "$\max z(\pi)$", and let $\mathcal{P}_l \subseteq \mathcal{S}$ be the Pareto optimal set to the constrained problem "$\max z(\pi)$ subject to $z(\pi) \geq l$", then $\mathcal{P}_l \subseteq \mathcal{P}$.*

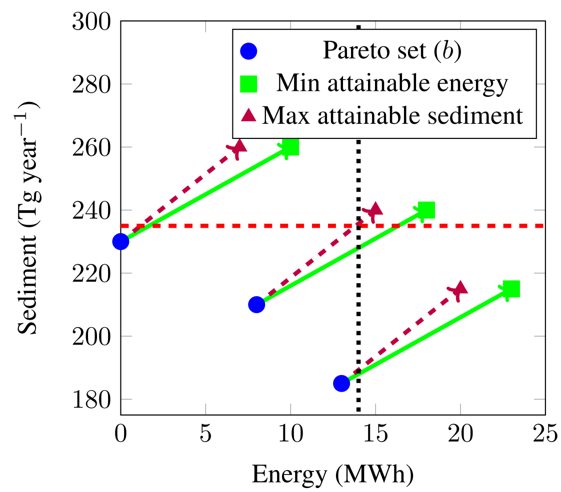
The proof is provided in the appendix. Combining Lemma 1 with Theorem 2 from (Wu et al. 2018), which shows Pareto optimal solutions can be recursively generated from partial Pareto optimal solutions, we can construct an extension of the algorithm that removes solutions that cannot meet the lower bound constraints. That is, for any node in $v \in V$ in the tree, and some Pareto optimal partial solution $\pi \in \mathcal{P}^v$ at v, we consider whether that solution can meet the lower bound constraints given the remaining decisions. If such a solution cannot possibly meet the lower bound constraint, it can be safely removed from $\mathcal{P}^v$.

Theorem 1. *Let $T = (V, E)$ with root r be a multi-edged tree and $v \in V$ be a node of the tree, and let $\mathcal{P}^v$ be the Pareto frontier at node v without considering lower bounds. Let $T_{\setminus v} \subseteq T$ be the tree excluding the sub-tree rooted at v, but including node v. Let $z_{\setminus v} \in \mathcal{R}^d$ be a vector of maximum attainable values for the d objectives in tree $T_{\setminus v}$. For any solution $\pi \in \mathcal{P}^v$, if $z_\pi + z_{\setminus v} < c_l$ for any of the d objectives, then any solution π' that contains the decisions in π will not exist in the constrained Pareto optimal set $\mathcal{P}$ of the root r.*

The proof is in the appendix. This is implemented in the pseudocode in the appendix in the ENFORCE_CONSTRAINTS method. As a consequence of Theorem 1, at node $v \in V$, for the Pareto frontier constrained by the lower bounds $\mathcal{P}_l^v$, and the unconstrained Pareto frontier $\mathcal{P}^v$, we have $\mathcal{P}_l^v \subseteq \mathcal{P}^v$ and $|\mathcal{P}_l^v| \leq |\mathcal{P}^v|$. Since the running time is directly proportional to the number of solutions considered, this results in a faster runtime when considering lower bounds. Figure 3b shows an example of using bounds to prune solutions based on the example in Figure 3a, with a lower bound on sediment.



(a) Example network rooted at r with four dam sites and two objectives: sediment and energy. Dam 2 has already been built, whereas 1, 3, and 4 have not.



(b) Pareto frontier points for b and the min/max attainable values. The red line is a lower bound on sediment, and the black line is an upper bound on energy. Point 1 is feasible.

Figure 3: Example network with constraints.

Upper Bounds

Now we consider upper bounds $z(\pi) \leq u$. Similarly to the lower bounds methods, we prune any solution that will necessarily break the upper bound by determining the minimum amount of a given objective that is forced to be accrued due to the nodes in the tree that have not been considered yet, namely for already built dams. For node $v \in V$, let $z_{(\setminus v)'}$ be the minimum value of an objective forced by the nodes outside the tree rooted at node v, then we can prune any solution π generated at each node such that $z(\pi) + z_{(\setminus v)'} > u$, which we introduce in the second pruning check in the ENFORCE_CONSTRAINTS method in the algorithm (provided in the appendix). However, by doing so, we can no longer guarantee the optimality of the solutions, nor can we guarantee a reduction in the number of solutions considered. In the appendix, we demonstrate a case where a Pareto solution for the constraint-aware problem is not found and a solution that is found is not Pareto optimal for the constrained

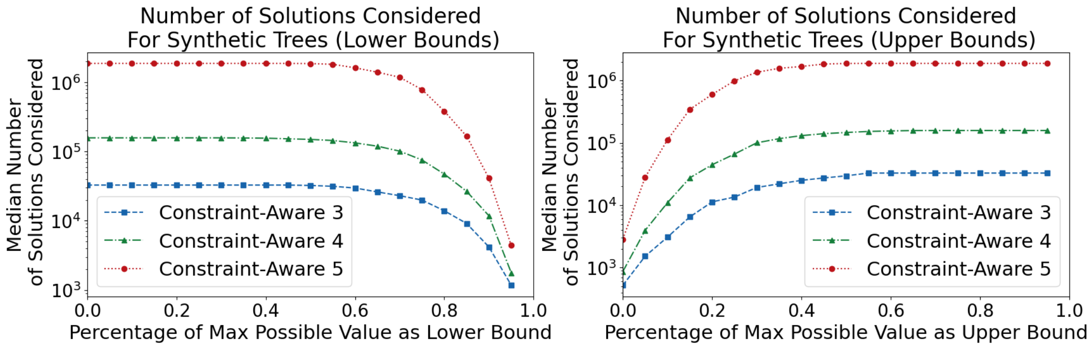


Figure 4: Median number of solutions considered on synthetic data as a function of the lower (left) and upper (right) bounds. As the number of objectives increases the number of solutions also increases. As the lower bound increases, the number of solutions considered decreases, and as the upper bound increases the number of solutions considered decreases.

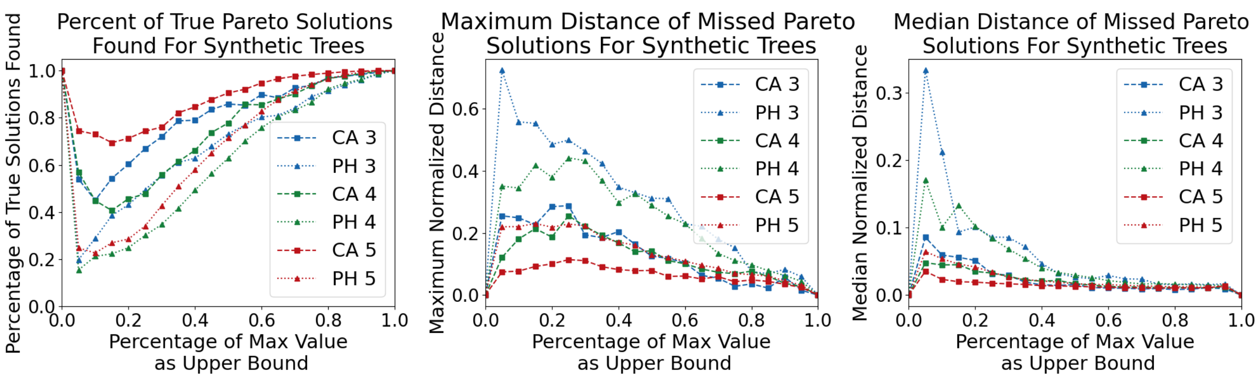


Figure 5: (Left) Percentage of true constrained Pareto optimal solutions (middle) maximum, and (right) median normalized distances from missed true constrained Pareto optimal solutions for synthetic trees at different upper bounds and numbers of objectives. Our constraint-aware method (CA) [squares] always finds more true Pareto solutions than post hoc removal (PH) [triangles], with this gap increasing with the number of objectives. Our constraint-aware method consistently finds closer points to the true constrained Pareto optimal solutions in the constrained region.

problem. While the Pareto solutions generated by the algorithm are not guaranteed to be optimal, doing so allows us to generate a frontier that covers all the solutions produced by post-processing the unconstrained Pareto frontier.

Theorem 2. *Let $T = (V, E)$ with root node r be a multi-edged tree, $\mathcal{S} = \{\pi \subseteq E\}$ be the set of all feasible solutions where each π is a spanning tree of the edges E, $z(\pi) \in \mathbb{R}^d$ be the d objective values for a given solution π, and $u \in \mathbb{R}^d$ be a set of upper bounds on each objective. For the optimization problem $\max z(\pi)$ subject to $z(\pi) \leq u$, let $\mathcal{P} \subseteq \mathcal{S}$ be the Pareto frontier found by the algorithm and $\mathcal{P}' \subseteq \mathcal{S}$ be the Pareto frontier found by solving the unconstrained optimization problem first and then removing all the solutions that violate the constraints. Then $\mathcal{P}' \subseteq \mathcal{P}$.*

The proof is provided in the appendix. Given Theorem 2, $\mathcal{P}$ has a hypervolume that is at least as large as $\mathcal{P}'$. The full pruning step of ENFORCE_CONSTRAINTS in the algorithm (provided in the appendix) considers both lower and upper bound constraints. Figure 3b depicts solution pruning based on the example in Figure 3a, where the upper bound on energy production results in only point 1 being feasible.

Empirical Evaluation

Synthetic Experiments

To methodically test our method, we generate a synthetic dataset consisting of 10 trees with 26 nodes that resemble real-world instances. For each tree, we calculated the maximum attainable value z^a for objective a, and constructed 21 uniform intervals from $[0, z^a]$ used as the bound. Instance generation details are in the appendix. We evaluate models for 3, 4, and 5 objectives and compare against previous state-of-the-art Pareto methods for tree-structured problems.

Lower Bounds As post hoc filtering for lower bounds is guaranteed to find the set of Pareto optimal solutions, we look at how lower bounds affect the number of solutions considered. For each tree and each lower bound interval on objective a, we ran the constraint-aware algorithm and tracked the number of solutions considered for 3, 4, and 5 objectives, showing the distribution of median values in Figure 4. A lower bound of 0 is equivalent to the unconstrained problem as they generate the same solution set, and at the lower bound of z^a there is only a single feasible so-

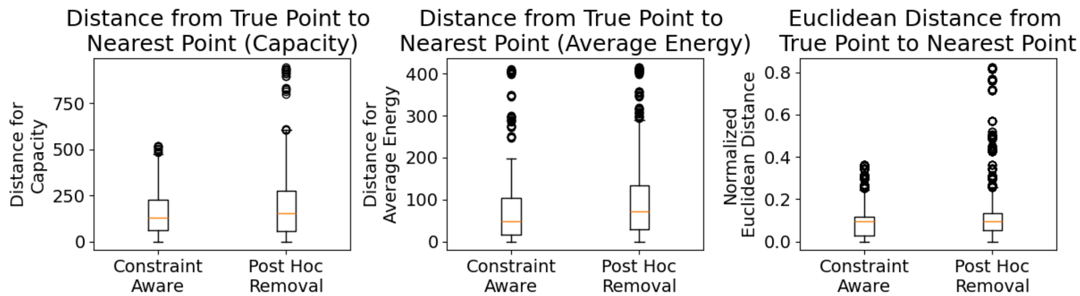


Figure 6: Constraint-aware vs. post hoc removal of solutions for the Cauca River compared with the true constrained Pareto frontier. The left and middle boxplots show the distances in the constrained objectives to the nearest Pareto optimal solution. The right plot shows normalized Euclidean distances across all 7 objectives to the nearest Pareto optimal solution.

Bounds	**Model**	**Runtime (sec)**	**# Pareto Solutions**	**# Pareto Solutions ≥ 99% UB**	**Normalized Hypervolume**
BAU-NZ	Post Hoc	41,907	22,420,662	15,181	0.318
BAU-NZ	Constraint-Aware	26,375	22,469,399	17,773	0.324
NZ-1%	Post Hoc	41,907	132,159	21,473	0.585
NZ-1%	Constraint-Aware	4,779	135,283	23,175	0.589

Table 1: Results for constraint-aware versus post hoc removal of solutions with ϵ rounding of 0.01 on the full Magdalena, considering Business-As-Usual (BAU) and Net Zero (NZ) 2050 targets, and within 1% (NZ-1%) of NZ targets. Constraint-aware runs faster and finds more solutions near the upper bound, with larger hypervolumes.

lution: building every dam. Overall, Figure 4 demonstrates the importance of using known lower bounds in reducing the search space and thus, runtime.

Upper Bounds We evaluate the impact of upper bounds on the resulting solution set compared to the true Pareto set. For each tree and each objective interval as an upper bound on objective a, we ran an exhaustive search, our constraint-aware algorithm, and an unconstrained optimizer with post hoc removal of infeasible solutions. An upper bound of 0 has one feasible solution and an upper bound of z^a is unconstrained. Tests consider 3, 4, and 5 objectives.

Results We compare the constraint-aware algorithm to post hoc filtering of an unconstrained Pareto frontier using the tree-structured Pareto optimization from (Grimson et al. 2024; Wu et al. 2018; Gomes-Selman et al. 2018), which was shown to outperform other methods in scale and quality of solutions by using the tree network. Figure 5 shows the median percentage of true solutions found, as well as the median and maximum distances from each of the missed true Pareto solutions to the nearest solution. In all cases, the constraint-aware algorithm finds at least as many solutions as the post hoc removal, and finds up to 50% more of the true Pareto solutions. Lastly, our method consistently finds points closer to missed optimal solutions.

Magdalena Hydropower Expansion

For the Magdalena River basin, which has a total of 143 proposed hydropower dam locations, we consider six objectives: average energy output, river connectivity, greenhouse gas emissions, sediment transport, number of displaced people, and area of inundated farmland. In this scenario, we consider constraints on average energy as well as installed capacity (maximum output) without treating the latter as an objective to optimize. The lower bounds are from Business-As-Usual 2050 scenario values (UPME 2018) of 15,900 MW and 8,700 MW equivalent for capacity and energy respectively, and the upper bounds were obtained from Colombia's Net Zero 2050 agenda (Gobierno de Colombia 2021) which estimates upper bound targets of 23,900 MW and 13,100 MW equivalent for capacity and energy respectively.

In all Magdalena experiments, we compare solutions from our constraint-aware method against post hoc removal of solutions based on the previous state-of-the-art work on unconstrained Pareto frontiers for tree-structured problems(Wu et al. 2018; Gomes-Selman et al. 2018; Grimson et al. 2024) which is able to scale up to larger numbers of objectives and decisions than other methods for the unconstrained problem.

Due to the large number of unbuilt dams in the Magdalena, it is infeasible to identify the true Pareto frontier for comparisons. Thus, we first consider a sub-basin of the Magdalena, the Cauca, consisting of 19 built and 28 unbuilt dams, where the search space is small enough to exhaustively identify the true frontier. We evaluate constraint-aware and post hoc removal against exhaustive search for the Cauca with scaled constraints: upper bounds of 11,010 MW and 6,400 MW and lower bounds of 6,660 MW and 3,860 MW for capacity and average energy respectively. Figure 6 shows the distances between missed solutions from the true frontier to the nearest solution in the other frontiers, where we see the constraint-aware method performs best.

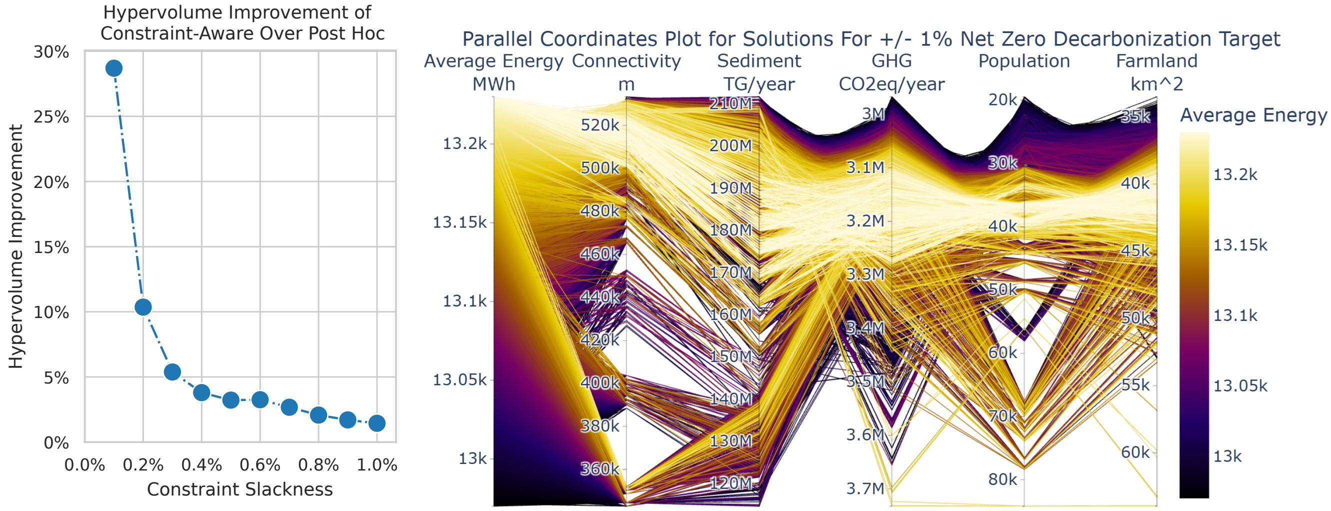


Figure 7: **Left:** Hypervolume improvement in regions close to the upper bounds for the Magdelena basin for the solutions in the constraint-aware algorithm vs. post hoc removal, focusing on solutions that fall into the region that is {Constraint Slackness} away from the upper bounds. The constraint-aware algorithm consistently gives higher coverage, and the improvement greatly increases close to the upper bound. **Right:** Parallel plot visualizing results within +/- 1% of the net zero targets for the six objectives: average energy (max), river connectivity (max), sediment (max), GHG emissions (min), population displacement (min), and inundated farmland (min). Each parallel coordinate corresponds to an objective and each line to a solution.

Finally, we run the post hoc removal and constraint-aware methods for the Magdalena River basin with an ϵ rounding of 0.01. In the original work (Wu et al. 2018), this ϵ value was an approximation. However, with the upper bounds, this is no longer a true approximation, but a rounding mechanism to reduce the solution set size. Table 1 shows the running times, the number of Pareto solutions found near the upper bound, and the normalized hypervolume coverage of the two methods. We see that in all metrics, the constraint-aware method performs better. Figure 7 shows the improvement of hypervolume in the regions close to the upper bounds. We see that the constraint-aware algorithm consistently provides better coverage for the regions close to the upper bounds and the closer we are to the bound, the more improvement we observe from using the constraint-aware algorithm. Figure 7 also visualizes the objective values of various solutions in a parallel plot, part of a tool provided to policymakers (https://www.cs.cornell.edu/gomes/udiscoverit/magdalena-ecovistas), for solutions found within 1% of hydropower expansion needed to meet Net Zero Emission targets for Colombia by 2050. While we see a smooth set of solutions across all average energy targets, we see clusters of solutions around different values for the ecosystem services. Specifically, distinct groupings are observed around different GHG and river connectivity values. These results underscore the importance of evaluating the full set of objectives and ensuing tradeoffs under reasonable constraints in decision making.

Future Work

For future work, it would be important to consider constraints beyond just those applied to the objective space, but also the decision space - for example if two planned hydropower dams are mutually exclusive. Additionally, while the strong theoretical guarantees of the solution quality along the upper bound no longer exists, there may yet exist guarantees on the proximity by considering whether the largest valued solution may fit in the gap between the current value and the upper bounds.

Conclusion

Unconstrained Pareto frontiers are highly unrealistic for many multi-objective optimization problems, as they contain many infeasible solutions that policymakers would never consider. As countries work towards decarbonization targets via hydropower dam expansion, they face realistic constraints based on expected energy needs. By considering constraints, we can leverage the reduced search space to run more efficiently via our constraint-aware algorithm, which identifies more Pareto solutions close to the upper bound constraints that are missed by an unconstrained solver, resulting in a broader distribution of solutions. More importantly, our constraint-aware Pareto algorithm performs well and provides valuable insights into a real-world case study on the Magdalena River basin in Colombia, with 143 dams in consideration, based on the energy needs of Colombia for the business-as-usual (UPME 2018) and the desired decarbonization targets (Gobierno de Colombia 2021) by 2050.

Our work highlights the importance of constraining the solution space to assist policymakers in investigating and finding meaningful solutions by considering realistic scenarios. We hope the research community continues improving constraint-aware Pareto optimization approaches, which are vital for addressing sustainable development goals.

Acknowledgements

This project is partially supported by Schmidt Sciences programs, a AI2050 Senior Fellowship and two Eric and Wendy Schmidt AI in Science Postdoctoral Fellowships; the National Science Foundation (NSF); the National Institute of Food and Agriculture (USDA/NIFA); the Air Force Office of Scientific Research (AFOSR); and the Cornell Atkinson Center for a Sustainable Future (ACSF).

References

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.

Deb, K.; and Jain, H. 2013. An evolutionary many-objective optimization algorithm using reference-point-based nondominated 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.

Doerr, B.; and Qu, Z. 2023a. A first runtime analysis of the NSGA-II on a multimodal problem. *IEEE Transactions on Evolutionary Computation*.

Doerr, B.; and Qu, Z. 2023b. From understanding the population dynamics of the NSGA-II to the first proven lower bounds. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 12408–12416.

Doerr, B.; and Qu, Z. 2023c. Runtime analysis for the NSGA-II: Provable speed-ups from crossover. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 12399–12407.

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*, 817–850. New York, NY: Springer New York. ISBN 978-1-4939-3094-4.

Gobierno de Colombia. 2021. Evaluación de las vías de neutralidad de carbono para Colombia a través de la metodología de toma de decisiones robustas (RDM) en varios escenarios futuros inciertos utilizando el modelo GCAM. Retrieved October 14, 2023 from https://e2050colombia.com/wp-content/uploads/estudios/EstudioAEvaluacionDe\\LasViasDeNeutralidad.pdf.

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.

Grimson, M.; Almeida, R.; Shi, Q.; Bai, Y.; Angarita, H.; Siqueira Pacheco, F.; Schmitt, R.; Flecker, A.; and Gomes, C. P. 2024. Scaling Up Pareto Optimization for Tree Structures with Affine Transformations: Evaluating Hybrid Floating Solar-Hydropower Systems in the Amazon. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(20): 22067–22075.

Hayes, C. F.; Rădulescu, R.; Bargiacchi, E.; Källström, J.; Macfarlane, M.; Reymond, M.; Verstraeten, T.; Zintgraf, L. M.; Dazeley, R.; Heintz, F.; et al. 2022. A practical guide to multi-objective reinforcement learning and planning. *Autonomous Agents and Multi-Agent Systems*, 36(1): 26.

Jiang, R.; Deng, Y.; Chen, Y.; Luo, H.; and An, B. 2024. Deep reinforcement learning for multi-objective game strategy selection. *Computers & Operations Research*, 106683.

Lin, X.; Zhen, H.-L.; Li, Z.; Zhang, Q.; and Kwong, S. 2019. Pareto Multi-Task Learning.

Ma, P.; Du, T.; and Matusik, W. 2020. Efficient Continuous Pareto Exploration in Multi-Task Learning.

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

Mısır, M.; and Cai, X. 2023. Algorithm Selection for Large-Scale Multi-objective Optimization. In *International Conference on Optimization and Learning*, 36–47. Springer.

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

Qu, Z.; Grimson, M.; Mao, Y.; Heilpern, S.; Miqueleiz, I.; Pacheco, F.; Flecker, A.; and Gomes, C. P. 2024. Strategies for Compressing the Pareto Frontier: Application to Strategic Planning of Hydropower in the Amazon Basin. In Dilkina, B., ed., *Integration of Constraint Programming, Artificial Intelligence, and Operations Research*, 141–157. Cham: Springer Nature Switzerland. ISBN 978-3-031-60599-4.

Sierra-Altamiranda, A.; Charkhgard, H.; Dayarian, I.; Eshragh, A.; and Javadi, S. 2024. Learning to project in a criterion space search algorithm: an application to multi-objective binary linear programming. *Optimization Letters*, 1–28.

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

United Nations. 2015. Transforming our world: the 2030 Agenda for Sustainable Development. Retrieved May 17, 2024 from https://sdgs.un.org/2030agenda.

UPME. 2018. Plan Energético Nacional 2020-2050. Retrieved May 17, 2024 from https://www1.upme.gov.co/DemandayEficiencia/Documents/PEN_2020_2050/Plan_Energetico_Nacional_2020_2050.pdf.

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

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).

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

Zheng, W.; Liu, Y.; and Doerr, B. 2022. A first mathematical runtime analysis of the Non-Dominated Sorting Genetic Algorithm II (NSGA-II). In *Conference on Artificial Intelligence, AAAI 2022*. AAAI Press. Preprint at https://arxiv.org/abs/2112.08581.

Zitzler, E.; Thiele, L.; Laumanns, M.; Fonseca, C. M.; and Da Fonseca, V. G. 2003. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on evolutionary computation*, 7(2): 117–132.