Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

Expanding Connected Components from Alternative Terminals: Global Optimization for Freshwater Fishes Under the UN's 30x30 Conservation Goal

Yue Mao[1], **Zhongdi Qu**[1], **Imanol Miqueleiz**[2], **Aaron Ferber**[1], **Sami Wolf**[1], **Marc Grimson**[1], **Sebastian Heilpern**[3], **Felipe S. Pacheco**[4], **Alexander S. Flecker**[4], **Peter B. McIntyre**[2], **Carla P. Gomes**[1]
[1]Department of Computer Science, Cornell University
[2]Department of Natural Resources and the Environment, Cornell University
[3]Department of Public and Ecosystem Health, Cornell University
[4]Department of Ecology and Evolutionary Biology, Cornell University
{ym277, zq84, im298, amf272, smw362, mg2425, s.heilpern, felipe.pacheco, asf3, pbm3}@cornell.edu, gomes@cs.cornell.edu

Abstract

Climate change and biodiversity loss are among humanity's most pressing challenges. In 2022, under the auspices of the United Nations, over 190 countries reached a historic agreement to address the alarming loss of biodiversity and restore natural ecosystems. Target 3, often referred to as "30x30", seeks to effectively protect and manage 30% of the world's terrestrial, inland water, coastal, and marine areas by 2030. In this work, we address the UN 30x30 target in the context of global freshwater fish conservation. Freshwater ecosystems are disproportionately unprotected, and their biota are declining at an alarming rate. Our goal is to select new protected areas that protect freshwater fish species as much as possible without exceeding total coverage of 30% of land area. To support this goal, we introduce the *Expansion of Connected Components from Alternative Terminals Problem*, a graph-based optimization problem that captures ecological priorities and connectivity constraints. We analyze its computational complexity, propose novel integer programming formulations, and develop scalable solution methods. We further evaluate its typical-case complexity under diverse settings and demonstrate that our approach scales to a global real-world scope, encompassing approximately 200,000 freshwater basins and 13,000 species, paving the way for implementing the 30x30 target on a worldwide scale.

1 Introduction

Humans share the Earth with millions of species, but our activities are threatening the planet's biodiversity at an unprecedented scale and intensity. In 2022, 196 UN countries signed the Kunming-Montreal Global Biodiversity Framework to address biodiversity loss and the restoration of natural ecosystems. Comprised of 23 targets, this agreement seeks to both halt biodiversity loss and benefit people. Target 3, often referred to as "30x30" seeks to protect and manage 30% of the world's terrestrial, inland water, coastal, and marine areas by 2030. Protected Areas (PAs) are central to the 30x30 approach to biodiversity conservation, yet currently encompass only 8% of ocean area and 12% of land and inland waters [UNEP-WCMC and IUCN, 2024]). Identifying a set of additional conservation areas that maximizes the protection of freshwater fish species within a 30% area limit is a pressing challenge that can only be met through the efficient use of computational tools.

With one in four freshwater species now at risk of extinction, preserving freshwater ecosystems is critical [Sayer *et al.*, 2025]. Unfortunately, past conservation efforts have primarily focused on terrestrial biodiversity, allowing many of the most species-rich rivers and lakes to remain unprotected by the current network of PAs [Miqueleiz *et al.*, 2023]. With the 2030 deadline approaching, integrating freshwater biodiversity into the plans for PA expansion is both urgent and increasingly complex, particularly considering various spatial scales.

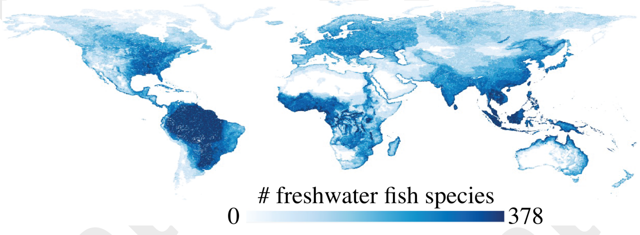

Figure 1: **The global distribution of freshwater fish species richness** (total of 13,740 species), where color saturation indicates the number of species in each hydrological unit (total of 190,675 units).

Our contributions: We seek to advance the UN's 30x30 goal by maximizing the inclusion of the world's freshwater fishes within an expanded network of protected areas. This endeavor is driven by a team of computer scientists and ecologists from Cornell University. Our multidisciplinary team jointly worked towards identifying priority regions where PA expansion would maximize the conservation outcomes for freshwater fishes. More specifically: **(1)** We introduce the *Expansion of Connected Components from Alternative Terminals Problem*. **(2)** We provide a formal model for the problem and a mixed integer program formulation. **(3)** We characterize its computational complexity for general graphs. **(4)** We characterize its empirical complexity for planar and tree-structured graphs. **(5)** We provide a detailed study of the real-world UN 30x30 worldwide freshwater conservation goal, from an ecological point of view, considering two scenarios for enforcing the 30% target. Remarkably, our methods scale globally, despite the large problem size of approximately 200,000 river basins and 13,000 fish species, paving the way for developing a 30x30 freshwater target on a worldwide scale (see Figure 4).

2 Problem Description

Our analyses are designed to support the UN's 30x30 goal for global conservation from a freshwater fish biodiversity perspective. While we acknowledge that other sociopolitical perspectives are important and we discuss them in Section 5, they fall beyond the scope of this paper. The ecological perspective serves as a crucial baseline for conservation planning, providing a foundational understanding of biodiversity protection before integrating socio-economic and policy dimensions. In this context, the overall goal is to identify basin units (sub-regions of river basins), where PAs should be established to maximize the conservation of freshwater fish species. The utility for protecting a region is measured by Rarity-Weighted Richness (RWR) [IUCN, 2025b]. The rarity score of a basin unit for a given species present in the basin is calculated as the basin's area divided by the species' total habitable area. Summing the rarity scores of all species in a basin unit yields the basin's RWR. The goal is to select basins to protect that maximize the total RWR. The UN protection target can be encoded as a constraint where the total area of PAs should be at most 30% of the world's freshwater area. Moreover, to create connected habitats where the fish can migrate freely, the PAs should form connected components, expanded from existing PAs.

To formally define the problem, we construct an undirected graph G where vertices represent basin units, and an edge exists between two vertices if the river flows from one basin unit to the other. For the 30x30 problem, since we are working with river networks where the river flows from higher to lower elevation without any cycles, the underlying graph G is a forest where each tree represents a connected river network. For each basin unit, the protection cost is its currently unprotected area and the protection utility is its RWR. The existing PAs act as terminal vertices from which we expand the connected protected components. Formally, the problem can be described as:

Definition 1 (Expansion of Connected Components from Alternative Terminals Problem (ECCAT)).
Input: *Given an undirected graph* $G = (V, E)$ *with terminal vertices* $T \subseteq V$*, vertex costs* $c : V \to \mathbb{R}_{\geq 0}$*, vertex utilities* $u : V \to \mathbb{R}_{\geq 0}$*, and a cost bound* $C \in \mathbb{R}_{\geq 0}$.
Output: *Find a vertex-induced subgraph* H *of* G *that maximizes* $\sum_{v \in H} u(v)$*, subject to the constraints that (i) each connected component in* H *contains at least one terminal vertex, and (ii)* $\sum_{v \in H} c(v) \leq C$.

To better understand the ECCAT problem and its hardness, we also consider the general case where the underlying G is any graph, including those with cycles. It is clear that the decision version of the problem is NP-complete for any class of graph G (even when G is a tree), since it can be reduced from the Knapsack problem [Kellerer *et al.*, 2004]. We also show its strong NP-completeness in the theorem below:

Theorem 1 (Strong NP-completeness)**.** *The decision version of the Expansion of Connected Components from Alternative Terminals Problem (*ECCAT*) is strongly* NP*-complete when the maximum degree of the graph is at least four.*

Proof. See Appendix A. ☐

3 Model Formulation

To solve the ECCAT problem, we formulate it as a mixed integer program (MIP). While the ECCAT problem is a general problem, for a more intuitive explanation, we describe its MIP formulation in the context of the 30x30 freshwater conservation problem. In the experimental section, we also consider more general settings. Given the river network G, we denote the set of basin units as $B = [n]$, the area of basin unit i as a_i, and the currently protected area of basin unit i as a'_i. We define the set of currently protected basin units as $P = \{i | a'_i = a_i\}$. Moreover, we denote the area budget ratio as β, which is 30% for the 30x30 problem. Each basin unit i is also associated with an r_i denoting its RWR. We introduce a decision variable $x_i \in \{0, 1\}$ for each basin unit i, where $x_i = 1$ represents protecting basin unit i. For all $i \in P$, x_i is fixed as 1.

We use a flow-based encoding for the connectivity constraint that all the proposed PAs are expanded from existing PAs. To do that, we turn the underlying graph $G = (B, E)$ into a directed graph $G' = (B', E')$, such that $B' = B \cup \{s\}$, where s is an additional source vertex. For every undirected edge $ij \in E$, there are two directed edges ij and ji in E'. Also, we add a directed edge si from the source s to every $i \in P$, the existing PAs. We introduce a decision variable $f_{ij} \in \mathbb{R}_{\geq 0}$ for each directed edge $ij \in E'$, representing the amount of flow along edge ij. The problem can now be encoded as:

$$
\begin{aligned}
\text{maximize} \quad & \sum_{i\in B} r_i x_i \\
\text{subject to} \quad & \sum_{i\in B}(a_i - a_i')x_i \le \beta \sum_{i\in B} a_i - \sum_{i\in B} a_i' && (1)\\
& 0 \le f_{ij} \le n x_j \qquad \forall ij \in E' && (2)\\
& \sum_{ki\in E'} f_{ki} = \sum_{ij\in E'} f_{ij} + x_i \qquad \forall i \in B && (3)\\
& x_i \in \{0,1\} \qquad \forall i \in B \\
& x_i = 1 \qquad \forall i \in P
\end{aligned}
$$

While the objective maximizes the overall RWR of all fully protected basins, constraint (1) ensures that the total area of protection is under the area budget. Constraints (2) and (3) encode the connectivity constraint. Together, they ensure that flow goes into basin i if and only if basin i is protected, i.e., $x_i = 1$. Flow is injected from the source into each existing PA and then propagates outward to expand the connected component of protected areas.

4 Related Work

Establishing protected areas for biodiversity conservation has fostered much interdisciplinary research. Previous work has proposed readily usable but restrictive frameworks for prioritizing protected areas that leverage simulated annealing [Ball *et al.*, 2009], integer linear programs [Beyer *et al.*, 2016; Hanson *et al.*, 2024; Deléglise *et al.*, 2024], iterative refinement [Moilanen *et al.*, 2005; Hamonic *et al.*, 2023], constraint programming [Justeau-Allaire *et al.*, 2021], and recently reinforcement learning [Silvestro *et al.*, 2022]. These approaches were developed for generally selecting protected areas without regard for the intricacies present in freshwater networks. Case studies have explored using these frameworks in specific freshwater conservation domains considering connectivity [Hermoso *et al.*, 2011], river flow [Hermoso *et al.*, 2012], protecting river stretches [Gomes-dos Santos *et al.*, 2019], inter-species interactions [Decker *et al.*, 2017; Nogueira *et al.*, 2023], local regulations [Howard *et al.*, 2018], and various other criteria [Nogales *et al.*, 2023]. These approaches, while raising interesting considerations, often sacrifice fidelity to the challenge of protecting river networks to fit in the confines of the restrictive general frameworks.

Previous work has also investigated tailoring optimization models to incorporate additional considerations such as enforcing connectivity of terrestrial corridors [Dilkina and Gomes, 2010; Beger *et al.*, 2022; Wang *et al.*, 2022; Dickson *et al.*, 2019; Gonzalez-Saucedo *et al.*, 2021; Lai *et al.*, 2011; Justeau-Allaire *et al.*, 2018; Conrad *et al.*, 2012], incorporating wildfire mitigation [Yemshanov *et al.*, 2023], and designing spatial conservation networks [Beger *et al.*, 2022]. These approaches leverage techniques in graph theory, circuit theory, network optimization, mathematical programming, and constraint programming to encode intricate problem-specific constraints and objectives to model ecological, social, behavioral, and environmental considerations. Several survey articles provide overviews on related aspects such as operations research methods for corridor optimization [Alagador and Cerdeira, 2022], circuit theory applications in conservation planning [Dickson *et al.*, 2019], AI applications in landscape ecology [Frazier and Song, 2025], and the broader area of computational sustainability [Gomes *et al.*, 2019]. None of these works address the same problem as ours, although some adopt MIP models with related themes.

Variations of knapsack incorporating graph-related constraints have been explored in the literature. One of the most relevant formulations is the Connected Knapsack problem [Dey *et al.*, 2024], a standard knapsack problem optimizing the value of selected items subject to a budget constraint but also requiring that the chosen subset forms a connected component in a graph. This problem has been shown to be strongly NP-complete. In [Dey *et al.*, 2024], the authors also proposed an approximation algorithm and an exact algorithm. In Tree Knapsack Problem (TKP), the items also form a tree structure but the subgraph of the selected items must include a single fixed terminal. Here, previous work proposed exact algorithms using dynamic programming [Johnson and Niemi, 1983], and branch-and-bound with specialized bounds [Shaw and Cho, 1998], while other work explored applications in network design [Shaw *et al.*, 1997]. While these graph-constrained variants of knapsack are thematically similar to our work, the consideration that we expand from alternative existing terminals prevents us from readily applying the polynomial or pseudo-polynomial algorithms from previous works.

In a similar vein, previous work has explored knapsack variants incorporating dependency constraints between items and presented pseudo-polynomial algorithms [Lalou and Kheddouci, 2023]. In [Aghezzaf *et al.*, 1995], the authors investigate the problem of finding optimal unweighted constrained subtrees rooted at specified nodes. Finally, the knapsack problems with item conflicts and forcing constraints have been extensively studied [Pferschy and Schauer, 2017; Zhou *et al.*, 2024], using dynamic programming [Gurski and Rehs, 2019] in special graph classes, branch-and-bound [Bettinelli *et al.*, 2017]. A survey extensively covers knapsack variants [Cacchiani *et al.*, 2022]. However, none of these previous approaches resemble our problem formulation or can be directly modified to solve it, due to the structural and combinatorial properties unique to our setting.

5 Experimental Results

In this section we present empirical results for the ECCAT problem. Our main focus is to address the real-world application of the UN 30x30 target for freshwater fish conservation. Additionally, we seek to deepen our understanding of the empirical complexity of general instances of the problem, particularly by identifying its critical complexity parameters. To this end, we conducted a series of experiments to assess its typical-case complexity. We examined both general planar graphs, which well capture the structure of land conservation problems, and trees, which model the configuration of freshwater ecosystems. This dual approach allows us to explore the factors influencing the problem complexity across different but relevant graph topologies.

We first discuss the empirical complexity of the ECCAT problem instances for planar and tree-structured graphs, followed by results concerning the real-world UN 30x30 target for worldwide freshwater fish conservation. For all experiments, we used Gurobi 11.0.0 on a cluster with Intel Xeon 3.47 GHz CPUs and 1.5TB of memory.

5.1 Empirical Easy-Hard-Easy Complexity Patterns

Planar Graphs

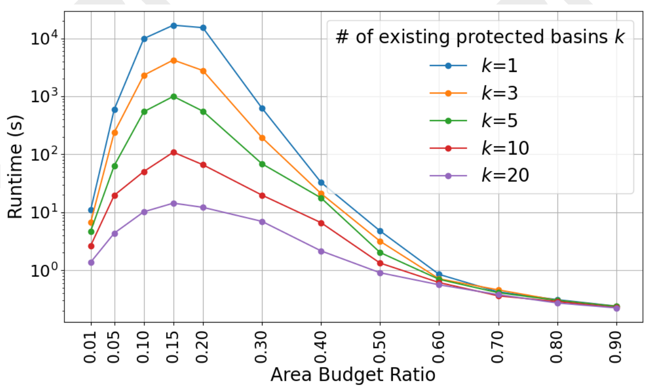

Figure 2: **Runtime vs. Area Budget Ratio for a** 20×20 **grid (400 decision variables).** We observe an *easy-hard-easy pattern* in the empirical complexity as we vary the area budget ratio, with the hardest instances occurring when the area budget ratio is around 15%. The easy-hard-easy pattern is observed for different numbers of existing protected areas, but interestingly, the overall computational complexity decreases as the number of protected areas increases.

We start by considering planar graphs modeled as $l \times l$ lattices (grids) of basin nodes, in which we randomly place k existing protected basins, varying the budget ratio with respect to the unprotected area[1]. These graph structures capture the topology of maps and are therefore suitable for land conservation studies.

For each target budget ratio in our experiments, we generate 20 instances with randomly placed protected basins. For each instance, we perform 10 runs where the basin area and score values are uniformly sampled from $(0, 100)$. The runtime for an instance is computed as the median of the 10 runs. The overall runtime is then determined as the median across the 20 instances.

Figure 2 summarizes the computational time for 20×20 grid instances (400 decision variables per instance). We observe a clear "easy-hard-easy" pattern in the empirical computational complexity (see e.g., [Mitchell *et al.*, 1992; Conrad *et al.*, 2012; Yadav *et al.*, 2018]), with the hardest instances occurring around a 15% budget ratio. While the easy-hard-easy pattern is consistent across different numbers of existing protected areas (PAs), interestingly, the overall computational complexity decreases as the number of existing PAs increases. Intuitively, having more existing PAs reduces the number of effective choices and constrains the solution space, making the problem "easier" to solve.

Trees

The 30x30 worldwide freshwater conservation problem concerns river networks, where the underlying graph G is a forest, with each tree representing a connected river system. Given this structure, we also examined the empirical complexity of tree instances, specifically in a "star" shape, where each instance consists of l branches extending from a center node, with l additional nodes along each branch (except for one branch with $l-1$ additional nodes), denoted as $l \times l$ stars.

For runtime comparisons, we generated instances the same way as for the planar graphs and analyzed instances of the same size. Figure 3 summarizes the runtime results for 20×20 star instances (400 decision variables per instance), matching the instance sizes plotted in Figure 2. The most striking takeaway from Figure 3 is that the runtimes for trees are exponentially lower than those for planar graphs, with a three-order-of-magnitude speedup. We hypothesize that this efficiency stems from the network-flow nature of our encoding, which effectively leverages the hierarchical structure of trees.

As we will discuss in the next section, the tree structure of real-world instances combined with additional problem-specific structures (as opposed to randomly generated data), further enhances computational efficiency—leading to surprisingly fast runtimes. Further research is needed to identify the critical parameters of trees.

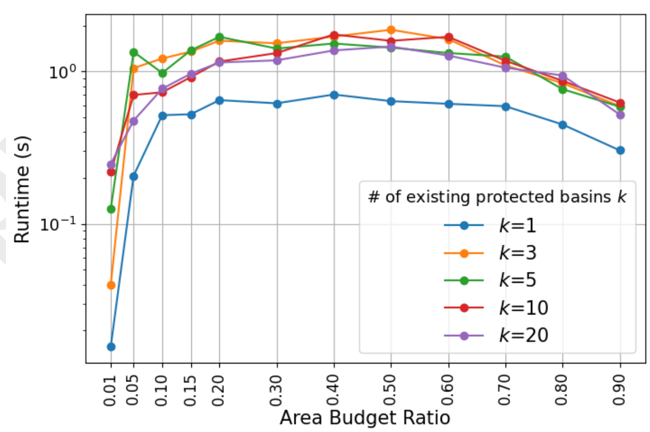

Figure 3: **Runtime vs. Area Budget Ratio for** 20×20 **star instances (400 decision variables per instance).** The key message from this plot is that the runtimes for trees are exponentially faster than those for planar graphs of the same size (see Figure 2), with a three-order-of-magnitude speedup.

5.2 Adapting the UN's 30x30 Framework for Global Conservation of Freshwater Fishes

To address the UN 30x30 target for worldwide freshwater fish conservation, we optimized the PA expansion of the world by solving the ECCAT problem using the formulation described in Section 3 with a global 30% area budget ratio. Apart from

[1]For example, a budget ratio of $= 0.2$ means that 20% of the unprotected area can be selected for protection, independently of the area of protected basin units, which varies across instances.

this, we also performed a second experiment where we imposed the 30% area target per *main basin* (a connected, independent river network, e.g., the Amazon). The first experiment with global 30% target is referred to as the *Global* approach, and the second experiment with main-basin-wise 30% area budget ratio is referred to as the *Per Main Basin* approach (details in Appendix B). We compare the current status with the outcomes of both approaches and discuss their tradeoffs, focusing on the effectiveness of species protection and practical feasibility.

Dataset

We combined two key datasets: the HydroBASINS dataset [Lehner *et al.*, 2022] and the IUCN Red List [IUCN, 2025a]. The HydroBASINS dataset partitions the world's freshwater area into 190,675 basin units across 23,996 main basins. Each main basin is a connected river network, a tree in our model, consisting of connected basin units. For each basin unit, we calculate its total area and currently protected area [UNEP-WCMC and IUCN, 2024]. Current PAs cover 11.81% of the world (Figure 4, top panel). The IUCN Red List provides data on the habitats of 13,740 freshwater fish species. By merging these datasets, we calculate the total Rarity-Weighted Richness (RWR) of each basin.

Metrics

We calculate several metrics to evaluate the conservation success and feasibility of the two approaches. For each fish species, we compute a) the percentage of its habitat that is protected, referred to as the "protection score", and b) its protection score counting only connected components of protected basin units with a total area of at least the species' Minimum Viable Range (MVR), referred to as the "effective protection score". The MVR is the smallest geographic area required to support a self-sustaining long-term population of the species. Out of the 13,740 species, [IUCN, 2025a] provided MVR on 11,451 of them, for which we calculate the effective protection score. Additionally, we computed the fraction of the protected land within each country.

Results and Discussion

The runtime of the Global approach experiment was 1h 20min — a remarkable result given the scale of the problem, which involves 190,675 decision variables corresponding to the basin units. This efficiency suggests that the rich tree-structure of the real-world instance is well-suited to the network-flow nature of our encoding, leading to a relatively fast runtime, despite the large number of decision variables.

The solution under the Global approach suggests the expansion of densely connected PAs from existing ones (Figure 4, middle panel). Notably, due to the ecological goal of maximizing the conservation of freshwater fish species, significant PA expansions occur in some of the world's most biodiverse rivers, such as the Mississippi and Amazon rivers in the Americas, the African Great Lakes, and the Western Ghats and Mekong River in southern Asia (Figure 1).

The Per Main Basin approach (Figure 4, bottom panel) produces a more spatially distributed PA expansion plan compared to the Global Approach. Allocating protections up to the 30% target within every major river basin avoids concentrating PAs in only the most species-rich basins, thereby increasing protection in regions that might otherwise receive less priority under a global optimization strategy. Since some species are endemic to these smaller basins and do not exist in the large, highly biodiverse river systems, this Per Main Basin approach also provides a more balanced protection for a broader range of species, as we will discuss more in the paragraphs below.

Species Protection Figure 5 shows histograms of the protection score and effective protection score of the species for the current status, the Global approach, and the Per Main Basin approach. Notably, both approaches shift the distributions towards higher protection score compared to the current status where protection scores are highly skewed towards low values. Currently, the aggregate species protection score is 2189.33, whereas the Global approach increases it to 8462.28 (a 286.5% increase) and the Per Main Basin approach increases it to 6930.44 (a 216.6% increase). For the aggregate effective protection score, the value is increased from the current 213.38 to 6552.37 (about 31 times) by the Global approach, and to 4732.42 (about 22 times) by the Per Main Basin approach. Notably, both approaches lead to a drastic increase in the effective protection scores, since enforcing connectivity results in large, contiguous protected areas. These areas create protected regions that are sufficiently extensive to support self-sustaining species populations.

While the Per Main Basin approach affords smaller aggregate protection to freshwater fish species, it achieves more spatially balanced protection. As is shown in Figure 5, for both the protection score and effective protection score, the Global approach results in U-shape distributions, indicating a bimodal and uneven allocation of species protection where many species receive either very low or very high scores. This imbalance is mitigated to some extent by the Per Main Basin approach, which results in a more convex frequency distribution. Currently, 1,282 species have zero protection score (i.e., their habitats are completely outside of existing PAs), where the Global approach reduces the count to 558, and the Per Main Basin Approach further reduces it to 345. As for the effective protection score, the number of species with a zero score is reduced from the current 8,031 to 2,885 by the Global approach, and further to 1,690 by the Per Main Basin approach. We observe that most of the unprotected species only inhabit areas with low species richness. For example, for each species, we calculate the total RWR across all the basin units it inhabits. The median is 21.23 across all species, while among the 1,690 species that observe zero effective protection score by the Per Main Basin approach, 1,507 of them have the total habitat RWR below this median. Thus, these species are left out from protection because assigning their habitats as PAs would not provide efficient protection across all species.

Protected Areas per Country As shown in Figure 6 middle panel, the Global approach imposes highly uneven protection requirements on different countries (e.g. Brazil would have to protect 79% of its territory), making global conservation efforts less balanced and potentially challenging to im-

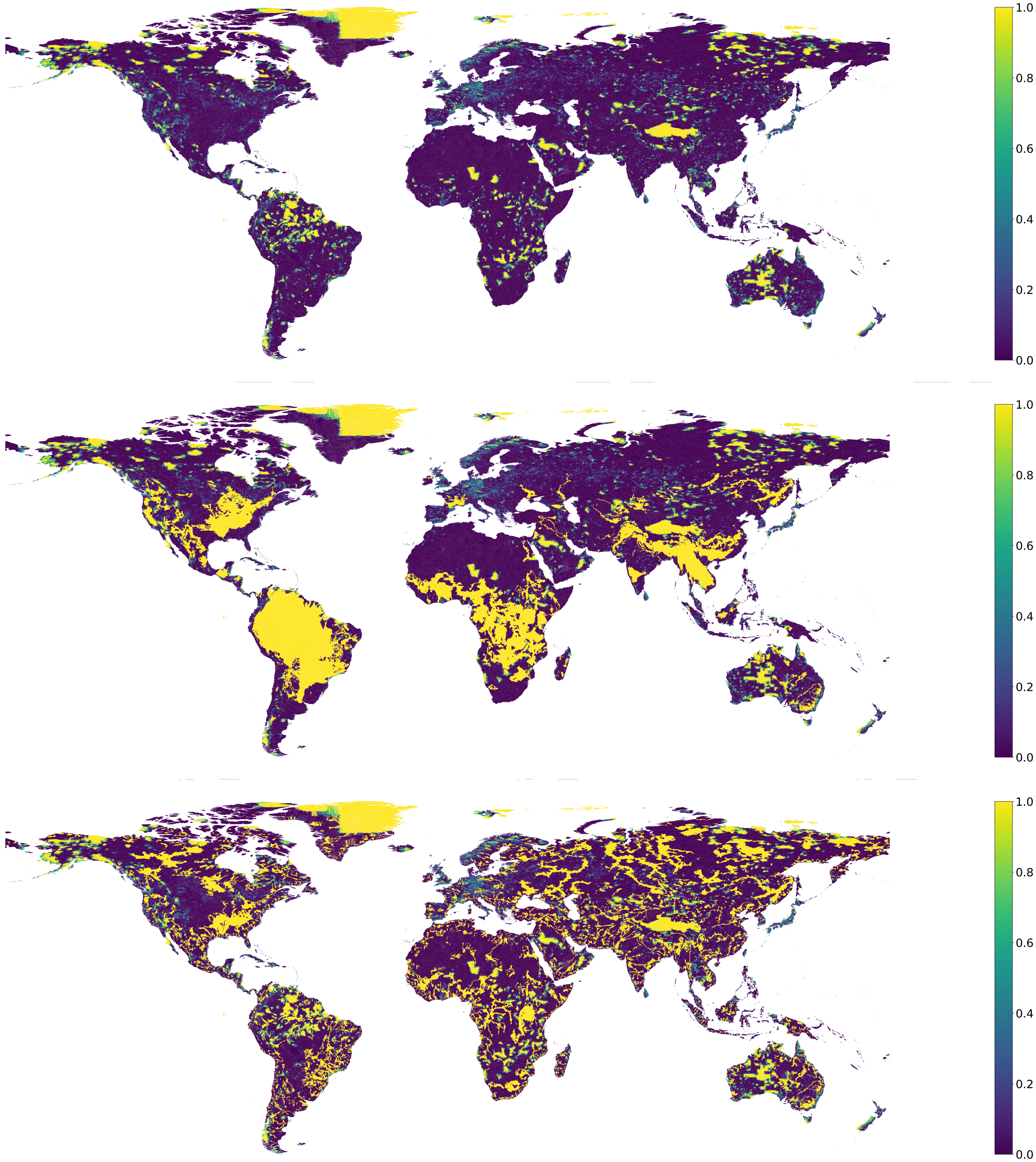

Figure 4: **World map colored by the freshwater protection level.** The color coding reflects the proportion of each basin unit protected. **Top panel: current status.** Basins colored in yellow are currently fully protected and provide the starting points from which new protected areas expand. **Middle panel: Global approach.** Solution enforcing the 30% protection target globally. **Bottom panel: Per Main Basin approach.** Solution enforcing the 30% protection target per main basin, rather than globally, distributing the conservation efforts more equitably than the global target approach. Details in Section 5.2.

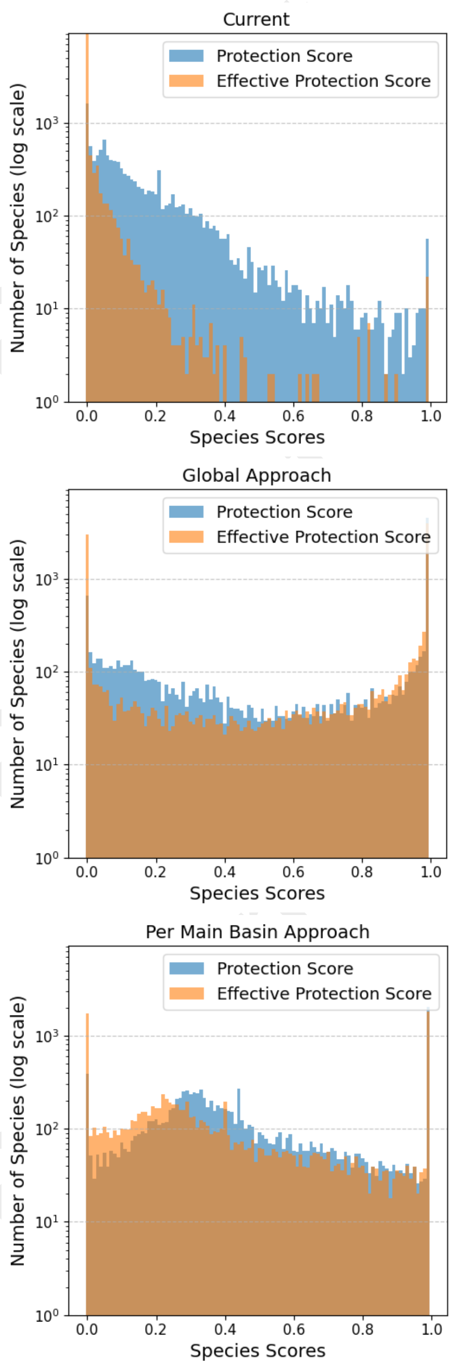

Figure 5: **Histograms of species protection scores and effective protection scores. Top panel:** current status. **Middle panel:** Global approach. **Bottom panel:** Per Main Basin approach. Both approaches significantly increase the total protection of the species and the effective protection score. While the Per Main Basin approach is less optimal in terms of the aggregate protection, it provides a more even protection across species. Details in Section 5.2.

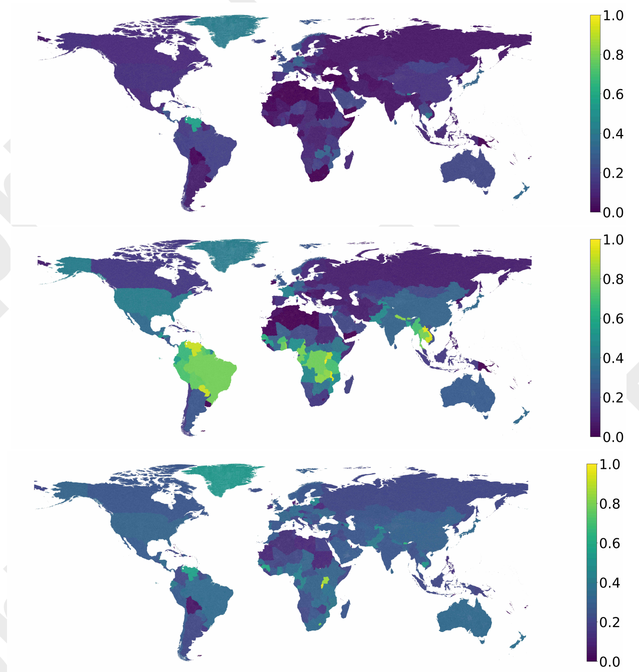

Figure 6: **Protected Areas per Country. Top panel:** current status. **Middle panel:** Global approach. **Bottom panel:** Per Main Basin approach. Compared to the Global approach, the Per Main Basin approach results in a more equitable distribution of conservation efforts across different countries.

plement. This issue is alleviated by the Per Main Basin approach, which results in a more equitable distribution of conservation efforts across countries (Figure 6, bottom panel) .

6 Conclusions

We introduced the *Expansion of Connected Components from Alternative Terminals Problem*, characterized its typical-case and worst-case complexity, and proposed novel formulations and algorithms to solve it. Our experimental results on planar graphs reveal an easy-hard-easy pattern in their typical complexity, and substantial faster runtimes for trees, which is the topology of our motivational application: the UN 30x30 target for worldwide freshwater fish conservation. Remarkably, our methods scale to a global level, despite the large problem size — approximately 200,000 decision variables and 13,000 species. In this study, we emphasized an ecological perspective, focusing on maximizing freshwater fish species protection, while enforcing PA connectivity, under the 30% global conservation target. Our algorithm naturally identified conservation expansions in some of the world's most biodiverse rivers, yet creating an imbalance conservation effort for different regions in the world. We also explored a variant of the problem in which the 30% protection constraint was applied per main basin rather than globally. While this approach is suboptimal from a total protection score perspective, it offers two main advantages: it redistributes conservation efforts more equitably across different basins and countries, and provides effective protected habitats for more species.

As with most sustainability challenges, real-world progress relies not only on ecological perspectives but also integration of socio-economic and geopolitical factors that are central to conservation outcomes, like countries' uneven responsibility in freshwater biodiversity protection and possible offsetting mechanisms. While these considerations are beyond the scope of this paper, they are critical for the practical implementation of large-scale conservation policies. We will further explore these issues in future multi-objective studies to develop more balanced and actionable conservation strategies. Neverthess, our ecological framework provides a critical foundation for freshwater conservation planning, serving as a baseline before integrating broader socio-economic and policy considerations. By demonstrating scalable methods for optimizing connected conservation areas, we hope this research paves the way for other researchers to contribute to the implementation of the UN 30x30 target, expanding its applicability to diverse ecosystems and policy contexts worldwide.

A Proof of Theorem 1

Proof. Clearly, ECCAT is in NP, as one can verify in polynomial time that the cost and utility of the connected components satisfy the decision bounds and that each connected component contains at least one terminal. To show NP-hardness, we reduce from CONNECTED KNAPSACK, which was shown to be strongly NP-complete, when the maximum degree of the graph is at least four [Dey *et al.*, 2024].

Given an instance $(\mathcal{G}, (w(u))_{u\in\mathcal{V}}, (\alpha(u))_{u\in\mathcal{V}}, s, d)$ of CONNECTED KNAPSACK [Dey *et al.*, 2024], we can solve for each $v \in V(\mathcal{G})$ an instance $(G = \mathcal{G}, T = \{v\}, c = w, u = \alpha, C = s, U = d)$ of ECCAT (where U is introduced to turn ECCAT into a decision problem where one needs to find whether an H exists such that $\sum_{v\in H} u(v) \geq U$).

If any of the ECCAT instances is a YES instance, then there is a connected subgraph H of G such that $\sum_{v\in H} c(v) \leq C$ and $\sum_{v\in H} u(v) \geq U$. Then we know that the CONNECTED KNAPSACK instance is a YES instance. If none of the ECCAT instances is a YES instance, then there is no vertex $v \in V(G)$ such that there exists a connected subgraph H where $\sum_{v\in H} c(v) \leq C$ and $\sum_{v\in H} u(v) \geq U$. Thus the CONNECTED KNAPSACK instance is a NO instance. □

B Per Main Basin Approach

In Section 5.2, we introduced the *Per Main Basin* approach to apply a 30% area budget ratio within each main basin, as an alternative of the Global approach. Herein, we explain the details of this Per Main Basin approach.

Among the 23,996 main basins, 277 have an Rarity-Weighted Richness (RWR) of zero. These main basins are left unchanged, as establishing PAs in these regions would provide no conservation benefit. Additionally, 4,713 main basins have already designated at least 30% of their area as PAs, meeting the protection target. Therefore, these main basins are also left unchanged. Ultimately, after enforcing the 30% area budget ratio for main basins with nonzero RWR and less than 30% current protection, the final solution allocates 27.04% of the global area for protection, slightly below the 30.00% protected area with the Global approach.

Currently, 22,118 main basins have nonzero RWR but no existing protected basins (i.e., basin units that are currently fully protected). This suggests that with our model formulation in Section 3, these main basins would get zero protection because there is no existing PA to expand from. However, many of them have non-negligible species richness (e.g. 11,455 of them have RWR above 0.0056, which is the median across all main basins), and provide habitat to some endemic and rare species. Therefore, it is crucial to establish PAs in these main basins. To address this, we modified our model to allow each of these main basins to choose one unprotected basin unit (the "seed" of expansion) to be protected and expand it to one connected component of PAs, while still making sure the total protected area is below the area constraint. Specifically, in our encoding, we introduce binary decision variables $z_i, \forall i \in B$, indicating if basin unit i should be the "seed" of expansion. Formally:

$$\text{maximize} \sum_{i\in B} r_i x_i$$

$$\text{s.t.} \sum_{i\in B} (a_i - a'_i) x_i \leq \beta \sum_{i\in B} a_i - \sum_{i\in B} a'_i \tag{4}$$

$$0 \leq f_{ij} \leq n x_j, \qquad \forall ij \in E' \tag{5}$$

$$\sum_{ki\in E'} f_{ki} = \sum_{ij\in E'} f_{ij} + x_i, \qquad \forall i \in B \tag{6}$$

$$\sum_{i\in B} z_i = 1 \tag{7}$$

$$f_{si} \leq n z_i, \qquad \forall i \in B \tag{8}$$

$$x_i \geq z_i, \qquad \forall i \in B \tag{9}$$

$$\begin{aligned} &x_i \in \{0,1\}, && \forall i \in B \\ &x_i = 1, && \forall i \in P \\ &z_i \in \{0,1\}, && \forall i \in B \end{aligned} \tag{10}$$

Constraint (7) ensures that there is exactly one "seed", and constraint (8) ensures that flow can only be injected into the graph through the "seed" (i.e., the basin with $z_i = 1$). Constraint (9) ensures that the "seed" is protected. Same as the encoding in Section 3, constraint (4) ensures that the total area of protection is under the area budget, and constraints (5) and (6) encode the connectivity by ensuring that flow goes into basin i if and only if basin i is protected, i.e., $x_i = 1$.

Acknowledgements

This project is partially supported by Schmidt Sciences programs, a AI2050 Senior Fellowship and three 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

[Aghezzaf *et al.*, 1995] EH Aghezzaf, TL Magnanti, and LA Wolsey. Optimizing constrained subtrees of trees. *Mathematical Programming*, 1995.

[Alagador and Cerdeira, 2022] Diogo Alagador and Jorge Orestes Cerdeira. Operations research applicability in spatial conservation planning. *Journal Of Environmental Management*, 2022.

[Ball *et al.*, 2009] Ian R Ball, Hugh P Possingham, and Matthew Watts. Marxan and relatives: software for spatial conservation prioritisation. *Spatial conservation prioritisation: Quantitative methods and computational tools*, 2009.

[Beger *et al.*, 2022] Maria Beger, Anna Metaxas, Arieanna C. Balbar, Jennifer A. McGowan, Remi Daigle, Caitlin D. Kuempel, Eric A. Treml, and Hugh P. Possingham. Demystifying ecological connectivity for actionable spatial conservation planning. *Trends In Ecology & Evolution*, 2022.

[Bettinelli *et al.*, 2017] Andrea Bettinelli, Valentina Cacchiani, and Enrico Malaguti. A branch-and-bound algorithm for the knapsack problem with conflict graph. *INFORMS Journal On Computing*, 2017.

[Beyer *et al.*, 2016] Hawthorne L Beyer, Yann Dujardin, Matthew E Watts, and Hugh P Possingham. Solving conservation planning problems with integer linear programming. *Ecological Modelling*, 2016.

[Cacchiani *et al.*, 2022] Valentina Cacchiani, Manuel Iori, Alberto Locatelli, and Silvano Martello. Knapsack problems - an overview of recent advances. part i: Single knapsack problems. *Computers & Operations Research*, 2022.

[Conrad *et al.*, 2012] Jon M. Conrad, Carla P. Gomes, Willem-Jan van Hoeve, Ashish Sabharwal, and Jordan F. Suter. Wildlife corridors as a connected subgraph problem. *Journal of Environmental Economics and Management*, 2012.

[Decker *et al.*, 2017] Emilia Decker, Simon Linke, Virgilio Hermoso, and Juergen Geist. Incorporating ecological functions in conservation decision making. *Ecology and evolution*, 2017.

[Deléglise *et al.*, 2024] Hugo Deléglise, Dimitri Justeau-Allaire, Mark Mulligan, Jhan-Carlo Espinoza, Emiliana Isasi-Catalá, Cecilia Alvarez, Thomas Condom, and Ignacio Palomo. Integrating multi-objective optimization and ecological connectivity to strengthen peru's protected area system towards the 30* 2030 target. *Biological Conservation*, 2024.

[Dey *et al.*, 2024] Palash Dey, Sudeshna Kolay, and Sipra Singh. Knapsack: Connectedness, path, and shortest-path. In *LATIN*, 2024.

[Dickson *et al.*, 2019] Brett G. Dickson, Christine M. Albano, Ranjan Anantharaman, Paul Beier, Joe Fargione, Tabitha A. Graves, Miranda E. Gray, Kimberly R. Hall, Josh J. Lawler, Paul B. Leonard, Caitlin E. Littlefield, Meredith L. McClure, John Novembre, Carrie A. Schloss, Nathan H. Schumaker, Viral B. Shah, and David M. Theobald. Circuit-theory applications to connectivity science and conservation. *Conservation Biology*, 2019.

[Dilkina and Gomes, 2010] Bistra Dilkina and Carla P. Gomes. Solving connected subgraph problems in wildlife conservation. In *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*, 2010.

[Frazier and Song, 2025] Amy E Frazier and Lei Song. Artificial intelligence in landscape ecology: recent advances, perspectives, and opportunities. *Current Landscape Ecology Reports*, 2025.

[Gomes *et al.*, 2019] Carla Gomes, Thomas Dietterich, Christopher Barrett, Jon Conrad, Bistra Dilkina, Stefano Ermon, Fei Fang, Andrew Farnsworth, Alan Fern, Xiaoli Fern, et al. Computational sustainability: Computing for a better world and a sustainable future. *ACM*, 2019.

[Gomes-dos Santos *et al.*, 2019] André Gomes-dos Santos, Elsa Froufe, Duarte V Gonçalves, Ronaldo Sousa, Vincent Prié, Mohamed Ghamizi, Hassan Benaissa, Simone Varandas, Amílcar Teixeira, and Manuel Lopes-Lima. Freshwater conservation assessments in (semi-) arid regions: Testing river intermittence and buffer strategies using freshwater mussels (bivalvia, unionida) in morocco. *Biological conservation*, 2019.

[Gonzalez-Saucedo *et al.*, 2021] Zaira Y. Gonzalez-Saucedo, Alejandro Gonzalez-Bernal, and Enrique Martinez-Meyer. Identifying priority areas for landscape connectivity for three large carnivores in northwestern mexico and southwestern united states. *Landscape Ecology*, 2021.

[Gurski and Rehs, 2019] Frank Gurski and Carolin Rehs. Solutions for the knapsack problem with conflict and forcing graphs of bounded clique-width. *Operations Research*, 2019.

[Hamonic *et al.*, 2023] François Hamonic, Cécile Albert, Basile Couëtoux, and Yann Vaxès. Optimizing the ecological connectivity of landscapes. *Networks*, 2023.

[Hanson *et al.*, 2024] Jeffrey O Hanson, Richard Schuster, Matthew Strimas-Mackey, Nina Morrell, Brandon PM Edwards, Peter Arcese, Joseph R Bennett, and Hugh P Possingham. Systematic conservation prioritization with the prioritizr r package. *Conservation Biology*, 2024.

[Hermoso *et al.*, 2011] V Hermoso, S Linke, J Prenda, and HP Possingham. Addressing longitudinal connectivity in the systematic conservation planning of fresh waters. *Freshwater Biology*, 2011.

[Hermoso *et al.*, 2012] Virgilio Hermoso, Mark J Kennard, and Simon Linke. Integrating multidirectional connectivity requirements in systematic conservation planning for freshwater systems. *Diversity and Distributions*, 2012.

[Howard *et al.*, 2018] Jeanette K Howard, Kurt A Fesenmyer, Theodore E Grantham, Joshua H Viers, Peter R Ode, Peter B Moyle, Sarah J Kupferburg, Joseph L Furnish, Andrew Rehn, Joseph Slusark, et al. A freshwater conservation blueprint for california: prioritizing watersheds for freshwater biodiversity. *Freshwater Science*, 2018.

[IUCN, 2025a] IUCN. The iucn red list of threatened species. version 2024-2. https://www.iucnredlist.org/, 2025. Accessed: 2025-02-04.

[IUCN, 2025b] IUCN. Other spatial data downloads – iucn red list of threatened species. https://www.iucnredlist.org/resources/other-spatial-downloads, 2025. Accessed: 2025-02-04.

[Johnson and Niemi, 1983] D. S. Johnson and K. A. Niemi. On knapsacks, partitions, and a new dynamic programming technique for trees. *Operations Research*, 1983.

[Justeau-Allaire *et al.*, 2018] Dimitri Justeau-Allaire, Philippe Birnbaum, and Xavier Lorca. Unifying reserve design strategies with graph theory and constraint programming. In *Constraint Programming*, 2018.

[Justeau-Allaire *et al.*, 2021] Dimitri Justeau-Allaire, Ghislain Vieilledent, Nicolas Rinck, Philippe Vismara, Xavier Lorca, and Philippe Birnbaum. Constrained optimization of landscape indices in conservation planning to support ecological restoration in new caledonia. *Journal of Applied Ecology*, 2021.

[Kellerer *et al.*, 2004] H. Kellerer, U. Pferschy, and D. Pisinger. *Knapsack Problems*. Springer, 2004.

[Lai *et al.*, 2011] Katherine Lai, Carla Gomes, Michael Schwartz, Kevin McKelvey, David Calkin, and Claire Montgomery. The steiner multigraph problem: wildlife corridor design for multiple species. In *AAAI*, 2011.

[Lalou and Kheddouci, 2023] Mohammed Lalou and Hamamache Kheddouci. Pseudo-polynomial algorithms for solving the knapsack problem with between items. *Computers & Operations Research*, 2023.

[Lehner *et al.*, 2022] Bernhard Lehner, Mathis L Messager, Maartje C Korver, and Simon Linke. Global hydro-environmental lake characteristics at high spatial resolution. *Scientific Data*, 2022.

[Miqueleiz *et al.*, 2023] Imanol Miqueleiz, Arturo H Ariño, and Rafael Miranda. Spatial priorities for freshwater fish conservation in relation to protected areas. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 2023.

[Mitchell *et al.*, 1992] David Mitchell, Bart Selman, and Hector Levesque. Hard and easy distributions of sat problems. In *AAAI*, 1992.

[Moilanen *et al.*, 2005] Atte Moilanen, Aldina MA Franco, Regan I Early, Richard Fox, Brendan Wintle, and Chris D Thomas. Prioritizing multiple-use landscapes for conservation: methods for large multi-species planning problems. *Royal Society Biological Sciences*, 2005.

[Nogales *et al.*, 2023] Jonathan Nogales, Carlos Rogéliz-Prada, Miguel A Cañon, and Andres Vargas-Luna. An integrated methodological framework for the durable conservation of freshwater ecosystems: a case study in colombia's caquetá river basin. *Frontiers in Environmental Science*, 2023.

[Nogueira *et al.*, 2023] Joana Garrido Nogueira, Manuel Lopes-Lima, Pedro Beja, Ana Filipa Filipe, Elsa Froufe, Duarte V Gonçalves, Janine P da Silva, Ronaldo Sousa, Amílcar Teixeira, Simone Varandas, et al. Identifying freshwater priority areas for cross-taxa interactions. *Science of The Total Environment*, 2023.

[Pferschy and Schauer, 2017] Ulrich Pferschy and Joachim Schauer. Approximation of knapsack problems with conflict and forcing graphs. *Journal Of Combinatorial Optimization*, 2017.

[Sayer *et al.*, 2025] Catherine A Sayer, Eresha Fernando, Randall R Jimenez, Nicholas BW Macfarlane, Giovanni Rapacciuolo, Monika Böhm, Thomas M Brooks, Topiltzin Contreras-MacBeath, Neil A Cox, Ian Harrison, et al. One-quarter of freshwater fauna threatened with extinction. *Nature*, 2025.

[Shaw and Cho, 1998] Dong X. Shaw and Geon Cho. The critical-item, upper bounds, and a branch-and-bound algorithm for the tree knapsack problem. *Networks*, 1998.

[Shaw *et al.*, 1997] DX Shaw, G Cho, and HS Chang. A depth-first dynamic programming procedure for the extended tree knapsack problem in local access network design. *Telecommunication Systems*, 1997.

[Silvestro *et al.*, 2022] Daniele Silvestro, Stefano Goria, Thomas Sterner, and Alexandre Antonelli. Improving biodiversity protection through artificial intelligence. *Nature Sustainability*, 2022.

[UNEP-WCMC and IUCN, 2024] UNEP-WCMC and IUCN. Protected planet report 2024. Technical report, UNEP-WCMC and IUCN, 2024.

[Wang *et al.*, 2022] Yicheng Wang, Peng Qin, and Hayri Onal. An optimisation approach for designing wildlife corridors with ecological and spatial considerations. *Methods In Ecology And Evolution*, 2022.

[Yadav *et al.*, 2018] Nitin Yadav, Carsten Murawski, Sebastian Sardina, and Peter Bossaerts. Phase transition in the knapsack problem. *arXiv preprint arXiv:1806.10244*, 2018.

[Yemshanov *et al.*, 2023] Denys Yemshanov, Denyse A. Dawe, Amanda Bakalarczyk, Ning Liu, Yan Boulanger, Jonathan Boucher, Alexandre Beauchemin, Dominique Arseneault, Mathieu Leblond, and Marc-Andre Parisien. Balancing wildlife protection and wildfire threat mitigation using a network optimization approach. *Frontiers In Forests And Global Change*, 2023.

[Zhou *et al.*, 2024] Qing Zhou, Jin-Kao Hao, Zhong-Zhong Jiang, and Qinghua Wu. An effective hybrid search method for the quadratic knapsack problem with conflict graphs. *Operations Research*, 2024.