

# Efficiently Optimizing for Dendritic Connectivity on Tree-Structured Networks in a Multi-Objective Framework

Qinru Shi  
Center for Applied Mathematics,  
Cornell University  
qs63@cornell.edu

Jonathan M. Gomes-Selman  
Department of Computer Science,  
Stanford University  
jgs8@stanford.edu

Roosevelt García-Villacorta  
Department of Ecology and  
Evolutionary Biology,  
Cornell University  
rg676@cornell.edu

Suresh Sethi  
U.S. Geological Survey, New York  
Cooperative Fish and Wildlife  
Research Unit, Cornell University  
suresh.sethi@cornell.edu

Alexander S. Flecker  
Department of Ecology and  
Evolutionary Biology,  
Cornell University  
asf3@cornell.edu

Carla P. Gomes  
Department of Computer Science,  
Cornell University  
gomes@cs.cornell.edu

## ABSTRACT

We provide an exact and approximation algorithm based on Dynamic Programming and an approximation algorithm based on Mixed Integer Programming for optimizing for the so-called dendritic connectivity on tree-structured networks in a multi-objective setting. Dendritic connectivity describes the degree of connectedness of a network. We consider different variants of dendritic connectivity to capture both network connectivity with respect to long and short-to-middle distances. Our work is motivated by a problem in computational sustainability concerning the evaluation of trade-offs in ecosystem services due to the proliferation of hydropower dams throughout the Amazon basin. In particular, we consider trade-offs between energy production and river connectivity. River fragmentation can dramatically affect fish migrations and other ecosystem services, such as navigation and transportation. In the context of river networks, different variants of dendritic connectivity are important to characterize the movements of different fish species and human populations. Our approaches are general and can be applied to optimizing for dendritic connectivity for a variety of multi-objective problems on tree-structured networks.

## ACM Reference Format:

Qinru Shi, Jonathan M. Gomes-Selman, Roosevelt García-Villacorta, Suresh Sethi, Alexander S. Flecker, and Carla P. Gomes. 2018. Efficiently Optimizing for Dendritic Connectivity on Tree-Structured Networks in a Multi-Objective Framework. In *COMPASS '18: ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS)*, June 20–22, 2018, Menlo Park and San Jose, CA, USA. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3209811.3209878>

## 1 INTRODUCTION

Many real-world problems in computational sustainability involve multiple, often conflicting objectives concerning environmental, economic, and societal issues [9]. In multiple objective optimization,

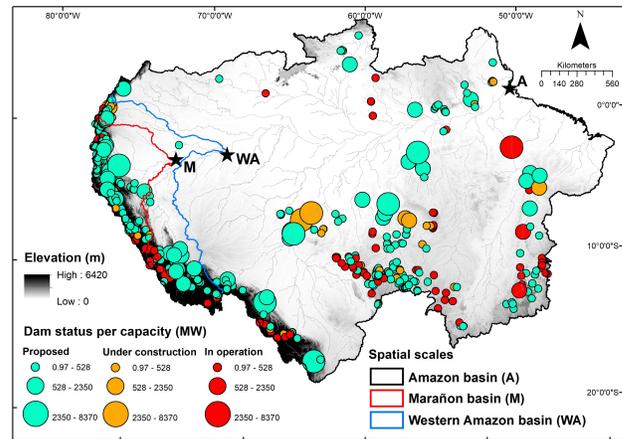
ACM acknowledges that this contribution was authored or co-authored by an employee, contractor, or affiliate of the United States government. As such, the United States government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for government purposes only.

*COMPASS '18, June 20–22, 2018, Menlo Park and San Jose, CA, USA*

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-5816-3/18/06...\$15.00

<https://doi.org/10.1145/3209811.3209878>



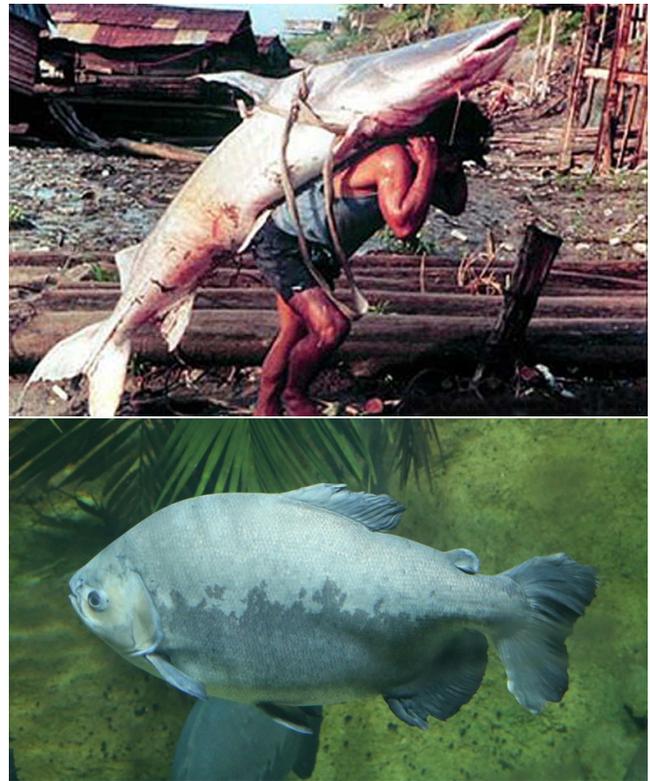
**Figure 1: Dams throughout the Amazon basin: around 300 dams are proposed or planned. The Marañón basin includes 128801 river segments; the Western Amazon basin includes 455156 river segments, and the Amazon basin includes 4083059 river segments.**

we are interested in understanding the so-called Pareto frontier, which captures the trade-offs between the objectives for different solutions. Recently there has been considerable interest in the study of multi-objective optimization problems (see e.g., [3, 4, 8, 15, 16, 18–21, 23–25, 30]). The work presented in this paper is motivated by multi-objective optimization problems that arise in the context of the management of river networks. River networks are good examples of physical systems that are naturally captured by a tree-structured network. In particular, we consider the problem of placement of hydropower dams throughout the Amazon basin: around 300 dams are proposed or planned with the potential of dramatically affecting a variety of ecosystem services, in particular concerning freshwater fisheries [5, 11, 26] (Fig. 1). We are interested in studying the trade-offs between energy production and river connectivity due to the proliferation of hydropower dams throughout the Amazon basin.

River infrastructure provides an array of ecosystem services including power generation (e.g. hydropower dams), water storage (e.g. potable water reservoirs), and water conveyance (e.g. road culverts). Yet, such projects often present physical barriers to the movement of aquatic organisms. When scaled up across landscapes, river connectivity barriers have driven profound changes in aquatic ecosystems [6, 7]. Fisheries represent a key ecosystem service from river systems, playing foundational roles in food security, recreation, and cultural identity. Almost all fish species move during the course of their life cycle, transiting between habitats to breed, forage, or to escape predators [14]. Thus, fisheries-based ecosystem services are particularly vulnerable to river connectivity loss. Planning to restore connectivity in impacted systems, or to pro-actively site infrastructure problems to minimize connectivity impacts, is a top challenge in maintaining sustainable ecosystem service production from rivers. The migrations of fish species span a continuum of distances, and are often divided into two major groups: (1) long-distance migrants that utilize the full river network length, such as *diadromous* species which move between the ocean to headwaters of rivers (1000s of kms) (Fig. 2, top panel) [1] and (2) species that move within more localized regions of river networks (e.g. 10s-100s of kms), such as *potadromous* species which exclusively utilize freshwater habitats throughout their life cycle. Cote et al. [2] developed the *Dendritic Connectivity Index (DCI)* to describe the degree of river network connectedness, which can be mapped onto fish life histories. One version, DCI for longitudinal connectivity, describes the probability that the path between a root reference location and any other point in the network is unimpeded. This connectivity index is suited to describe how long-distance migrants use river networks, such as Pacific salmon (*Oncorhynchus spp.*) in North America, or Amazonian goliath catfish (*Brachyplatystoma spp.*) (Fig. 2, top panel) which make migrations over 5000 km from river mouth estuaries to headwater spawning habitats, and as such is labeled DCI<sub>l</sub> for *diadromy*. Alternatively, DCI<sub>p</sub> for *potadromy*, quantifies the probability that the path between any two locations in river network is unimpeded by a barrier. The DCI<sub>p</sub> metric focuses on localized river connectivity, and thus better describes the life history needs for fish that move short to middle distances (Fig. 2, bottom panel).

Contemporary watershed planning efforts utilize DCI metrics as objective criteria to measure habitat fragmentation [12, 13, 27–29]. Planning problems involving siting barriers to minimize connectivity losses can in some cases involve decisions with 100s of potential dams, however, typical planning problems for barrier removals to maximize connectivity involve much greater complexity. For example, optimization efforts to evaluate hydropower proliferation in the Amazon basin [26] have involved a similar number of barriers as a recent optimization exercise to mitigate connectivity impacts from road crossing culverts in a single small watershed in Alaska [22]. Furthermore, connectivity criteria may be only one of multiple objectives in watershed planning. For these reasons, efficient algorithms to optimize over DCI<sub>p</sub> are needed to support river connectivity planning.

In [26] and [10], we proposed different approaches for computing the exact and approximate Pareto frontier, with application to the problem of hydropower placement throughout the Amazon



**Figure 2: Amazonian fish species. (Top panel) Migratory goliath catfish (*Brachyplatystoma*): species in this genus travel throughout much of the Amazon basin during their life cycle, and make the longest migrations known for freshwater fishes, including measured distances between spawning and nursery areas exceeding 5700 km (Photo credit:USGS). (Bottom panel) Short-distant migrant (*Colossoma macropomum*): a commercially important fruit-eating fish that migrates from rivers to flooded Amazon forests during periods of high water. These fish species play vital roles in food security, recreation, and cultural identity, throughout the Amazon basin.(Photo credit: Wikipedia:Rufus46)**

basin. In particular, we proposed a multi-objective dynamic programming algorithm to compute the exact Pareto frontier on tree-structured networks, which is motivated by the single-objective algorithms in [27–29]. We also proposed a rounding technique applied to the exact dynamic programming algorithm that provides a fully polynomial-time approximation scheme (FPTAS). 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  $1/\epsilon$ . Our FPTAS for approximating the Pareto frontier is inspired by the FPTAS proposed in [27–29]. A key difference is the fact that the FPTAS algorithms proposed in [27–29] do not consider the Pareto frontier, but rather a single objective subject to a constraint. To complement the DP algorithms, we formulated the problem of optimizing the placement of dams as a mixed integer programming problem

(MIP) and used it to approximate the Pareto frontier. In [26], we considered **only longitudinal connectivity (i.e.,  $DCI_D$ )**. In this paper, we consider **both longitudinal connectivity (i.e.,  $DCI_D$ ) as well as localized connectivity (i.e.,  $DCI_P$ )**, which requires substantially different encodings, both when considering MIP and DP formulations.

**Our Contributions:** (1) We provide an exact algorithm based on dynamic programming (DP) to compute the Pareto frontier on tree-structured networks, for  $DCI_P$  (and other criteria). Note that for the longitudinal connectivity, we only need to consider the movement upwards from the root of the tree. The complexity of dealing with localized connectivity arises from the fact that we have to consider the *potential movement between any two possible locations throughout the network*. (2) We developed a DP-based fully polynomial-time approximation scheme (FPTAS) for approximating the Pareto frontier for tree-structured networks, for  $DCI_P$  (and other criteria). (3) We also provide a MIP encoding for the problem of optimizing the placement of hydropower dams using  $DCI_P$  (and other criteria). We use it to approximate the Pareto frontier following the scheme proposed in [17, 26]. (4) We provide empirical results on different basins of the Amazon, when optimizing for energy,  $DCI_P$ , and  $DCI_D$ , showing interesting trade-offs between the DP and MIP approaches and also the different connectivity metrics.

**Preview of Results:** Our DP and MIP Pareto frontier algorithms are complementary. The DP approach can compute the exact Pareto frontier, which is not possible with MIP. While the MIP approach can provide faster results, the DP provides a much better coverage of the Pareto frontier in number of solutions. The MIP is also more flexible, allowing the incorporation of additional constraints. Finally, we observe significantly different solutions when optimizing for  $DCI_P$  vs  $DCI_D$ : solutions optimized for  $DCI_P$  tend to prefer fewer, but bigger, dams than the solutions optimized for  $DCI_D$ .

## 2 PROBLEM FORMULATION

In this section, we first present the formal definition of  $DCI_P$ . Then, we will demonstrate how to encode  $DCI_P$  as one of the objectives in the hydropower dam placement problem.

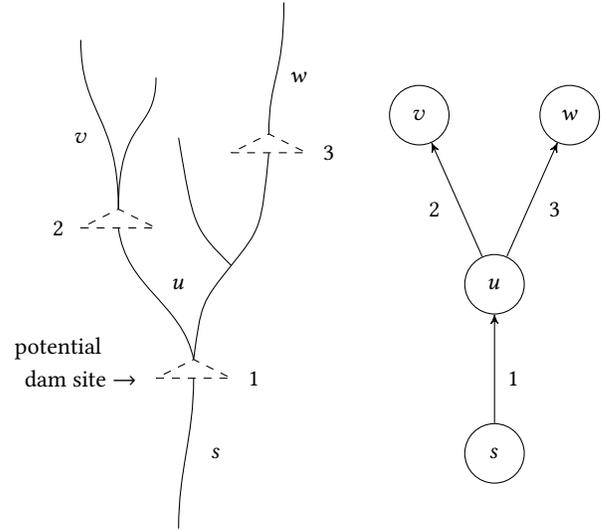
### 2.1 Formal Definition of $DCI_P$

The formal definition of Dendritic Connectivity Index (DCI) proposed in [2] captures *the probability that a fish can successfully migrate from its starting point to its destination in a river network*. When we are considering diadromous fish that migrate between marine and freshwater environments, we only need to compute the probability that a fish can successfully migrate from the mouth of the river to its destination in the river network. Let  $l_1$  be the length of the connected section from the mouth of the river and  $L$  be the length of the whole river network. Under the assumption that the possible locations of the destination are distributed evenly across the entire network, the DCI for diadromous fish is defined as

$$DCI_D = \frac{l_1}{L} \times 100,$$

which is essentially the longitudinal connectivity objective used in [26].

For potadromous fish that migrate within the freshwater system, we make the assumption that the fish migrate between two



**Figure 3: Abstracting a river network (left) into a directed rooted tree (right).  $s$  is the root node (river mouth).**

randomly chosen locations in the river network. Let  $s_1, s_2, \dots, s_k$  be the contiguous sections in the river network uninterrupted by dams or other barriers, and let  $l_1, l_2, \dots, l_k$  be the lengths of these sections, respectively. Let  $L$  be the length of the whole river network. Let  $p_{ij}$  be the passability coefficient between  $s_i$  and  $s_j$ , i.e. the probability that a fish can move between  $s_i$  and  $s_j$ . Then, the DCI for potadromous fish is defined as

$$DCI_P = \sum_{1 \leq i, j \leq k} \Pr[\text{move from } s_i \text{ to } s_j] \cdot p_{ij} \times 100 \quad (1)$$

$$= \sum_{1 \leq i, j \leq k} \frac{l_i}{L} \frac{l_j}{L} p_{ij} \times 100 \quad (2)$$

In this paper, we only consider the case where the barriers that divide the river network are absolutely non-passable, which is often true for most Amazonian dams. This assumption implies that

$$p_{ij} = \begin{cases} 0, & \forall i \neq j, \\ 1, & \text{otherwise.} \end{cases}$$

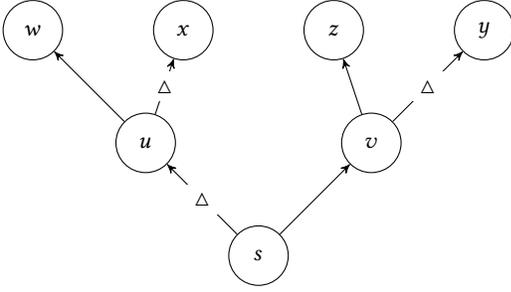
Then, we can simplify  $DCI_P$  to

$$DCI_P = \sum_{i=1}^k \frac{l_i^2}{L^2} \times 100. \quad (3)$$

The new formula of  $DCI_P$  in (3) makes it possible for us to apply a dynamic programming algorithm. In the MIP formulation of the DCI problem, we still use the definition of  $DCI_P$  in (2).

### 2.2 Encoding of $DCI_P$

The hydropower dam placement problem can essentially be described as: we are given a river network with a set of proposed dams, and we need to find an optimal subset  $\pi$  of the proposed dams to build considering certain objectives. To encode this problem as a multi-objective optimization problem, we first need to abstract the river network into a directed rooted tree as suggested by [26, 28]



**Figure 4: Example of a solution: The triangle means a dam is built at the location of the edge**

to reduce the size of the problem. We represent each contiguous section of the river network uninterrupted by existing or potential dam locations as a hypernode and each potential dam location as a directed edge pointing from downstream to upstream. See Fig. 3 for an example of the conversion.

A solution (or policy)  $\pi$  is a subset of the proposed dams we decide to build. We can encode each objective as a function of  $\pi$ . For instance, the energy objective ( $E$ ) can be encoded as

$$E(\pi) = \sum_{e \in \pi} h_e,$$

where  $h_e$  represent the proposed hydropower output of dam  $e$ .

In the case of  $\text{DCI}_P$ , we assume that  $V$  is the set of all hypernodes and  $c_v$  is total length of river segments in node  $v \in V$ . We still assume  $L$  to be the total length of the entire river network. When we use the definition of  $\text{DCI}_P$  in (2), let  $x_{uv}$  be the indicator variable of whether  $u$  and  $v$  are connected. Then, we have

$$\text{DCI}_P(\pi) = \left( \sum_{u,v \in V} c_u c_v x_{uv}(\pi) \right) / L^2 \times 100.$$

When we use the definition of  $\text{DCI}_P$  in (3), we assume that  $|\pi| = k-1$  and these dams divide  $V$  into  $k$  subset  $s_1, s_2, \dots, s_k$ . Then, we have

$$\text{DCI}_P(\pi) = \left( \sum_{i=1}^k \left( \sum_{v \in s_i} c_v \right)^2 \right) / L^2 \times 100$$

For instance, in Fig. 4, we decide to build dams  $(s, u)$ ,  $(u, x)$ , and  $(v, y)$ . The  $\text{DCI}_P$  value of this solution is then

$$\text{DCI}_P(\pi) = \frac{(c_s + c_v + c_z)^2 + (c_u + c_w)^2 + c_x^2 + c_y^2}{L^2} \times 100.$$

### 3 DP ALGORITHM

In [26], we proposed a dynamic programming (DP) based algorithm shown in Algorithm 1 for solving multi-objective optimization problems on a tree. The DP algorithm can compute the Pareto frontier exactly for value functions defined on a tree-structured network and find all Pareto-optimal policies. For faster pruning of sub-optimal solutions, the original tree is converted into an equivalent binary tree.

The main idea of the algorithm is to only consider the network above the current node and recursively compute the Pareto-optimal

partial solutions at each node from the Pareto-optimal partial solutions of its children. The validity of the DP algorithm depends on the following theorem:

**THEOREM 1.** *Let  $l$  and  $r$  be the children of  $u$ . Any Pareto-optimal partial solution at  $u$  can be constructed by combining one Pareto-optimal partial solution from node  $l$ , one from node  $r$  and four different joint states of edge  $(u, l)$  and  $(u, r)$ .*

However, this theorem does not hold when one of the objectives is  $\text{DCI}_P$ . A simple counterexample is shown in Fig. 5. If we directly apply the DP algorithm to optimize for energy and  $\text{DCI}_P$  on the river network shown in Fig. 5, we will miss a Pareto-optimal solution.

---

**Algorithm 1:**  $\text{Pareto}_{\mathcal{T}}(u)$ : compute the Pareto frontier for the value function defined on the subtree of  $\mathcal{T}$  rooted at node  $u$ .

---

```

1 if is_leaf( $u$ ) then
2   | return  $\{(r_u^1, \dots, r_u^m)\}$ ; // values of objectives
3 else
4   |  $l \leftarrow u.\text{left\_child}$ ;
5   |  $r \leftarrow u.\text{right\_child}$ ;
6   |  $P_{\text{left}} \leftarrow \text{Pareto}_{\mathcal{T}}(l)$ ;
7   |  $P_{\text{right}} \leftarrow \text{Pareto}_{\mathcal{T}}(r)$ ;
8   |  $S_u \leftarrow$  the set of all possible partial solutions at  $u$  obtained
   |   by combining solutions from  $P_{\text{left}}$  and  $P_{\text{right}}$  and
   |   possible policies on  $(u, l)$  and  $(u, r)$ ;
9   | return  $\text{Non\_Dominated}(S_u)$ ;
10 end
    
```

---

### 3.1 Applying DP Algorithm to $\text{DCI}_P$

We observe that the main conflict between  $\text{DCI}_P$  and the DP algorithm is that when the DP algorithm is computing the Pareto frontier at a node, it only considers the part of the network above the current node. However, the length of the section connected to the current node can still be expanded. In order to address this problem, we need to keep track of two numbers: (1) the sum of squares of the lengths of all closed sections (sections that are separated from the current node by dams) above the current node, which we will call  $Q$ ; (2) the total length of upstream river segments connected to the current node, which we call  $C$ . (Notice that  $C$  is essentially the numerator of  $\text{DCI}_D$  or the longitudinal connectivity in [26], so when we are computing  $\text{DCI}_P$  using the DP algorithm, we also compute  $\text{DCI}_D$  as a byproduct.)

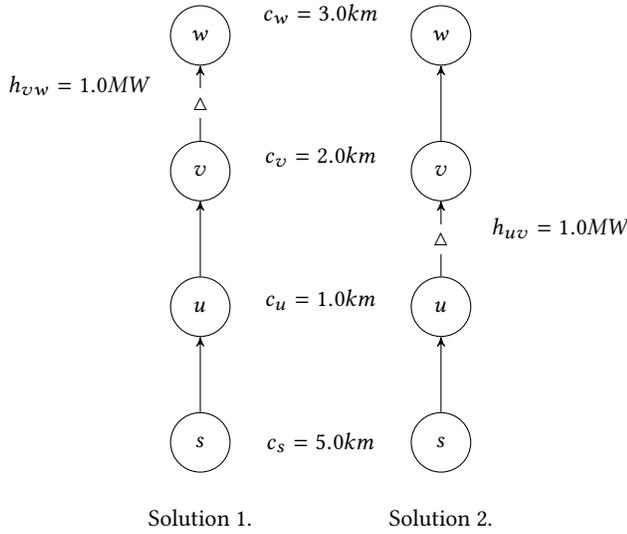
For instance, in Fig. 4, at node  $s$ , we have

$$Q_s = (c_u + c_w)^2 + c_x^2 + c_y^2,$$

and

$$C_s = c_s + c_v + c_z.$$

**Combination Step:** (line 8 in Algorithm 1.) Assume the tree is binary. For a non-leaf node  $u$ , denote its children as  $l$  and  $r$ . Given fixed partial schemes on  $l$  and  $r$ , we now show how to combine these partial schemes at  $u$  for computing  $\text{DCI}_P$ . The main idea is that whenever we decide to build a dam, we add the square of total length of the cut-offed branch to  $Q$ . Specifically, if we decide not to build  $(u, l)$  and  $(u, r)$ , then  $Q_u = Q_l + Q_r$  and  $C_u = C_l + C_r + c_u$ ;



**Figure 5: Counterexample of Theorem 1 when the objectives are DCIP and energy output.  $s$  is the root (river mouth) of the river network. Triangles represent dams we decide to build. Solution 1 provides better DCIP than solution 2. However, when only considering the part of the network above  $u$ , solution 1 is dominated by solution 2.**

if we decide to build  $(u, l)$  but not  $(u, r)$ , then  $Q_u = Q_l + Q_r + C_l^2$  and  $C_u = C_r + c_u$ ; if we decide to build both  $(u, l)$  and  $(u, r)$ , then  $Q_u = Q_r + Q_l + C_r^2 + C_l^2$  and  $C_u = c_u$ .

At the root node  $s$ , we have

$$DCIP = \frac{Q_s + C_s^2}{L^2} \times 100.$$

**Pruning Step:** (line 9 in Algorithm 1.) At the pruning step, we treat  $C$  and  $Q$  as two separate objectives. For instance, assume we are optimizing for Hydropower  $H$  and DCIP. Then, at a node  $u$ , a partial scheme dominates another partial scheme if and only if it has larger  $H$ ,  $C$ , and  $Q$  values. We have the following theorem.

**THEOREM 2.** *Assume  $\pi$  is a solution on the Pareto frontier of the hydropower dam placement problem optimizing for DCIP and other objectives that satisfy Theorem 1, then  $\pi$  will not be eliminated in the pruning steps if we keep track of both  $C$  and  $Q$ .*

**PROOF.** We use a proof by contradiction. We use  $\pi_u$  to represent the part of the solution  $\pi$  on the subtree rooted at  $u$ . Assume for contradiction that, at a node  $u$ ,  $\pi_u$  is dominated by  $\pi'_u$ , then we can modify  $\pi$  by replacing the part  $\pi_u$  with  $\pi'_u$  and get a new solution  $\pi'$ . Note that  $\pi'$  is superior to  $\pi$  in terms of DCIP and all other objective, i.e.  $\pi'$  dominates  $\pi$ . This is a contradiction.  $\square$

The above theorem ensures that, by splitting the DCIP objective into  $C$  and  $Q$ , we can now apply the DP algorithm to successfully obtain the entire Pareto frontier when one of the objective is DCIP.

### 3.2 Rounding Scheme for DCIP

In [26], we also proposed a fully polynomial-time approximation scheme (FPTAS) for approximating the Pareto frontier on tree-structured networks using a rounding technique for the DP algorithm. The main idea is to bound the number of solutions in every partial Pareto frontier by rounding the objective values to multiples of  $K$ , where the value of  $K$  depends on the objective and the current node. Our FPTAS for approximating the Pareto frontier is inspired by the FPTAS proposed in [27–29]. A key difference is the fact that the FPTAS algorithms proposed in [27–29] do not consider the Pareto frontier, but rather a single objective subject to a constraint. We will show that we can also apply this rounding technique to DCIP related problems and obtain a FPTAS.

Notice that  $C$  is essentially the same objective as longitudinal connectivity, so we will apply a similar rounding scheme as longitudinal connectivity. At node  $u$ , assume that the total length of river segments in node  $u$  is  $c_u$ , then the longitudinal connectivity value is rounded to

$$\lfloor \frac{C}{\frac{\epsilon}{4}c_u} \rfloor \cdot \frac{\epsilon}{4}c_u.$$

As shown in [26], this rounding scheme incurs at most  $\frac{\epsilon}{4}C$  error.

When rounding the objective  $Q$ , we notice that the minimum possible value of  $Q$  at the root node  $s$  is

$$Q_{\min} = \sum_{v \in V, v \neq s} c_v^2,$$

which is reasonably large compared to maximum possible value of  $Q$ . Therefore, here we apply a fixed error rounding scheme. Note that the total number of rounding steps is at most  $n$  ( $n$  is the total number of nodes). Thus, we let

$$K = \frac{\epsilon}{2n} Q_{\min}.$$

After each combination step, we round  $Q$  to

$$\lfloor \frac{Q}{K} \rfloor \cdot K.$$

The error incurred by the rounding of  $Q$  is at most  $\frac{\epsilon}{2} Q_{\min}$ .

One complication of the above rounding scheme is that  $Q$  is computed based on  $C$ . During the recursion process, whenever we decide to build a dam, we add the square of the  $C$  value of the cut-off branch to  $Q$ , so the rounding error of  $C$  will be carried over to  $Q$ . Let  $\hat{C}$  represent the rounded value of  $C$ . Let  $\tilde{Q}_s$  denote the value of  $Q$  at the root computed with rounding for  $C$  but no rounding for  $Q$ . Also, let  $\hat{Q}_s$  denote the value of  $Q$  at the root computed with rounding for both  $C$  and  $Q$ . We know that

$$\tilde{Q}_s - \frac{\epsilon}{2} \sum_{v \in V, v \neq s} c_v^2 \leq \hat{Q}_s \leq Q_s.$$

Assume that  $D$  is the set of nodes under which the dams are built. Then, we have

$$Q_s = \sum_{v \in D} C_v^2$$

and

$$\tilde{Q}_s = \sum_{v \in D} \hat{C}_v^2 \geq \sum_{v \in D} [(1-\epsilon/4)C_v]^2 \geq (1-\epsilon/2) \sum_{v \in D} C_v^2 = (1-\epsilon/2)Q_s$$

$$\begin{aligned}
 & \text{maximize: } \text{DCI}_P \\
 & \text{subject to: } \text{DCI}_P = \sum_{u \neq v \in V} \frac{c_u c_v}{L^2} \cdot x_{uv} + \frac{\sum_{v \in V} c_v^2}{L^2} \\
 & H = \sum_{e \in E} \pi_e h_e \\
 & H \geq \hat{H} \\
 & \pi_e \in \{0, 1\}, \quad \forall e \in E \\
 & x_{uv} \in \{0, 1\}, \quad \forall u \neq v \in V \\
 & x_{uv} \leq 1 - \pi_e, \quad \forall e \in D_{uv}, \quad \forall u \neq v \in V \\
 & x_{uv} \geq 1 - \sum_{e \in D_{uv}} \pi_e, \quad \forall u \neq v \in V.
 \end{aligned}$$

**Figure 6: MIP formulation for optimizing  $\text{DCI}_P$  with constraint on  $H$ . We ignore the  $\times 100$  term in  $\text{DCI}_P$ .**

Let  $C_s$  denote the longitudinal connectivity value at the root node  $s$ . Then,

$$\begin{aligned}
 \hat{Q}_s + \hat{C}_s^2 & \geq \tilde{Q}_s - \frac{\epsilon}{2} \sum_{v \in V, v \neq s} c_v^2 + [(1 - \epsilon/4)C_s]^2 \\
 & \geq (1 - \epsilon/2)Q_s - \frac{\epsilon}{2}Q_{\min} + (1 - \epsilon/2)C_s^2 \\
 & \geq (1 - \epsilon)(Q_s + C_s^2)
 \end{aligned}$$

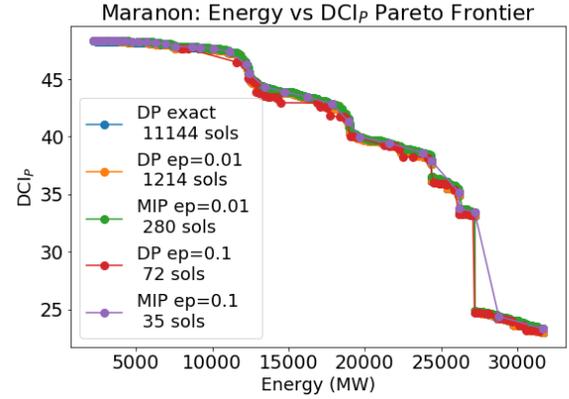
Hence, we may conclude that the  $\text{DCI}_P$  computed by the rounding scheme is within  $\epsilon$  of the actual  $\text{DCI}_P$  value. Using the same proof idea as in [26], we have the following theorem:

**THEOREM 3.** *DP algorithm with rounding is a FPTAS for multi-objective optimization problems on tree structured networks with  $\text{DCI}_P$  as one of the objectives.*

## 4 MIXED INTEGER PROGRAMMING FORMULATION

In this section, we provide a Mixed Integer Programming (MIP) formulation for the problem of optimizing the placement of hydropower dams using  $\text{DCI}_P$  and energy production ( $H$ ) as objectives. Note that we only consider the case of binary passability and assume that a dam will block all fish. We use the same notations as used in [26], specifically:

- $V$ : the set of all nodes.
- $E$ : the set of all edges (dams).
- $x_{uv}$ : the indicator variable of whether fish can move between  $u$  and  $v$ . ( $x_{vv} = 1$  for all  $v \in V$ .)
- $\pi$ : the set of dams we plan to build.
- $\pi_e$ : indicator variable of  $e \in \pi$ .
- $D_{uv}$ : the set of all edges on the unique path between  $u$  and  $v$ .
- $H$ : total energy production of the dams.
- $\hat{H}$ : the bound for energy production. In practice,  $\hat{H}$  will range over  $\{(1 + \epsilon)^i H_{\min} : i \geq 0, (1 + \epsilon)^i H_{\min} \leq H_{\max}\}$ .
- $h_e$ : hydropower output of dam  $e$ .



**Figure 7: The Pareto frontier for energy vs.  $\text{DCI}_P$ . The MIP and DP approximations for  $\epsilon = 0.01$  nearly overlap the exact Pareto curve (blue). The DP approximations for  $\epsilon = 0.01$  (red) lie slightly farther from the exact Pareto-Frontier than the corresponding MIP approximation (purple). The DP approximations has a much better coverage of the exact Pareto frontier in number of solutions. (Note: the number of solutions in the DP results are different from Table 1 because we eliminated the solutions that have better longitudinal connectivity but worse  $\text{DCI}_P$ .)**

By definition of passability, we have

$$x_{uv} = \prod_{e \in D_{uv}} (1 - \pi_e).$$

In order to linearize this constraint, we utilize the property that  $x_{uv}$ s and  $\pi_e$ s are binary variables. When  $\pi_e = 0$  for all  $e \in D_{uv}$ , we know that  $\sum_{e \in D_{uv}} \pi_e = 0$  and the constraint

$$x_{uv} \geq 1 - \sum_{e \in D_{uv}} \pi_e$$

ensures that  $x_{uv} = 1$ . When  $\pi_e = 1$  for at least one  $e \in D_{uv}$ , the set of constraints

$$x_{uv} \leq 1 - \pi_e, \quad \forall e \in D_{uv}$$

ensure that  $x_{uv} = 0$ . The complete MIP formulation of the problem is shown in Fig. 6.

To approximately map out the entire Pareto frontier, we solve the MIP for every  $\hat{H} \in \{(1 + \epsilon)^i H_{\min} : i \geq 0, (1 + \epsilon)^i H_{\min} \leq H_{\max}\}$ . The set of all solutions  $\epsilon$ -approximates the actual Pareto frontier.

## 5 EXPERIMENTAL RESULTS

To evaluate the performance of the new methods at different scales, we used three datasets: the Marañon, Western Amazon, and Amazon basins, with 107, 219, and 467 hypernodes, respectively (corresponding to 128801, 455156 and 4083059 river segments, respectively). We considered three criteria: energy production, longitudinal connectivity (numerator of  $\text{DCI}_D$ ) and localized connectivity ( $\text{DCI}_P$ ). Note that the DP algorithm computes the longitudinal connectivity as a byproduct when computing for  $\text{DCI}_P$ . For better comparison of the performances, we tested it against two versions of

B	$\epsilon$	DP (sec) E, DCI <sub>P</sub> , DCI <sub>D</sub>	MIP (sec) E, DCI <sub>P</sub>	MIP (sec) E, DCI <sub>P</sub> , DCI <sub>D</sub>	DP #Sol E, DCI <sub>P</sub> , DCI <sub>D</sub>	MIP #Sol	MIP #Sol E, DCI <sub>P</sub> , DCI <sub>D</sub>
M	0	1076	–	–	55364	–	–
M	0.001	566	2627	1day+	37998	2055	1day+
M	0.005	202	599	1day+	10377	518	1day+
M	0.01	80	297	23323	4422	280	868
M	0.05	4	61	972	317	63	114
W	0.005	1day+	5052	1day+	1day+	530	1day+
W	0.05	6287	477	14761	1385	61	72
W	0.1	588	275	4285	511	34	41
A	0.025	1day+	5140	1day+	1day+	89	1day+
A	0.1	6616	1510	14742	2547	23	34

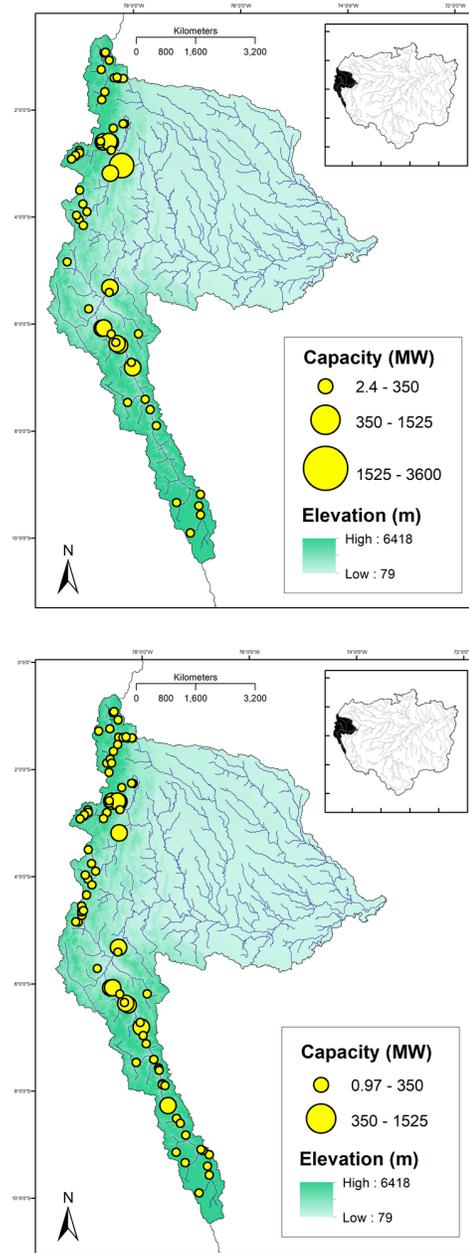
**Table 1: Examples of runtimes and number of solutions for the criteria E (energy), DCI<sub>P</sub> and DCI<sub>D</sub> for Marañon (M; 109 hypernodes), Western Amazon (W; 219 hypernodes) and Amazon (A; 468 hypernodes) showing the trade-offs between the different methods. Note that DP computes simultaneously for energy, DCI<sub>P</sub>, and DCI<sub>D</sub>, so we consider two versions for MIP: one optimizing for energy and DCI<sub>P</sub> and the other one optimizing for energy, DCI<sub>P</sub>, and DCI<sub>D</sub>, so the time comparison is more fair.**

the MIP formulation: The first one is the 2-dimensional MIP approximation scheme that optimizes for energy production and DCI<sub>P</sub>, as shown in the previous section; the second one is a 3-dimensional MIP approximation scheme that optimizes for energy production, longitudinal connectivity (numerator of DCI<sub>D</sub>), and localized connectivity (DCI<sub>P</sub>). (See [26] for the MIP formulation of longitudinal connectivity and the MIP approximation scheme for higher dimensions). The performance of the three approaches is shown in table 1. We observe that when we consider only 2 objectives, energy production and DCI<sub>P</sub>, MIP is generally faster, especially in larger networks. When we also consider longitudinal connectivity, DP performs much better in terms of speed. In both cases, the DP approach provides better coverage of the Pareto frontier. (See also Fig. 7.)

We also compared the results of optimizing for DCI<sub>P</sub> with the results of optimizing for DCI<sub>D</sub> (longitudinal connectivity) under the same energy production constraint. See Fig. 8 for an example. We observe that the DCI<sub>D</sub>-based optimization solutions tend to build dams in series on single rivers. Also, if some dams already exist in the river network, the DCI<sub>D</sub>-based solutions will build all dams upstream of the existing ones. In contrast, DCI<sub>P</sub>-based optimization solutions tend to build fewer but larger dams, sacrificing some lengths of long-distance migration corridors for a less fragmented upstream area.

## 6 CONCLUSION

We provide an exact and approximation algorithm based on dynamic programming and an approximation algorithm based on MIP



**Figure 8: The top and bottom panels depict two Pareto-optimal solutions for the Marañon basin, with similar total energy output (around 14000MW): the top panel shows a solution using localized connectivity (DCI<sub>P</sub>) as the second objective while the bottom panel shows a solution using longitudinal connectivity (DCI<sub>D</sub>) as the second objective. The solution optimized for DCI<sub>D</sub> tries to build all dams upstream of existing dams (84 dams in total), while the solution optimized for DCI<sub>P</sub> replaced many small dams with a larger one (3600MW) downstream (56 dams in total).**

for optimizing for dendritic connectivity on tree-structured networks in a multi-objective setting. Our DP and MIP Pareto frontier algorithms are complementary. The DP approach can compute the exact Pareto frontier, which is not possible with MIP. While the MIP approach can provide faster results, the DP provides a much better coverage of the Pareto frontier in number of solutions. The MIP is also more flexible, allowing the incorporation of additional constraints.

Our work is motivated by an important problem in computational sustainability concerning the evaluation of trade-offs in ecosystem services due to the proliferation of hydropower dams throughout the Amazon basin. In particular, we consider trade-offs between energy production and river connectivity. River fragmentation can dramatically affect fish migrations and other ecosystem services, such as navigation and transportation, and therefore it is important to consider different variants of dendritic connectivity to characterize the movements of different fish species and human populations. While perhaps not obvious from their formulation, the choice of connectivity criterion influences the nature of dam siting solutions when optimizing for connectivity and energy (Figure 8). For instance, in the solution example from the Marañon, dam siting under DCI<sub>D</sub>-based optimization tends to build dams in series on single rivers, 'sacrificing' some rivers to hydropower proliferation while preserving other long-distance migration corridors as unimpeded. In contrast, the DCI<sub>P</sub>-based optimization example solution concentrates connectivity losses to a set of localized 'impact hotspots' spread spatially disparately across the river network. This example solution requires fewer but larger dams, siting them lower down in the basin to take advantage of greater potential energy for hydropower generation.

Ultimately, hydropower proliferation decisions will depend on a greater number of criteria than just connectivity and energy objectives, requiring decision makers to navigate multiple tradeoffs in ecosystem services when deciding the number and location of dam projects. The advances in computational methods such as those provided here will facilitate this multi-objective planning of high complexity for decision makers.

## REFERENCES

- [1] Ronaldo B Barthem, Michael Goulding, Rosseval G Leite, Carlos Cañas, Bruce Forsberg, Eduardo Venticinque, Paulo Petry, Mauro L de B Ribeiro, Junior Chuctaya, and Armando Mercado. 2017. Goliath catfish spawning in the far western Amazon confirmed by the distribution of mature adults, drifting larvae and migrating juveniles. *Scientific reports* 7 (2017), 41784.
- [2] David Cote, Dan G Kehler, Christina Bourne, and Yolanda F Wiersma. 2009. A new measure of longitudinal connectivity for stream networks. *Landscape Ecology* 24, 1 (2009), 101–113.
- [3] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 6, 2 (2002), 182–197.
- [4] Matthias Ehrgott and Xavier Gandibleux. 2000. A survey and annotated bibliography of multiobjective combinatorial optimization. *Or Spectrum* 22, 4 (2000), 425–460.
- [5] Matt Finer and Clinton N Jenkins. 2012. Proliferation of hydroelectric dams in the Andean Amazon and implications for Andes-Amazon connectivity. *Plos one* 7, 4 (2012), e35126.
- [6] Alexander S Flecker, Peter B McIntyre, Jonathan W Moore, Jill T Anderson, Brad W Taylor, and Robert O Hall Jr. 2010. Migratory fishes as material and process subsidies in riverine ecosystems. In *American Fisheries Society Symposium*, Vol. 73. 559–592.
- [7] Mary C Freeman, Catherine M Pringle, EFFIE A Greathouse, and BYRON J Freeman. 2003. Ecosystem-level consequences of migratory faunal depletion caused by dams. In *American Fisheries Society Symposium*, Vol. 35. 255–266.
- [8] J.R. Kasprzyk G. Fu, Z. Kapelan and P.M. Reed. 2013. Optimal design of water distribution systems using many-objective visual analytics. *Journal of Water Resources Planning and Management* 139 (2013), 624–633.
- [9] Carla P Gomes. 2009. Computational sustainability: computational methods for a sustainable environment, economy, and society. *The Bridge* 39, 4 (2009), 5–13.
- [10] Jonathan M. Gomes-Selman, Qinru Shi, Yexiang Xue, Roosevelt Garcia-Villacorta, Alexander Flecker, and Carla P. Gomes. 2018. Boosting efficiency for computing the Pareto frontier on tree structured networks. *to appear in CPAIOR 2018* (2018).
- [11] Peter M Kareiva. 2012. Dam choices: analyses for multiple needs. *Proceedings of the National Academy of Sciences* 109, 15 (2012), 5553–5554.
- [12] Steven King and Jesse R O'Hanley. 2016. Optimal fish passage barrier removal-revisited. *River Research and Applications* 32, 3 (2016), 418–428.
- [13] Steven King, Jesse R O'Hanley, Lynda R Newbold, Paul S Kemp, and Matthew W Diebel. 2017. A toolkit for optimizing fish passage barrier mitigation actions. *Journal of Applied Ecology* 54, 2 (2017), 599–611.
- [14] MC Lucas and E Baras. 2001. Migration of freshwater fishes. (2001).
- [15] E.S. Matrosov, I. Huskova, J.R. Kasprzyk, J.J. Harou, C. Lambert, and P.M. Reed. 2015. Many-objective optimization and visual analytics reveal key trade-offs for London's water supply. *Journal of Hydrology* 531 (2015), 1040 – 1053. <https://doi.org/10.1016/j.jhydrol.2015.11.003>
- [16] Frank Neumann. 2007. Expected runtimes of a simple evolutionary algorithm for the multi-objective minimum spanning tree problem. *European Journal of Operational Research* 181, 3 (2007), 1620 – 1629.
- [17] C. H. Papadimitriou and M. Yannakakis. 2000. On the approximability of trade-offs and optimal access of web sources. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science (FOCS '00)*.
- [18] Chao Qian, Ke Tang, and Zhi-Hua Zhou. 2016. Selection hyper-heuristics can provably be helpful in evolutionary multi-objective optimization. In *International Conference on Parallel Problem Solving from Nature*. Springer, 835–846.
- [19] Chao Qian, Yang Yu, and Zhi-Hua Zhou. 2015. Pareto ensemble pruning. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15)*. 2935–2941.
- [20] P.M. Reed, D. Hadka, J.D. Herman, J.R. Kasprzyk, and J.B. Kollat. 2013. Evolutionary multiobjective optimization in water resources: The past, present, and future. *Advances in Water Resources* 51 (2013), 438 – 456. <https://doi.org/10.1016/j.advwatres.2012.01.005> 35th Year Anniversary Issue.
- [21] P.M. Reed and J.B. Kollat. 2013. Visual analytics clarify the scalability and effectiveness of massively parallel many-objective optimization: A groundwater monitoring design example. *Advances in Water Resources* 56 (2013), 1 – 13. <https://doi.org/10.1016/j.advwatres.2013.01.011>
- [22] Suresh Andrew Sethi, Jesse R O'Hanley, Jonathan Gerken, Joshua Ashline, and Catherine Bradley. 2017. High value of ecological information for river connectivity restoration. *Landscape Ecology* 32, 12 (2017), 2327–2336.
- [23] Wanxing Sheng, Yongmei Liu, Xiaoli Meng, and Tianshu Zhang. 2012. An improved strength Pareto evolutionary algorithm 2 with application to the optimization of distributed generations. *Computers and Mathematics with Applications* 64, 5 (2012), 944 – 955.
- [24] Miguel Terra-Neves, Inês Lynce, and Vasco Manquinho. 2017. Introducing Pareto minimal correction subsets. In *International Conference on Theory and Applications of Satisfiability Testing*. Springer, 195–211.
- [25] Margaret M Wiecek, Matthias Ehrgott, Georges Fadel, and José Rui Figueira. 2008. Multiple criteria decision making for engineering.
- [26] Xiaojuan Wu, Jonathan M. Gomes-Selman, Qinru Shi, Yexiang Xue, Roosevelt Garcia-Villacorta, Suresh Sethi, Scott Steinschneider, Alexander Flecker, and Carla P. Gomes. 2018. Efficiently approximating the Pareto Frontier: hydropower dam placement in the Amazon basin. In *AAAI*.
- [27] Xiaojuan Wu, Daniel Sheldon, and Shlomo Zilberstein. 2014. Rounded dynamic programming for tree-structured stochastic network design. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*. 479–485.
- [28] Xiaojuan Wu, Daniel Sheldon, and Shlomo Zilberstein. 2014. Stochastic network design in bidirected trees. In *Advances in Neural Information Processing Systems*. 882–890.
- [29] Xiaojuan Wu, Daniel Sheldon, and Shlomo Zilberstein. 2016. Optimizing resilience in large scale networks. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*.
- [30] Michael Yukish and Timothy W Simpson. 2004. Analysis of an algorithm for identifying Pareto points in multi-dimensional data sets. In *10th ALAA/ISSMO Multidisciplinary Analysis and Optimization Conference*. 4324.