MARITIME TRANSPORTATION RESEARCH AND EDUCATION CENTER
TIER 1 UNIVERSITY TRANSPORTATION CENTER
U.S. DEPARTMENT OF TRANSPORTATION

Efficient Dredging Strategies for Improving Transportation Infrastructure Resilience

August 1, 2014 – December 31, 2016

Kelly Sullivan, Ph.D. (PI, University of Arkansas, ksulliv@uark.edu)
Khatereh Ahadi (Ph.D. Candidate)

February 1, 2017

FINAL RESEARCH REPORT
Prepared for:
Maritime Transportation Research and Education Center

University of Arkansas
4190 Bell Engineering Center
Fayetteville, AR 72701
479-575-6021

ACKNOWLEDGEMENT

This material is based upon work supported by the U.S. Department of Transportation under Grant Award Number DTRT13-G-UTC50. The work was conducted through the Maritime Transportation Research and Education Center at the University of Arkansas.

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Abstract

The inland navigation system is highly dependent on uncertain natural factors such as shoaling that can render waterways unnavigable. In order to ensure waterway navigability, maintenance dredging must be completed. We consider the problem of selecting a budget-limited subset of maintenance dredging projects to maximize the expected commodity tonnage that can be transported through the inland waterway system. Our model incorporates uncertainty due to unpredictable amount of budget required for emergency dredging. This problem is modeled as a two-stage stochastic program and a genetic algorithm is developed as a solution approach. The model and heuristic is implemented using data obtained for the U.S. inland waterway network.

Contents

1 Project Description

The inland navigation system is an important part of the maritime transportation system. Transporting products by water requires less fuel and is significantly cheaper per ton-mile than land-or air-based transportation modes. Inland waterways are responsible for transporting nearly 600M tons of freight each year [15], resulting a system-wide net savings of over $8B each year over the transportation costs that would result from using the next-cheapest alternative. Sustained operation of this infrastructure depends on its maintenance.

This research investigates allocation of limited budget resources to inland dredging projects to optimize system-level performance of the navigation system. Dredging, or excavating, is a costly maintenance operation that is necessary to ensure waterways maintain a navigable nine-foot depth. This depth is difficult to achieve reliably because water levels change over time; that is, a period of dry weather or localized accumulation of sediment (known as *shoaling*) could cause this depth to fall below nine feet. As a result, insufficient or untimely dredging in a particular channel can result in water levels below the required nine-foot depth. This may cause immediate damages and/or delays (e.g., if a ship runs aground), and it can also result in more significant system-wide costs (see, e.g., [10]) due to draft restrictions and/or closure of portions of the inland waterway transportation system.

In the United States, dredging operations (and the planning of these operations) fall within the purview of the U.S. Army Corps of Engineers (USACE). We estimate[1] that USACE spends on the order of $100M annually on inland dredging. Motivation for our research stems from the opportunity for reducing the risk associated negative effects described in the previous paragraph while operating under a limited budget. **Optimization enables the identification of strategies for allocating inland dredging resources that are efficient with respect to both cost and risk.**

As discussed in the previous paragraph, inland dredging is a relatively small percent of the overall USACE O&M budget. The USACE also oversees lock maintenance and coastal dredging, both of which are also vital to sustained transportation. We do not explicitly consider coastal dredging in this report, but it is our opinion that the models in this paper could be adapted for coastal dredging, too. Unfortunately, the problems dynamics (e.g., shoaling and water levels) differ drastically between the coastal and inland settings. Because the data requirements are somewhat extensive, we opted to focus initially on the inland side.

[1]Based on the 2015 dataset at [17], we estimate that around 3.5% of dredging expenditures come from inland-specific USACE districts, which corresponds to roughly $53M of the annual $1.5B dredging budget [16]; however, this excludes some of the districts (e.g., New Orleans, Galveston, and Detroit) that have both inland and coastal dredging projects.

We also excluded lock O&M activities, which are important for ensuring inland waterway navigability, from our initial research focus for reasons we now discuss. Preventive maintenance is routinely performed on most locks, which results in *scheduled* downtime. The objective of performing preventive maintenance is to reduce unscheduled downtime (causing slowed barge traffic and/or increased congestion) that results when lock outages occur. Preventive maintenance is already pervasive, and furthermore, this maintenance can usually be scheduled in a way that does not have a negative effect on transportation (see, e.g., [14]). Whereas scheduled downtime seems to have only minor effects on transportation, unscheduled downtime could pose a much more serious threat. Unscheduled lock downtime rates are currently around 2% [20], which will likely increase over time (due to aging locks) if not addressed. Major upgrade and/or rehabilitation projects can help—and will be necessary over time to keep the system functional—but each project of this type may cost tens to hundreds of millions of dollars, which is comparable to the national inland dredging budget. Furthermore, whether or not historical downtime has had a significant negative economic effect remains the subject of ongoing research (see, e.g., [1]).

To summarize the above discussion, selecting locks for rehabilitation is a strategic problem that involves an extended planning horizon (e.g., locks may go 30 years between rehabilitation projects) and significant investments. By contrast, allocating resources for inland dredging involves a shorter time scale (e.g., sites may be dredged once every three years) and smaller investments. We initially focus on the inland dredging problem because it comprises a relatively small portion of the budget yet has the potential for significant impacts (i.e., there is a potential for significant risk reduction without substantial additional expenditures).

Our interactions with professionals at USACE revealed that the inland dredging problem is, in itself, quite complex. Some distinguishing features of this problem are outlined below:

1. Budgets are allocated to USACE districts for inland dredging *months in advance* of the yearly dredging season, before all of that year's dredging requirements are known.

2. Districts do not necessarily coordinate with respect to selecting and scheduling dredging projects. That is, each district's dredging projects are selected and scheduled by a dredge manager at the district level using assets (i.e., dredge vessels) that are contracted specifically to that district.

3. Some locations' dredging requirements are predictable (e.g., dredging once every three years is always sufficient to ensure full waterway navigability). Henceforth, we will refer to this type of dredging as *routine* dredging. Other locations—particularly stretches of open river, including the lower Ohio and Mississippi—have needs for dredging that

emerge dynamically due to unpredictable shoaling, and can cause disruptions to navigation if not remedied immediately. We will refer to this type of dredging as *reactive* dredging.

4. The underlying shoaling process is not well understood, nor is there a reliable source of data from which to draw inferences. (This is however, the subject of ongoing research at USACE, and so there is some hope that this data can be used to a greater extent to support decision models of of the variety examined herein.) It is known, however, that shoaling patterns (i) differ by location and (ii) depend on if/how nearby waterways are dredged. For instance, material dredged from one location may actually cause shoals downstream as sediment deposited into the river re-settles.

Given the items discussed above, we have opted to focus our model on the complexities arising from Items 1–3 (i.e., those issues for which we have a reasonable source of data), and we have excluded Item 4. That is, we do not attempt to include a "forecasted shoaling pattern" into our model for each waterway segment; rather, in view of Item 3, our model incorporates information (in the form of a probability distribution) about the requirements for reactive dredging in each district—this can be estimated more easily using available data. Our preliminary investigation revealed that yearly reactive dredging requirements are highly uncertain, comprising anywhere between 40 and 80 percent of inland maintenance dredging expenditures. The uncertainty associated with reactive dredging forces dredge managers to trade off immediate versus long term benefit. For instance, dredge managers may have to postpone a routine dredge job in order to ensure availability of resources to complete a reactive job that has more immediate consequences. Postponing a routine job, however, may not cause significant impacts unless it happens several years in a row.

Given the motivation expressed above, the primary contribution of this project is a new optimization model (and associated solution approach) that (i) allocates inland dredging budgets to USACE districts and (ii) selects projects for routine dredging—based upon uncertain requirements for reactive dredging. In selecting dredging projects, we seek to maximize network-wide throughput of a set of origin-destination commodity flows. Thus, by considering the impact of dredging projects on network-wide flows, our model incorporates *dependence* among projects. We now relate our work to the literature.

Within recent years, several researchers have considered strategic allocation of resources for maintenance or improvement of transportation infrastructure. We focus on those that are most relevant to the proposed work, i.e., mathematical modeling approaches that consider either uncertainty or dependence among projects (such as that described in the preceding paragraph).

Resource allocation models that select transportation improvement projects under uncertainty are somewhat scarce in the literature. A preliminary literature review identified a few examples of existing work in this area (see, e.g., [12, 13]) and one paper [6] that was directly relevant to the dredging resource allocation problem considered herein. We distinguish our work from this research in the coming paragraphs.

Project dependence has been incorporated more frequently into resource allocation models for selecting transportation improvement projects. The problems of allocating budget resources over time to (i) expand a series of locks [5] and (ii) rehabilitate a tree-structured network of locks [21] have been modeled under the consideration of interconnected queueing effects. Waterway project dependence has also been considered within the context of complex logical relationships between projects (e.g., precedence relationships and mutual exclusivity) [22], concepts that have received much attention within an abstract setting [4, 7]. Tao and Schonfeld [12, 11] model dependence in roadway network improvement projects via considering traffic flow equilibrium.

Dependent project selection models have also been applied to the specific application of dredging. Mitchell et al. [8] utilize a mixed integer linear program and heuristic methods to select dredge projects to enable maximum flow of a set of origin-destination (o-d) commodities through the waterway network. In [8], o-d flow for a particular commodity may be restricted if insufficient dredging is completed on the route from origin to destination. Khodakarami et al. [6] extend [8] to consider the effect of shoaling after dredging projects and provide two heuristics were proposed to find approximate solutions for this problem.

The work of Khodakarami et al. [6] is closest to our research. Like [6], we select budget-limited maintenance projects for an o-d commodity flow network where uncertainty arises due to unpredictable natural/hydrologic conditions in the waterway segments. By our understanding, however, the model in [6] is a deterministic program that condenses, for the purposes of optimization, each location's shoaling probability distribution into an expected value. Our model, which incorporates uncertainty in how much budget is required for reactive maintenance in a district, is a true two-stage stochastic program that enables weighing the risk associated with planning dredging operations. Furthermore, our research enhances previous models—in view of the challenge described in Item 2 of the above list— by examining the allocation of budget into districts.

The remainder of this paper is organized as follows: In the following section a scenario-based stochastic program is proposed, and a Genetic Algorithm (GA) is provided to solve the problem. Section 3 presents our dataset and summarizes preliminary computational results used to tune algorithm performance. Based on these results, we apply the GA to a realistic, large-scale instance in Section 4, and we discuss benefits obtained via applying our modeling and solution approach. Section 5 concludes.

2 Methodological Approach

In this section, we model the effect of inland dredging on transportation of different commodities through waterway network. Our model incorporates uncertainty due to unpredictable reactive maintenance that occurs over the planning horizon. A scenario-based stochastic program is developed to select maintenance projects and maximize the total value of flow that is not disrupted. We begin developing the model by defining notation in the following paragraph.

We represent the system as a network which river segments (defined by the set N) are the nodes and each node has a specific capacity to transport different commodities (let K denote the set of commodities). Commodity $k \in K$ is defined by an origin segment $o^k \in N$ and a destination segment $d^k \in N$. We represent the network's topology by defining $P(k) \subseteq N$ as the set of segments on the origin-destination path of commodity k (Note: Origin-destination flow always moves along a unique path in the waterway network even though the network is not tree). Let Z denote the set of districts that includes several waterway segments.

The overall structure of our model is a two-stage stochastic program in which the first-stage allocates maintenance funds to districts in preparation for the upcoming dredging season (typically beginning in early spring). Because funding is typically allocated to districts several months before the dredging season, certain parameters (namely, how much reactive dredging will be required in each district) are assumed to be unknown at the time funds are allocated to districts. Uncertain reactive dredging requirements are modeled via a discrete set of scenarios, and the probability of each scenario is assumed to be known. After funds are allocated to districts, the reactive dredging scenario is revealed; then, second-stage decisions variables are incorporated to select an optimal subset of routine dredging projects given the budget allocation and observed reactive dredging requirements.

Towards developing a mathematical model for this problem, we first define some terminology. We refer to the *flow* of a commodity k as the amount of commodity transported via waterways from o^k to d^k. This flow is moved, generally speaking, over thousands of barge trips in which several different commodities may be combined in one shipment. As a result, each barge exhibits its own set of characteristics. Most importantly, depending on the cargo loaded onto a barge, the barge has a characteristic *draft* (vertical distance from the bottom of the barge to the water surface) that determines a minimum channel depth necessary for the barge to be able to pass. There are various factors that change the required draft for barges to transport the cargo such as shoaling along the channels. If the required draft is unavailable (due to shoaling and/or insufficient dredging), flow will be *disrupted* and there will be unsatisfied demand in the waterway network. To decrease the disrupted flow, maintenance projects (dredging the channels) should be selected among all possible projects which seek to maximize the total *value* of the flow, i.e., over of all commodities successfully transported from origin to destination. In the traditional network flow sense, channel depth (determined by dredging and the amount of shoaling) determines for each commodity an effective capacity—the maximum amount of a commodity (in tons) that can be transported using barges with a draft of no more than the channel depth—for undisrupted flow; however, we hereafter refer to this quantity as *availability* due to the fact that *capacity* has alternative, conflicting meanings in regards to the flow of water. The following sections develop mathematical models for the maintenance project selection.

2.1 Optimization Model

We now develop our mathematical model using the following notation:
Sets and Indices:

- K: Set of commodities (k represents a commodity)
- $P(k)$: Set of all segments in the origin-destination path of commodity k
- Ω: Set of scenarios (ω represents a scenario)
- Z: Set of all districts (z represents a district)
- N: Set of all segments (i represents a segment)
- N_z: Set of all segments in district z

Parameters:

- c_z^ω: Cost of reactive maintenance at district z in scenario ω
- v^k: Value per unit flow of commodity k
- D^k: Amount of commodity k to be transported from o^k to d^k
- B: Total budget available for dredging (routine and reactive)
- p^ω: Probability of scenario ω

- σ_z: Penalty for lack of budget to do reactive maintenance at district z
- l_i: The maximum draft (ft) of vessels that can pass through segment i before routine dredging

Decision Variables:

- B_z: The required budget for routine and reactive maintenance at district z
- $Y_0^{k\omega}$: Flow of commodity k (tons) in scenario ω passing through all of the segments on the o-d path
- X_i^{ω}: Feet of routine dredging at segment $i \in N$ in scenario ω
- P_z^{ω}: The lack of budget for reactive maintenance at district z in scenario ω

In accordance with the above, let $B_z,\ z \in Z$, denote the budgeted dollars for dredging in district z. Uncertain requirements for reactive dredging are incorporated via a scenario-based approach: Let Ω denote the set of scenarios (which may be sampled or constructed based on dredging records data), and let c_z^{ω} denote the dollars required for reactive dredging in district $z \in Z$ in scenario $\omega \in \Omega$. Given a realization of uncertainty (i.e., a value of ω), we then select routine dredging jobs for completion: Let X_i^{ω} denote the depth (in feet) of routine dredging at segment $i \in N$ in scenario $\omega \in \Omega$. For each district $z \in Z$ and scenario $\omega \in \Omega$, the routine dredging decisions must be feasible within the allocated budget B_z; else, the objective is penalized σ_z units per dollar of budget shortage in each district $z \in Z$.

Variable X_i^{ω} plays a role in determining the availability of segment i after routine maintenance. (For instance, not completing adequate dredging could result in barges running aground if shoaling causes river depth to fall below the required draft for navigation.) Let $f^k(X_i^{\omega})$ denote the availability (in tons) of segment $i \in N$ after dredging (X_i^{ω}) unit in scenario ω for commodity k. If D^k is the amount of commodity k we wish to transport from o^k to d^k, the actual amount of commodity (tons) transported in scenario ω is given by $\min\{\min\{f^k(X_i^{\omega}) : i \in P(k)\}, D^k\}$ (i.e., the flow of commodity k is limited by the demand of commodity k as well as the effective availability of each segment on the path from o^k to d^k). We model this via introducing a variable $Y_0^{k\omega}$ to represent the flow of commodity k from o^k to d^k after routine dredging, along with constraints

$$Y_0^{k\omega} \leq D^k,\ \forall k \in K, \forall \omega \in \Omega, \tag{1}$$

$$Y_0^{k\omega} \leq f^k(X_i^{\omega}),\ \forall i \in P(k), \forall k \in K, \forall \omega \in \Omega. \tag{2}$$

Any objective rewarding flow of commodity k, will cause one of Constraints (1)–(2) to be binding, thereby enforcing the desired relationship on $Y_0^{k\omega}$.

The stochastic program for our problem is given as

$$\max \sum_{\omega\in\Omega}\sum_{k\in K} p^{\omega} v^{k} Y_0^{k\omega} - \sum_{\omega\in\Omega}\sum_{z\in Z} p^{\omega}\sigma_z P_z^{\omega}, \tag{3}$$

$$\text{s.t. } c_z^{\omega} + \sum_{i\in N_z} \phi(X_i^{\omega}) \le B_z + P_z^{\omega},\ \forall z \in Z, \forall \omega \in \Omega, \tag{4}$$

$$\sum_{z\in Z} B_z \le B, \tag{5}$$

$$l_i + X_i^{\omega} = F_i^{\omega},\ \forall i \in N, \forall \omega \in \Omega, \tag{6}$$

$$Y_0^{k\omega} \le D^k,\ \forall k \in K,\ \forall \omega \in \Omega, \tag{7}$$

$$Y_0^{k\omega} \le f^k(F_i^{\omega}),\ \forall i \in P(k)\cap N, \forall z \in Z,\ \forall k \in K, \forall \omega \in \Omega, \tag{8}$$

$$X_i^{\omega} \ge 0 \text{ and integer},\ \forall i \in N,\ \forall \omega \in \Omega, \tag{9}$$

$$B_z \ge 0, \forall z \in Z, \tag{10}$$

$$Y_0^{k\omega} \ge 0,\ \forall \omega \in \Omega,\ \forall k \in K. \tag{11}$$

The objective function (3) maximizes the expected value of increased flow after dredging from origin to destination of all of the commodities. Constraints (4) require the cost of reactive and routine dredging to be no more than the assigned budget to each district. If the assigned budget is not enough for reactive maintenance at a district, a penalty is added to the objective function. Constraints (5) allocate the dredging budget to each district based on the total available budget. Flow of commodity k is limited by its demand (Constraint (7)) and by the availability of segments after maintenance on the path from o^k to d^k (Constraint (8)). Constraints (9)–(11) enforce the non-negativity and integrality.

Let l_i denote the maximum draft (in feet) of vessels that can pass through segment i if no dredging is performed. (Note: In practice, this value depends on shoaling conditions and water levels during the upcoming dredging season that are not known with certainty. Our model could be extended, via changing l_i to l_i^{ω}, to incorporate uncertainty in this value; however, our experience suggests that data to support generating a probability distribution on l_i is not currently available through the USACE, and we have therefore opted to utilize the deterministic estimates l_i, which could be generated based on industry expertise.) Let F_i^{ω} denote the maximum level of vessel draft (in feet) after routine maintenance that can pass through segment i in scenario ω. If the minimum and maximum possible drafts at segment i are defined as u_i and m_i respectively, $f^k(j)$, $\forall j \in \{u_i, u_i+1, \ldots, m_i\}$, is defined as the tonnage for commodity k at a draft of j feet or less. The demand, D^k, of each commodity

k is obtained by summing its o-d flow tonnages across all draft levels. Define $b_j^k = f^k(j)$, $\forall j \in \{u_i, u_i+1, \ldots, m_i\}$ as the average tonnage of commodity k when the draft is less than or equal to j. Note that the f^k-functions are nonlinear in general. We therefore introduce variables

$$I_{ij}^{\omega} = \begin{cases} 1 & \text{if the maximum draft of segment } i \text{ is } j \text{ in scenario } \omega, \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

to linearize these functions, yielding a mixed integer linear program. Using these variables, Model (3)–(11) is linearized by adding constraints

$$l_i + X_i^{\omega} = F_i^{\omega}, \ \forall i \in N, \ \forall \omega \in \Omega, \tag{13}$$

$$F_i^{\omega} \geq \sum_{j=u_i}^{m_i} j I_{ij}^{\omega}, \quad \forall i \in N, \ \forall \omega \in \Omega, \tag{14}$$

$$\sum_{j=u_i}^{m_i} I_{ij}^{\omega} = 1, \quad \forall i \in N, \ \forall \omega \in \Omega, \tag{15}$$

$$Y_0^{k\omega} \leq \sum_{j=u_i}^{m_i} b_j^k I_{ij}^{\omega}, \ \forall k \in K, \ \forall i \in P(k) \cap N, \ \forall \omega \in \Omega, \tag{16}$$

in place of Constraint (8).

To obtain the maintenance dredging cost data, we stipulated a dredging cost model that put the "full dredging" amount at each location equal to the historical total of all requested dredging funds at each location (from dredging record in Navigation Data Center [18]), and we assumed that full dredging would add three feet draft to each segment at the beginning of the dredging year. We assume, as in [6], the cost of dredging of X_i^{ω} feet is specified by the function

$$\phi_i(X_i^{\omega}) = \begin{cases} 0 & \text{if } X_i^{\omega} = 0, \\ 0.25 m_i + 0.75 m_i (\frac{X_i^{\omega}}{L})^n & \text{otherwise,} \end{cases} \tag{17}$$

where m_i is the largest historical cost of dredging at segment i that is assumed to restore $L = 3$ feet of depth. The first term in the $X_i^{\omega} > 0$ case ($0.25 m_i$) is the fixed mobilization cost of dredging equipment and the second term represents cost of dredging X_i^{ω} feet at segment i and n is the degree of nonlinearity. We assume $n = 2$ and define binary variables

$$\lambda_{ij}^{\omega} = \begin{cases} 1 & \text{if segment } i \text{ is dredged for } j \text{ units at scenario } \omega, \\ 0 & \text{otherwise,} \end{cases} \tag{18}$$

in order to linearize ϕ_i. Letting $h_{ij} \equiv \phi_i(j)$ denote the cost of dredging segment i for j units, Model (3)–(11) is linearized upon replacing Constraint (4) with the constraint set

$$c_z^\omega + \sum_{i \in N(z)} \sum_{j=0}^{3} h_{ij} \lambda_{ij}^\omega \leq B_z + P_z, \ \forall z \in Z, \forall \omega \in \Omega, \tag{19}$$

$$\sum_{j=0}^{3} \lambda_{ij}^\omega = 1, \ \forall i \in N, \ \forall \omega \in \Omega, \tag{20}$$

$$X_i^\omega \leq \sum_{j=0}^{3} j \lambda_{ij}^\omega, \ \forall i \in N, \ \forall \omega \in \Omega. \tag{21}$$

The resulting linear model is given as

$$\max \ \sum_{\omega \in \Omega} \sum_{k \in K} p^\omega v^k Y_0^{k\omega} - \sum_{\omega \in \Omega} \sum_{z \in Z} p^\omega \sigma_z P_z^\omega, \tag{22}$$

$$\text{s.t.} \ \sum_{z \in Z} B_z \leq B, \tag{23}$$

$$Y_0^{k\omega} \leq D^k, \ \forall k \in K, \ \forall \omega \in \Omega, \tag{24}$$

Constraints (13)–(16) and (19)–(21),

$$X_i^\omega \text{and } F_i^\omega \geq 0 \text{ and integer}, \ \forall i \in N, \ \forall \omega \in \Omega, \tag{25}$$

$$B_z \geq 0, \forall z \in Z, \tag{26}$$

$$\lambda_{ij}^\omega = 0 \text{ or } 1, \ \forall i \in N, \ \forall \omega \in \Omega, \ \forall j \in \{0, 1, 2, 3\}, \tag{27}$$

$$I_{ij}^\omega = 0 \text{ or } 1, \ \forall i \in N, \ \forall j \in \{u_i, u_i + 1, ..., m_i\}, \ \forall \omega \in \Omega, \tag{28}$$

$$Y_0^{k\omega} \geq 0, \ \forall \omega \in \Omega, \ \forall k \in K. \tag{29}$$

The following section details our solution approach for this model.

2.2 Solution Methodology

After some initial experimentation with our model on a realistic data set, it became clear that a heuristic solution approach would be necessary. We settled on a genetic algorithm (GA), which we describe in the following subsections. For a more detailed introduction on GA, the reader is referred to [2, 9].

2.2.1 Solution Encoding and Initial Solution

The chromosomes are defined by a string of weights W_z, $z \in Z$, assigned to each district in which budget for each district can be computed according to its weight. Binary digits (i.e., 0 or 1) are used to encode weights of all districts and weights are uniformly distributed on the set $\{1, ..., 2^g\}$, where g (an input parameter) is a positive integer that represents the number of bits needed to encode each weight. (Note: Larger values of g correspond to a more resolute budget encoding that is more nearly continuous whereas smaller values of g lead to budget encodings that are characteristically discrete. We utilize $g = 20$ in our experiments.) Based on the weights $W_z, z \in Z$, budgets are assigned to districts according to

$$B_z = \left(\frac{W_z}{\sum_{z \in Z} W_z} \right) B. \tag{30}$$

Since the problem is a two-stage stochastic program, we could have opted to encode the variables of both stages; however, doing so is computationally expensive because the number of variables grows in $|N|$ and $|\Omega|$. Furthermore, preliminary experiments suggested that it was preferable to encode only the $|Z|$ first stage variables, utilizing a subroutine to evaluate (or estimate) the fitness of each $|Z|$-length chromosome by exploring second-stage solutions.

A set of random solutions is generated as the initial population. For each chromosome, we assigned each bit equal to one (independently) with probability 0.5 and equal to zero otherwise.

2.2.2 Fitness Function and Selection Procedure

For a given set of B_z-values, the fitness function is defined as the optimal objective value of Model (19)–(29) that results when these B_z-values are fixed. Thus, our solution approach decouples decisions by stage in the stochastic program: The GA handles the first-stage (budget allocation) decisions, and a separate optimization subroutine evaluates the fitness of the first-stage decisions by solving for the second-stage variable values. Unfortunately, the second-stage problem remains a mixed integer program and, while it solves much faster than the full Model (19)–(29) (without fixing B_z-values), it requires too much time to incorporate this into the GA in each iteration. As a result, a greedy heuristic method is developed to find a solution for X_i^ω ($\forall i \in N, \omega \in \Omega$), that approximates the fitness for each chromosome. The second-stage heuristic ranks the river segments based on an estimated benefit-to-cost ratio that is scenario independent (thus yielding savings in computational effort). In the estimated benefit-to-cost ratio, we calculate the benefit associated with dredging $x \in \{1, 2, 3\}$ feet at segment i as the potential increase, summed across all commodities, in the availability of segment i due to dredging x feet. This benefit is calculated as $\sum_{k:i \in P(k)} (b^k_{l_i+x} - b^k_{l_i})$. The cost associated with dredging x feet is given by $h_{i,x}$, and the estimated benefit-to-cost ratio is therefore given by

$$r(i,x) = \frac{\sum_{k:i \in P(k)} (b^k_{l_i+x} - b^k_{l_i})}{h_{i,x}}, \ i \in N, \ x \in \{1,2,3\}. \tag{31}$$

Note that $h_{i,0}$ need not be subtracted from the denominator because $\phi_i(0) = 0$. The estimated benefit-to-cost ratio $r(i,x)$ is calculated for each segment $i \in N$ and dredging level $x \in \{1,2,3\}$ and sorted in decreasing order. Then, for each scenario $\omega \in \Omega$, pairs (i,x) are selected (and the corresponding X^ω_i-value set equal to x) greedily (i.e., largest $r(i,x)$ value first) until the budget for scenario ω (i.e., Constraint (19)) has been exhausted. If there is a remaining budget at the end of the first round of selection, we start the second round (using the same sorted ordering of (i,x)-pairs but excluding any pairs (i,x') corresponding to a segment i that has already been selected for dredging at x' feet or more) to see if there is an opportunity to increase the level of dredging at some segments.

The selection criteria of chromosomes to produce offspring is based on a roulette-wheel procedure in which the probability of selecting a particular chromosome is computed by its fitness value divided by the total fitness value summed over all chromosomes in the population.

2.2.3 Genetic Operators

The operators in genetic algorithm influence the performance of the heuristic and include three simple types: selection, crossover and mutation. The selection operator selects chromosomes for reproduction. The probability of selecting a chromosome depends on the chromosome's fitness value. We apply two crossover subroutines, which take as input two chromosomes selected for reproduction: One-point crossover randomly chooses one position $z \in \{1, \ldots, Z\}$ and generates two new offspring by interchanging the substring of each parent chromosome that appears after position z; Two-point crossover randomly chooses two positions $z, z' \in \{1, \ldots, Z\}, \ z < z'$, and generates two new offspring by interchanging the substrings of each parent chromosome that appear between positions z and z'. The crossover rate λ ($0 \leq \lambda \leq 1$) determines the proportion of each generation that is created by crossover. If the rate is one, all of the population in the new generation is produced by crossover, but if it is zero all new children are directly a copy of previous population. The rates λ_1 and λ_2 specify the proportion of each generation generated by one-point and two-point crossover, respectively. The summation of these rates should be one.

After crossover has been completed and (if $\lambda < 1$) chromosomes copied from the previous generation to ensure the number of chromosomes in each generation remains constant,

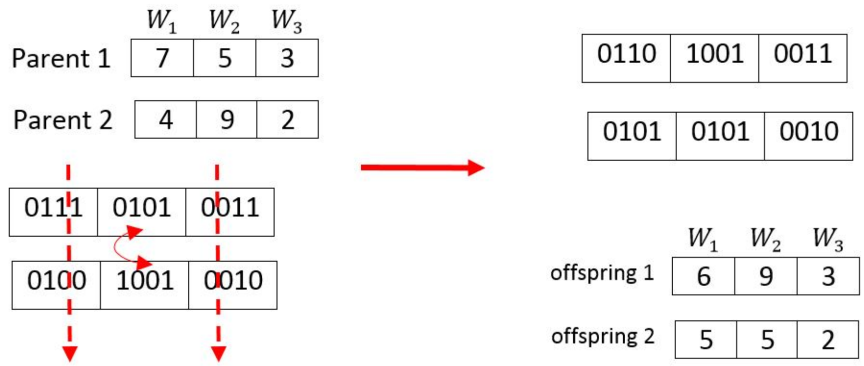

Figure 1: Example of two-point crossover

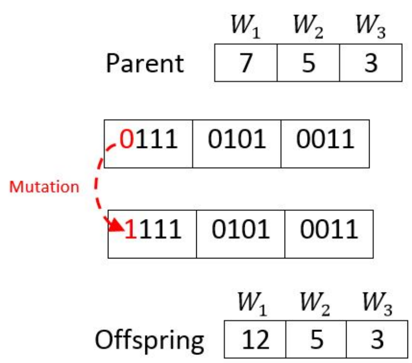

Figure 2: Example of mutation

mutation is applied to each chromosome. Mutation selects one or more genes randomly and swaps their values (i.e., from 0 to 1 or from 1 to 0). We restrict the mutation operator to apply only to genes that change the weight and hence the budget assignment drastically. Specifically, given the binary encoding of weights W_z, we restrict single-gene mutation to randomly choose a single district $z \in Z$ and randomly alter only one of the first six bits of its W_z-value. In multi-gene mutation, we repeat this process four times.

The mutation probability μ $(0 \leq \mu \leq 1)$ is the proportion of chromosomes selecting for mutation in the new generation. If mutation is applied to a chromosome, single-gene mutation is used with probability μ_1 $(0 \leq \mu_1 \leq 1)$; otherwise, multi-gene mutation is used. After performing crossover and mutation, the fitness of each new chromosome is calculated and a proportion e $(0 \leq e \leq 1)$ of the highest-fitness chromosomes from the previous generation are introduced to replace the worst chromosomes of the new population. Two examples are illustrated in Figures 1 and 2 to represent the procedure of crossover and mutation in our problem. Pseudocode for the GA is provided in Figure 3.

Algorithm: $GA\ (\boldsymbol{p}, \boldsymbol{t}, \boldsymbol{\lambda}, \boldsymbol{\lambda_1}, \boldsymbol{\mu}, \boldsymbol{\mu_1}, \boldsymbol{e})$

// Initialize generation 0:

$i = 0;$

P_i := *Generate random chromosomes;*

// Evaluate P_k:

- *Compute the fitness value for each chromosome;*

For k=1 to t:

// Create generation k

// 1. Copy:

- *Select $(1 - \lambda) * p$ members of P_{k-1} and insert into P_k;*

//2. Crossover:

- *Select $\lambda * p$ members of P_{k-1}:*
    - *Select $\lambda_1 * \lambda * p$ members, do one-point crossover and insert them into P_k;*
    - *Select $(1 - \lambda_1) * \lambda * p$ members, do two-point crossover and insert them into P_k;*

// 3. Mutation:

- *Select $\mu * p$ members of P_k:*
    - *Select $\mu_1 * \mu * p$ members, and flip a randomly- selected gene in each;*
    - *Select $(1 - \mu_1) * \mu * p$ members, and flip four randomly- selected genes in each;*

// 4. Evaluate P_k:

- *Compute the fitness value for each chromosome;*

// 5. Elitism:

- *Select $e * p$ best members of P_{k-1} base on their fitness values and replace them with the worse members of P_k.*

//6. Return the best solution of P_k;

Return the fitness value of the best solution found among all generations.

Figure 3: Genetic Algorithm Pseudocode

3 Results/Findings

In this section, we demonstrate (and tune) the performance of our model and algorithm using data obtained for the U.S. inland waterway network. A total of 7703 commodities are defined based on different paths in this network by extracting average origin-destination (o-d) tonnages—along with the associated draft, in feet—over 2009–2015 from the Waterborne Commerce Statistics Center [19]. Functions $f_k(j)$ are defined with respect to this data as the average tonnage (across all years) passing at a draft of no more than j feet, where j ranges from 6 to 12. The value of demand (D^k) for each commodity is assumed as the largest average tonnage that could be transported at each path. Also, the value of each commodity (v^k) is assumed to be one so that the model maximizes the expected tonnage that can be transported through the network.

There are 12 different districts (depicted in Figure 5) that have historical inland dredging record from 1990 to 2015 and the total number of segments in these districts is 440 (depicted in Figure 4). With a few exceptions that are explained in the following sentences, we assume that $l_i = 7,\ \forall i \in N$; thus, each segment has an initial depth of 7 feet, which can be improved by completing routine dredging. Exceptions to this rule were made based upon a search of the dredging records data set, provided in [18], which was executed for each of the 440 segments in our data set in attempt to identify evidence of historical dredging at that location. This was particularly cumbersome (and likely subject to some error) because the dredging records data set does not include a unique identifier for the river segment associated with each dredge job; rather, the data set includes a text string associated with each completed dredge job that provides some reference to the job location but may not include enough detail to pinpoint the job to one of our segments. (And furthermore, different strings of text may be used to identify jobs completed at the same location.) Still, we were able to identify some evidence of historical dredging for most of our 440 segments. For those segments at which at least 3 dredging records (362 of the 440 segments) were identified over 1990–2015, we set $l_i = 7$. For the remaining segments, we set $l_i = 10$, which discourages dredging via allowing the passage of traffic at the 8-, 9-, and 10-foot drafts even without completing any dredging. Naturally, we also removed availability constraints of the form (7)–(8) associated with coastal/deep-draft segments at which a lack of dredging is not likely to cause a disruption for barges meant to travel over inland waterways.

Using the same mapping of dredging records to segments in our data set, we then applied Equation (17) to assign each segment a cost of dredging 0, 1, 2, or 3 feet (i.e., the h_{ij}-values). In completing this process, we encountered many of the same obstacles as described in the previous paragraph. Additionally, cost data are missing for many of the dredging records that we were able to map back to an individual location. We estimated h_{ij}-values using only those records that both (i) could be mapped to a specific location and (ii) included cost data. The result is that our estimates of h_{ij} are subject to error due to both small sample sizes and the potential for our own misclassification of dredging records to waterway segments.

For this data set, we then generate $|\Omega|$ scenarios which describe the uncertain requirements for reactive dredging. Our conversations with industry experts suggests that reactive dredging is highly variable, comprising anywhere between 40% and 80% of total inland maintenance dredging costs. Thus, we generate parameters $c_z^\omega,\ z = 1, \ldots, Z$, according to a continuous uniform distribution that spans the interval beginning at 40% of historical average expenditures in district z and ends at 80% of historical average expenditures in district z. We assume c_z^ω-values in each scenario are generated independently, and we take these scenarios to be equally likely (i.e., $p^\omega = 1/|\Omega|,\ \forall \omega \in \Omega$). The assumptions underlying the instance generation process were necessary due to the challenges in obtaining historical data to support generation of a more realistic instance (e.g., modeling dependence among reactive dredging requirements in different districts).

Initially, for the purposes of demonstrating the performance of our GA against a known optimal solution, we examine smaller instances with $|\Omega| \in \{2, 5, 10\}$ scenarios and a total available budget $B \in \{\$140M, \$160M, \$180M, \$200M\}$. We take the budget shortfall penalty σ_z to be equal to the summation of total demands for all of the o-d paths. (Note: For this preliminary set of instances, σ_z turns out not to have any effect on the objective value because the selected levels of total budget are large enough to prevent any chance of shortfall. However, for the problems with low budget level, we will demonstrate a sensitivity analysis for the value of penalty in Section 4.) Additionally, because there are a large number of o-d pairs represented in the commodity flow data set, the result is a large-scale optimization problem which is difficult to solve optimally. Pareto analysis (see Figure 6) reveals that 86% of total tonnage of commodities corresponds to 21% of o-d paths. Based upon this analysis, we have used the reduced set of 1690 o-d pairs to construct our problem instances. Preliminary computational results verified that this would allow us to obtain more sensible results from the optimization. Table 1 displays the results obtained from solving each of these instances using both CPLEX (rows labeled **CPLEX**) and our GA (rows labeled **GA**) with population size 50, 400 iterations, and parameter values $\lambda = 0.6$, $\lambda_1 = 0.6$, $\mu = 0.2$, $\mu_1 = 0.2$ and $e = 0.4$. Both algorithms are implemented in JAVA using the University of Arkansas High Performance Computing Center, and the CPLEX implementation employs version 12.4 callable libraries. For each instance and algorithm (i.e., either CPLEX or GA), we report the optimal objective value (in expected tonnage) and the optimality gap (Gap

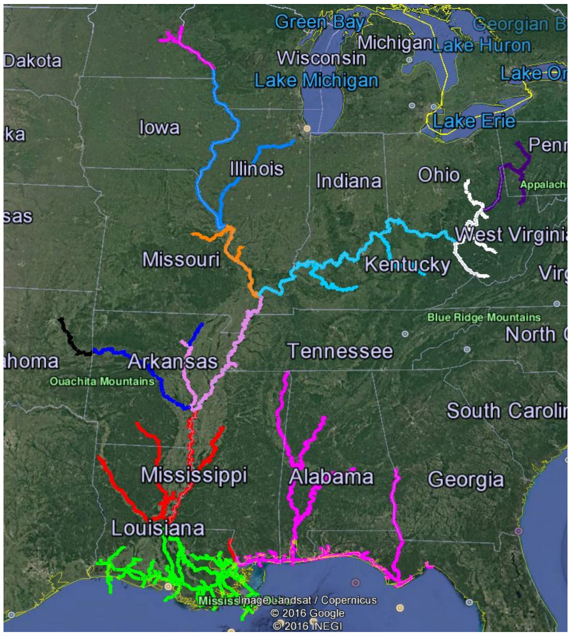

Figure 4: Districts

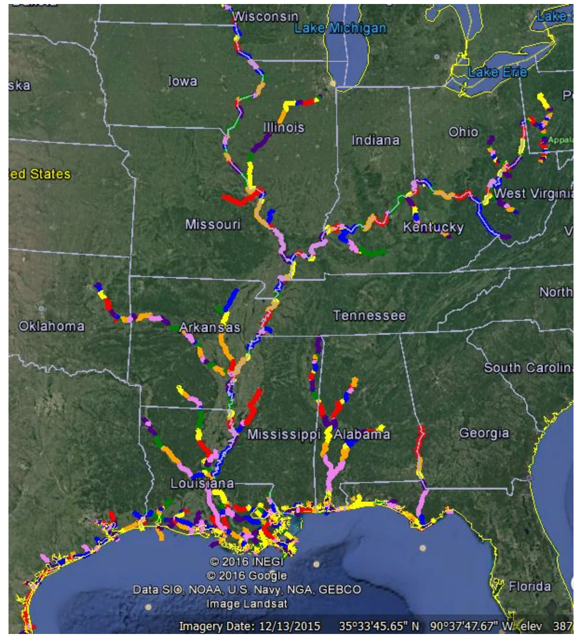

Figure 5: Waterway segments

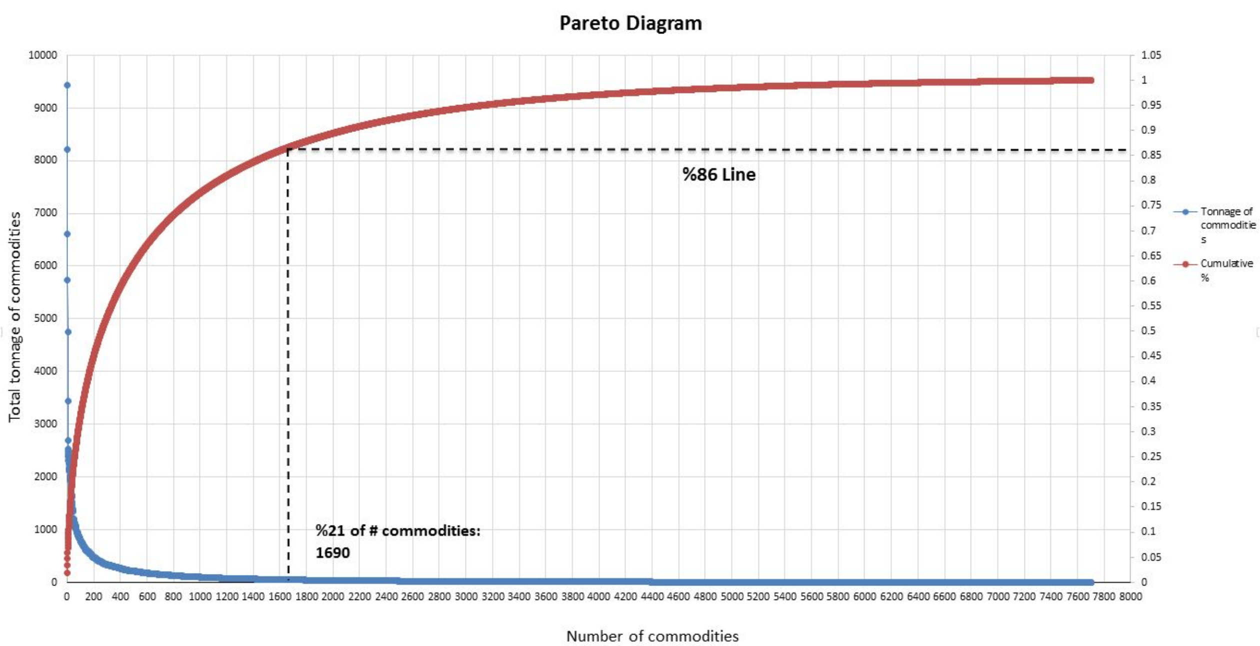

Figure 6: Pareto Analysis

%, defined as the absolute difference between the obtained objective and the objective value upper bound attained with CPLEX.) Even in these preliminary instances, which have a relatively small number of scenarios, we were forced to terminate CPLEX before optimality due to memory issues. Specifically, we terminate the instances with $|\Omega| = 5$ (respectively, $|\Omega| = 10$) as soon as CPLEX is able to prove it has obtained a solution within 5% (respectively, 10%) of optimality. This is why the rows labeled **CPLEX** contain nonzero gaps for the instances with $|\Omega| \in \{5, 10\}$.

The results demonstrate that the optimality gap increases by increasing the number of scenarios. In order to enhance the performance of our GA for instances with more scenarios, we analyzed the optimal budget allocation (provided by CPLEX) for the smaller instances. This analysis revealed that greater budgets tend to be assigned to districts containing many segments with relatively large benefit-to-cost ratios, as calculated by Equation (31). Therefore, we defined two new budget assignments (to replace Equation 30) based on the weights of each district to drive towards budget allocations possessing this structure:

1. The estimated benefit-to-cost ratio $r(i, x)$ is calculated based on Equation (31) for each segment of each district. The weight W_z for each district $z \in Z$ is then adjusted by a factor of $\sum_{i \in N_z} \sum_{x=1}^{3} r(i, x)$, and we use the adjusted weights to assign budget to each district $z \in Z$ as

$$B_z = \left(\frac{W_z \sum_{i \in N_z} \sum_{x=1}^{3} r(i, x)}{\sum_{z \in Z} W_z \sum_{i \in N_z} \sum_{x=1}^{3} r(i, x)} \right) B. \tag{32}$$

2. The number of segments that have a benefit-to-cost ratio ($\sum_{i \in N} \sum_{x=1}^{3} r(i, x)$) in excess of one percent is counted for each district (let M_z denote the count for district $z \in Z$) and the budget for district $z \in Z$ is assigned as

$$B_z = \left(\frac{W_z \ \max\{M_z, 1\}}{\sum_{z \in Z} W_z \ \max\{M_z, 1\}} \right) B, \tag{33}$$

   where weights are adjusted by a factor of $\max\{M_z, 1\}$ instead of M_z to prevent fixing a district's budget to zero.

Utilizing these budgeting strategies, we solved the instances from Table 1 again, this time (i) applying all three of the approaches (i.e., Equations (30), (32), and (33)) to convert the W_z-values in to an allocated budget to each district and (ii) selecting, for each chromosome, the approach that yields the largest fitness value. The new results and the optimality gap are represented in Table 1, where rows labeled $\mathbf{GA'}$ correspond to the improved GA. These

		Objective value in kton (Gap %)			
$\|\Omega\|$	Algorithm	$B = \$140$M	$B = \$160$M	$B = \$180$M	$B = \$200$M
2	CPLEX	171,497.32 (0.0%)	181,924.51 (0.0%)	189,681.59 (0.0%)	212,876.80 (0.0%)
	GA	134,296.65 (21.6%)	166,217.88 (8.6%)	168,275.93 (11.3%)	178,784.10 (16.1%)
	GA′	163,396.87 (4.7%)	166,602.72 (8.4%)	171,388.86 (9.6%)	187,331.74 (11.9%)
	GA″	165,556.357 (3.5%)	172,187.314 (6.9%)	178,924.38 (5.7%)	192,456.40 (9.6%)
5	CPLEX	170,561.17 (2.2%)	180,485.83 (5.2%)	196,343.55 (4.1%)	217,005.51 (4.2%)
	GA	135,445.55 (22.3%)	136,203.43 (28.4%)	166,895.48 (18.4%)	177,252.76 (21.7%)
	GA′	163,723.64 (6.1%)	166,217.68 (12.6%)	168,941.00 (17.4%)	178,717.50 (21.1%)
	GA″	164,300.35 (5.7%)	172,187.314 (9.5%)	172,017.092 (15.6%)	193,758.093 (14.5%)
10	CPLEX	161,723.52 (6.7%)	176,498.93 (8.2%)	188,761.00 (9.3%)	213,816.44 (5.6%)
	GA	135,445.55 (22.5%)	136,203.43 (28.9%)	166,895.48 (19.8%)	177,252.76 (21.5%)
	GA′	163,323.40 (5.7%)	165,978.12 (13.6%)	169,321.82 (18.6%)	177,940.58 (21.4%)
	GA″	165,329.95 (4.6%)	170,211.68 (11.4%)	171,092.64 (17.8%)	190,789.01 (15.8%)

Table 1: Objective value and optimality gap for preliminary experiments

results suggest that the optimality gap is reduced by adding new budget allocation strategies to the genetic algorithm; however, we also made one additional improvement to GA′ in which, after GA′ terminates, we then use CPLEX to solve the second-stage problem exactly using the budget allocation obtained by GA′. Results of the modified GA are also included in Table 1 (with rows labeled as **GA″**). This helps to find the optimal objective value for the best solution found by GA and hence reduce the optimality gap.

Table 2 reports the CPU time for each algorithm and instance from our preliminary experimentation. Results in Tables 1 and 2 reveal that GA″ provides a nice balance of obtaining (as compared to GA and GA′) a high-quality solution and (as compared to CPLEX) running quickly. (Note: In some cases, the GA″ time actually exceeds the full CPLEX time; however, this is due to the fact that we terminated the full CPLEX implementation early, whereas we ran the second-stage-only CPLEX implementation runs to full optimality.)

3.1 Tuning the GA

In order to find an effective setting for parameters in our GA, we perform a procedure based on statistical design of experiments proposed by Coy et al. [3]. The first step is selecting a subset of instances to analyze the results which is called *analysis set*. This set should cover different sizes of problems. The size of our problem is define based on the number of scenarios. We choose three problems with 20, 30 and 40 number of scenarios and solve them with total budget level of \$160M, 50 population size and 100 number of iterations. In the second step, we determine the initial set of parameters (design center), the size of incremental change of each parameter (Δ) and the limits of each parameter's value. These

		Computation time (s)			
$\|\Omega\|$	Algorithm	$B = \$140$M	$B = \$160$M	$B = \$180$M	$B = \$200$M
2	CPLEX	648.8	798.6	6575.7	127.1
	GA	46.0	47.2	47.5	47.6
	GA$'$	116.1	112.9	384.7	117.0
	GA$''$	132.0	526.0	2903.02	2114.0
5	CPLEX	104.0	1103.6	2008.8	1147.5
	GA	102.4	101.2	94.4	100.0
	GA$'$	198.7	203.4	190.5	227.3
	GA$''$	279.0	903.0	209.0	3307.0
10	CPLEX	198.8	135.5	25449	275.5
	GA	192.8	188.8	191.6	189.2
	GA$'$	357.9	365.3	358.8	404.1
	GA$''$	373.0	1368.3	375.0	3095.0

Table 2: CPU time for preliminary experiments

Parameter	Min value	Design center	Δ	Max value
Crossover rate (λ)	0	0.6	0.2	1
One-point crossover (λ_1)	0	0.6	0.2	1
Mutation probability (μ)	0	0.2	0.2	1
Mutation one-gene (μ_1)	0	0.2	0.2	1
Elitism (e)	0	0.4	0.2	1

Table 3: Min value, design center, Δ, and max value for parameters in GA

values are obtained by conducting a pilot study for a few number of trials and select the values yielding the best solution. Since, $\lambda_1 + \lambda_2$ and $\mu_1 + \mu_2$ should be one, we conduct the experiment for only five parameters instead of seven parameters. In Table 3, we show these values for parameters of our GA.

The third step is performing a factorial experimental design. We choose a full factorial design with $2^5 = 32$ runs. We convert this design to a matrix of coded variables shown in Table 6. To obtain the parameter's settings for each run, the Δ-value is added or subtracted to the design center depending on the sign of the coded variables in Table 6 and five replications is conducted for each parameter vector using different seeds to initialize the heuristic's random number generator in each case. Then, a linear regression model is fitted to the average results of five runs to find an estimate of response surface. The results of the regression analysis for three instances are represented in Tables 4. Based on the analysis, all models are significant at the 0.05 level. The minimum adjusted R^2 is 0.5224 and the maximum value is 0.6475. Also, each parameter is statistically significant in at least one

Problem	Adj. R^2	Intercept	λ	μ	λ_1	μ_1	e
20 scenarios	0.6475	165691	-463	4638	-185	-86	-3048
30 scenarios	0.6381	165305	41	4249	-107	212	-2730
40 scenarios	0.5224	164015	2356	5030	-412	636	-4119

Table 4: Coefficients of linear regression model

Problem	λ	μ	λ_1	μ_1	e
20 scenarios	-0.0199	0.2	-0.0079	-0.0037	-0.1314
30 scenarios	0.0019	0.2	-0.005	0.0099	-0.1285
40 scenarios	0.0936	0.2	-0.0164	0.0253	-0.1638

Table 5: Step size for each parameter

regression model based on their p-values.

Since we are maximizing, we find the path of steepest ascent on the response surface. In order to move along this path, we need to find the step size for each parameter within each instance. At first, We find the maximum of the absolute values of each parameter coefficient for a particular instance, and then divide the regression coefficient of each parameter by this maximum value. The step size is multiplying this ratio by Δ. The step sizes of each parameter at each instance are represented in Table 5.

The next step is evaluating parameter settings along the regression equation's steepest ascent direction from the design center. The procedure starts at the design center, and then, we add the step size of each parameter to its previous level. In the case of violating the upper or lower levels of any parameter, we hold that parameter constant and continue making steps with the other parameters. Then, we perform five experiments at each step and calculate the average objective value to determine the performance of GA for each set of parameter values. We continue making steps along the path until all of the parameters reach their maximum or minimum level. Tables 7–9 represent first 20 steps for three problems. The maximum average values for problems with 20, 30, and 40 scenarios are respectively obtained in steps 3, 3, and 17. Finally, we select the parameter settings associated with the maximum objective value. In the final step, the average of parameter values for the three instances is calculated to obtain the final set of parameters (Table 10).

3.2 Selecting Budget Shortfall Parameter Values

Using the GA parameters calculated in Table 10, we then solved a set of instances to determine an appropriate value for the budget shortfall penalty parameters σ_z. We apply

Run	λ	μ	λ_1	μ_1	e
1	-1	-1	-1	-1	-1
2	$+1$	-1	-1	-1	-1
3	-1	$+1$	-1	-1	-1
4	$+1$	$+1$	-1	-1	-1
5	-1	-1	$+1$	-1	-1
6	$+1$	-1	$+1$	-1	-1
7	-1	$+1$	$+1$	-1	-1
8	$+1$	$+1$	$+1$	-1	-1
9	-1	-1	-1	$+1$	-1
10	$+1$	-1	-1	$+1$	-1
11	-1	$+1$	-1	$+1$	-1
12	$+1$	$+1$	-1	$+1$	-1
13	-1	-1	$+1$	$+1$	-1
14	$+1$	-1	$+1$	$+1$	-1
15	-1	$+1$	$+1$	$+1$	-1
16	$+1$	$+1$	$+1$	$+1$	-1
17	-1	-1	-1	-1	$+1$
18	$+1$	-1	-1	-1	$+1$
19	-1	$+1$	-1	-1	$+1$
20	$+1$	$+1$	-1	-1	$+1$
21	-1	-1	$+1$	-1	$+1$
22	$+1$	-1	$+1$	-1	$+1$
23	-1	$+1$	$+1$	-1	$+1$
24	$+1$	$+1$	$+1$	-1	$+1$
25	-1	-1	-1	$+1$	$+1$
26	$+1$	-1	-1	$+1$	$+1$
27	-1	$+1$	-1	$+1$	$+1$
28	$+1$	$+1$	-1	$+1$	$+1$
29	-1	-1	$+1$	$+1$	$+1$
30	$+1$	-1	$+1$	$+1$	$+1$
31	-1	$+1$	$+1$	$+1$	$+1$
32	$+1$	$+1$	$+1$	$+1$	$+1$

Table 6: Augmented matrix of coded variables

Step	λ	μ	λ_1	λ_2	μ_1	μ_2	e	average
1	0.6	0.2	0.6	0.4	0.2	0.8	0.4	166,119.570
2	0.5801	0.4	0.5921	0.4079	0.1963	0.8037	0.2686	166,106.738
3	**0.5602**	**0.6**	**0.5842**	**0.4158**	**0.1926**	**0.8074**	**0.1372**	**166,137.529**
4	0.5403	0.8	0.5763	0.4237	0.1889	0.8111	0.0058	166,020.249
5	0.5204	1	0.5684	0.4316	0.1852	0.8148	0.0058	166,105.279
6	0.5005	1	0.5605	0.4395	0.1815	0.8185	0.0058	166,053.415
7	0.4806	1	0.5526	0.4474	0.1778	0.8222	0.0058	166,076.680
8	0.4607	1	0.5447	0.4553	0.1741	0.8259	0.0058	166,044.733
9	0.4408	1	0.5368	0.4632	0.1704	0.8296	0.0058	166,078.417
10	0.4209	1	0.5289	0.4711	0.1667	0.8333	0.0058	166,084.452
11	0.401	1	0.521	0.479	0.163	0.837	0.0058	166,000.877
12	0.3811	1	0.5131	0.4869	0.1593	0.8407	0.0058	165,982.361
13	0.3612	1	0.5052	0.4948	0.1556	0.8444	0.0058	166,046.475
14	0.3413	1	0.4973	0.5027	0.1519	0.8481	0.0058	166,071.494
15	0.3214	1	0.4894	0.5106	0.1482	0.8518	0.0058	166,074.446
16	0.3015	1	0.4815	0.5185	0.1445	0.8555	0.0058	166,025.181
17	0.2816	1	0.4736	0.5264	0.1408	0.8592	0.0058	166,100.2591
18	0.2617	1	0.4657	0.5343	0.1371	0.8629	0.0058	166,085.404
19	0.2418	1	0.4578	0.5422	0.1334	0.8666	0.0058	166,068.941
20	0.2219	1	0.4499	0.5501	0.1297	0.8703	0.0058	166,012.146
⋮								

Table 7: Parameter settings at each step (20 scenarios)

Step	λ	μ	λ_1	λ_2	μ_1	μ_2	e	average
1	0.6	0.2	0.6	0.4	0.2	0.8	0.4	165,754.758
2	0.6019	0.4	0.595	0.405	0.21	0.79	0.2715	166,055.983
3	**0.6039**	**0.6**	**0.5899**	**0.4101**	**0.22**	**0.78**	**0.143**	**166,088.644**
4	0.6058	0.8	0.5849	0.4151	0.2299	0.7701	0.0145	165,900.029
5	0.6077	1	0.5799	0.4201	0.2399	0.7601	0.0145	166,037.384
6	0.6096	1	0.5748	0.4252	0.2499	0.7501	0.0145	165,911.144
7	0.6116	1	0.5698	0.4302	0.2599	0.7401	0.0145	165,980.612
8	0.6135	1	0.5647	0.4353	0.2699	0.7301	0.0145	166,027.542
9	0.6154	1	0.5597	0.4403	0.2798	0.7202	0.0145	165,947.292
10	0.6174	1	0.5547	0.4453	0.2898	0.7102	0.0145	165,908.780
11	0.6193	1	0.5496	0.4504	0.2998	0.7002	0.0145	166,063.777
12	0.6212	1	0.5446	0.4554	0.3098	0.6902	0.0145	165,940.895
13	0.6232	1	0.5396	0.4604	0.3197	0.6803	0.0145	165,969.434
14	0.6251	1	0.5345	0.4655	0.3297	0.6703	0.0145	166,020.721
15	0.627	1	0.5295	0.4705	0.3397	0.6603	0.0145	166,072.009
16	0.6289	1	0.5245	0.4755	0.3497	0.6503	0.0145	165,916.764
17	0.6309	1	0.5194	0.4806	0.3597	0.6403	0.0145	166,010.784
18	0.6328	1	0.5144	0.4856	0.3696	0.6304	0.0145	166,067.654
19	0.6347	1	0.5093	0.4907	0.3796	0.6204	0.0145	165,942.106
20	0.6367	1	0.5043	0.4957	0.3896	0.6104	0.0145	165,954.107
⋮								

Table 8: Parameter settings at each step (30 scenarios)

Step	λ	μ	λ_1	λ_2	μ_1	μ_2	e	average
1	0.6	0.2	0.6	0.4	0.2	0.8	0.4	165812.497
2	0.6936	0.4	0.5836	0.4164	0.2253	0.7747	0.2362	165,871.898
3	0.7872	0.6	0.5672	0.4328	0.2506	0.7494	0.0724	165,837.506
4	0.8808	0.8	0.5508	0.4492	0.2759	0.7241	0.0724	165,867.396
5	0.9744	1	0.5344	0.4656	0.3012	0.6988	0.0724	165,883.435
6	0.9744	1	0.518	0.482	0.3265	0.6735	0.0724	165,871.857
7	0.9744	1	0.5016	0.4984	0.3518	0.6482	0.0724	165,865.739
8	0.9744	1	0.4852	0.5148	0.3771	0.6229	0.0724	165,873.022
9	0.9744	1	0.4688	0.5312	0.4024	0.5976	0.0724	165,824.245
10	0.9744	1	0.4524	0.5476	0.4277	0.5723	0.0724	165,854.621
11	0.9744	1	0.436	0.564	0.453	0.547	0.0724	165,885.148
12	0.9744	1	0.4196	0.5804	0.4783	0.5217	0.0724	165,889.192
13	0.9744	1	0.4032	0.5968	0.5036	0.4964	0.0724	165,869.559
14	0.9744	1	0.3868	0.6132	0.5289	0.4711	0.0724	165,884.340
15	0.9744	1	0.3704	0.6296	0.5542	0.4458	0.0724	165,871.127
16	0.9744	1	0.354	0.646	0.5795	0.4205	0.0724	165,861.755
17	**0.9744**	**1**	**0.3376**	**0.6624**	**0.6048**	**0.3952**	**0.0724**	**165,895.704**
18	0.9744	1	0.3212	0.6788	0.6301	0.3699	0.0724	165,876.327
19	0.9744	1	0.3048	0.6952	0.6554	0.3446	0.0724	165,838.856
20	0.9744	1	0.2884	0.7116	0.6807	0.3193	0.0724	165,852.752
⋮								

Table 9: Parameter settings at each step (40 scenarios)

Problem	λ	μ	λ_1	λ_2	μ_1	μ_2	e
20 scenarios	0.5602	0.6	0.5842	0.4158	0.1926	0.8074	0.1372
30 scenarios	0.6039	0.6	0.5899	0.4101	0.22	0.78	0.143
40 scenarios	0.9744	1	0.3376	0.6624	0.6048	0.3952	0.0724
Average	0.71	0.73	0.51	0.49	0.34	0.66	0.11

Table 10: Selected parameters

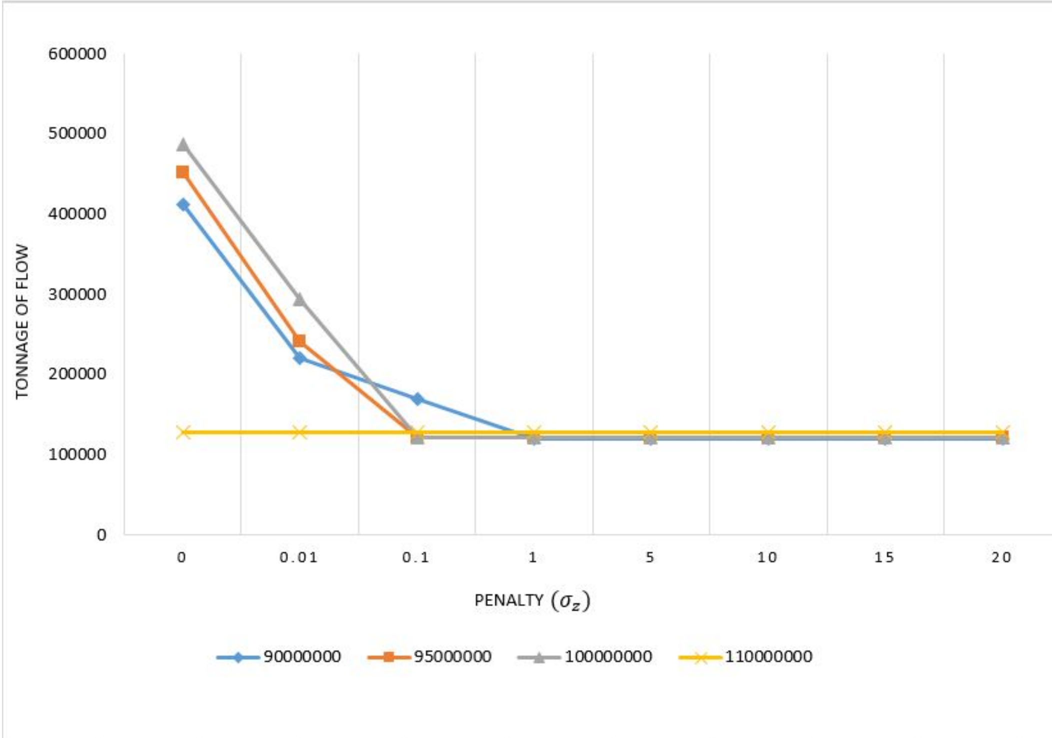

Figure 7: Objective value vs. penalty for different budget levels

GA″ (this time terminating the final CPLEX run at an optimality threshold of 10% because solving to optimality is too time consuming) to a set of instances corresponding to each combination of $B \in \{\$90M, \$95M, \$100M, \$110M\}$ and $\sigma_z \in \{0, 0.01, 0.1, 1, 5, 10, 15, 20\}$. (Note: We set σ_z equal to the same value for all $z \in Z$.) Each instance has $|\Omega| = 100$ scenarios in order to provide a more realistic representation of uncertainty. Results from this set of experiments are displayed in Figure 7. Based upon these experiments, we settled on $\sigma_z = 1$ as an acceptable penalty value because it seems to be a sufficient deterrent for exceeding each district's budget; however, it avoids some of the numerical issues that come with setting an arbitarily large (i.e., big-M type) penalty value.

4 Impacts/Benefits of Implementation

This section demonstrates the capabilities of our optimization model based upon applying GA″ (again terminating the final CPLEX at a 10% optimality threshold) to a set of instances corresponding to $|\Omega| = 100$ scenarios and $B \in \{\$90M, \$95M, \$100M, \$110M, \$120M, \$140M, \$160M, \$180M, \$200M\}$. We use the penalty value $\sigma_z = 1,\ z \in Z$, as explained in the previous section. The instances are constructed using the same approach as described at the beginning of Section 3.

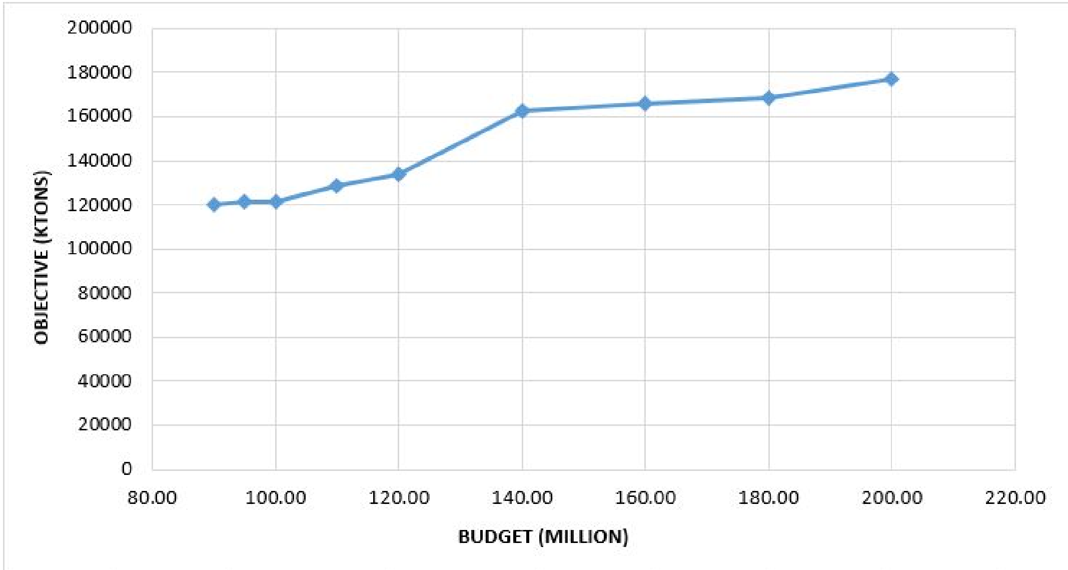

Figure 8: Objective value vs. budget

Figure 8 displays the GA$''$ objective value associated with each of the nine different budget levels. Naturally, the expected tonnage of commodities increases as more resources are made available for dredging. Furthermore, the increase seems to be sharpest before the $B = \$140M$ level, indicating that a significant portion of the impact can created without fully dredging all segments.

Figure 9 depicts each district's *shortfall probability* (i.e., the proportion of P_z^ω-variables that take on a positive value in each district). This probability represents the likelihood that district z will not have enough funds to complete the year's emergency/reactive jobs and therefore represents a measure of the risk of a significant disruption to navigation. In resource-limited instances (e.g., $B \in \{\$90M, \$95M, \$100M\}$), districts have shortfall probabilities as large as 0.27. This risk is reduced via increasing district budgets until (at $B = \$110M$) none of the district budgets are ever exceeded.

Figure 10 summarizes the first-stage decision variables, i.e., the percent of available budget that is allocated to each district. Results demonstrate that that resources are focused on several key districts (New Orleans, Mobile, Pittsburgh, Louisville, and Huntington). As the budget level is increased, the budget allocated to New Orleans and Mobile is increasing slowly (and the percentage thus decreases). This is presumably due to the fact that these districts are so critical to commodity flow that (even in the low budget instances) sufficient budget is allocated to these districts to perform a critical set of routine dredging jobs while simultaneously ensuring availability of funds for completing reactive jobs. In addition, the dredging costs for these districts tend to be higher, which serves to increase the budget al-

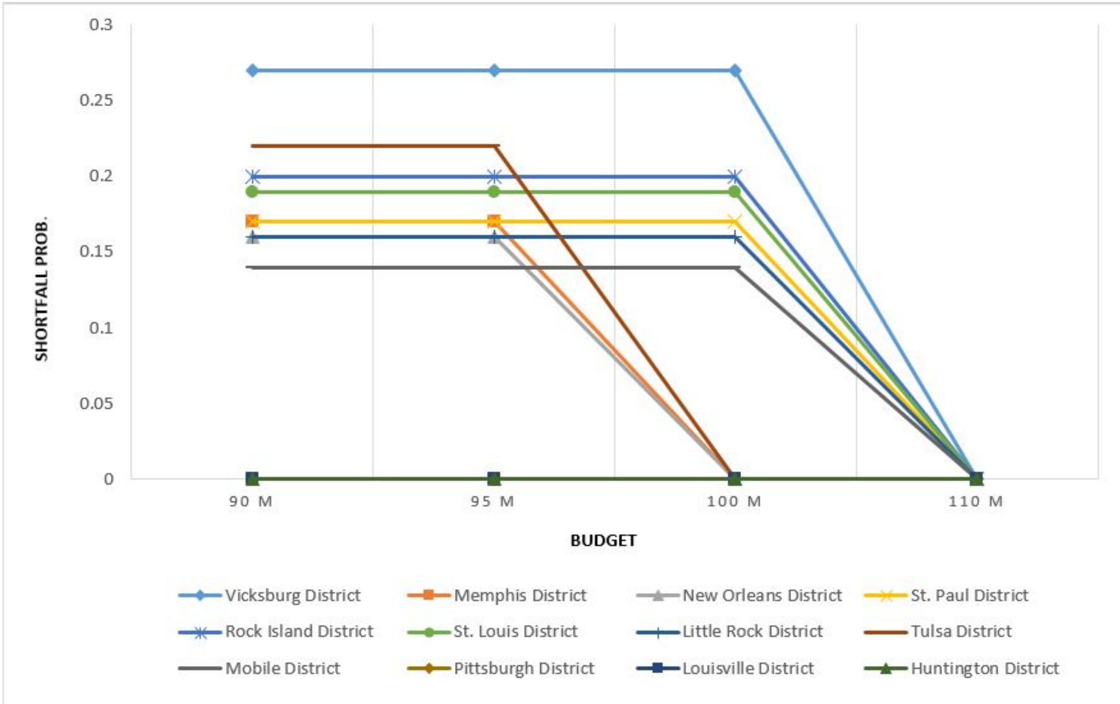

Figure 9: Probability of shortfall vs. budget for each district

located to these districts. Significant budgets are also allocated to Huntington, Louisville, and Pittsburgh. In these districts, a larger portion of the budget is used for routine dredging because the commodity tonnages passing through these districts are large relative to other districts while the routine dredging cost is relatively inexpensive. The Memphis, St. Paul, Rock Island, St. Louis, and Little Rock districts allocate enough budget to cover only reactive dredging. This is consistent with the dredging record data, which suggests that these districts are infrequently dredged, and dredging in these districts is almost always reactive in nature. Vicksburg contains a few sites selected for routine dredging; however, this district also appears less frequently in the dredging records, which is reflected by the smaller budget allocation in Figure 10.

We depict the second-stage solutions (i.e., routine maintenance dredging levels) in Table 11. Specifically, because each scenario's second-stage (routine dredging) variables (X_i^ω) may be different, we calculate the average dredging depth for each segment i as

$$\bar{X}_i = (1/|\Omega|) \sum_{\omega \in \Omega} X_i^\omega. \tag{34}$$

We then round $\bar{X}_i$ to the nearest integer (e.g., the rounded value is 2 if $1.5 < \bar{X}_i \leq 2.5$, but the rounded value is 3 if $2.5 < \bar{X}_i$) and report in Table 11 a count, by district, of the number of segments whose rounded average dredging depths equals 0, 1, 2, or 3. The table

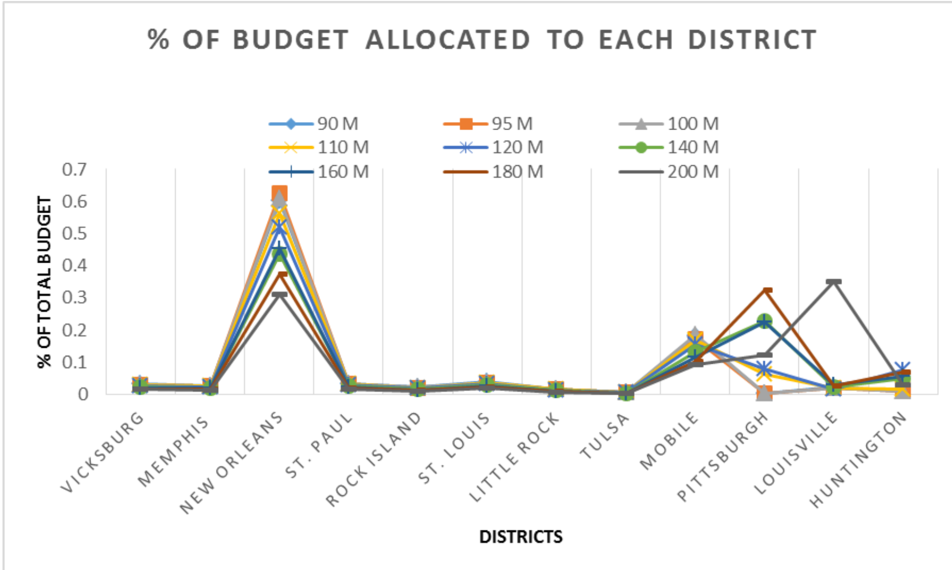

Figure 10: Percentage of budget assigned to each district

verifies some of the discussion in the preceding paragraph (e.g., regarding relatively stable dredging levels in New Orleans and Mobile as the budget increases and the fact that higher budget levels tend to concentrate resources on routine dredging in Pittsburgh, Louisville, and Huntington).

5 Recommendations and Conclusions

In this project, we consider a problem in which limited resources are allocated to districts for completing inland dredging projects. Uncertainty arises in this problem due to the underlying dynamics of the inland waterway system, in which natural events (e.g., weather or shoaling) can cause water levels to fall below the required depth for barges to pass. For this system, a scenario-based stochastic programming model is developed to maximize the expected value of origin-destination flow transporting through the network. We develop a genetic algorithm to solve realistically sized instances, and we report results obtained from an instance constructed for the U.S. waterway network.

Results demonstrate the tradeoff between investment in maintenance dredging and both (i) the network's overall capacity for transporting commodities (ii) risk associated with having insufficient budget to complete emergency projects. They also exhibit some intuitive characteristics such as budget allocations that drive towards more significant dredging at locations that transport larger amounts of freight and/or have cheaper dredging project costs. Still, we would not recommend implementing the model's output at this point. Rather, we

Table 11: Count of waterway segments by rounded average dredging depth for each budget level

B		Vicksburg	Memphis	New Orleans	St. Paul	Rock Island	St. Louis	Little Rock	Tulsa	Mobile	Pittsburgh	Louisville	Huntington	Total
90	Segments at 0 ft	37	22	94	20	12	2	17	10	103	39	32	24	412
	Segments at 1 ft	6	0	11	0	0	0	0	0	3	0	0	4	24
	Segments at 2 ft	0	0	3	0	0	0	0	0	0	0	1	0	4
	Segments at 3 ft	0	0	0	0	0	0	0	0	0	0	0	0	0
95	Segments at 0 ft	37	22	93	20	12	2	17	10	103	39	32	24	411
	Segments at 1 ft	4	0	6	0	0	0	0	0	3	0	0	4	17
	Segments at 2 ft	2	0	9	0	0	0	0	0	0	0	1	0	12
	Segments at 3 ft	0	0	0	0	0	0	0	0	0	0	0	0	0
100	Segments at 0 ft	38	22	90	20	12	2	17	10	102	39	32	24	408
	Segments at 1 ft	5	0	8	0	0	0	0	0	3	0	0	4	20
	Segments at 2 ft	0	0	10	0	0	0	0	0	1	0	1	0	12
	Segments at 3 ft	0	0	0	0	0	0	0	0	0	0	0	0	0
110	Segments at 0 ft	37	21	90	20	12	2	17	10	102	27	32	20	390
	Segments at 1 ft	6	1	10	0	0	0	0	0	3	0	0	1	21
	Segments at 2 ft	0	0	8	0	0	0	0	0	1	12	0	7	28
	Segments at 3 ft	0	0	0	0	0	0	0	0	0	0	1	0	1
120	Segments at 0 ft	37	22	91	20	12	2	17	10	102	21	32	0	366
	Segments at 1 ft	6	0	7	0	0	0	0	0	3	0	0	13	29
	Segments at 2 ft	0	0	10	0	0	0	0	0	1	18	0	11	40
	Segments at 3 ft	0	0	0	0	0	0	0	0	0	0	1	4	5
140	Segments at 0 ft	36	22	90	20	12	2	17	10	102	10	32	4	357
	Segments at 1 ft	7	0	10	0	0	0	0	0	3	1	0	10	31
	Segments at 2 ft	0	0	8	0	0	0	0	0	1	28	0	14	51
	Segments at 3 ft	0	0	0	0	0	0	0	0	0	0	1	0	1
160	Segments at 0 ft	36	21	88	20	12	2	17	10	102	11	31	0	350
	Segments at 1 ft	5	1	9	0	0	0	0	0	2	0	1	14	32
	Segments at 2 ft	2	0	11	0	0	0	0	0	2	28	0	11	54
	Segments at 3 ft	0	0	0	0	0	0	0	0	0	0	1	3	4
180	Segments at 0 ft	38	22	87	20	12	2	17	10	101	0	31	0	340
	Segments at 1 ft	5	0	7	0	0	0	0	0	3	10	1	9	35
	Segments at 2 ft	0	0	14	0	0	0	0	0	2	5	0	4	25
	Segments at 3 ft	0	0	0	0	0	0	0	0	0	24	1	15	40
200	Segments at 0 ft	37	22	89	20	12	2	17	10	102	8	16	6	341
	Segments at 1 ft	3	0	9	0	0	0	0	0	2	2	0	8	24
	Segments at 2 ft	3	0	10	0	0	0	0	0	2	29	16	14	74
	Segments at 3 ft	0	0	0	0	0	0	0	0	0	0	1	0	1
	Total	43	22	108	20	12	2	17	10	106	39	33	28	440

feel that a more thorough assessment of the cost/effect functions (i.e., $f(\cdot)$ and $\phi(\cdot)$) *associated with each segment* would lead to improved results that may have differences from those reported herein, and would recommend revisiting this before proceeding with the implementation of our model's results. Further experimentation with the model could also lend additional insights (e.g., what are the potential benefits of combining the dredging operations in two or more districts?) that may be useful in developing effective maintenance dredging strategies.

A subtle result that we found interesting was the impact—sometimes reducing the overall optimality gap by as much as 6.6%—of solving the second-stage project selection problem exactly (i.e., in algorithm GA$''$) as opposed to using a greedy approach (as in algorithm GA$'$). (Note: The reduction in the optimality gap would be a great deal more if we could afford to solve the second-stage problem exactly at every step.) This suggests there is significant value to be obtained via developing more effective solution techniques for large-scale resource allocation models (i.e., that include thousands of o-d pairs, hundreds of waterway segments, and hundreds of scenarios) such as ours—as opposed to allocating based on a simple rule of thumb. With this in mind, our research motivates further study into how to obtain high-quality solutions quickly for these large-scale problems. Future research may seek to improve upon our approach by drawing from the literature on stochastic programs with a discrete second stage.

Given that routine dredging jobs are completed on a multi-year schedule, a dynamic (multi-year) extension of this problem is also worthy of future research; however, the challenges (both in terms of acquiring the necessary data and solving the resulting optimization problem) become extreme in this case.

References

[1] J. R. Chimka. Economic impacts of lock usage and unavailability. Technical report, Maritime Transportation Research and Education Center (MarTREC) Final Research Report, 2016.

[2] D. A. Coley. *An introduction to genetic algorithms for scientists and engineers*. World scientific, 1999.

[3] S. P. Coy, B. L. Golden, G. C. Runger, and E. A. Wasil. Using experimental design to find effective parameter settings for heuristics. *Journal of Heuristics*, 7(1):77–97, 2001.

[4] G. A. Jones and J. G. Wilson. Optimal scheduling of jobs on a transmission network. *Management Science*, 25:98–104, 1979.

[5] J.-C. Jong and P. Schonfeld. Genetic algorithm for selecting and scheduling interdependent projects. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 127(1):45–52, 2001.

[6] M. Khodakarami, K. N. Mitchell, and X. B. Wang. Modeling maintenance project selection on a multimodal transportation network. *Transportation Research Record: Journal of the Transportation Research Board*, 2409(1):1–8, 2014.

[7] J. K. Lenstra and A. H. G. Rinnooy Kan. Complexity of scheduling under precedence constraints. *Operations Research*, 26:22–35, 1978.

[8] K. N. Mitchell, B. X. Wang, and M. Khodakarami. Selection of dredging projects for maximizing waterway system performance. *Transportation Research Record: Journal of the Transportation Research Board*, 2330(1):39–46, 2013.

[9] M. Mitchell. *An introduction to genetic algorithms*. MIT press, 1998.

[10] J. Sanburn. As barges sit idle along the mississippi, the economic costs grow. `http://business.time.com/2012/08/22/as-barges-sit-idle-along-the-mississippi-the-economic-costs-grow/`, August 2012. Accessed: 2014-06-17.

[11] X. Tao and P. Schonfeld. Selection and scheduling of interdependent transportation projects with island models. *Transportation Research Record: Journal of the Transportation Research Board*, 1981(1):133–141, 2006.

[12] X. Tao and P. Schonfeld. Island models for stochastic problem of transportation project selection and scheduling. *Transportation Research Record: Journal of the Transportation Research Board*, 2039(1):16–23, 2007.

[13] C.-J. Ting and P. Schonfeld. Optimization through simulation of waterway transportation investments. *Transportation Research Record: Journal of the Transportation Research Board*, 1620(1):11–16, 1998.

[14] U.S. Army Corps of Engineers. Corps implements new preventative maintenance program at three lock and dams. http://www.swt.usace.army.mil/Media/News-Stories/Article/491412/corps-implements-new-preventative-maintenance-program-at-three-lock-and-dams/. Accessed: 2017/01/24.

[15] U.S. Army Corps of Engineers. Inland navigation national level report. http://www.corpsresults.us/navigation/reports/navinationalreport.cfml. Accessed: 2016/07/11.

[16] U.S. Army Corps of Engineers. Navigation data center: Dredging information system. http://www.navigationdatacenter.us/dredge/dd14cos1.htm. Accessed: 2016/07/07.

[17] U.S. Army Corps of Engineers. Navigation data center: Dredging program. http://www.navigationdatacenter.us/dredge/dredge.htm. Accessed: 2016/07/07.

[18] U.S. Army Corps of Engineers. U.S. Waterway Data. http://www.navigationdatacenter.us/data/datadrgsel.htm.

[19] U.S. Army Corps of Engineers. U.S. Waterway Data: State to State and Region to Region Commodity Tonnages. http://www.navigationdatacenter.us/data/datapdom.htm.

[20] U.S. Army Corps of Engineers. The u.s. waterway system: Transportation facts and information. www.navigationdatacenter.us/factcard/factcard13.pdf. Accessed 2016/07/07.

[21] S.-L. Wang and P. Schonfeld. Scheduling interdependent waterway projects through simulation and genetic optimization. *Journal of Waterway, Port, Costal, and Ocean Engineering*, 131:89–97, 2005.

[22] S.-L. Wang and P. Schonfeld. Scheduling of waterway projects with complex interrelations. *Transportation Research Record: Journal of the Transportation Research Board*, (2062):59–65, 2008.