Technical Report: Large-scale Multi-objective Optimization for Watershed Planning and Assessment.
Full paper link: https://ieeexplore.ieee.org/document/10452414

Gregorio Toscano, Hoda Razavi, A. Pouyan Nejadhashemi, Kalyanmoy Deb, and Lewis Linker

***Abstract*—Selecting the appropriate Best Management Practices (BMPs) is crucial for reducing pollution levels and improving the watershed's water quality. However, identifying cost-effective BMP combinations for various locations is challenging, especially when using computationally expensive evaluation procedures like the Chesapeake Assessment Scenario Tool (CAST).**

This study presents a customized and hybrid evolutionary multi-objective optimization (EMO) algorithm aimed at enhancing the water quality in the Chesapeake Bay Watershed for two conflicting objectives: cost of BMP implementation and the amount of resulting nitrogen loading to streams. First, we present a surrogate model-based optimization approach and evaluate its accuracy and execution time against the CAST evaluation system. Then, we present a hybrid two-stage EMO procedure, which is initialized with solutions obtained from a point-based ϵ-constraint procedure and works with a repair operator to satisfy equality constraints. The hybrid EMO procedure yields a set of non-dominated trade-off solutions for problems with as few as 1,012 variables (West Virginia's Tucker County) to as large as 153,818 variables (the whole state of West Virginia). Alternate trade-off solutions provide a knowledge of different possible options and also importantly provide a flexible method of arriving at a single preferred solution for deployment.

The EMO procedure is then integrated with CAST using recent RESTful API approaches, and interesting accuracy versus computational trade-offs are discussed. Finally, a number of interesting insights of the scale-up optimization study reveal promising strategies to scale the application to multiple counties and the watershed level.

***Index Terms*—Large Scale Optimization, Evolutionary Multi-objective Optimization, Watershed Optimization, Hybrid Approach, Best Management Practices, Chesapeake Bay Watershed**

I. Introduction

MANY anthropogenic activities are directly or indirectly affecting water quality and aquatic habitats. For example, agriculture requires the application of fertilizers and manures that produce crops, but when applied in excess, can ultimately discharge into waterbodies, resulting in water quality degradation in the form of eutrophication and hypoxia in the Chesapeake Bay, Gulf of Mexico [2], and other coastal waters [20]. This is a concern as most nutrient pollution in the Chesapeake and Gulf of Mexico is from nonpoint sources, and their control and regulation are challenging [11]. Nutrient export from agricultural lands can be due to many reasons, including the heterogeneity of the agricultural landscape, the lack of regulatory enforcement to control fertilizer or manure application rate and amount, climate variability, and groundwater and surface water interactions [10], [11].

In order to control agrochemical discharge to waterbodies, both structural and non-structural measures known as Best Management Practices (BMPs) have been introduced and standardized by many agencies, including the Natural Resources Conservation Service (NRCS) [14]. However, their performance level varies by both physiographical (e.g., soil type, slope) and climatological (e.g., rainfall intensity, dry spell) factors. Therefore, BMP performance not only depends on the design characteristics but is also influenced by the location of the implementation site. Meanwhile, the level of complexity can exponentially increase as many types of BMPs can be implemented on the same parcel of land and in thousands of locations throughout a watershed [21]. Furthermore, considering all these factors in developing a watershed restoration plan can be a massive undertaking as numerous factors must be simultaneously considered [21].

To address these issues, watershed and water quality models have been widely adopted by water resource managers. However, despite their effectiveness in handling large and complex watershed planning through model scenarios, each scenario can only consider a single large-scale BMP implementation strategy. Evaluating the most cost-effective strategy would require a multitude of time-consuming evaluations and, in many cases, must be manually implemented. On the other hand, optimization methods can be integrated with watershed models to address these problems [22].

Prior research on watershed optimization has centered on agricultural management [13], [25], [31], urban restoration [3], [32], stormwater management [8], [33], and model calibration [15], [17]. In most BMP allocation studies, a watershed

Gregorio Toscano, and Kalyanmoy Deb are with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48863 U.S.A. (e-mail: toscano5@msu.edu; kdeb@egr.msu.edu)

Hoda Razavi and A. Pouyan Nejadhashemi are with the Department of Biosystems and Agricultural Engineering, Michigan State University, East Lansing MI 48863 U.S.A. (email: {razavi,pouyan}@msu.edu)

Lewis Linker is with the Environmental Protection Agency, Annapolis, U.S.A. (e-mail: Linker.Lewis@epa.gov)

model, a cost evaluation model, and an optimization algorithm aimed at pollutant reduction and implementation costs have been utilized [31]. However, current optimization algorithms are falling short of addressing this type of problem due to the large size of the optimization search space and the time-consuming process. Therefore, in this study, we used a surrogate model to improve the optimization computational time and narrow the search space to achieve the most cost-effective BMP allocation scenarios for the CBW.

A. Problem description

Chesapeake Bay Watershed (CBW) is located in the Mid-Atlantic region of the United States, with a drainage area of 165,760 km^2. The three major sources of pollution in the Bay are nitrogen, phosphorus, and sediment, primarily from agricultural activities followed by atmospheric nitrogen deposition, stormwater, wastewater, and failing septic tanks [26]. In order to address water quality issues in the Bay and its tributaries, thousands of variables and constraints need to be considered for the CBW optimization problem due to the size of the watershed and the complexity of land use and pollutant sources.

In our previous studies [16], [27], [29], we developed an evolutionary multi-objective optimization (EMO) technique to overcome the challenge of solving a large-scale watershed problem. The approach combines EMO with an ϵ-constraint approach and a new repair operator that satisfies the maximum number of constraints. In addition, a surrogate model of the objectives and constraints was used to verify the method to investigate the performance as the problem size increased. The proposed method was evaluated using the Chesapeake Assessment Scenario Tool (CAST) and the surrogate model.

Generic optimization techniques, such as random initialization, penalty-based constraint handling, and standard recombination and mutation operators, could be more efficient for large-scale practical problems [7]. To improve efficiency, these techniques must be tailored to specific problem classes. The constrained balanced optimization problem (CBOP) is a class of optimization problems that involve many issue instances over several counties, states, or a cluster of states with similar land-river segments (LRSs), BMPs for implementation, and techniques for computing objective and constraint functions (e.g., cost, nitrogen, phosphorus, and sediment loads). A generic optimization technique for the CBOP class will be suitable for most issue situations at a county level. However, as the problem's size increases, the optimization algorithm's complexity (mostly variables and constraints) will also increase. This research aims to provide an efficient optimization technique that works well in various variants of the problem class, regardless of whether the CBOP is formulated for a single county, several counties, or the state level. The proposed methodology is evaluated based on its operating principles, scalability to large areas, and practicality to combine with computationally cheaper surrogate models.

In the remainder of this paper, we first provide a background of the current tools and problems in Section II. Then, Section III presents different customized components to improve our selected evolutionary algorithm to optimize the CBW optimization problem. Next, we present the proposed approach and a study balancing the use of the CAST system and surrogate models in Section IV. Finally, conclusions are discussed in Section V.

II. Background

A. The Chesapeake Assessment Scenario Tool (CAST)

CAST is widely used as the primary watershed modeling tool for evaluating the impact of management scenarios on overall water quality conditions, measured by the reduction of pollution from point and non-point sources in the Chesapeake Watershed.

Decision variables are presented as a vector of all possible combinations of Land River Segments (LRSs), load sources, jurisdictions, and efficiency BMPs to optimize the model. Each element in the vector represents the percentage of the corresponding BMP relative to the total LRS area, the load source, and the political/agency jurisdiction. However, the number of decision variables increases with LRSs, particularly in larger areas. Constraints ensure that BMPs allocated to each combination of LRS, load source, and jurisdiction sum up to 100% of the available area without overlapping.

This optimization model is presented in a generic formulation, where the number of involved variables and constraints is contingent on CBW's focal area, be it county-level, statewide, or encompassing the entire watershed, as referenced in [19], [28]:

$$
\begin{aligned}
&\text{Min. } f_1(\mathbf{x}) = \sum_{s\in S}\sum_{h\in H_s}\sum_{u\in U}\sum_{b\in B_u} \tau_b x_{s,h,u,b},\\
&\text{Min. } f_2(\mathbf{x}) = \sum_{s\in S}\sum_{h\in H_s}\sum_{u\in U}\left[\alpha_{s,h,u}\phi_{s,h,u}\prod_{G^B\in\mathcal{G}^B}\left(1-\sum_{b\in G^B}\eta^N_{s,h,b}\frac{x_{s,h,u,b}}{\alpha_{s,h,u}}\right)\right],\\
&s.t. \quad \sum_{b\in G^B} x_{s,h,u,b} = \alpha_{s,h,u}, \quad \forall s\in S,\ h\in H_s,\ u\in U_s,\ G^B\in\mathcal{G}^B,\\
&\qquad x_{s,h,u,b}\geq 0, \quad \forall s\in S,\ h\in H_s,\ u\in U_s,\ b\in B_u.
\end{aligned}
\tag{1}
$$

The number of acres (real-valued variables) used to implement a specific BMP b on an agency h, load-source u, and land-river-segment s is denoted by $x_{s,h,u,b}$. Note that not all combinations of the four tuples are allowed due to practical reasons, making another level of difficulty in directly using the above formulation within an optimization algorithm. EMO methods allow a relatively easy way to allow restrictive nature of variables. The first objective function $f_1(\mathbf{x})$ represents the overall cost of implementing all BMPs, while the second objective function $f_2(\mathbf{x})$ calculates the nitrogen load reduction. The parameter τ_b denotes the cost per unit acre of implementing BMP b, and $\eta^N_{s,h,b}$ represents the efficiency of BMP b in removing nitrogen when applied to agency h and land-river-segment s. The parameter α indicates the total available acres, and $\mathcal{G}$ includes all the groups of BMPs G that can be applied to a given (s,h,u). All parameter values are selected from CBW's practice.

The primary objective is to determine the cost of implementing all allocated BMPs, and the secondary objective is to assess the nitrogen load after implementing the BMPs. Therefore, our optimization goal is to design BMP allocations that minimize both the cost of implementation and the nitrogen loading to waterbodies (i.e., edge of the stream).

This study examines all 11 counties in the West Virginia Chesapeake Bay Watershed (CBW), including Berkeley,

TABLE I
BASE NITROGEN, NUMBER OF CONSTRAINTS, AND VARIABLES FOR THE STUDIED COUNTIES.

County	Base N_2 (f_2^{base})	#Constraints	#Variables
Berkeley	977,896	1,813	14,090
Grant	1,049,450	3,448	25,228
Hampshire	1,012,797	1,700	12,783
Hardy	1,344,295	2,491	18,607
Jefferson	1,018,012	1,606	12,303
Minral	763,864	2,698	20,260
Monroe	48,655	399	3,102
Morgan	271,134	1,665	11,880
Pendleton	1,133,327	4,352	33,083
Preston	4,683	193	1,470
Tucker	1,702	144	1,012
Total	7,625,818	20,509	153,818

Grant, Hampshire, Hardy, Jefferson, Mineral, Monroe, Morgan, Pendleton, Preston, and Tucker. County-specific variables and constraints are presented in Table I. Even the smallest county in the study has over 1,000 variables, making it a large-scale optimization problem for any algorithm, including evolutionary algorithms. However, when considering the entire state, as executed in this paper, the scale of the problem becomes even larger, involving 153,818 real-parameter variables to be optimized. In addition, the base nitrogen values in the table are derived from existing BMP usage in each county and state, which are used to show the results of our proposed algorithm.

In summary, this paper develops a multi-objective decision optimization program that must identify a set of BMPs chosen from a large number of options covering states and the watershed, satisfying a large number of constraints, minimizing the cost of implementation and simultaneously minimizing nitrogen loading to the environment.

B. Optimizing Multiple Objectives

A large number of variables in an optimization formulation of the overall problem poses a challenge for any optimization method. However, in this study, besides having a large number of variables, the nature of the problem in CBW is multi-objective since it requires the simultaneous minimization of two contradictory objectives: (i) lowering the costs of BMP implementation and (ii) decreasing pollutant loads to the environment.

In contrast to a single optimal solution, a multi-objective optimization problem yields a set of trade-off Pareto-optimal solutions [6], [23]. The Pareto-optimal solutions lie on the boundary of the feasible objective space. No other feasible solution outperforms these Pareto-optimal solutions in all objectives in the search space. The Pareto-optimal solution set usually contains multiple trade-off solutions among objectives. It is worth noting that the set also contains individual optimal solutions for each single-objective constraint problem.

Usually, solving a multi-objective optimization problem involves scalarizing numerous objectives into a single parameterized objective function and solving several scalarized problems by systematically varying the parameters. Typically, a weighted-sum approach [5] with a weight vector $\mathbf{w}$ as a user-defined parameter set is employed to minimize $\sum_{k=1}^{M} w_k f_k(\mathbf{x})$. Nonetheless, there are at least two issues with this strategy. First, the objectives must be normalized so that their weighted total gives both objectives an equivalent value. Second, when set out on a f_1-f_2 plot, the best solutions may not present decision-makers with well-distributed spots from which to choose.

In this work, we opted for the Non-dominated Sorting Genetic Algorithm III (NSGA-III) [18], a population-based multi-objective optimization method that identifies multiple sets of Pareto-optimal solutions in a single run. The selected method is based on evolutionary optimization and employs a non-dominated sorting strategy and a reference vector-based diversity-preserving operator to prioritize diverse objective solutions within an evolving population. It is essential to highlight that our current framework can be readily adapted to minimize additional pollutants, such as phosphorus or sediments. Given the potential complexity of managing multiple objectives, choosing an approach like NSGA-III, which has shown effectiveness in handling problems with three or more objectives, becomes important. Coupled with this is the algorithm's reliance on Pareto-dominance, which seamlessly handles objectives of different magnitudes without scalarization. These two features ensure that solutions are evaluated and prioritized based on their dominance in the multi-objective space rather than on the inherent scale of any single objective.

III. DESIGN OF CUSTOMIZED EMO PROCEDURE

This paper employs an evolutionary algorithm (EA), an adaptable and modifiable technique successfully used to solve many other real-world problems. Using problem-specific knowledge, it is possible to (i) tailor an EA's effective representation scheme, (ii) develop a simple and computationally efficient assessment technique, (iii) work with a biased initial population, (iv) customize its operators, and (v) improve the termination criterion.

We apply these characteristics to create the algorithm's components. For this reason, we devise three improvements to customize our proposed method's components: (i) We adopt a surrogate model to reduce the computational burden for evaluating a solution, ii) we use a scalarization approach with the surrogate model to find a biased initial population or good solutions. By leveraging the ϵ-constraint approach, we can efficiently harness problem-specific data, such as gradients/Jacobians and the Hessian matrix, to swiftly pinpoint efficient solutions within the realm of constrained optimization, and iii) we devise a repair method for linear constraints to deal with predominantly possible solutions, rather than dealing with complex and generic methods of handling severely constrained space. Similar to other real-world applications [30], our method unfolds in two stages. In the first stage, the ϵ-constraint method swiftly identifies an initial approximation of a well-distributed set of near Pareto solutions, leveraging a surrogate model to optimize computational efficiency. Following this, in Stage 2, the NSGA-III strategy finds well-distributed and well-converged trade-off solutions by integrating evaluations from both the surrogate model and CAST. Since

CAST evaluation tool is not differentiable, the use of a direct search method, such as NSGA-III, is justified. Moreover, given that we are working on a large-scale optimization problem, we will measure how well our strategy performs when the problem is scaled. It is also worth highlighting that, apart from the ϵ-constraint approach, alternative methods such as the Achievement Scalarization Function, among other scalarizing functions, can be integrated for problem-solving. Furthermore, given our venture into large-scale optimization, we aim to evaluate the robustness of our strategy, especially in scenarios with scaled complexities.

The experimental conditions are described in detail below.

- **Large-run front**: By injecting (already described) solutions into the NSGA-III, we construct a non-dominated reference set. The NSGA-III algorithm is run for 1,000 generations with a population size of 1,000 and 600 reference points. We adopted standard values for the remaining parameters and carried out the identical approach for each county and the computed merged counties.
- **Performance measure**: We employ the Hypervolume (HV) ratio, which divides the HV of each execution by the HV of the PF. We determined each Pareto front's maximum and lowest values for a fair comparison across all problems. We utilized these numbers to normalize the output of the NSGA-III. Therefore, we adopt $(1.1, 1.1)$ as a reference point.

A. Surrogate Model for Efficient BMPs in the CBW

To analyze a vast number of data points and variables, large-scale optimization problems require significant computational power, making them challenging to solve. Furthermore, the quantity of data and variables to be processed can make it more difficult to identify a suboptimal solution. In addition, it can be arduous to discover patterns and relationships between variables due to the volume of the data. Finally, the problem's intricacy can make it challenging to build efficient algorithms to address it.

CAST's scenario execution is a time-consuming process. Consequently, building components and the execution of optimization algorithms employing CAST to evaluate scenarios may demand significant time.

Utilizing surrogate models is a systematic method for addressing computationally expensive problems. In this instance, the execution of CAST adds layers of complexity due to the lengthy evaluation time required by CAST.

Among the various forms of BMP, efficiency BMPs are the ones that can be modeled most effectively using surrogate models. Therefore, in this study, we will focus on these BMPs.

For our surrogate model, we adopted Equation 1. For one to have confidence in the results evaluated by a surrogate model, it is imperative that the model be precise. Consequently, we must maintain the Pareto dominance relation when dealing with multi-objective optimizations.

With this experiment, we want to determine whether the surrogate model described in Equation 1 is comparable to the CAST model. Below, we compare the execution time of the proxy model to that of CAST. In addition, we also assess the accuracy of the surrogate model and determine the effect of its use regarding Pareto dominance.

To validate the execution time, we randomly generated 10,000 scenarios for each county and counted the milliseconds that the surrogate model and CAST require to evaluate them. Then, the data in Table II presents the evaluation time.

The surrogate model requires an average of 18 milliseconds, whereas CAST requires an average of 1.5 minutes (the surrogate model requires 0.01% of the time CAST requires to evaluate the same scenario). The CAST assessment time is affected by a number of variables, most significantly network speed and job saturation. Currently, CAST can simultaneously run five scenarios. However, we would like to highlight that currently, CAST is running in an environment where the computational power can be increased by software (Hardware as a service, or HaaS for short); thus, it is possible to increase the computational power such that the evaluation of all our scenarios could be computed concurrently if more hardware is deployed. Despite this, the distinction is so substantial that, from a computational standpoint, utilizing the surrogate model will always be more advantageous than using CAST.

TABLE II
TIME COMPARISON IN MILLISECONDS OF THE SURROGATE MODEL AND CAST.

Statistic	Surrogate model (ms)	CAST (ms)
Best	1	10,881
Worst	59	175,857
Avg	18.44	93,022.81
STD	14.34	42,173.00

As the BMP implementation cost (the first objective of our CBOP) is computationally straightforward to determine, no surrogate model is necessary. Therefore, our surrogate model focuses mainly on calculating the nitrogen load (second objective).

To verify the accuracy of the surrogate model, we simulate other 10,000 scenarios for each of the eleven study counties. Then, we execute each scenario on both the surrogate model and the CAST system. Finally, we compute the percentage absolute error using Equation 2.

$$E = 100.0 \times \frac{|load - load'|}{load} \tag{2}$$

where $load$ refers to the load value obtained by evaluating the function using the CAST system. $load'$ refers to the evaluation of the solution using the surrogate model. In this case, the second objective of our formulation f_2.

Table III shows the results of the computed error. The table's rows show basic statistics for the computed error, while the columns link to each studied county. The results indicate that our adopted surrogate model performs similarly to the CAST system. To ease the analysis of results, we also show a boxplot comparison in Figure 1. This figure shows us graphically that the surrogate model has very stable performance, always presenting less than 2% of error regarding the CAST system.

Although the results indicate that our surrogate model is accurate regarding the CAST system when working with

TABLE III
SURROGATE MODEL EVALUATION ERROR (IN PERCENTAGE) FROM THE CAST SYSTEM EVALUATION

	Berkeley	Grant	Hampshire	Hardy	Jefferson	Mineral	Monroe	Morgan	Pendleton	Preston	Tucker
Best	0.047	0.068	0.111	0.113	0.036	0.082	0.126	0.081	0.087	0.013	0.045
Worst	0.844	1.230	1.629	1.557	0.724	1.631	1.708	1.661	1.269	0.354	1.101
AVG	0.497	0.703	0.863	0.863	0.381	0.893	0.957	0.949	0.700	0.178	0.611
STD	0.239	0.350	0.422	0.439	0.205	0.460	0.433	0.470	0.355	0.097	0.302

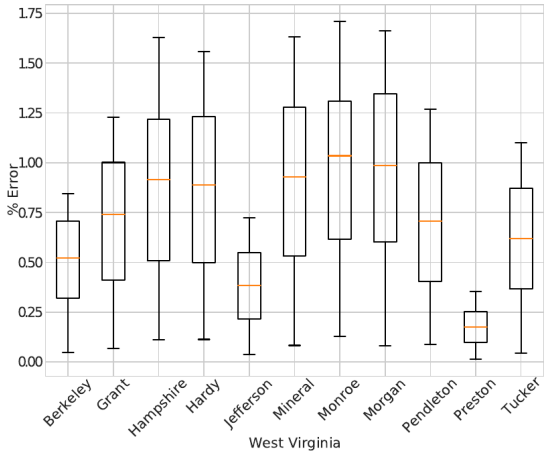

Fig. 1. Surrogate model percentage error regarding the CAST system.

multi-objective problems, we want the surrogate model to preserve the Pareto dominance relation [1]. That is when we compare two solutions on the surrogate-model space. Here, we expect to have the same Pareto dominance relation as if they were evaluated using the real CAST system. For this task, we perform a Pareto-dominance comparison of pairs of solutions using the surrogate model and the CAST system.

We generated 10,000 random scenarios for each of the 11 counties of this study. For each generated scenario (solution), we followed the following steps:

- Select a different scenario (solution) at random.
- Evaluate both solutions in the surrogate model and compare their objectives using Pareto dominance.
- Evaluate both solutions in the CAST system and compare their objectives using Pareto dominance.
- In the event of a discrepancy in the Pareto dominance relation, we increase a counter that keeps track of these disagreements.
- Finally, we compute the disagreement ratio: counter/10,000. We favor low values for this ratio because high values can harm our assurance of the surrogate model's conformity with the Pareto dominance compliance in CAST.

Figure 2 presents the ratio of Pareto dominance disagreements regarding the 10,000 performed comparisons. The figure shows that the ratio is small in all cases and even zero for one county (Tucker). These findings imply that the surrogate model can serve as a substitute for the CAST system in the design of our approach's components. However, to make the final solutions acceptable to the real users, their evaluation using the CAST system either for the final solutions or partially during optimization is warranted.

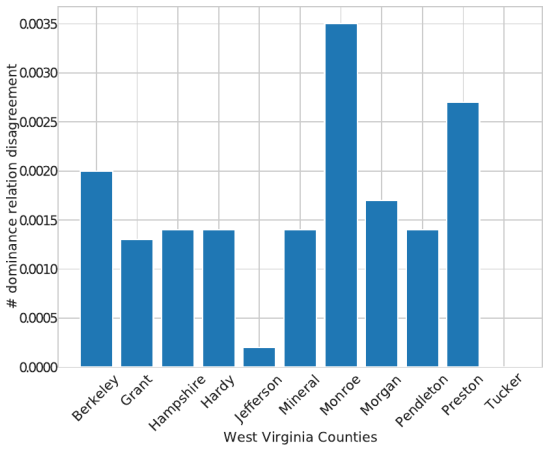

Fig. 2. Bar plot showing the proportion of Pareto dominance disagreements regarding the 10,000 performed comparisons.

B. Biased Initialization

We use a surrogate model to expedite the search process. The surrogate model adoption lets us propose a hybrid evaluation strategy. The initial population of the multi-objective evolutionary algorithm is seeded with a point-based interior-point optimization (IPOPT) method [4]. As IPOPT is a single objective optimization method, we reformulate the problem as a constrained optimization problem, as follows: minimize $f_1(\mathbf{x})$, subject to $f_2(\mathbf{x}) \leq \epsilon_k f_2^{\text{base}}$, where f_2^{base} is the nitrogen loading associated with current practice, and $\mathbf{X}$ denotes the feasible variable set associated with any other constraints of the original formulation. The objective function is restated as a constraint using ϵ, which transforms the initial single-objective problem into a single-point multi-objective technique that must be executed k times to obtain k representative Pareto solutions. ϵ_k is used to find a respective solution $\mathbf{x}^{(k)}$, which is added to the initial population. We use the IPOPT approach to build an epsilon-constraint method [6], [23], and we have used 11 values of ϵ_k from 1.0 to 0.70 in decreasing steps of 0.03. However, independent executions to reach a single point make this strategy inefficient and time-consuming for a large k. That explains the small k value selected for this biased initialization method.

The next experiment tries to determine whether incorporating knowledge into the NSGA-III can enhance its performance. In order to achieve this, we select the sets of points created by our epsilon-constraint method and inject them into the NSGA-III.

Figure 3 illustrates the performance of injecting various points in Berkeley, West Virginia. The points for this county

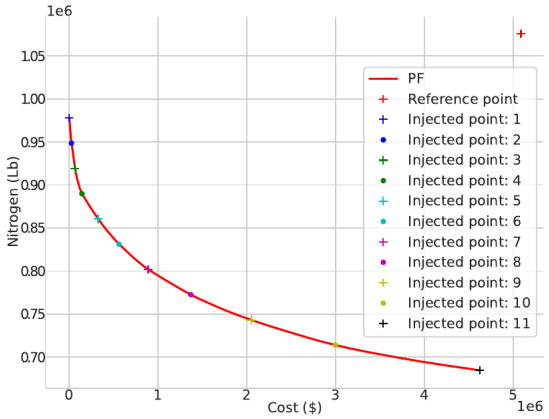

Fig. 3. Plot showing the effect of injected points into NSGA-III procedure for the Berkeley county.

and the other counties are determined as follows:

- 0 injected points: No point is injected into NSGA-III.
- 1 injected point: Point 1 (best for cost objective).
- 2 injected points: Points 1 and 11 (best for cost and nitrogen objectives).
- 3 injected points: Points 1, 6 (a near 50%-50% compromise point) and 11.
- 4 injected points: Points 1, 4, 8, and 11.
- 5 injected points: Points 1, 3, 6, 9, and 11.
- 6 injected points: Points 1 3, 5, 7, 9, and 11.
- 11 injected points: All 11 points.

We modified the termination criterion of NSGA-III to evaluate its performance with various injected point sets. To achieve this, we normalize each Pareto front and compute its hypervolume using $(1.1, 1.1)$ as a reference point. In every generation, we normalize the current population using the lower and upper bound of the Pareto front and remove solutions that exceed the HV's reference point in any dimension. Finally, we have computed the HV of the filtered population. The algorithm terminates when either the maximum number of generations, i.e., 1,000, is reached, or a trade-off solution set with 90% of the large-front run's HV is obtained.

Our findings underline the necessity of augmenting the initial population with external solutions to achieve competitive results using NSGA-III. The results indicate that the NSGA-III requires the injection of solutions in its initial population to produce competitive results. Notably, the algorithm could consistently yield competitive solutions across all counties until five solutions were injected into the initial population. The comprehensive results of this strategy, achieved through 31 runs with varying initial populations, are detailed in Table IV.

Figure 4 graphically displays the generation where NSGA-III stopped while optimizing the Berkeley county. This plot highlights the importance of injecting solutions into NSGA-III. On the one hand, when zero or one solution was injected, NSGA-III failed to achieve 90% of the PF's HV. On the other hand, providing 11 solutions almost automatically met the 90% goal. Upon analyzing Figure 4, it can be inferred that the benefits of point injection in NSGA-III become evident when two or more points are injected. Figure 5, thereby highlighting that NSGA-III can yield satisfactory results when more than one solution is utilized in the initial population.

Although the data may give us a false sense of simplicity for this problem, as NSGA-III only needs two generations in many cases to achieve 90% HV when 11 points are injected, exploring the extent of improvement that NSGA-III can achieve with additional iterations will be interesting. However, it is important to note that these results are only possible to obtain with injecting solutions.

Table V shows the results of running the NSGA-III for 100 generations with 2, 5, and 11-point injections (see boxes 2 (R), 5 (R), and 11 (R)). NSGA-III continues to improve solutions as more iterations are performed.

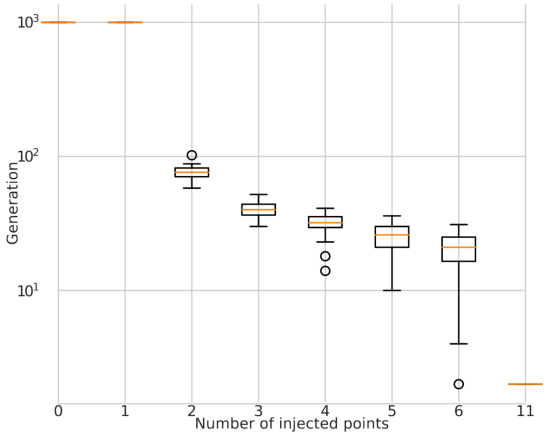

Fig. 4. Boxplots display the generation at which NSGA-III achieved 90% of the hypervolume of the large run. The results show that injecting more points leads to faster attainment of the desired hypervolume (Berkeley county).

C. Repair Operator

Equation 1 shows that each LRS is subject to an equality constraint: the sum of all BMP proportions must be equal $\alpha_{s,h,u}$. Suppose we select all BMPs randomly in the interval $[0, \alpha_{s,h,u}]$, where the value represents the implementation ratio. It would be improbable that the sum across all BMPs

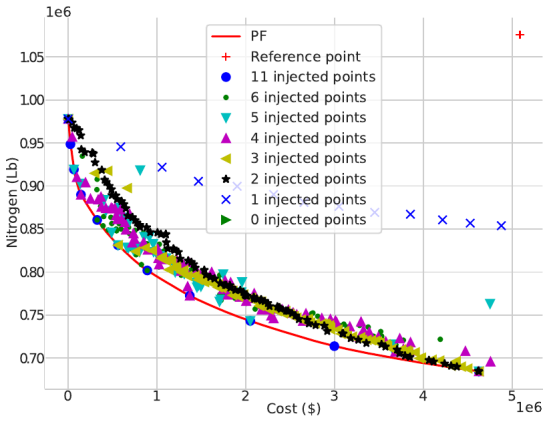

Fig. 5. The non-dominated points of NSGA-II were obtained using various numbers of injected IPOPT points in the initial population. The performance of NSGA-III improves as the number of injected points increases. The 'PF' line represents the Pareto front obtained from the large-run NSGA-III (Berkeley county).

TABLE IV
NSGA-III'S HYPERVOLUME ON THE EXIT GENERATION FOR EVERY COUNTY IN WEST VIRGINIA AFTER INJECTING A DIFFERENT NUMBER OF POINTS.

Counties	0 points AVG ±SD	1 point AVG ±SD	2 points AVG ±SD	3 points AVG ±SD	4 points AVG ±SD	5 points AVG ±SD	6 points AVG ±SD	11 points AVG ±SD
Berkeley	0.0 ±0.0	0.3842 ±0.080	0.9013 ±0.000	0.9045 ±0.002	0.9067 ±0.005	0.9082 ±0.006	0.9078 ±0.005	0.9355 ±0.005
Grant	0.0 ±0.0	0.5118 ±0.050	0.9019 ±0.001	0.9053 ±0.004	0.9073 ±0.005	0.9103 ±0.006	0.9101 ±0.005	0.9364 ±0.002
Hampshire	0.0 ±0.0	0.5259 ±0.066	0.5520 ±0.095	0.8896 ±0.018	0.9037 ±0.003	0.9039 ±0.003	0.9037 ±0.003	0.9345 ±0.008
Hardy	0.0 ±0.0	0.4532 ±0.095	0.9011 ±0.000	0.9041 ±0.002	0.9066 ±0.004	0.9105 ±0.007	0.9098 ±0.007	0.9346 ±0.005
Jefferson.	0.0 ±0.0	0.5750 ±0.120	0.8753 ±0.048	0.9015 ±0.001	0.9046 ±0.003	0.9069 ±0.005	0.9072 ±0.008	0.9259 ±0.004
Mineral	0.0 ±0.0	0.4345 ±0.083	0.5577 ±0.150	0.8712 ±0.035	0.9015 ±0.007	0.9044 ±0.003	0.9042 ±0.003	0.9092 ±0.009
Monroe	0.0 ±0.0	0.7350 ±0.040	0.7491 ±0.050	0.9018 ±0.002	0.9029 ±0.002	0.9030 ±0.004	0.9034 ±0.003	0.9138 ±0.004
Morgan	0.0 ±0.0	0.4219 ±0.069	0.5446 ±0.210	0.8426 ±0.048	0.8719 ±0.031	0.9049 ±0.004	0.9073 ±0.006	0.9164 ±0.004
Pendleton	0.0 ±0.0	0.4830 ±0.029	0.9011 ±0.000	0.9056 ±0.003	0.9072 ±0.005	0.9097 ±0.006	0.9112 ±0.008	0.9459 ±0.004
Preston	0.0 ±0.0	0.7889 ±0.035	0.8773 ±0.034	0.9021 ±0.002	0.9025 ±0.001	0.9055 ±0.004	0.9059 ±0.003	0.9324 ±0.005
Tucker	0.0 ±0.0	0.8144 ±0.039	0.8355 ±0.043	0.8994 ±0.005	0.9054 ±0.007	0.9012 ±0.001	0.9047 ±0.004	0.9084 ±0.006

TABLE V
COMPARISON OF RESULTS OF THE ORIGINAL CONSTRAINED FORMULATION WITHOUT ANY REPAIRS, DENOTED AS NR, AND THE APPROACH USING A VIOLATION REPAIR TECHNIQUE REFERRED TO AS R.

Counties	2 (NR) AVG ±SD	2 (R) AVG ±SD	5 (NR) AVG ±SD	5 (R) AVG ±SD	11 (NR) AVG ±SD	11 (R) AVG ±SD
Berkeley	0.5377 ±0.110	0.9299 ±0.007	0.5421 ±0.082	0.9815 ±0.000	0.5003 ±0.050	0.9882 ±0.000
Grant	0.4767 ±0.110	0.9412 ±0.005	0.4952 ±0.094	0.9843 ±0.000	0.4143 ±0.061	0.9888 ±0.000
Hampshire	0.5967 ±0.130	0.3997 ±0.150	0.5786 ±0.110	0.9500 ±0.018	0.5510 ±0.066	0.9527 ±0.040
Hardy	0.5183 ±0.120	0.9292 ±0.007	0.6266 ±0.064	0.9825 ±0.001	0.6435 ±0.053	0.9882 ±0.000
Jefferson	0.7088 ±0.120	0.6246 ±0.140	0.7836 ±0.075	0.9455 ±0.026	0.7374 ±0.082	0.9540 ±0.042
Mineral	0.4988 ±0.099	0.3877 ±0.190	0.5427 ±0.082	0.9492 ±0.021	0.5347 ±0.061	0.9603 ±0.022
Monroe	0.8445 ±0.058	0.5690 ±0.072	0.8408 ±0.064	0.9327 ±0.031	0.8486 ±0.049	0.9728 ±0.012
Morgan	0.5865 ±0.140	0.4059 ±0.240	0.6566 ±0.098	0.9435 ±0.034	0.6156 ±0.100	0.9635 ±0.024
Pendleton	0.4451 ±0.140	0.9370 ±0.004	0.4472 ±0.090	0.9850 ±0.001	0.4003 ±0.088	0.9898 ±0.000
Preston	0.8599 ±0.025	0.7490 ±0.081	0.9442 ±0.009	0.9584 ±0.012	0.9498 ±0.012	0.9776 ±0.006
Tucker	0.9011 ±0.037	0.5806 ±0.150	0.9044 ±0.032	0.9254 ±0.039	0.9278 ±0.024	0.9743 ±0.008

is equal to or even close to $\alpha_{s,h,u}$. For every equality constraint, however, every variable $x_{s,h,u,b}$ is replaced with $\alpha_{s,h,u}\left(x_{s,h,u,b}/\sum_{b=1}^{B_u} x_{s,h,u,b}\right)$. Modifying each BMP element automatically satisfies the respective equality criterion, allowing the AG framework to focus its efforts on the feasible search space.

This study employs a fixed maximum number of generations and hypervolume-based termination criteria to compare various techniques, as outlined in the following section.

Similar to our prior experiment, we utilize several injection points, and the same selection method described previously. We only consider three alternatives: 11 points, 5 points, and 1 point. However, in this experiment, we evaluate the effectiveness of our mending strategy. To this end, we compute the NSGA-III using the original formulation, indicated as no-repair (NR), and our efficient yet simple repair approach, represented as R in our graphics.

Table V compares the HV metric of the original limited formulation with the proposed repair operator. The constraints add extra strain to the optimization process. The results are significantly improved by removing these constraints and allowing NSGA-III to operate within a feasible search space.

Figure 6 displays the non-dominated sets of Berkeley county obtained with and without the repair operator [7]. The repair operator enhances the performance of the NSGA-III. Furthermore, boxplots in Figure 7 show that the repair operator outperforms the same NSGA-II without it. In this instance, the repair operator assists the NSGA-III in reaching the non-dominated set.

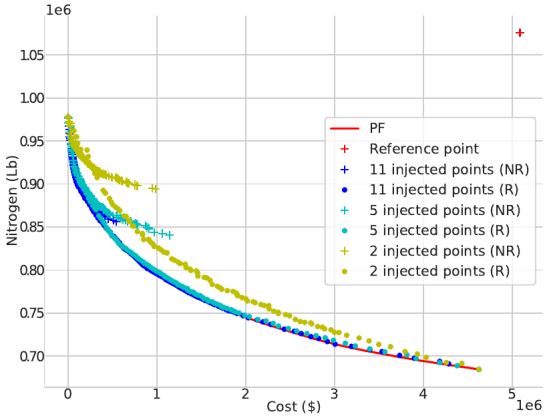

Fig. 6. Comparing the approaches of not repairing (NR) versus utilizing the repair operator (R). The data shows that the repair operator plays a crucial role in achieving a more varied set of points (Berkeley county).

D. Scale-up Study

Our focus now shifts to testing the efficacy of the customized NSGA-III algorithm when applied to optimization problems ranging from a single county to the entire state. In light of this, we have made adjustments to the termination criteria of NSGA-III. The algorithm will now stop either (i) when the maximum generation count of 1,000 is reached or (ii) when a trade-off solution set with 97% HV of the large-front-run is achieved.

Figure 8 shows a general scheme of our scale-up study. We begin by optimizing each county independently. Next, we conduct the scale-up study by increasing the size of the problem. Then, we group at most three counties and optimize the formed group. We kept grouping counties until we formed

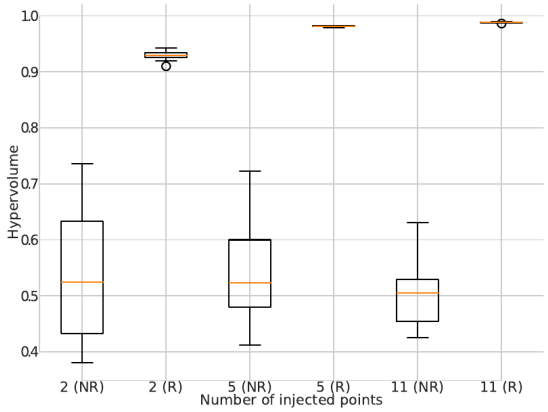

Fig. 7. Boxplots compare the original constrained method (NR) and the violation repair operator (R). Results show that using NSGA-III with the repair operator significantly outperforms the method without repairs (Berkeley county).

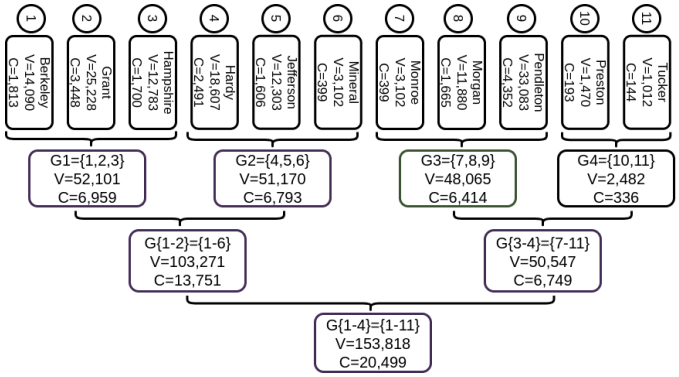

Fig. 8. Clustering scheme for the scale-up study.

a single scenario containing all counties together. In total, we explored four scaled groups. The "individual" group is comprised of each county; the "small-size" groups (G1 to G4) comprised of groups of three counties; the "medium-size" groups (G1-2 and G3-4) are comprised of groups of five or six counties, and the "big-size" group (G1-4) is comprised of all 11 counties with 153,818 variables and 20,509 constraints.

Figure 9 shows that all groups are able to achieve the target HV pf 97% of the HV of their individual group, except for a few outliers.

Meanwhile, Figure 10 illustrates the number of generations required to reach this objective. Notably, the processing effort does not rise monotonically from left to right along the x-axis as the number of variables increases from 1,012 to 153,818.

The findings indicate a noticeable shift in the optimization problem. Initially, the task becomes more difficult as the targeting areas become scarce. However, as more counties are included, a turning point is reached where there are adequate target areas for applying BMPs, resulting in a more manageable problem to solve.

The scale-up study showed that when we combine counties in general, the approach requires fewer generations to achieve the exit criterion (e.g., G1, G3, and G4). However, when we combine Hardy, Jefferson, and Mineral into the G2 group, we see a significant increase in the number of generations. One possible explanation is that Hardy and Jefferson counties are

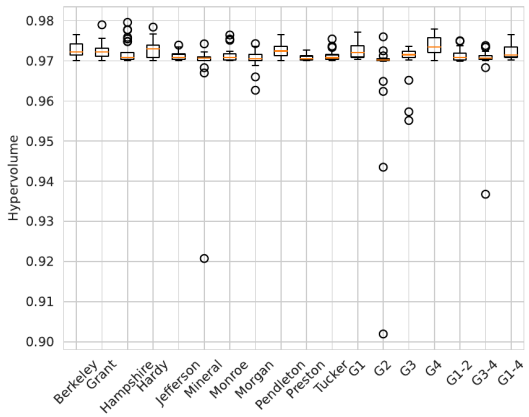

Fig. 9. Obtained hypervolume for different size groups of the scale-up study.

mostly agricultural-dominated. In general, significantly more BMPs can be selected in agricultural areas compared to urban areas. Then when multiple agriculturally-dominated counties are added, the search space is significantly increased, resulting in a higher number of generations to meet the exit criterion.

The demographic factors in Hampshire present challenges in optimizing nitrogen level reduction in an efficient manner. However, combining it with other counties reduces the difficulty and hastens the resolution of the combined issue. Interestingly, this tendency is also observed in other groups.

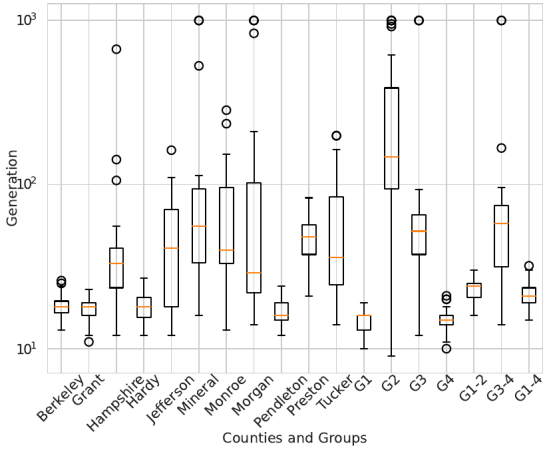

Fig. 10. Termination generation of the NSGA-III in the scale-up study. The algorithm stops at either: (i) 1,000 generations, or (ii) NSGA-III scale-up study termination generation. Stops at: (i) 1,000 generations, or (ii) achieving a 97% HV of that observed in the large-front-run.

Nonetheless, the ability of our proposed NSGA-III method to solve a 153,818 variable problem in an average of 21 generations is outstanding, and it makes our method promising to solve the watershed level problem.

IV. Proposed Hybrid NSGA-III-CAST Approach

The experiments presented in Section III gave us insightful information regarding the EMO components that we have developed. The main findings are: i) The adopted surrogate model provides a computationally fast evaluation procedure that captures most of the CAST interactions (as it presents

a low accuracy error) and preserves the Pareto dominance relation between pairs of solutions. ii) The interior-point-based approach (IPOPT) that our ϵ-constraint uses is highly effective in finding solutions in promissory zones. From our experiments, it was clear that the injection of more solutions produces faster convergence of our NSGA-III procedure. However, such an improvement reduces as the number of points increases (there was little difference between five and 11). iii) The repair approach allows the NSGA-III to focus on finding efficient solutions rather than dealing with constraints. iv) The scale-up study provides valuable information as it is clear that grouping counties together for optimization helps find efficient solutions faster than optimizing them independently.

The above insights let us build up a better evolutionary multi-objective optimization approach. Therefore, we will use the repair approach and the ϵ-constraint method to feed the NSGA-III. Furthermore, as the difference between five and 11 points is negligible, we will keep the five points injection. Finally, after careful assessment, we realize we can still use the surrogate approach and the CAST system. Therefore, we propose a hybrid evaluation, where the surrogate will produce a set of non-dominated solutions and then use the CAST system to fine-tune the values for the objectives. We utilize CAST due to its trustworthy output and the fact that it is based on a physically-based model instead of surrogate models' approximations. Therefore, an additional experiment having hybrid evaluation (surrogate model and CAST) will be performed below.

Figure 11 shows a coarse-grained flowchart of our final algorithm. We first compute n solutions with our ϵ-constraint approach, which uses the surrogate model to evaluate. Then, these solutions are injected into the initial population of the NSGA-III. The NSGA-III uses the surrogate model to evaluate solutions until we reach a certain number of generations. Once the algorithm reaches a number of generations, it selects CAST to evaluate the remaining evaluations. Finally, the algorithm stores the final results and ends.

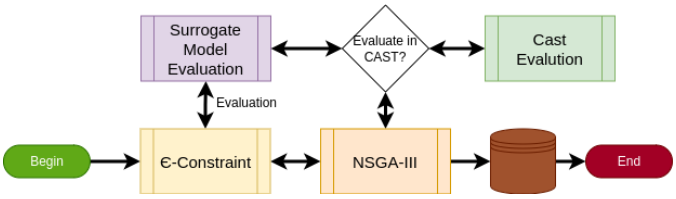

Fig. 11. Proposed customized evolutionary multi-objective approach.

A. *Efficient combination of surrogate and CAST models*

We selected two pairs of West Virginia counties for this experimentation: Berkeley-Mineral, Hardy-Jefferson. The first two counties are urban, while the last two are agricultural.

As the evaluation in CAST is computationally expensive, we reduced the population size to 20, and we selected different configurations where the proposed approach will use the surrogate model and CAST. This is possible since we have observed in Section III-A that the surrogate evaluation of solutions produces a small error compared to their evaluation using CAST.

- 50/50: NSGA-III evaluates 50 generations on the surrogate model and 50 generations using CAST.
- 95/5: NSGA-III evaluates 95 generations using the surrogate model and five generations using CAST.
- 1: We took the output of our ϵ-constraint approach and evaluate it using CAST. This configuration will serve as a control configuration.

To keep a low number of evaluations on CAST, we executed our proposed approach 11 times for each of the previously mentioned configurations. To have better insights regarding the use scenarios of the configurations, we measured the time and computed the convergence according to the ratio of the hypervolume.

Figure 12 displays the ratio of the HV results. From this figure, it is clear that the control execution, which consists of five points obtained by the ϵ-constraint method, does not achieve competitive results regarding the two configurations of the proposed approach. On the other hand, the results obtained by the proposed approach were quite similar with both configurations. When the proposed approach used the 50/50 configuration, it achieved slightly better results for Mineral and Hardy. However, such a difference is questionable, as the medians are similar. Finally, the 95/5 configuration achieved slightly better results on Berkeley and Jefferson counties.

We decided to apply Friedman's χ^2 test to validate our results [12]. The test evaluates the position of the observed values within each group and informs us if there are significant disparities between the groups. The results of this test are shown in Table VI. The results indicate that there is a significant difference between the groups. Therefore, we calculated pairwise comparisons using the Nemenyi post hoc test to identify the different configurations [24]. Figure 13 shows the results obtained by applying the Nemenyi post hoc test to the data. The plot is separated by county. The rows indicate the configuration used, and the columns refer to the compared configuration. The main diagonal of each county is empty. We included the p-value on each square and colored p-values less than 0.05. The results indicate that both proposed approaches' configurations differ from the control. However, there is insufficient information to indicate that both configurations are statistically different.

TABLE VI
FRIEDMAN'S χ^2 ON THE RESULTS FOR THE DIFFERENT CONFIGURATIONS.

Statistics	Berkeley	Mineral	Hardy	Jefferson
χ^2	16.9	17.6	19.9	16.5
p-value	0.0002	0.0001	0.0002	0.0002

Table VII shows the time required by the different configurations. As we can see, when the CAST was used for 50 generations, the approach required around 143 minutes to complete one single execution. The time was greatly diminished with the reduction of the evaluations on CAST. It took around 18 minutes to evaluate five generations on CAST and around three minutes on average for the control execution.

It is observed that the use of 95/5 approach takes about eight times faster, on an average, and produces almost similar

(in some cases, better) performance. Hence, 95/5 approach shows promise for its practical deployment in large-sized CBW problems.

TABLE VII
TIME REQUIRED IN MINUTES BY THE DIFFERENT CONFIGURATIONS: 50/50 - 50 GENERATIONS EVALUATED ON THE SURROGATE MODEL AND THEN 50 GENERATIONS EVALUATED ON CAST. 95/5 - 95 GENERATIONS EVALUATED ON THE SURROGATE MODEL AND THEN 5 GENERATIONS EVALUATED ON CAST. 1 - THIS IS OUR CONTROL CONFIGURATION, WHICH CONTAINS THE SOLUTIONS PROVIDED BY OUR ϵ-CONSTRAINT APPROACH.

	50/50	**95/5**	**1**
Best	138.90	17.53	2.70
Worst	152.96	20.55	3.80
AVG	143.79	18.20	3.18
STD	2.61	0.69	0.33

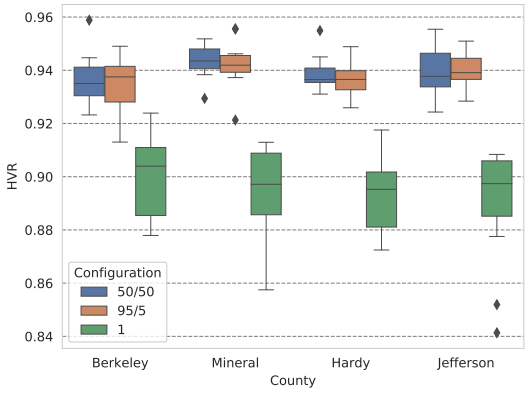

Fig. 12. Ratio of the hypervolume obtained by the different compared configurations: 50/50: 50 generations evaluated on the surrogate model and then 50 generations evaluated on CAST. 95/5: 95 generations were evaluated on the surrogate model, and then five generations were evaluated on CAST. 1: This is our control configuration, which contains the solutions provided by our ϵ-constraint approach.

B. Analysis of Trade-off Solutions

The proposed approach generated three non-dominated solutions from the Pareto front, and Figure 14 displays the spatial allocation of BMPs in West Virginia associated with these solutions. However, it is worth emphasizing that the color assigned to a BMP in an LRS does not indicate its location or degree of implementation. Instead, the color is solely used to signify the proportion of implementation of that particular BMP in relation to all other BMPs implemented within the same LRS.

Solution 1, selected from the top left corner of the Pareto front, represents the BMP implementation scenario characterized by the lowest cost and the highest nitrogen load. Out of hundreds of BMPs that are eligible for implementation, only two were selected (Barnyard Runoff Control and Nutrient Management). However, Solution 1's most prevalent and applied BMPs belong to Nutrient Management. Out of the qualified BMPs, this particular BMP has the lowest unit cost and is responsible for 95% of the nitrogen reduction load. Solution 2 sits in the middle of the Pareto front and is closest to the reference point. The solution is characterized

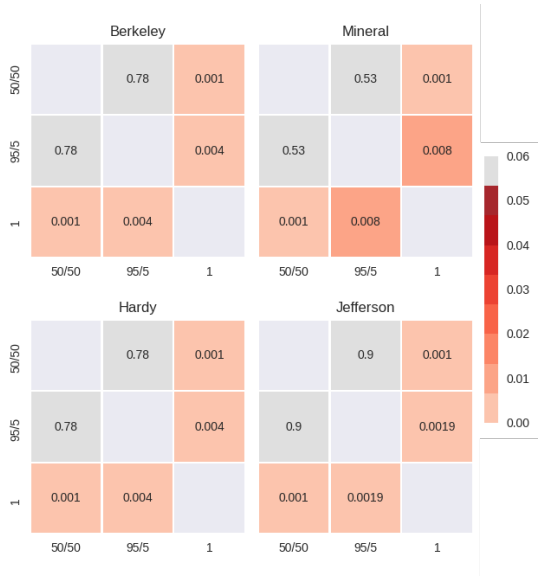

Fig. 13. The Nemenyi post hoc test applied to the hypervolume ratio from different configurations: 50/50: 50 generations on the surrogate model, then 50 on CAST; 95/5: 95 on the surrogate model, then 5 on CAST; and 1: the control with the ϵ-constraint approach, all on CAST. Squares display p-values, with red scaling for p-values < 0.5, indicating a statistical difference.

by the average implementation cost and modest nitrogen load reduction. In general, solutions from this portion of the Pareto front are popular with both stakeholders and producers, as they offer cost-effective watershed management plans. Therefore, a larger group of BMPs is selected to represent this intermediate solution comprised of six BMPs (Agricultural Stormwater Management, Barnyard Runoff Control, Cover Crop, Forest Harvesting Practices, Nutrient Management, and Off-Stream Watering Without Fencing). Similar to the Solution 1 implementation strategy, Nutrient Management was the most dominant BMP and responsible for approximately 94% of the total nitrogen load reduction. Nevertheless, the Off-Stream Watering Without Fencing BMP has been implemented more extensively than other BMPs in several LRSs. Finally, Solution 3 falls within the bottom right corner of the Pareto front. This point represents BMP implementation strategies that are most efficient but cost the most among all strategies. Compared to Solution 2, a lower number of BMPs (four) were chosen, including Agricultural Stormwater Management, Barnyard Runoff Control, Nutrient Management, and Off-Stream Watering Without Fencing. Again, the most dominant BMPs here are Nutrient Management and Off Stream Watering Without Fencing. The BMP Nutrient Management aids in approximately 83% of the total nitrogen load reduction, while the three others assist in reducing around 17% of the total nitrogen load. In summary, Nutrient Management and Off-Stream Watering Without Fencing are the most selected BMPs, while the first can be applied to both developed (e.g., urban) and undeveloped (e.g., agriculture) areas. In contrast, the Off-Stream Watering Without Fencing BMP can only be implemented on pasture-lands.

In regards to the overall cost of implementing BMPs, the cost rose significantly from about $353,883 for Solution 1 to about $999,442 for Solution 3. However, it is worth noting

that under Solution 1, 97% of the total cost was related to Nutrient Management, whereas, in Solution 3, this portion was decreased to 47% for the same BMPs. Meanwhile, for Off-Stream Watering Without Fencing, the distribution of the overall cost was 0%, 42%, and 51% for Solutions 1, 2, and 3, respectively.

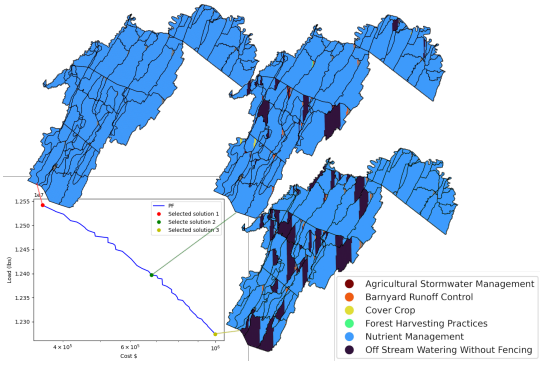

Fig. 14. Example of solutions produced by our proposed approach for the entire West Virginia state.

C. Deployment of NSGA-III-CAST with RESTful API

In a major application area serving thousands of users, it becomes crucial to have code versatility and the software to be readily available. Since the CAST evaluation system is managed by different individuals and hosted in a secured location different from the developers and the location of the customized NSGA-III code, developing our framework became challenging. We created an API to deploy our approach, facilitating new algorithm development and enabling researchers to test, debug, and validate them. The two-level API accesses CAST for scenario evaluation and execution retrieval. Implemented in C++20, it centralizes problem communication and error handling, primarily for internal use.

Our second level of API is to expose our optimization approaches to users and decision-makers. Since this API will be made accessible to users, we adopted a RESTful approach using Django 4.1.1. [9].

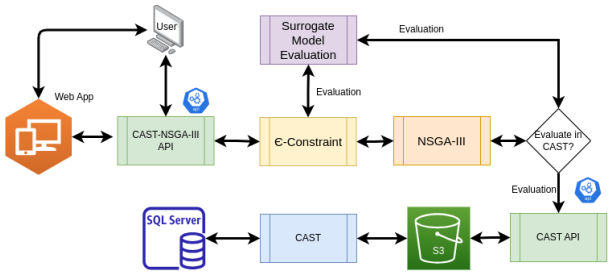

Fig. 15. A flowchart of the proposed approach showing NSGA-III-CAST integration.

RESTful APIs offer a popular, scalable, and reliable solution for web APIs. They ensure a standardized, secure, and easy-to-maintain interaction with web services. Being language-agnostic, RESTful APIs allow developers to use their preferred language, catering to the diverse programming preferences of CBPO research groups.

Figure 15 shows a diagram of our implemented system. It is worth noting that we added Redis to our deployment to access data much faster. Redis is an in-memory data structure store often used as a message broker, cache, and database. It is a popular choice for developers because of its speed, flexibility, and scalability.

In addition, we kept flexibility and scalability in mind while developing our solution. Therefore, for the first API, the so-called CAST-Interaction API, we added RabbitMQ to our structure to make asynchronous calls to the system. RabbitMQ is a messaging software that implements the Advanced Message Queuing Protocol (AMQP). It sends, receives, and processes messages between systems and applications.

Finally, the CAST system stores their evaluations using the Amazon Simple Storage Service (or Amazon S3). Therefore, we implemented an Amazon S3 monitor to gather the scenarios that have been evaluated by CAST. A more detailed architecture of our proposed integration can be found at https://api4opt-dev.chesapeakebay.net/api4opt.

V. Conclusions

Through an extensive study, we have shown that carefully designed customized optimization algorithm components can result in a practical solution to complex problems. Our approach has involved utilizing surrogate models, a biased initialization technique, and a repair operator to efficiently generate sub-optimal scenarios while minimizing computational expenses and guaranteeing the viability of involving 1,012 to 153,818 real-parameter variables exhibiting the effectiveness of the proposed approach in solving practical problems using an evolutionary multi-objective optimization algorithm.

We have developed a customized and computationally effective methodology that can assess scenarios using either the surrogate model or the CAST system, or a combination of both, through state-of-the-art API and storage management techniques. As a result, we have conducted a study to acquire insights into how our methodology performs with each evaluation system. Our investigation has revealed that the approach could leverage the surrogate model for most of the search process, with only a small number of generations in the CAST at the end producing competitive results. Our developed algorithm and corresponding results have provided evidence that customizing the genetic algorithm's components is essential for effectively addressing large-scale problems to be solved with a limited computational time.

To the contrary to general belief, we have also observed for this practical problem that increasing the problem size does not necessarily make the problem more complex to solve. In this specific watershed management problem, including more counties to find a desired environment-friendly solution has been found to be relatively easier than trying to reduce a similar percentage of pollutants from certain individual counties. This observed phenomenon gives us the promise to extend the scope of optimization to multi-state to the whole watershed level involving millions of variables. Nevertheless, this study remains a hallmark optimization study for handling a real-world optimization problem having over 150,000 variables integrating an optimization algorithm with

a real-world computationally expensive evaluation tool and hybridizing with the surrogate-assisted and fast point-based method. Lessons learned from this extensive large-scale study can be extended to tackle other such real-world problems.

The importance of this research comes from the fact that it aims to determine the most cost-effective watershed management plan that improves water quality at the lowest cost. Due to the size of the study area and types of BMPs, thousands of variables needed to be considered. Researchers can evaluate and compare various BMP implementation plans at a large scale through the methods introduced in this study. By finding several alternate solutions to the proposed approach, decision-makers can identify the most suitable BMPs that offer maximum environmental benefits at the lowest cost, making it an invaluable tool for watershed managers. The multi-objective approach allows decision-makers to select the most-preferred solution by considering stakeholders' preferences, such as financial constraints, meeting specific water quality targets, and other logistics-related constraints. Overall, this enables us to make informed and effective decisions when implementing water quality initiatives.

In addition to nitrogen, we plan to incorporate other pollutants like phosphorus and sediments as objective functions in our future work. This broadening of objectives will enhance the complexity of the optimization problem, transitioning it into a many-objective optimization framework. Such a transition may likely introduce specific challenges, including the presence of dominance resistance solutions. Our future research work will address these complexities by developing and refining our methodology with cone dominance and other means.

Acknowledgements

The authors would like to thank the U.S. Environmental Protection Agency for their support through Grant Number 96373801.

References

[1] Ali Ahrari, Julian Blank, Kalyanmoy Deb, and Xianren Li. A proximity-based surrogate-assisted method for simulation-based design optimization of a cylinder head water jacket. *Engineering Optimization*, 53(9):1574–1592, 2021.

[2] Naseem Akhtar, Muhammad Izzuddin Syakir Ishak, Showkat Ahmad Bhawani, and Khalid Umar. Various natural and anthropogenic factors responsible for water quality degradation: A review. *Water*, 13(19), 2021.

[3] Nasrin Alamdari and David J. Sample. A multiobjective simulation-optimization tool for assisting in urban watershed restoration planning. *Journal of Cleaner Production*, 2019.

[4] R. H. Byrd, M. E. Hribar, and J. Nocedal. An interior point algorithm for large scale nonlinear programming. *SIAM Journal on Optimization*, 9:877–900, 1997.

[5] Vira Chankong and Yacov Y. Haimes. *Multiobjective Decision Making: Theory and Methodology*. Dover, 1983.

[6] Kalyanmoy Deb and Deb Kalyanmoy. *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley & Sons, Inc., USA, 2001.

[7] Kalyanmoy Deb and Christie Myburgh. A population-based fast algorithm for a billion-dimensional resource allocation problem with integer variables. *Eur. J. Oper. Res.*, 261:460–474, 2017.

[8] Michael Di Matteo, Holger R. Maier, and Graeme C. Dandy. Many-objective portfolio optimization approach for stormwater management project selection encouraging decision maker buy-in. *Environmental Modelling & Software*, 111:340–355, 2019.

[9] Django Software Foundation. Django.

[10] Feifei Dong, Jincheng Li, Chao Dai, Jie Niu, Yan Chen, Jiacong Huang, and Yong Liu. Understanding robustness in multiscale nutrient abatement: Probabilistic simulation-optimization using Bayesian network emulators. *Journal of Cleaner Production*, 378:134394, 2022.

[11] P.M. Fleming, K. Stephenson, A.S. Collick, and Z.M. Easton. Targeting for nonpoint source pollution reduction: A synthesis of lessons learned, remaining challenges, and emerging opportunities. *Journal of Environmental Management*, 308:114649, 2022.

[12] Milton Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200):675–701, 1937.

[13] Runzhe Geng, Peihong Yin, and Andrew N. Sharpley. A coupled model system to optimize the best management practices for nonpoint source pollution control. *Journal of Cleaner Production*, 2019.

[14] Subhasis Giri, A. Pouyan Nejadhashemi, and Sean A. Woznicki. Evaluation of targeting methods for implementation of best management practices in the saginaw river watershed. *Journal of Environmental Management*, 103:24–40, 2012.

[15] Matthew R. Herman, Amir Pouyan Nejadhashemi, Mohammad Abouali, Juan Sebastian Hernandez-Suarez, Fariborz Daneshvar, Zhen Zhang, Martha C. Anderson, Ali M. Sadeghi, Christopher R. Hain, and Amirreza Sharifi. Evaluating the role of evapotranspiration remote sensing data in improving hydrological modeling predictability. *Journal of Hydrology*, 556:39–49, 2018.

[16] Juan Hernández-Suárez, Gregorio Toscano, Pouyan Nejadhashemi, and Kalyanmoy Deb. Development of an Efficient Optimization Framework for Improving Water Quality in the Chesapeake Bay Watershed, 12 2021.

[17] Juan Sebastian Hernandez-Suarez, Amir Pouyan Nejadhashemi, and Kalyanmoy Deb. A novel multi-objective model calibration method for ecohydrological applications. *Environ. Model. Softw.*, 144:105161, 2021.

[18] Himanshu Jain and Kalyanmoy Deb. An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part ii: Handling constraints and extending to an adaptive approach. *IEEE Transactions on Evolutionary Computation*, 18:602–622, 2014.

[19] Daniel E. Kaufman, Gary W. Shenk, Gopal Bhatt, Kevin W. Asplen, Olivia H. Devereux, Jessica R. Rigelman, J. Hugh Ellis, Benjamin F. Hobbs, Darrell J. Bosch, George L. Van Houtven, Arthur E. McGarity, Lewis C. Linker, and William P. Ball. Supporting cost-effective watershed management strategies for chesapeake bay using a modeling and optimization framework. *Environmental Modelling & Software*, 144:105141, 2021.

[20] David G. Kimmel, William C. Boicourt, James J. Pierson, Michael R. Roman, and Xinsheng Zhang. A comparison of the mesozooplankton response to hypoxia in chesapeake bay and the northern gulf of mexico using the biomass size spectrum. *Journal of Experimental Marine Biology and Ecology*, 381:S65–S73, 2009. Ecological Impacts of Hypoxia on Living Resources.

[21] Jincheng Li, Mengchen Hu, Wenjing Ma, Yong Liu, Feifei Dong, Rui Zou, and Yihui Chen. Optimization and multi-uncertainty analysis of best management practices at the watershed scale: A reliability-level based bayesian network approach. *Journal of Environmental Management*, 331:117280, 2023.

[22] Shaobin Li, Kevin Wallington, Sundar Niroula, and Ximing Cai. A modified response matrix method to approximate swat for computationally intense applications. *Environmental Modelling & Software*, 148:105269, 2022.

[23] Kaisa Miettinen. *Nonlinear Multiobjective Optimization*, volume International Series In Operations Research & Management Science. Kluwer, Boston, 1999.

[24] P.B. Nemenyi. Distribution-Free Multiple Comparisons, 1963.

[25] Anna Raschke, Juan Sebastian Hernandez-Suarez, Amir Pouyan Nejadhashemi, and Kalyanmoy Deb. Multidimensional aspects of sustainable biofuel feedstock production. *Sustainability*, 2021.

[26] The Chesapeake Bay Foundation. What is killing the bay. https://www.cbf.org/how-we-save-the-bay/chesapeake-clean-water-blueprint/what-is-killing-the-bay.html.

[27] Gregorio Toscano, Juan Hernández-Suárez, Julian Blank, Pouyan Nejadhashemi, and Kalyanmoy Deb. Large-scale Multi-objective Optimization for Water Quality in Chesapeake Bay Watershed. In Alessandro Sperduti and Marco Gori General Co-Chairs of IEEE WCCI 2022, editor, *2022 IEEE Congress on Evolutionary Computation (CEC'2022)*, pages 1–10. IEEE Press, Padua, Italy, 07 2022.

[28] Gregorio Toscano, Juan Hernández-Suárez, Julian Blank, Pouyan Nejadhashemi, and Kalyanmoy Deb. Large-scale Multi-objective Optimization for Water Quality in Chesapeake Bay Watershed. In A. Sperduti and M.

Gori, editor, *2022 IEEE Congress on Evol. Computation (CEC'2022)*, pages 1–10. IEEE Press, Padua, Italy, 07 2022.

[29] Gregorio Toscano, Hoda Razavi, A. Pouyan Nejadhashemi, Kalyanmoy Deb, and Lewis Linker. Utilizing innovization to solve large-scale multi-objective chesapeake bay watershed problem. In *2023 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–8, 2023.

[30] Zhenkun Wang, Hui-Ling Zhen, Jingda Deng, Qingfu Zhang, Xijun Li, Mingxuan Yuan, and Jia Zeng. Multiobjective optimization-aided decision-making system for large-scale manufacturing planning. *IEEE Transactions on Cybernetics*, 52(8):8326–8339, 2022.

[31] Lin Yang, Shujiang Pang, Xiaoyan Wang, Yi Du, Jieyun Huang, and Charles S. Melching. Optimal allocation of best management practices based on receiving water capacity constraints. *Agricultural Water Management*, 2021.

[32] Eun Joo Yoon, Bomi Kim, and Dong Kun Lee. Multi-objective planning model for urban greening based on optimization algorithms. *Urban Forestry & Urban Greening*, 2019.

[33] Kun Zhang and Ting Fong May Chui. A comprehensive review of spatial allocation of lid-bmp-gi practices: Strategies and optimization tools. *The Science of the total environment*, 621:915–929, 2018.