AN ABSTRACT OF THE THESIS OF

Kenneth Richard Cannady-Shultz for the degree of Master of Science in Civil Engineering presented on September 5, 2017.

Title: Are Watershed Management Plans Selected and Preferred by Stakeholders Considering Current Climate Conditions Robust against Climate Change Scenarios? A Sensitivity Study of Stakeholders Spatially-Explicit Preferences

Abstract approved: ____

Meghna Babbar-Sebens

In the light of the changing climate, the importance of designing effective watershed management plans that are likely to be implemented is becoming ever more important. This research introduces a new concept, consensus, for incorporation into stakeholder-guided interactive optimization of watershed management plans. User preferences were mathematically simulated based upon scenarios of possible stakeholder attitudes in sub-basins of an agricultural watershed in Indiana, USA, and incorporated into an

existing interactive genetic algorithm (GA) framework. These simulated users along with the watershed hydrologic model were used to evaluate overall preference for and performance of hundreds of different possible distributions of wetlands throughout the Eagle Creek Watershed, weighing cost and environmental concerns on and off of their property. Solutions generated using the interactive GA with the consensus measure performed at least as well as the non-interactively generated baseline solutions, and many out-performed the baseline solutions, with higher peak flow reductions for similar total wetland areas. This result is opposite of what was expected. Previous research has characterized adding stakeholders to the optimization process as a "tradeoff" process, where users sacrifice performance for certain intangible factors. In addition to adding a consensus measure to the interactive GA as an additional objective function, this research also developed a method to select short climate model realizations that best represent extreme flow events arising from climate extremes in the projected future. When the interactively and non-interactively generated solutions were subjected to these extreme climate years, their performance was reduced, even when adjusted for the different magnitudes of expected maximum peak flows. Data issues arising from an interruption to the interactive optimization at generation 30 likely led to some irregularities in the results of this research. Nevertheless, it appears that designing watershed management plans that perform well in the present does not necessarily lead to strong performance in the projected future. Any attempts to address climate change in management plans must do so explicitly.

Are Watershed Management Plans Selected and Preferred by Stakeholders Considering Current Climate Conditions Robust against Climate Change Scenarios? A Sensitivity Study of Stakeholders Spatially-Explicit Preferences

by
Kenneth Richard Cannady-Shultz

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented September 5, 2017
Commencement June 2018

Master of Science thesis of Kenneth Richard Cannady-Shultz presented on September 5, 2017

APPROVED:

Major Professor, representing Civil Engineering

Head of the School of Civil and Construction Engineering

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Kenneth Richard Cannady-Shultz, Author

ACKNOWLEDGEMENTS

First and foremost, I must thank my advisor, Dr. Meghna Babbar-Sebens, for her constant assistance (and prodding). Without her help, I am not certain whether or not I could have finished this thesis. Regardless, her input and experience strongly shaped the work that I completed, and I cannot thank her enough for her support.

Another important group of people I should thank are at Indiana University Purdue University Indianapolis (IUPUI). Dr. Snehasis Mukhopadhyay of the Computer Science department at IUPUI, as well as the many, many graduate students who worked with him on the WRESTORE project, have been invaluable to the development of the optimization algorithm I used in this research. Dr. Mukhopadhyay was also a major contributor to the beginning of my research, and helped give me direction and feedback on my first literature reviews. In a way, it is easy to take all the work that Dr. Mukhopadhyay and his students did to make my research possible. They, along with Dr. Babbar-Sebens, worked tirelessly to develop an algorithm that worked, and could be adapted to a wide variety of problems. One student in particular, Siva Sandeep Kumar Kamadana, revised several fitness functions that the genetic algorithm I worked with used. Thanks to his work, I could just make some small adjustments to his completed code, and incorporate them seamlessly into my work.

Finally, I must acknowledge that none of this would have been possible without the love and support of my family. My wife prodded me endlessly to stay committed and on task. Almost certainly, I would not have completed this thesis without her encouragement and strength. In many ways, she was more dedicated than I. My parents were also invaluable and supportive. They reminded me that I am not the only person to struggle to complete a thesis, and could relate to my difficulties on a personal level. They modelled for me what a successful student looks like – I can only hope that I have successfully followed in their footsteps.

To everyone who contributed to my research, whether in a major way or in passing, thank you. Everyone's support meant a lot to me, and made this thesis possible.

CONTRIBUTION OF AUTHORS

Dr. Meghna Babbar-Sebens was integral for all phases of this research, from inception through to the end. Her assistance was invaluable in altering the algorithms used for this research to include my contributions, providing initial sources for information in the literature, and reviewing all of my findings.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

Page

LIST OF FIGURES

LIST OF FIGURES (CONTINUED)

LIST OF FIGURES (CONTINUED)

LIST OF TABLES

Chapter 1. Introduction

As the climate changes, best management practices for any given location will need to adapt, and thus change as well. Practices that are effective now at managing runoff and preventing high peak flows and pollutant/nutrient loads in streams may not work with similar expected performance fifty years from now. Furthermore, the exact nature of how the climate will change going forward is unknown at this time. In order for watershed planners, in partnership with government agencies, local not-for-profits, and concerned citizens, to manage a watershed effectively and sustainably, they need to know how the changing climate will affect the expected benefits of best management practices going forward. This thesis will address the uncertainty of how best management practices selected for today's conditions will fare in the future.

1.1 Background and Motivation

Earth's climate is ever changing and evolving. The planet cycles through periods of warmth and cold, and has been doing so for millennia without help from humanity. However, recent changes in the climate have been occurring at a rate that likely cannot be explained by natural processes alone; the anthropogenic causes cannot be ignored. "Global Warming," which describes the recent steep trend in global average temperatures, and "Climate Change," which describes changes in the mean or variability of weather patterns or other climatic factors (and is often described as an umbrella term that includes global warming) are occurring now: this is broadly agreed upon (IPCC 2012 & 2014, NASA 2015). The effects of climate change will be felt by people across the world, as food security, fresh water availability, and human health will all be negatively impacted. These impacts will differ from person to person according to their geographical location and socio-economic condition, but no-one will be entirely immune to what climate change will bring (IPCC 2012 & 2014).

Climate Change will likely lead to an increase in extreme weather events. The current functioning of hydrological systems will be altered, and heavy precipitation events will become more common globally. This may also lead to a change in fluvial flooding patterns and timing, though the exact nature of those changes is difficult to predict, and requires detailed modelling of a specific watershed or region to assess (IPCC 2012 & 2014). Usually, the change in flooding pattern includes a shift of the flood starting time and an increase to the magnitude of the flood peak (Simonovic et al 2003). However, there is some evidence that the magnitude of fluvial flooding may be changing. Milly et. al. (2002) demonstrated that the frequency of flood events that have a recurrence interval of 100 years or greater (the recurrence interval is calculated using historic flood prevalence), has been increasing for some time. The authors considered 29 large basins with a history of at least 30 years. Floods with 100-year or greater recurrence interval occurred 21 times within their sample of large basins. One half of the observed floods of any magnitude within the basins occurred after 1953; however, 16 of the 23 observed major floods occurred after 1953. The probability of a stationary process producing this temporal distribution of major floods is 1.3% (2002). Wobus (2013) found that some locations in the United States, such as the Great Basin, may experience an increase of damaging floods (floods that cause monetary damage to persons or) up to 114% more than present values by the year 2100, with the national average increase being approximately 29% (2013). Thus, there is evidence that climate change may also correspond to increased severe flooding.

Flooding can impact the functionality of agricultural land in several ways. Flood water can remove soil and destroy crops, leaving nothing behind for harvest. Even if the crops are not destroyed, flood water or damage may prevent their successful harvest, which functionally leads to the same end results from the point of view of food security. Livestock, equipment, and buildings essential to agriculture can also be damaged or destroyed (USDA 2011). Agricultural insurance is often used by farmers to offset monetary losses from natural disasters, such as flooding. According to the World Bank, the US dollar values of direct agricultural premiums written between 2005 and 2008 increased from $ 8 billion to $18.5 billion, which is an increase of 131% (Iturrioz 2009). During this same time period, the global economic value added for agriculture only grew by 49% (World Bank 2016). By the end of the century, gross flood damage (from all economic sources) is projected to increase by 30% across the United States; specific locations, such as the Upper Colorado River seeing increases as high as 94% (Wobus et al 2013). Combining the trends in premiums paid and flood damage predicted as time progresses into the future suggest that the damage caused by flooding within the agricultural sector will increase. Farmers are purchasing more crop insurance as losses to natural disasters increases. While insurance can offset the economic damage suffered by farmers due to flooding, the food that is washed away by the flood cannot be replaced. In this way, increasing flood damage to crops and livestock present a real and growing threat to global food security.

It is possible to manage watersheds efficiently to mitigate flood risk within its borders. Some flood protection systems, such as the Red River basin system in Manitoba, Canada (Simonovic et al 2003) are currently prepared mitigate increased flood risk associated with climate change, while others, such as an urban flood prevention system in Incheon, Korea (Kang et al 2016), the Taipei Flood Prevention System in Taiwan (Shih et al 2014), and the reservoir system for Salt Lake City in the United States (Goharian et al 2016) will likely not perform adequately under climate change conditions. There is a need to develop methodologies to design and improve current watershed management plans: the general consensus is that the best way to effectively manage watersheds is a method known as Integrated Water Resources Management (IWRM) (Tortajada et al 2003, Castelletti and Soncini-Sessa 2006). Put simply, IWRM is a process where a holistic view of a basin is used in the design effort. IWRM is driven by participation of a diverse collection of interested parties, including both decision makers and stakeholders (Castelletti and Soncini-Sessa 2006). However, IWRM is difficult to implement successfully. In situations where river basins cross political boundaries, or when several smaller basin management areas must be collected into a single large management unit, the process often stalls. Additionally, the participatory framework of IWRM necessitates buy-in from stakeholders, who may resist any plans that negatively impact their interests, even if the overall proposal will meet the stated goals (Tortajada et al 2003). Thus, there is a need to improve IWRM.

Castelletti and Soncini-Sessa (2006) suggests a method called Participatory and Integrated Planning (PIP), which introduces the stakeholder into all phases of the planning process, from the identification of goals through to final negotiations (2006). In the spirit of this new paradigm, where stakeholders are included in the full process, the emerging field of Human Computation provides a unique opportunity to improve engagement of stakeholders in PIP processes via innovative Web 2.0 technologies. WRESTORE is one such recently developed Human Computation technology that expands upon PIP, and includes stakeholders in the optimization process via the Web (Babbar-Sebens et al 2015).

WRESTORE is a joint venture between Oregon State University and Indiana University Purdue University-Indianapolis, and is funded by the NOAA. It is a "web-based, user-friendly, interactive, and participatory decision support system for helping land owners, government agencies, policy makers, planners, and other stakeholders design a distributed system of conservation practices in their watersheds" (WRESTORE 2017). The tool utilizes Genetic Algorithms (Babbar-Sebens et al 2013, Espinoza et al 2005, Goldberg 1989) to generate a variety of possible conservation land use options. WRESTORE currently supports strip cropping, crop rotation, cover crops, filter strips, grassed waterways, no-till, and wetlands as possible conservation land uses. Using the hydrology modelling program SWAT, the algorithm-generated conservation land use plans are subjected to certain climatic conditions, and the effects of implementing those practices on a variety of factors, such as stream peak flows and nutrient loading are generated. The program then selects the best designs based upon the SWAT outputs, and these best options are then evaluated, via a Web-based interface, by users representing a wide variety of interested parties, spanning the spectrum from federal government officials to farmers. The Genetic Algorithm uses the end-user's evaluations to generate new conservations practice layouts to better meet the needs of the user. The end result is a selection of preferred alternatives for each user that can serve as the beginning of a joint effort to identify synergies among stakeholder-generated solutions and manage their watershed effectively (Babbar-Sebens et al 2015).

WRESTORE has currently been developed and tested for a pilot site in Indiana: the Eagle Creek Watershed, outside of Indianapolis (Babbar-Sebens et al 2013 & 2015; Piemonti and Babbar-Sebens 2015). Work by Piemonti et al (2013) showed that including stakeholders in the search process for "optimal" best watershed management plans significantly alters the Pareto front of solutions. The performance of these different solutions do not perform as well as those generated by pure algorithmic optimization. Setbacks between 2-50% for nitrate reduction, 11-98% for peak flow reduction, and 20-77% for sediment reduction were observed. However, these setbacks are better described as tradeoffs for optimizing a non-quantifiable social objective function. These solutions were arrived at based upon the attitudes and desires of the stakeholders, and reflect the space of solutions that stakeholders may accept (2013). Piemonti titled the paper presenting these findings "Optimizing conservation practices in watersheds: Do community preferences matter?" This research seeks to expand upon the answer Piemonti found. When considering the changing climate, and the risks associated with it, how will community preferences alter the performance of optimal best watershed management plans? If these designs are designed solely based upon current climate conditions, how will they fare in the future, when the climate has changed?

1.2 Goals and Objectives

Given the rapidly changing climate, it is imperative to better incorporate stakeholders into the watershed management planning process to improve the likelihood of management plans being adopted and operated. Thus, the goal of this research is to research how stakeholder preferences towards wetlands as a conservation practice affect the flood reduction benefits of management plans. To support this goal, this research has three main goals. The first goal of this research is to determine how the watershed management plans selected by virtual "stakeholders" (algorithmic representations of stakeholders) based solely upon current climate conditions differ from solutions that these "stakeholders" do not favor. Each of these "stakeholders" will have one of two basic bias classes programmed into them: either they are willing to implement the project best management practice (wetlands) within their subcatchment, or they are not. Their rating of different designs will be based on a combination of two different factors – the total area of wetlands, acting as a proxy for cost, and peak flow reductions. These two different factors will be considered at two possible scales: locally (in their assigned subcatchment) and globally (for the entire watershed). A "stakeholder" will be assigned two scaling factors representing the relative importance of each design goal to that person. One will define the relative importance of peak flow reduction versus wetland area minimization; the other will represent the relative impact local and global performance have on their rating. Individual "stakeholders" will evaluate different management scenarios according to their biases and their specific combination of focus factors using current climate conditions. This research will then assess how the designs preferred by these surrogate users change when the users are included as part of the optimization process. Finally, all of the preferred solutions will be subjected to the projected future climate, and their performance assessed. The following specific objectives were completed as part of this project to achieve the three main goals:

- **Objective 1:** Investigate how to create representative ensemble, including no more than five separate, single years from a large collection of climate models that best represents extreme events (peak flows) in the mid-century (for instance, the precipitation data and temperature projections for 2055, which was selected from the MM5I-CCSM-LS model). These climate models years should not only produce high peak flows during the mid-century, but should also come from a specific bias-corrected model that shows a distinct change in the magnitude of extreme events between the past (1970-200) and mid-century that can be explained by some change between those two same time periods.
- **Objective 2**: Determine what the optimal non-interactive algorithmically generated best management practice distributions (watershed management scenarios) will be for current climate conditions.
- **Objective 3:** Determine user ratings for these non-interactively optimized scenarios for several different distributions of virtual users, and find the set of

management plans are most commonly rated highly with the least variability. Each considered user layout will have the same distribution of focus factors, but the biases will vary.

- **Objective 4:** Evaluate how well the set of preferred plans performs when subjected to projected future climate conditions.
- **Objective 5:** Incorporate the user rating functions into the optimization algorithm as an objective function, and generate best management practice distributions (watershed management scenarios) for current climate conditions.
- **Objective 6:** Determine user ratings for these interactively optimized scenarios for several different distributions of virtual users, and find the set of management plans are most commonly rated highly with the least variability using the same techniques used in objective 3.
- **Objective 7:** Evaluate how well the new set of preferred plans performs when subjected to projected future climate conditions.
- **Objective 8:** Compare the sets of preferred plans generated during objectives 3 and 6 for similarities and differences, considering both performance and design characteristics.

1.3 Outline

In this research, an NSGA-II search process was utilized to optimize watershed management plan designs. These designs specified the location and size of wetlands in each of the Eagle Creek watershed's 130 subbasins, and were evaluated according to their respective objective function values for current climate conditions. This research performed two types of algorithmic optimization: interactive and non-interactive. Both processes evaluated the individual generated designs according to their watershed scale peak flow reduction and wetland area objective functions, but the interactive process added input from simulated stakeholders. This input took the form of an accumulated user rating function and a user consensus function. Each set of optimized solutions was subjected to projected future climates that create extreme peak flows, and their behavior in both the projected and current climate conditions were evaluated and compared.

This thesis is split into six chapters. Chapter two provides a broad literature review to provide the reader with the basic background and prior research that this thesis is built upon. It has three main focuses: stakeholder involvement in planning processes, algorithmic optimization, and climate modelling. Chapter three contains the article that proposes and evaluates a proposed method for selecting climate models that result in extreme peak flows. The proposed method is a modification on a technique developed by Lutz et al (2016) for selecting a diverse climate ensemble. The focus of the proposed technique shifts from a precipitation and temperature focus to precipitation and peak flows, and the new method also introduces a novel ordinal approach to address the unique focus of this research. Chapter four details the methods used to incorporate simulated users into the optimization process, and compares and contrasts the response of each set of generated designs to present and projected climate conditions. This chapter introduces a method for simulating the agrarian stakeholders of the Eagle Creek watershed, as well as two new metrics for combining their disparate opinions together into measures of overall opinion and consensus for use in the interactive algorithm. Chapter five synthesizes and combines the conclusions presented in chapters three and four, and suggests further work. Finally, chapter six contains the references used in this thesis.

Chapter 2. Literature Review

This literature review presents the varied topics that provide the foundation for this research. The research focuses on how best management practice designs selected based on today's climate conditions will fare in the future. The platform stakeholders will use to select their preferred distribution of different conservation practices is WRESTORE (Watershed REstoration using Spatio-Temporal Optimization of Resources) tool. The research is a continuation of work done both by Oregon State University (OSU) and Indiana University Purdue University in Indianapolis (IUPUI), under the auspices and supervision of the funding agency, the National Oceanic and Atmospheric Administration (NOAA).

This research has two main components. The first is selecting an appropriate climate model ensemble to represent the changing climate – specifically how mid-century (2040-2070) climate patterns differ from past climate conditions (1970-2000). The second evaluating how different distributions of best management practices favored by virtual users considering only current climate data perform when subjected to the selected climate model realizations. This research utilizes the WRESTORE tool, which is an interactive decision support system (DSS) utilizing Adaptive Interactive Genetic Algorithms (Piemonti 2015). This research requires a review of the literature in the following areas:

- Stakeholder Attitudes and Inclusion in Decision Making
- Algorithmic Optimization
- Climate Model Ensemble Selection

2.1 Stakeholder Attitudes and Inclusion in Decision Making

2.1.1 Farmers' Conservation Attitudes and Behaviors

Factors that explain why certain farmers adopt agricultural practices and others do not, while widely explored, are not well understood at a global scale. Demographic data, such as farmer age, wealth, and education, are inconsistent predictors of conservation practice adoption (Reimer et al 2012). Characteristics of the farm, such as size and terrain, are also unable to consistently predict conservation behaviors. Not even farmer attitudes towards conservation (a feeling of social obligation, self-identifying as a steward of the land, or believing they are not a part of the problem are some example attitudes) were universal predictors. Indeed, it is entirely possible that attitudes in particular are not as important in the decision making process as is often assumed. Of the predictors listed previously, education and farm size are the most consistent, but both show a mixture of positive and negative correlations in the literature. This pattern is typical of most well-studied factors (Knowler and Bradshaw 2007; Ahnstrom et al 2008).

Examples of this inconsistency are abundant in the literature. For example, a study in Sweden found that perceived environmental benefits of the conservation practice both on and off the farm were significant factors in predicting whether or not the farmers would adopt a practice (Soderqvist 2003). This combination of on and off farm environmental considerations was also present in a study performed in the Florida panhandle (Lunne et al 1988). However, only on farm environmental concerns predicted behavior in the Pacific Northwest region of the United States and among US corn farmers generally, and these concerns often included a financial consideration, such as reducing tilling costs with low or no till practices (Luzar and Diagne 1999). Research in Ontario, Canada, as well as in the US Pacific Northwest, represented a middle ground, where farmers were largely concerned with the environmental impact of implementing the practice, but did not include finances as a major factor for consideration (Duff et al 1991, Tosakana et al 2010). Education level was not a significant predictor of conservation behavior in the US Pacific Northwest (Tosakana et al 2010), but was negatively correlated with adoption of conservation practices by farmers in Louisiana (Luzar and Diagne 1999). By contrast, most studies analyzed by Knowler and Bradshaw (2007) showed a positive correlation between education level and the likelihood of adopting a conservation practice (2007). Clearly, predicting farmer behavior from demographic and attitudinal data is not possible at a global scale. Even within the United States, different regions seem to make decisions based upon a widely divergent set of variables. Duff et al (1991) made this point best, when talking about the implications of his findings with respect to government regulations.

> *Implicit in the…discussion has been the notion that, because of differences in farms and farmers, inter-farm differences in soil conservation behavior and effort can be expected. Any government attempt to promote the use of soil-conserving farming methods should recognize these inter-farm differences, something that cannot be accomplished with the status quo of universal policies (1991).*

There has only been one study of the conservation behaviors and attitudes of farmers in the Eagle Creek Watershed to date. Reimer, Thompson, and Prokopy (2012) explored the relationship between conservation behaviors and qualitative socio-demographic, attitudinal, and awareness variables. They interviewed 32 participants in person, and led them through a lightly-structured interview that touched on behaviors and beliefs. Three attitudes were explored: farm as a business, off-farm environmental benefits, and stewardship. In general, those whose primary attitude towards conservation was that the farm is a business first and foremost, with all other concerns being secondary, did not implement many conservation practices. Those concerned more strongly with externalities, such as greater watershed health and environmental stewardship, generally implemented more practices. Furthermore, those who self-identified as stewards of the land rarely exhibited any of the other attitudinal dimensions, or mentioned them as secondary motivators. Those concerned with off-farm benefits often expressed an internal balance they strike between cost and perceived benefit. Approximately 60% of those farmers who identified off-farm benefits or stewardship as a motivator of their conservation behaviors were high adopters, while only 24% of individuals who only identified financial motivations were high adopters. Any conservation practices adopted by financially-motivated farmers were typically justified as a cost-savings measure (2012).

2.1.2 Participatory and Integrated Planning Procedures (PIP)

Attitudes are a significant contributor to stakeholder reluctance to adopt conservation practices, but another factor also influences their decision making processes. If stakeholders, such as farmers, believe that their judgement and experiential knowledge is being ignored, and their decision processes bypassed during the creation of a best management practice plan, they will resist implementing that plan, no matter how well conceived, explained, or technically sound it is. This has been a historical weakness of previous agricultural decision support systems (DSS), as well as watershed management plan optimization. Focus was given to having a technically sound and sophisticated model. DSS existed to apply rational and logical analysis to complicated problems with deep uncertainty, and provide the end user with a recommended course of action. The end user (the stakeholder) is required to trust the judgement of a "black box" type tool – regardless of the effort a DSS designer may put into stakeholder education or persuading them that the tool is legitimate, the adoption of a DSS relies upon the ability of a stakeholder to accept the judgment of the tool as better than their own. Stakeholders often have extensive practical knowledge at their disposal, and are understandably resistant to having it ignored (Babbar-Sebens 2015, Cox 1996, McCown 2002).

As a remedy to the previously mentioned issue, Castelletti and Soncini-Sessa (2006) proposed a new planning procedure for integrated water resources management problems. They termed the new procedure Participatory and Integrated Planning (PIP), and the goal of PIP is to create a framework where scientists and policy makers can work together to define problems, settle on methodologies for design and modelling, and create solutions (2006). PIP is "participatory" in that it requires stakeholder involvement in every phase of the planning process, not just during a final review phase. The stakeholder moves from being a person who is asked for feedback, and possibly also for information, to being a central part of the decision-making process from its inception to final implementation. "Integrated" means that all possible effects of the plans, both positive and negative, must be articulated and quantified (if quantifiable). Policy decisions for all parts of the management solution, including reservoir policies, must be analyzed to determine a complete set of positive and negative impacts. This is not a simple task, as many impacts may not be quantifiable, or are not well understood. These impacts may either be simplified, or, after much deliberation and discussion, ignored in the modelling, though they should still be included in the evaluation in any way that is feasible (Castelletti & Soncini-Sessa 2007).

PIP is a nine phase procedure (with an additional step for scoping prior to the rest of the process), which includes several recursions and loops to allow for the incorporation of new data or opinions that arise as the result of analysis or discussion. The phases are described as follows (Castelletti & Soncini-Sessa 2007):

Step 0: Preliminary Activities and Objectives. This step defines the objectives and stakeholders for the project. Once stakeholders are identified, their opinions should be solicited to further identify objectives, project scope, and stakeholders. The process for decision making going forward also needs to be addressed.

Step 1: Identification of Actions. This step identifies any possible actions and measures that could be incorporated into a final plan, with special care given to ensuring that stakeholder-favored and preferred options are considered. A key for successfully implementing this step is making sure that stakeholders believe their concerns and desires are taken seriously. The objective of this step is not to evaluate which options will serve the final objective the best, but rather brainstorm alternatives.

Step 2: Identification of Criteria and Indicators. This step identifies criteria that will be used to evaluate different alternatives going forward. Special care must be given to incorporating stakeholders in this identification. In many cases, the stakeholders may have particular skill or expertise that may directly influence the evaluation process, and not considering them could limit the usefulness of the ensuing design process.

Step 3: Model Identification. During this step, all interested parties should define what characteristics the model used to represent the watershed should be. The type and extent of the model will strongly depend upon the evaluation criteria and considered actions. In many cases, the modelling will need to occur in two steps. During the first step, a parsimonious model will be used during the design, then screen out unacceptable alternatives. After a set of alternatives has been designed and selected, a more complicated model can be used to evaluate their performance.

Step 4: Design of the Alternatives. During this step, all possible alternatives are analyzed. As opposed to relying upon expert opinion and stakeholder preferences to suggest alternatives, the design process should be an exhaustive search of the decision space defined by the actions available. The solutions selected must be Pareto optimal, meaning all dominated solutions are removed from the set of alternatives. Due to the considerable size of the decision space this method requires, a screening model should be used to quickly analyze the full set of considered solutions and remove dominated solutions.

Step 5: Estimate the Effects. Once a Pareto optimal set of solutions has been selected, they must be evaluated based upon the values of the different indicators. If the system is not dynamic, this step often occurs simultaneously with step 4. For models with a sizable time horizon, the effects may need to be aggregated in some fashion. An appropriate scenario to use in the estimation of the effects is a historic scenario. In the case of a historic scenario, the actual behavior of the system is known, so the difference between the alternatives and the current state of affairs can be directly calculated. Using actual data for comparison is also preferred by most stakeholders.

Step 6: Evaluation of the Alternatives. This step translates the estimates of the indicators for each alternative to the value stakeholders see in the alternatives. Each alternative is evaluated according to the degree of satisfaction stakeholders associate with the design. Often, this degree of satisfaction is represented on a continuous scale from zero to one to allow for direct comparison to other alternatives. In the case where there is only one decision maker and one stakeholder, this step results in the determination of the preferred alternative, and the process ends. Otherwise, the process continues to step seven.

Step 7: Comparison and Negotiation of the Alternatives. The goal of this step is to find a solution that no stakeholder is opposed to. If that is not possible, then the goal becomes to come to a consensus where the fewest stakeholders are opposed to it. To facilitate the process of consensus building, negotiation methods appropriate for the specific combination of decision makers and stakeholders involved should be employed. If a consensus solution can be found where no stakeholder is inconvenienced, or no accommodations can be reasonably made to reduce the negative impact of the consensus alternatives on the stakeholders opposing them, then the process progresses directly to step nine. Otherwise, the process proceeds to step eight.

Step 8: Mitigation and Compensation. In the case where there is a broad consensus among the majority, but not all, of the stakeholders, then PIP dictates that either compensation or mitigation of the negative effects must be offered to the stakeholders opposing the consensus. In order to achieve this, it may be necessary to identify new types of actions that act specifically to benefit the stakeholders whose interests are not being fully met. In that case, it is necessary to return to step one, and repeat the process with the new actions considered. It may be possible to add these new actions directly to existing solutions, in which case all that needs to occur again is estimation and evaluation of the alternatives (step 5 and 6), leading to a new round of negotiation (step 7).

Step 9: Political Choice. This final step is where the decision makers select the best compromise solutions from the set of consensus alternatives generated in step seven. While this deliberation among decision makers may take a form similar to the deliberation that occurred in step seven, it will not necessarily be the case.

The main purpose of PIP processes is to include stakeholder opinions in all phases of design, from scoping through to the selection of compromise alternatives. While the decision makers will make the final determination of which alternative is finally implemented, the process guarantees that whatever alternative is selected enjoys broad consensus among the interested stakeholders. As an added benefit, this process forces stakeholders and decision makers to work together and learn about one another. Stakeholders learn about each other and the decision makers, and also become more aware of how their watershed functions. As the PIP process proceeds, the stakeholders undergo continuous social learning, and are better able to participate in the process constructively (Castelletti & Soncini-Sessa 2007).

2.1.3 The WRESTORE Tool

WRESTORE currently allows for direct interaction between stakeholders and the optimization process, which represents an expansion upon the PIP framework described previously. Whereas the PIP procedure allows for stakeholders to evaluate alternatives after they have been created, WRESTORE incorporates the biases, needs, and preferences of stakeholders in the watershed into the optimization process as an additional indicator to optimize over (Piemonti 2013; Babbar-Sebens 2015; Babbar-Sebens 2014). This process, incorporating the outcomes of the Interactive Genetic Algorithm (IGA) with user preferences, allows for optimization to occur in a solution space that suits the user's subjective criteria that the IGA cannot directly quantify (Babbar-Sebens 2014).

The WRESTORE tool provides a platform for multiple stakeholders, such as government officials, land owners, or watershed planners to participate in a joint effort to select a distribution of conservation practices that best suits their collective needs (WRESTORE). The program currently includes seven different conservation practice land uses, which are: strip cropping, crop rotation, cover crops, filter strips, grassed waterways, no-till, and wetlands. These practices are designed to manage a variety of runoff quantity and quality parameters, such as peak flows, sediment concentrations, and nitrate concentrations. Additionally, the program considers the socio-economic impacts of the selected conservation practice distribution (Babbar-Sebens 2015). Finally, WRESTORE incorporates user preferences as a fitness function, allowing for optimization to occur in a solution space that includes the user's subjective criteria, which the IGA cannot directly quantify (Babbar-Sebens 2014). The biases of the stakeholders strongly influence the characteristics and performance of the alternatives generated. For instance, a stakeholder who is largely driven by economic concerns will force the IGA to develop more solutions that perform well economically, but this performance will often be achieved at the cost of other objective function performances (Piemonti 2014). Current experiments have restricted the number of conservation practices available to WRESTORE for the purpose of reducing experimental complexity. Research completed by Walters and Babbar-Sebens in 2016 focused on wetland distributions; Piemonti et al focused their research on cover crops and filter strips (2015). This research will focus on wetlands exclusively.

Wetlands are capable of providing numerous services to watersheds, including hydrologic, environmental, and ecological. Wetlands can buffer peak flows, thus reducing flooding. They can also remove nutrients via uptake and filtration, and serve as part of the Nitrogen and Phosphorus cycles. Wetlands supply vital habitat for riparian species and birds, and also provide food for local fauna (Zedler, 2003). Runoff from livestock farms contains elevated levels of sediment, Nitrogen, Phosphorus, biological oxygen demand (BOD), and pathogens. These constituents are removed from runoff by wetlands with varying degrees of effectiveness. In Alabama, a study found that 65% of BOD is removed and 53% of total suspended solids (TSS) are filtered out by wetlands, while Nitrogen and Phosphorus are reduced by approximately 40%. (Knight et. al., 2000). In Midwestern tile drained fields, Dinnes and others (2002) observed significant variation in Nitrate removal. Removal rates varied from 8% to over 95%. The difference between these different removal rates was highly dependent on a variety of factors. The level of bioavailable carbon, temperature, and loading rate all affected the removal rate of Nitrates. Generally, wetlands can remove almost all Nitrates from runoff when runoff volumes are low, but the removal efficiencies drop quickly as volumetric loading rates increase (Dinnes et. al., 2002).

The underlying methodology for designing conservation practice distributions is based upon work by Babbar and Minsker (2006). The process they developed, the Interactive Genetic Algorithm with Mixed Initiative Interaction (IGAMII), features a sequence of human (human DM) and computer (simulated DM) guided optimization. The MII portion of the IGAMII algorithm determines when to use human or computational resources in the design process, a central tenant to human computation (Fraternali et al 2012). Each optimization session is ended with an introspection session, where the human user can review the ranks assigned to specific alternatives by either the human or simulated DM, and make any changes to the ratings or confidences they see fit. The transition from one DM to another is based upon trends in the variance and mean of the user's confidence in their selection. If the variance of the confidence rating is falling while the mean is increasing, the algorithm will switch who has the initiative for that optimization session (Babbar and Minsker 2006; Babbar-Sebens et al 2015; Piemonti 2014 & 2015). The latest WRESTORE research introduced Adaptive Interactive Genetic Algorithms to the optimization process utilized by the program. These modified IGA's improved the convergence speed to optimal decisions for users with well-defined objective functions (consistent criteria for rating alternatives), but did not improve convergence speed for stochastic (inconsistent criteria) users. The research team will pursue potential improvements to the Adaptive IGA, but this has not yet happened to date (Piemonti 2015).

Figure 2 is a schematic of the WRESTORE program architecture. The program has four main components: The Web Server, the Database Server, the Program Server, and the High Performance Computing (HPC) Infrastructure. The web server supports the WRESTORE site, and is what the users interact with and manipulate. The web server sends this information to the database server, which stores the user profile and real-time experiment run data. This database relays information from the web server to the program server, and also relays program information back from the program server to the web server. The

WRESTORE program server itself has several components. These components are all written in Java to carry out the various functions of the WRESTORE program (Babbar-Sebens et al 2015).

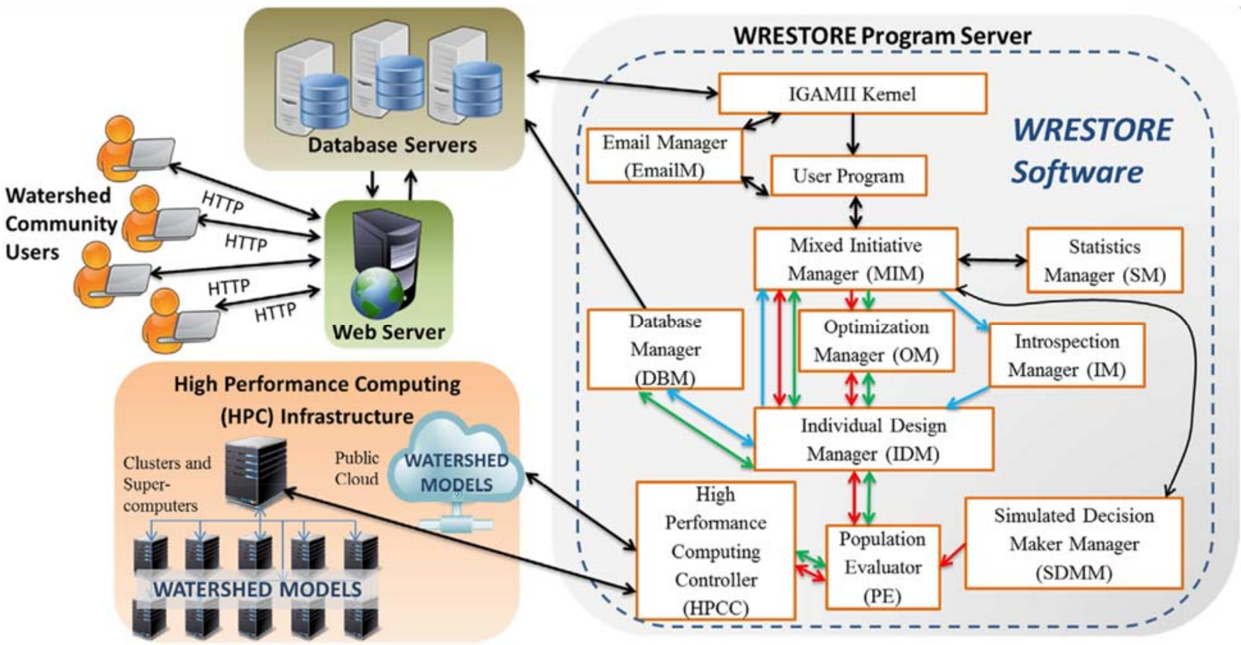

Figure 1 Schematic of WRESTORE Architecture (Babbar-Sebens et al 2015)

The following list is a brief overview of the specific functions of each software package or manager (Babbar-Sebens et al 2015):

1. IGAMII Kernel: Starts or stops real-time search experiments for any users registered to use WRESTORE.
2. User Program: Associates the user with the new experiment and allocates resources and software packages for the user.

3. Email Manager (Email M): Sends email notifications to users when session data is ready for review.
4. Mixed Initiative Manager (MIM): Manages the mixed initiative interaction portion of the IGAMII framework.
5. Statistics Manager (SM): Conducts all statistical tests required for the WRESTORE program to run.
6. Optimization Manager (OM): Manages the optimization algorithm used by WRESTORE.
7. Introspection Manager (IM: Manages introspection sessions.
8. Individual Design Manager (IDM): Communicates the details of each individual alternative to the other managers.
9. Simulated Decision Maker Manager (SDMM): Trains and tests various learning algorithms that simulate the human user, then selects the best available model to become the simulated DM for that session.
10. Population Evaluator (PE): Evaluates alternatives according to objective functions (currently, there are four objective functions: the Cost-Revenue Function, the Peak Flow Reduction Function, the Sediment Reduction Function, and the Nitrates Reduction Function), utilizing the High Performance Computing Infrastructure.
11. High Performance Computing Controller (HPCC): Connects the WRESTORE program to the High Performance Computing Infrastructure to reduce program run time.

12. Database Manager (DBM): Collects all data from previous software packages and managers, and relays them back to the database for display via the web server.

The workflow of the WRESTORE program is currently as follows. First, the user logs into the WRESTORE web portal, and selects appropriate program options for the current optimization session. Once the user submits the selected options, the database receives this information from the web server; the database sends the user data to the IGAMII Kernel, which initializes the WRESTORE program components mentioned previously under the IGAMII description. Once the program is initialized, the user goes through several interactive sessions; the very beginning of each new experiment begins with an introspection session, where the user reviews alternatives found either during a previous interactive optimization run, or a non-interactive optimization run if there are no previous interactive optimization runs in the database. After the introspection session, the MIM utilizes the SM to calculate statistics that will determine whether to the human or simulated user will have the initiative for the upcoming search session. The search session begins with the program generating a population of potential optimal solutions, then using either the human or simulated user to rate these solutions. The user cycles through several generations of solutions until it is time for another introspection session. The cycle detailed previously (determination of initiative to guided optimization to introspection) continued for a predetermined number of introspection phases, then the experiment ends (Babbar-Sebens et al 2015).

2.2 Algorithmic Optimization

2.2.1 Watershed Management Plan Optimization Techniques

Most water resources optimization problems involve the optimization of several variables at once, where no individual variable is considered more important than the others. Often, there is uncertainty in the values some of these variables, such as peak flows or precipitation, will adopt at any given time, which adds a layer of complexity to the problem. For decades, most optimization techniques used to optimize water resource designs utilized one objective function and several constraints. This has an obvious weakness, as most water resources problems involve several objective functions, and they all need to be optimized concurrently. Vectorization of the objective functions allows for simultaneous optimization to occur, and the outcome of that optimization is usually a set of non-inferior (non-dominated) solutions. However, the equations created often have a sizable set of decisions and considerations, which must be reduced in order for the problem to be tractable. Great care must be taken during this process to ensure that the outcome of the simplified problem will not deviate substantially from the outcome of the original problem (Haimes & Hall 1974).

Several approaches are possible when considering vectorized objective functions. One possible solution is applying economic pricing theory. Any non-commensurable objectives must be monetized in some fashion, which allows for the whole problem to be treated like a pseudo free-market using economic theory. In practice, this has not been successful, and most attempted monetizations of essential environmental considerations have not been successful. Another method uses expert judgement to assign an appropriate scale of relative importance the set of objectives. This method cannot account for dynamic interactions between different objectives, where the relative value of one value of an objective is strongly dependent upon the value of another objective. Both of the previously mentioned methods fail to recognize a simple truth about the nature of water resources optimization variables: relative values are often more important than absolute values (Haimes & Hall 1974).

An example of optimization utilizing relative values as opposed to absolute values is the surrogate worth tradeoff method. Tradeoff functions comparing the marginal improvement or detriment of one variable that results from a unit change of another variable, are optimized. Surrogate worth functions that relate the value of these tradeoff functions to the actual utility perceived by the stakeholder are applied to these tradeoff functions are constructed and utilized in the function space, then undergo transformation to the decision space. The assumption behind the decision to apply surrogate worth functions in the objective space is simple: stakeholders and decision makers care more about the values of the variables used to create and evaluate the alternatives more than the specifics of what the decision is. In this way, the stakeholder or decision maker is allowed to make subjective judgements using a familiar set of variables (Haimes & Hall 1974).

Transforming performance measures to utility is an essential feature of many optimization methods currently, but creating an optimization model that adequately explores many tradeoffs at once is still challenging. Maxted and others (2009) suggested a different method: defining the total project utility as the product of several different variable-specific utility functions, and using a generalized reduced gradient algorithm to find the options with the greatest utility. This bears a close resemblance to the single objective function method that Haimes and Hall moved away from in their 1974 study. For a non-point source (NPS) pollution reduction problem in Wisconsin, the authors considered three main utility factors, all scaled from zero to one: the marginal gain in statistical power (the reduction of uncertainty in the effectiveness of the NPS pollution control plan) resulting from the addition of the watersheds to the plan, the watershed area, and pollution reduction gained per unit expenditure (management effort). The decision variable in this context was whether or not a specific watershed would join the Wisconsin Buffer Initiative (WBI). The WBI is focused on reducing phosphorus NPS from specific watersheds by implementing active watershed management methods in that watershed. The goal of the research was to determine what characteristics the optimal set of actively managed watersheds would have for different program budgets (Maxted et al 2009).

The results of the study found that as program budgets increase, the number of watersheds included in the program increases most quickly, followed by watershed size, then finally by management effort. However, the authors warn us that this solution is far from a panacea, and will likely vary from one watershed to another (Maxted et al 2009). This study is only meant to explore another method for optimization, and add it to the list of possible approaches to solving a watershed optimization problem. The solution is unique to the area of study, but the method should be transferrable. This concept relates back to the concept explored in Section 2.1.1: each watershed is different, and a single, unified watershed management plan is neither present nor desirable.

Another optimization tool that is often applied to watershed management plan design is the evolutionary algorithm, also known as a genetic algorithm (GA). WRESTORE uses a GA in its optimization routine. The solutions generated by the GA are analyzed by the hydrologic program SWAT, and the results of the analysis are used in the GA to calculate the fitness function values used in the optimization (Babbar-Sebens 2015). Several authors have explored how to used GA's and SWAT concurrently to optimize watershed management plan designs. Arabi et al (2006) utilized a GA to perform a spatial search procedure to determine a set of optimal sediment and nutrient control plans for the Black Creek watershed in Indiana. They considered two objectives: economic cost and environmental benefit. The two objectives were combined into a singular objective function that the GA maximized: the ratio of pollutant load reduction to cost. Regulatory requirements for pollutant loadings and project budget were included as constraints. The main goal of the study was to examine the behavior of different parts of the GA. A sensitivity analysis where the GA population size and replacement rates were altered found that the algorithm performed best with smaller population sizes and replacement rates. This was likely the result of the convergence of the objective function to an approximately optimal value once one of the constraints was consistently met. The management plans selected by the GA were also typically more effective and cost efficient than plans that were either randomly generated or created via targeting strategies. This study showed that the use of GA's in watershed management plan optimization was appropriate, and superior to several other methods available for use (2006).

Kaini et al (2012) performed a similar analysis in the Kaskaskia watershed in Illinois. This approach modified the previous approach used by Arabi et al in 2006. The ratio objective function representing a cost to benefit ratio was replaced with a simple cost function, and the "benefits" (the environmental gains) were added as constraints. This optimization was structured such that the user could specify a desired reduction of the different nutrient and sediment loadings, and solve for a set of optimal solutions based upon those goals. The GA structure was also altered. The number of generations became dynamic: the algorithm terminates when the change in the search process (the solutions found) from one generation to the next is negligible. In general, the final solutions employed relatively few low cost best management practices (BMPs) in each subcatchment (2012). When viewed as a set, the studies by Kaini et al and Arabi et al represent optimization with two different users in mind. Kaini's work focuses on a group of users who have a fixed budget, and want to do the most with it they possibly can. By contrast, the users Arabi considers do not have budgetary constraints, but are rather looking for specific performance characteristics from the system. In reality, most stakeholders cannot be placed simply in one of these groups, limiting the usefulness of either GA model.

Artita et al (2008) suggested a refinement upon the GA and SWAT optimization method. The GA was modified to include speciation, where the decision space is divided into several subpopulations, or “species.” This modified GA was termed a Species Conserving Genetic Algorithm (SCGA). The overall population is fixed in an SCGA, but the species are not. The overall structure of the SCGA is as follows (2008):

1. The current generation is sorted in decreasing order of fitness. Each individual is tested to see if it is a species seed by testing whether or not the individual is within the speciation distance (determined a priori to the analysis) from an existing species seed (by default, the most fit individual is a species seed). If the individual is far enough away from the other species seeds, it is added to the set of species seeds.
2. Following the creation of the next generation via selection, crossover, and mutation, the diversity of the new generation is preserved by allowing for species seeds to be directly copied into the next generation. This is done by identifying all members of the species centered on a specific seed, then replacing the least fit individual with the seed. If no members of the species survived into the next generation, the least fit individual from that generation is replaced by the seed.
3. The “best” solutions are selected from the set of species seeds in the current generation. Any seeds that are within a set percentage of the maximum fitness function value are added to this set, and considered the “best design alternatives.”

The case study considered in the research by Artita et al (2008) was in the Silver Creek watershed in Illinois. They considered any solutions that were within 25% of the maximum objective function value after optimization was complete. The single objective function considered was cost, which was minimized subject to the following constraints: the watershed water balance, a maximum detention pond size, land use, peak flow and sediment reduction criteria, and BMP placement requirements. The set of possible management practices considered was detention ponds, infiltration ponds, field borders, grade stabilization, and grassed waterways. From a population of 200 individuals, the SCGA selected 12 best design alternatives. In general, these designs did not include costly ponds, and focused on grade stabilization and field borders. The study noted that the quality and quantity of selected alternatives are strongly related to factors such as modelled BMP's, SWAT model resolution, and the different SCGA characteristics selected (2008). This methodology was not directly compared to other GA structures, so it is difficult to assess whether or not this new type of GA represents an improvement on more traditional algorithms.

2.2.2 Interactive Algorithmic Optimization

The previously considered watershed optimization methods considered human preferences in a passive way, but their input was not actively sought as part of the search process. By contrast, WRESTORE falls into the broad category of human computation, where humans are explicitly included in the search process. A system is considered to be in the category of human computation when the human collaboration is facilitated by the system itself as opposed to by human initiative. The defining characteristic of a human computation system is the delegation of tasks between humans and computers that play to their respective strengths. The computational portion of the system takes care of any computational requirements, as well as splitting tasks, coordinating the system's components, and compiling and communicating results. The human portion of the system provides intuition and decision-making capabilities. The two components of the system, the computer and the human, complement each other: one provides what the other cannot. Human computation approaches can be broadly categorized into four groups, as detailed below (Fraternali et al 2012):

- Crowdsourcing: Distributes individual work tasks among a community of users. Crowdsourcing platforms usually have two types of users: work providers and work performers. The platform provides the medium for providers to find performers, and can manage workflow from inception to final implementation and payment. If necessary, the crowdsourcing platform can even split large tasks into smaller sub-tasks, which can be addressed individually.
- Games with a Purpose (GWAPs): Distills complex problems into a game-like format, so that users will solve these problems without realizing they are actually performing work. This approach capitalizes on the recent ubiquitination of computer games. The goal is to present the problem in an enjoyable and engaging form in order to make use of the critical reasoning capabilities of the general public in an unobtrusive fashion.
- Social Mobilization: Similar to Crowdsourcing, but with a time constraint. The focus of this modality is maximizing the efficiency of task spreading in such a way as to increase the speed of solution convergence.
- Human Sensors: Utilizes the widespread distribution of sensors that people carry around on a daily basis. With the increased distribution of mobile devices, the number of roving sensors available has increased greatly in recent times.

Human computation systems can also be classified according to two different human dimensions: humans involved and human faculties involved. Humans can be involved in different scales: the system may only support one user at a time, may support a small group of users, or may support a large group of users. These groups can either be closed or open; closed groups tend to be smaller, while open groups are typically larger. Human computation systems can make use of three different human capabilities as well. The system may utilize the emotional and/or perceptual capacity of human users, may take advantage of their judgment skills, or may encourage social interaction between users (Fraternali et al 2012).

In addition to the human dimension categorizations, the human computation tasks can be categorized according to the type of activity, the control of workflow and task allocation, the motivation mechanism, and the time requirements. These dimensions are presented in more detail below (Fraternali et al 2012):

- Type of Activity: Can either be a game or a task. A game provides an engaging challenge to a user, which produces a solution to a problem as a collateral outcome; a task is clearly work, and has well defined inputs and outputs. These tasks can either be singular and discrete, a composite of simple tasks, or implicit, where the user performs useful work without realizing it.
- Control: an activity can either be controlled in a central way or a distributed way. A third option is utilizing human computation to select the most appropriate method for activity control. Centralized controls place all components of the human computation system under the control of the system itself and, by extension, on the originator of the task. Distributed controls place the responsibility for recruiting people with appropriate skills on the current workers. Typically, this is a recursive process that incentivizes recruitment. Selection of the best control strategy for a particular task may itself be determined by a human computation system.

- Motivation: people may participate in a human computation system for a variety of reasons. Users may participate purely for enjoyment, as in a GWAP. Some users choose to participate for philanthropic reasons, while others look for monetary gain.
- Time: the human computation system may be solving a problem that is time sensitive. Individual tasks within the problem may be time sensitive as well. This dimension often drives the structure of the human computation system.

The WRESTORE tool is a task-oriented crowdsourcing application. The method it uses to coordinate the human user with the human computation system is broadly considered human-guided search. Human-guided search, often termed "human-in-the-loop optimization," includes a human interaction component to an optimization algorithm. Including human users in the optimization process has two benefits: people better trust optimized solutions they help generate, and people's skills in some areas actually exceed those of a computer, improving both the speed of convergence to optimality and the quality of the optimized solutions found (Klau et al 2009).

Work performed by Klau et al (2009) explored the concept of human guidance of optimization algorithms. They focused on developing a human guided heuristic algorithm, specifically a guided tabu search. The users could assign mobilities to individual elements of the structure being optimized. The article focused upon one task optimized with the guided tabu search: the packing problem. Users could set individual boxes to low mobility (the algorithm could not move them in subsequent trials for a fixed period of time), could overwrite algorithmically selected moves, and could allow the algorithm to run for a certain period of time. In the last case, when the user ended the automatic optimization process, they would be shown two outcomes – an optimal solution from a previous iteration, or the move being considered currently. The user could specify either alternative, and continue the guided algorithm from there (2009).

Including people in the algorithmic process greatly improved both the quality of the final optimal solution, as well as the time required to converge to this optimal solution. The experiment used two versions of the tabu search algorithm: one was guided and one was not. The unguided algorithm ran for two hours; the percent deviation from the algorithmically determined solution to the true optimal solution was recorded. The next portion of the experiment utilized the human guided tabu algorithm, and this algorithm was run until it found a solution that was 1% closer to the optimal value than the solution found using the non-interactive tabu algorithm. In every case, this benchmark was achieved in under half an hour; on average, the human-guided algorithm only needed fifteen minutes to find a more optimal solution than that found by running the un-guided algorithm for two hours (Klau et al 2009).

WRESTORE uses a form of human-guided optimization based upon the Interactive Genetic Algorithm Mixed Initiative Interaction (IGAMII) algorithmic framework developed by Babbar and Minsker in 2006. The IGAMII framework features two distinct phases: the optimization phase and the introspection phase. User fatigue is a major issue with interactive algorithmic optimization. To overcome this issue, Babbar and Minsker created a surrogate user, trained by the preferences of the user. This user could perform optimization on large populations of potential solutions. Interactive optimization with human users was limited to small populations generated by a micro IGA. Thus, the human user trained the surrogate user, who would optimize over a large sample space, allowing the IGA to migrate into a more favorable portion of the solution space more quickly. Each optimization session was ended with an introspection session – based upon trends in confidence that the human user exhibits during introspection sessions, the mixed initiative interaction portion of this optimization algorithm, the MIM, would select which user would have the initiative (perform the algorithmic optimization) for the next session. The algorithm uses Mann-Kendall statistics to track trends in user confidence. When the statistic reaches a specified level of significance, and there is a trend towards decreasing variance in user confidence, the MIM will switch initiative. In this way, the algorithm balances the need for adequate training of the surrogate user against the need for human user rest (Babbar and Minsker 2006; Babbar-Sebens and Minsker 2012).

The experiment validating the combination of the MIM and IGAMII framework centered upon discovering an optimal layout of monitoring wells that accurately depict the concentrations of a contaminant (BTEX) in an aquifer while keeping the number of wells to a minimum. The experiment considered four possible algorithmic initiative combinations. The first one did not utilize the surrogate user, the second and third included surrogate and human users with initiative assigned in an ad-hoc fashion, and the fourth allowed the MIM to transfer initiative from one user to the next whenever the sufficiency conditions were met. The sequences are shown below; "h" represents a human user, and "f" represents a simulated user (Babbar and Minsker 2006, Babbar-Sebens and Minsker 2012):

1. h-h-h-h
2. h-h-h-h-f-f-f
3. h-f-h-f-h-f-h
4. h-h-f-h-f-f-f

All sequences containing a simulated user out-performed the first sequence by a large margin. However, the fourth sequence performed better than the two ad-hoc selected IGAMII sequences (Babbar and Minsker 2006; Babbar-Sebens and Minsker 2012). WRESTORE still functions under this basic framework. However, the algorithm now includes an adaptive component. The IGA being used in WRESTORE at the time of the experiment was the Nondominated Sorting Genetic Algorithm II, or NSGA-II. The experiment used trends in the decision space of the preferred solutions selected by the user to modify the crossover and mutation rates of the NSGA-II algorithm to keep it in the neighborhood of the preferred solutions. For deterministic users, this new algorithm performed well, but for users who exhibit a large amount of noise in their decision making process, this new adaptation did very little. Further changes to this new adaptive IGA are needed to address the needs of a noisy user (Piemonti 2015).

2.3 Climate Model Ensemble Selection

In general, ensemble methods perform much better than individual models in creating accurate models (Block et al 2009; Knutti et al 2010; Luo et al 2007; Mote et al 2011; Tebaldi and Knutti 2007; Winkler 2012; Zhang et al 2009). Several methods exist for selecting model ensembles. Zhang et al (2009) explored using Bayesian methods to combine climate models, which SWAT used to generate flows. The authors compare the performance of single model generated flows to flows generated from a variety of ensembles generated using several methods, culminating with the application of Bayesian methods (2009).

Upon predicting the future climate using both individual models and several ensemble methods, the article reports two findings. First, any ensemble method far outperforms individual models. Second, Bayesian Model Averaging performs the best of any ensemble method according to all three statistics used to evaluate their performance. Utilization of Bayesian Model Averaging in conjunction with the SWAT tool trained via genetic algorithm appears to be a sound method to use moving forward (Zhang et al 2009). However, there is conflicting evidence suggesting that Bayesian model averaging actually does not produce a time series with better prediction power than any other method. The prior and posterior distributions found using this method are almost indistinguishable; this suggests that any modifications made to the bias of the time series ensemble is largely superficial (Luo et al 2007).

In general, Bayesian methods for combining precipitation ensemble models are highly promising (Luo et al 2007; Zhang et al 2007). For hydrologic processes, Block et al (2009) suggests using normal kernel density estimators. The author compared ensembles generated using simple arithmetic averaging, least squares linear regression, and a normal kernel density estimator. The ensemble combination method that performed the best was kernel estimation. All ensemble methods performed better than using a single model, but the normal kernel density estimator produced the highest correlation coefficient (2009).

However, most articles sampled for this analysis questioned the benefits of using more complicated methods, and implied or directly stated that utilization of a simple arithmetic mean was the best alternative (Knutti et al 2010; Mote et al 2011; Tebaldi and Knutti 2007; Winkler 2012). Precipitation data is extremely difficult to model accurately compared to other climate factors, such as temperature (Winkler 2012), compounding the difficulties associated with combining interrelated models (Tebaldi and Knutti 2007).

Recent studies have found that all weighting methods do not perform statistically better than simple arithmetic averaging (Mote et al 2011; Winkler 2012). Both Tebaldi and Knutti and Mote et al provide compelling reasons why these seemingly superior methods fail to outperform the simple average. A possible reason why weighting doesn't perform as well as expected is the lack of effective metrics and statistics to rate and weight model performance (Chen et al 2012; Knutti and Tebaldi 2010; Mote et al 2011). Additionally, there is evidence that climate models perform poorly with regards to precipitation (Luo et al 2007; Winkler 2012), which is of key interest to the WRESTORE development team. However, it seems likely that the best available method for combining ensemble models is to report both the arithmetic mean and some information about the ensemble variability as a whole (Mote et al 2012).

Large model ensembles also do not seem to perform much better than smaller model ensembles (Knutti et al 2010; Tebaldi and Knutti 2007; Zhang et al 2009). Three model ensembles produced good predictions for use in the SWAT tool (Zhang et al 2009). Bias reduction per added climate model is significantly reduced after the ensemble contains five models, and virtually disappears after ten models (Knutti et al 2010). The merits of a large ensemble appear limited at best, and a waste of valuable time at worst.

Chapter 3. Selecting a Robust Climate Ensemble to Capture Extreme Streamflow Events

Kenneth R. Cannady-Shultz[1], Meghna Babbar-Sebens[2]

[1]Department of Civil and Construction Engineering, Oregon State University, Corvallis, OR USA (shultzke@oregonstate.edu)

[2]Department of Civil and Construction Engineering, Oregon State University, Corvallis, OR USA (meghna@oregonstate.edu)

Corresponding Author: Meghna Babbar-Sebens

Journal TBD

3.1 Abstract

Numerous methods exist to work with large collections of climate models, focusing either on condensing all models in an ensemble to one averaged model, or selecting a sub-ensemble. Selecting an ensemble based upon climate factors may not result in scenarios with peak flows, which is a topic of interest in many studies. Studies of this type can also be computationally intensive, so reducing the number of models and length of the realizations is desirable. To this end, this research developed a four step method designed to select a small number of single years from multiple full-length climate model realizations that represent peak flows with diverse climatic conditions that considers past and future conditions, as well as changes from the past to the future. All factors used as part of the proposed process are compared ordinally, not chronologically. An ensemble selected using this method exhibited high flows with diverse climatic factors.

3.2 Introduction

There is general agreement that ensemble methods that combine multiple climate models outperform their individual component models in predictive ability (Block et al 2009, Knutti et al 2010, Luo et al 2007, Mote et al 2011, Tebaldi & Knutti 2007, Zhang et al 2009). Several different atmosphere-ocean general circulation models (AOGCM) exist, and each is based upon a different understanding of the physical processes that drive the climate. Regional climate models (RCM) downscale these global AOGCMs to apply to specific geographic locations, creating a plethora of AOGCM-RCM pairing available for use by the modelling community (Mearns et al 2009, Walters 2014). Each model then generates different predictions, even if all of them are supplied with the same data, resulting in a diverse set of climate predictions, none of which can really be viewed as superior to its peers. Anthropogenic climate change further complicates these relationship, as the human component of climate change is highly reliant on our behavior. Continued high emissions lead to extreme increases in global temperatures, and drastic alterations to existing climate patterns. Aggressive actions to combat climate change can heavily reduce the magnitude of the temperature increase, and will also affect how climate patterns change (IPCC 2012 & 2014). Thus, there are a staggering number of future climate projections available, each based upon a different combination of climate model and emission scenario. Even if a researcher focuses on a single emission scenario, it is not reasonable to assume that a single climate model realization will adequately represent the possible future, so ensemble methods seek to represent the general behavior of all possible climate models for the area (Mote et al 2011).

Ensemble creation methods proposed have largely fallen into one of two groups. The first group seeks to combine several different climate models into a single, representative model via some form of averaging. Bayesian methods are favored by researchers such as Luo et al (2007) and Zhang et al (2007), while other authors have favored more simple approaches, noting that the improvement in predictive ability from Bayesian averaging or other weighting procedures do not necessarily outperform the arithmetic average (Knutti et al 2010, Mote et al 2011, Tebaldi & Knutti 2007). Additionally, Mote et al (2011) note that the act of averaging several climate models removes variation and smooths out extreme values. The resulting muted representative model thus may not be appropriate as a sole representative of the future climate, especially where extreme values are concerned.

Alternately, the second set of methods don't seek to combine all the climate models to create a single representative model, but rather select a subset of models that adequately represent the behavior of the full set of models. Thober and Samaniego (2014) considered several methods for selecting these representative models. While their ultimate goal is to select a small sub-ensemble for eventual model averaging, the selection methods themselves still fall into this second set of methods. One selection method involved an exhaustive search for a subset of size n from a full set of climate models that have the lowest average rejection rate. Viable for small collections of climate models, this method quickly becomes computationally expensive as the number of climate models increases.

The next two methods they consider involve assigning each individual a score according to their bias relative to the maximum bias of the full set of climate models, and selecting the best n models as the ensemble. While they do differ on how they average the selected models together, they employ the same method for initial selection. The final two they considered were backwards and forwards elimination. Backwards elimination begins with a sub-ensemble equal to the full set of climate models, then iteratively removes single models to create successively smaller ensembles that have the lowest rejection rate. Alternatively, forwards elimination adds successive models to an ensemble of two models (a "seed," the pair of models that had the lowest rejection rate of all possible model pairs).

Ultimately, Thober and Samaniego (2014) preferred backwards elimination. Alternately, Lutz et al (2016) developed an envelope based approach that sought to preserve the variation of the larger set of climate models by performing the selection process on models that are near the extremes of the projected future, then evaluating those extreme models for changes in climate metrics from the past to the future, and the prediction skill of those models when subjected to past climate conditions.

The Lutz et al (2016) method is effective at selecting a small ensemble that represents the extremes of climate responses. However, the link between extreme precipitation and flooding is tenuous and inconsistent. Madsen et al (2014) performed an extensive review of papers analyzing extreme precipitation and hydrological floods in Europe. These studies reported findings based upon climate projections and historical trends. The review reported two general trends from the literature: extreme precipitation events are becoming more prevalent, but there are no indications of a similar trend in severe flooding. Some regions have observed increased flooding, but others have observed apparent decreases. These results were also observed in a global study that Dankers et al (2014) performed. Nine global hydrology and land surface models (seven that are part of the Water Model Intercomparison Project, as well as the PCRaster Global Water Balance Model and the Water Balance Model) were used in this research, each operating within a global 0.5° grid. Each of these hydrologic models were driven by five different Coupled Model Intercomparison Project Phase 5 (CIMP5) climate model realizations, and the 30-y return period stream flows were analyzed. Approximately half of the grid points showed an increase in extreme stream flows, but approximately one-third showed a decrease. Again, the areas that showed decreased extreme flows were largely in areas where streamflows are dominated by snowmelt, but this trend was not exclusive.

Several more localized studies have shown similar behaviors. Other locations with snowmelt dominated streamflow regimes, such a Minnesota (Novotny & Heinz 2007) and the Columbia River at the Dalles, Oregon (Mondal & Mujumdar 2016), exhibited similar flat to decreasing trends in extreme river flows. In Mondal & Mujumdar's (2016) study, however, some possible worst-case scenario simulations did show a possible detectible increase in extreme river flows by 2100. In one case, they noted that a detectable increase in the magnitude of the 75-y flow event could happen as early as 2027. Several studies also found an increase in flood risk. A study in the Kemptville Creek watershed in Ontario, Canada found that the exceedance probability for floods ranging from a 2-y to 100-y event increased 34.3% and 17.8% (respectively) for a dam with a 20-year service life (Seidou et al 2012). Jena et al (2014) found that recent increases in high floods in the Mahandi basin in India are linked to increased extreme precipitation in the middle of the basin. Wu and Huang (2015) observed a more nuanced relationship between extreme precipitation and high peak flows. Accounting for changes in peak flows induced by human activity, increasing precipitation lead to increased flooding in June and July, but September increases in precipitation did not. These studies, when viewed in conjunction with the larger studies by Dankers et al (2014) and Madsen et al (2014), all suggest that extreme precipitation events do not necessarily lead to flooding. Thus, if the goal is to select climate models that produce high peak flows, selecting a climate model ensemble based upon precipitation characteristics will not be sufficient. Stream flow characteristics must also be included in the selection process.

3.2.1 Purpose of Study

The research presented here seeks evaluate whether a proposed modification to the ensemble selection process developed by Lutz et al (2016) outperforms its parent method in selecting climate model years that result in high peak flows. These climate models are being selected for use within an Interactive Genetic Algorithm (IGA) that utilizes an NSGA-II optimization process coupled to a SWAT hydrologic model (Babbar-Sebens 2012, Babbar-Sebens et al 2015, Piemonti et al 2013). The run time of this algorithm is not insubstantial, and scales directly with the number of years modelled in SWAT. Thus, both the proposed method and the parent method are both structured to select single year climate realizations to represent these extreme peak flows.

3.3 Methodology

3.3.1 Study Area

Eagle Creek watershed is located northwest of Indianapolis, and features agricultural, urban, and undeveloped land uses. This watershed has a drainage area of approximately 419 km^2 and is part of the larger Upper White River Watershed. Eagle Creek drains to Eagle Creek reservoir, which is a major source of drinking water and recreation for Indianapolis and the surrounding areas. Over 60% of the watershed is agricultural, with the dominant crop being corn. In recent years, upstream agricultural areas have released pesticides, sediments, and fertilizers into nearby streams, resulting in the Eagle Creek reservoir becoming impaired (Babbar-Sebens et al 2013, Javaheri & Babbar-Sebens 2014, Piemonti 2013). Generally, the climate in Indiana is considered continental. As such, the area has high humidity, and highly variable temperature. Precipitation is high in this area as well, with typical cumulative annual precipitation between 965 and 1,016 mm. Average annual temperature is approximately 11°C (Clark et al 1980, Walters 2014).

3.3.2 Climate Model Selection

The climate model year selection methodology discussed in this section is predicated upon work completed by Walters and Babbar-Sebens (2016). They explored the performance of the potential wetlands specified by Babbar-Sebens et al (2013) in the Eagle Creek Watershed when subjected to the potential mid-century climate as predicted by six different North American Regional Climate Change Assessment Program (NARCCAP) regional climate model (RCM) atmosphere-ocean general circulation model (AOGCM) pairings, referred to from here on as the "climate models," developed for North America and the A2 emission scenario, detailed in Table 1 (Mearns et al 2009, Walters 2014). Each climate model was used to simulate precipitation and temperatures for two time periods: 1970-2000, and 2040-2070 ("mid-century"). Walters (2014) applied four different bias correction techniques to each climate model to calibrate their behavior to observed data. The methods were linear scaling (LS), local intensity scaling (LOCI), power transformation (PT), and distribution mapping (DM). The un-adjusted models (RAW) were also tested. Based on the predictive skill of the different bias corrected or un-adjusted models, 11 models (listed in appendix A) were selected to comprise a final climate model ensemble for the Eagle Creek watershed.

Table 1: Summary of Climate Models

RCM		AOGCM	
Model	**Supporting Agency**	**Model**	**Supporting Agency**
CRCM	Canadian Centre for Climate Modelling and Analysis	**CCSM** **CGCM3**	National Center for Atmospheric Research Canadian Climate Centre for Climate Modelling and Analysis
HRM3	Met Office Hadley Centre for Climate Science and Services	**GFDL**	Geophysical Fluid Dynamics Laboratory
MM5I	National Center for Atmospheric Research/Pennsylvania State University	**CCSM**	National Center for Atmospheric Research
RCM3	Abdus Salam International Center for Theoretical Physics	**CGCM3**	Canadian Climate Centre for Climate Modelling and Analysis
WRFG	National Center for Atmospheric Research	**CGCM3**	Canadian Climate Centre for Climate Modelling and Analysis

Two methods for selecting a climate ensemble that captures more extreme flow events are considered. Each involves separating the full thirty-year mid-century climate realization of each individual climate model into thirty individual climate realizations (each of which is one year long). The first ensemble approach evaluates each realization separately, and evaluates all changes in climate factors relative to the behavior of the observed past climate. The second takes these different realizations and ranks them with respect to the value of specific climatic factors. Thus, each model would be split into thirty realizations, of which 28 are usable (the first and last year of the model realizations are not full years). These 308 possible climate model realizations are ranked according to climatic factors (referred to as metrics going forward) specified by the researcher to create the set $MP_{ik} = \{mp_{ik}^1, mp_{ik}^2 \dots mp_{ik}^{28}\}$, where mp_{ik}^m is the calculated value for the i^{th} metric for the m^{th} ranked single year realization from climate model k. A similar procedure is performed for model validation period (1970-2000) on the eleven models and the observed climate to create the set $MV_{ik} = \{mv_{ik}^1, mv_{ik}^2 \dots mv_{ik}^{28}\}$ for the eleven models and set $OV_i = \{ov_i^1, ov_i^2 \dots ov_i^{28}\}$ for the observed climate. These ordered lists of climate model realizations and observed climate data are used for the analysis: changes between the past and projected climate are evaluated between projected and past realizations with the same rank. Work by Block et al (2009) and Seidou et al (2012) has focused upon improving forecasting accuracy for peak flows, but, to the authors' knowledge, no other research has selected climate model ensemble members based upon ordinal climate data in this fashion. After selecting two ensembles of climate models using each proposed method, the two ensemble selection methods will be evaluated using a SWAT model of the Eagle Creek Watershed calibrated by Walters (2014). The ability of each selected ensemble to select member models resulting in high peak flow will be evaluated, as well as the ensemble diversity. The SWAT model will be run for two years: one year for model initialization (whose results will not be evaluated), and one year for evaluation.

3.3.2.1 Basic Envelope-Based Selection Approach

The first method adapts the climate selection method proposed by Lutz et al (2016) to select individual model years. The overall structure of the method will remain unchanged, though some changes will be made. Lutz et al's (2016) method utilizes precipitation and temperature data from two different time spans: the projection period and the validation period. The projection period includes observed climate data, as well as model-generated climate data based upon historic climate conditions. While the projected time period will be split into individual years for comparison, the validation period will not. Hence, each projected singular year will be compared to an aggregated past climate, and the predictive skill will be evaluated for the full validation period of each parent model. The single year projections will each be assigned a predictive skill rating based upon its parent model (resulting in each year from the same parent model receiving an identical predictive skill score).

The basic structure of the procedure is as follows:

1. Each model year is evaluated with respect to the difference in the cumulative precipitation and average temperature values between it and the observed past. The data is split into four distinct groups centered about each possible pairing of the 90th and 10th percentiles of the precipitation and temperature metrics. The five models from each group (a total of 20) that are closest to their respective center (have the lowest Euclidian distance) are selected to proceed to the next step.
2. The remaining models are evaluated according to an appropriate pairing of the following metrics:
    a. Cumulative precipitation resulting from all days with rainfall intensity ≥ 95th rainfall intensity percentile for the model realization, CP_{95}
    b. Maximum number of consecutive dry days, *MCDD*
    c. Maximum number of consecutive days with maximum temperatures ≥ 90th percentile for daily maximum temperatures, *WSDI*
    d. Maximum number of consecutive days with minimum temperatures ≤ 10^{h} percentile for daily minimum temperatures, *CSDI*

    WSDI is used at all corners that are centered about the 90th percentile for annual average temperature, while the *CSDI* is used for the 10th percentile. Likewise, the CP_{95} measure is used for any corner centered about the 90th percentile for annual cumulative precipitation, while *MCDD* is used for corners centered about the 10th percentile. Each model is evaluated according to the difference between its metric values and those from the past, and the modes from each corner are ranked according to the magnitude of these differences separately (each model year is thus part of two ranked lists). For each metric, the models are assigned a score equal to their position in an ascending sorted list, with the largest value being in the 5th position, and the smallest in the 1st. All scores for each model are averaged, and the two models with the highest average from each corner progress to the next step (a total of 8 models).

3. The predictive skill of each parent model is evaluated for precipitation and temperature data according to the metrics proposed by Lutz et al (2016). The two predictive skill scores for each parent model are averaged, and the resulting value is assigned to each respective child model being considered in this step. The one model from each corner with the highest predictive skill rating is selected as an ensemble member, resulting in an ensemble *E* with four members.

3.3.2.2 Ordinal Flow and Precipitation Envelope-Based Selection Approach

The second method is adapted from the methodology presented by Lutz et al (2014) to incorporate flow and precipitation characteristics. This research added a step zero to their process, and modified the approach for evaluating the change from past to present climates, as well as the predictive skill of the past models, to use ordinal comparisons, not chronological comparisons.

0. Select a subset $MP5_{1k} \subset MP_{1k}$, where $MP5_{1k} = \left\{mp5_{1k}^{1}, mp5_{1k}^{2} \dots mp5_{1k}^{z} \dots mp5_{1k}^{5}\right\}$ ranked in decreasing order of PF_5 value calculated for each model *k*, leading to 55 models progressing to the next step as 11

distinct sets, referred to collectively as *MP*. Additionally, calculate the set $MP5_k = \{mpF_{k1}, mpF_{k2} \dots mpF_{k5}\}$, where $mpF_{zk} = \left\{mp_{1k}^{(k,y)}, mp_{2k}^{(k,y)} \dots mp_{ik}^{(k,y)}\right\}$. Each element $mp_{ik}^{(k,y)}$ of mpF_{zk} is a metric value from the set $\{PF_5, P_5, CP_{95}, MCDD, C_P\}$ calculated for model k realization year y that corresponds to the year used to calculate $mp5_{1k}^{z}$. During this 0th step, the subsets $MV5_{1k} \subset MV_{1k}$ and $OV5_1 \subset OV_1$ are similarly defined. Sets $MV5_K$ and $OV5$ contain the values of y that correspond to the year used to calculate $mv5_{1k}^{z}$ and $ov5_{1k}^{z}$, respectively. The following ETCCDI (Peterson 2005) and peak flow (Dankers et al 2014) metrics are calculated:

a. Maximum 5-day average peak flow, defined by equation 5.

$$PF_5 = max\left(\frac{\sum_d^{d+4} PeakFlow_d}{4}\right) \forall d \in \{1 \dots LengthYear - 4\} \qquad (1)$$

b. Maximum 5-day cumulative precipitation, as used by Thober and Samaniego (2014), defined by equation 6

$$P_5 = max\left(\sum_d^{d+4} PDepth_d\right) \forall d \in \{1 \dots LengthYear - 4\} \qquad (2)$$

c. Cumulative precipitation resulting from all days with rainfall intensity ≥ 95th rainfall intensity percentile for the model realization, CP_{95}

d. Maximum number of consecutive dry days, *MCDD*

e. Annual cumulative precipitation, *CP*

1. Select the 8 models from *MP* nearest to every possible combination of the 10th and 90th percentiles for *CP* and the annual cumulative streamflow, *CF*, and record what element *z* of which $MP5_k$ set corresponds to the specified value of C_p. Create four sets of 8 models, S_1 to S_4 for models near each combination of flow and precipitation percentiles that progress to step two, where they will be analysed independently. $S_v = \{ID_1, ID_2 \dots ID_{16}\}, v = \{1,2,3,4\}$, where each element $ID_m = (z, k)$ contains the element and set identifiers, respectively.

2. For each set S_v of 8 models, create sets $M_{iv} = \left\{dmp_{iv}^1, dmp_{iv}^2 \dots dmp_{iv}^b \dots dmp_{iv}^{16}\right\}$ for every i^{th} metric in the v^{th} corner, where $dmp_{iv}^b = |MP5_k\{z\}\{i\} - MV5_k\{z\}\{i\}|, (z,k) = S_v\{b\}$. From this data, create sets $R_{iv} = \{r_{iv}^1, r_{iv}^2 \dots r_{iv}^z \dots r_{iv}^8\}$, where each r_{iv}^z is the rank of dmp_{iv}^z relative to all elements of set M_{iv}. Consequently, the elements in R_{iv} range from 1 to 8, with the largest dmp_{iv}^z corresponding to $r_{iv}^z = 8$, and largest dmp_{iv}^z corresponding to $r_{iv}^z = 1$. The total score for each model in S_v defined as set $SC_{iv} = \{sc_{iv}^1, sc_{iv}^2 \dots sc_{iv}^z \dots sc_{iv}^8\}$, where sc_{iv}^z is defined in equation 7. The subsets $S'_1 \subset S_1$, $S'_2 \subset S_2$, $S'_3 \subset S_3$ and $S'_4 \subset S_4$ containing the 4 elements with the highest scores are selected, and combined to create one large set *S'* that progresses to the final step.

$$\boldsymbol{sc_{iv}^z = \frac{\sum_i r_{iv}^z}{5}} \tag{3}$$

3. All selected models from step two from each corner are tested for skill in predicting the observed climate during the validation period with respect to flow and precipitation. The score for precipitation predictive skill is the product *F* of f_1 to f_4, defined by equations 8 to 11, respectively:

$$\boldsymbol{f_1 = 1 - \left(\frac{|A_{GCM}^+ - A_{OBS}^+|}{2A_{OBS}^+}\right)^{0.5}} \tag{4}$$

$$\boldsymbol{f_2 = 1 - \left(\frac{|A_{GCM}^- - A_{OBS}^-|}{2A_{OBS}^-}\right)^{0.5}} \tag{5}$$

$$\boldsymbol{f_3 = 1 - \left(\frac{|\overline{P_{GCM}} - \overline{P_{OBS}}|}{2\overline{P_{OBS}}}\right)^{0.5} \; or \; f_3 = 1 - \left(\frac{|\overline{F_{GCM}} - \overline{F_{OBS}}|}{2\overline{F_{OBS}}}\right)^{0.5}} \tag{6}$$

$$\boldsymbol{f_4 = 1 - \left(\frac{|\sigma_{GCM} - \sigma_{OBS}|}{2\sigma_{OBS}}\right)^{0.5}} \tag{7}$$

Where A_{GCM}^+ and A_{OBS}^+ are the areas to the right of the 50th percentile under the cumulative distributions for the modelled and observed past precipitation or flow CDF, respectively. Replacing the "+" with a "–" superscript result in calculating the area to the left of the 50th percentile. $\overline{P_{GCM}}$ and $\overline{P_{OBS}}$ ($\overline{F_{GCM}}$ and $\overline{F_{OBS}}$) are the average annual cumulative precipitation (annual average daily peak flow) for the modelled and past climate realizations, and σ is the standard deviation for each PDF. All "GCM" subscript values are calculated from the b^{th} model year realization in set S' corresponding to $y = MV5_k\{z\}, (z, k) = S'\{b\}$, with k being the model ID and y the realization year. Similarly, all "OBS" subscript values are defined by the k, y pairing where $y = OV5_k\{z\}, (z, k) = S'\{b\}$. The four models with the highest score F are selected as the final ensemble E.

3.4 Results

Within their respective envelopes, both methods produced diverse ensembles that represent the variability of the full collection of climate model years. As expected, the proposed ordinal approach, hereafter referred to as the "ordinal method" produced ensembles that favored high peak flows relative to both the full collection of climate models and the Lutz et al (2016), hereafter referred to as the "baseline method". Details about the two ensembles are included in table 2 (baseline method) and table 3 (ordinal method).

Table 2: Model Data for Ensemble Selected Using Baseline Method

Baseline Method Ensemble

RCM	**AOGCM**	**Bias Correction**	**Year**	**Peak Flow (cms)**	ΣPrecipitation (mm)	$\mathbf{T_{avg}}$ (°C)
MM5I	CCSM	LOCI	2058	31	1411	13.7
MM5I	CCSM	LOCI	2066	41	882	14.8
WRFG	CGCM3	LOCI	2058	157	1402	11.3
WRFG	CGCM3	LOCI	2068	94	851	12.5

Table 3: Model Data for Ensemble Selected Using Ordinal Method

Proposed Ordinal Method Ensemble

RCM	AOGCM	Bias Correction	Year	Peak Flow (cms)	ΣPrecipitation (mm)	T_{avg} (°C)
MM5I	CCSM	DM	2057	336	1298	13.0
MM5I	CCSM	RAW	2055	165	1037	15.1
MM5I	CCSM	RAW	2065	136	1685	14.0
WRFG	CGCM3	LOCI	2058	157	1401	11.3

The average peak flow from the ordinal approach is 146% higher than the baseline method, and 113% higher than the full ensemble. However, neither of these differences are statistically significant, with two-tail t-test p-values of 0.210 and 0.187, respectively. Figures 2 and 3 show peak flow plotted as a function of cumulative precipitation, the one screening (step 1) metric both approaches shared. Both figures suggest that the two methods produced diverse ensembles with respect to both of these metrics, but that the ensemble selected by the ordinal approach selected models with higher peak flows than the baseline approach almost universally, with the exception of the one model that each ensemble shares. The centroid of the baseline method and the observed ensemble are very close to each other, with a deviation of 3.67% and 14.6% between the two with respect to cumulative precipitation and cumulative flow, respectively. By contrast, the ordinal approach deviates 23.7% and 75.2% from the full ensemble values of cumulative precipitation and flow, respectively.

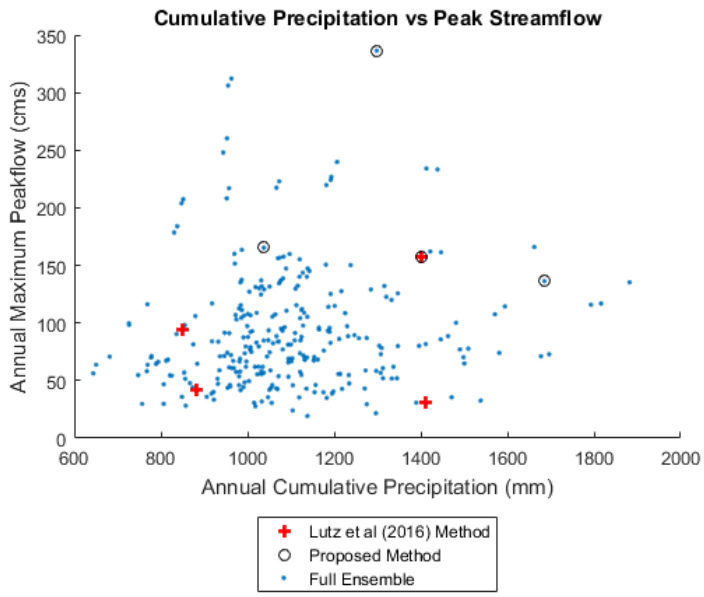

Figure 2: Scatterplot of Annual Cumulative Precipitation vs. Peak Streamflow for All Models and the Ensembles

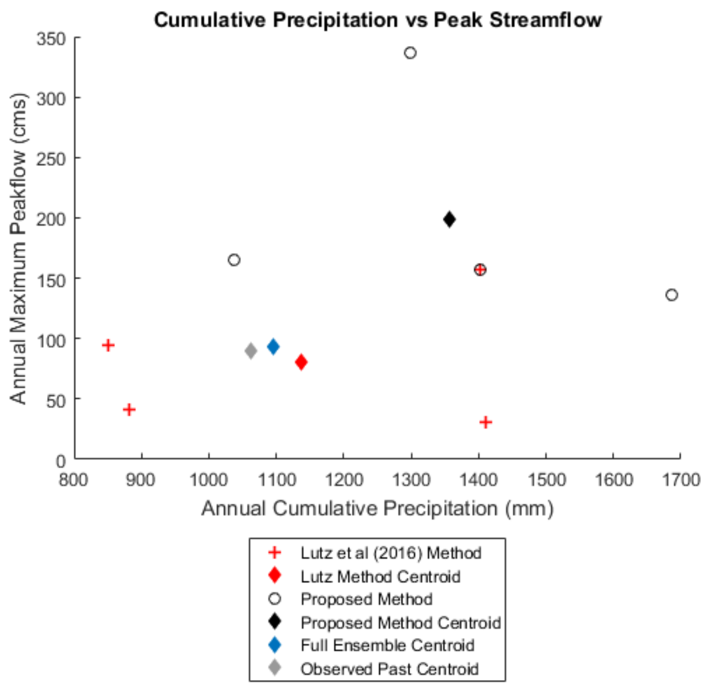

Figure 3: Peak Streamflow and Cumulative Precipitation Centroids for All Models and the Ensembles

Figure 3 strongly suggests that the ordinal method favors model years that occupy the corner defined by high peak flows and cumulative precipitation. Figures 4 and 5 plot cumulative precipitation against average temperature, which were the two metrics that the baseline method focused on maintaining diversity in. Figure 3 suggests that the baseline method did a decent job of maintaining the variability of the full ensemble with respect to these two factors, though it may exaggerate these values, as suggested by the variances for temperature and cumulative precipitation. The ordinal method favored high cumulative precipitation years, so it did not capture the full precipitation variation. However, it sampled a similar range of temperature values to those present in the baseline method, suggesting that the ordinal method did a respectable job of representing the temperature variability without directly incorporating it into its selection process. The centroid of the baseline method and the observed ensemble are very close to each other, with a deviation of 3.67% and 0.95% between the two with respect to cumulative precipitation and average temperature, respectively. The average temperature of the ordinal method is very similar to the other average values, with the exception of the past observed climate, which has a drastically lower average temperature, as shown in figure 5.

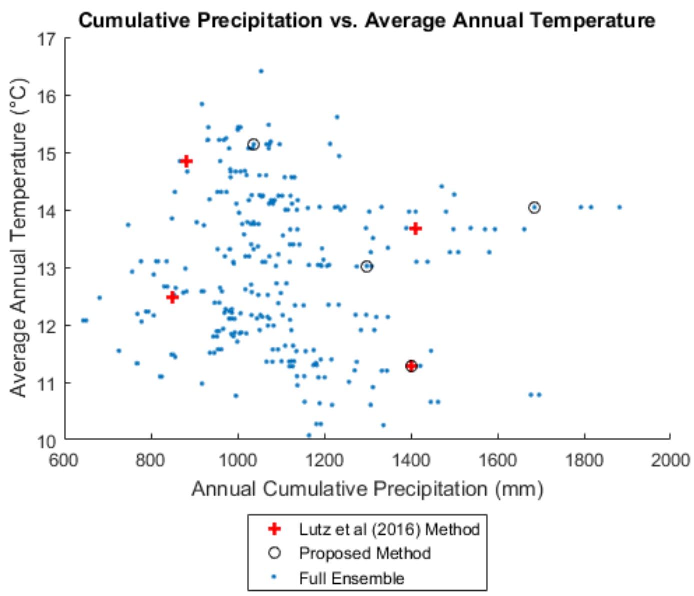

Figure 4: Scatterplot of Annual Cumulative Precipitation vs. Average Temperature for All Models and the Ensembles

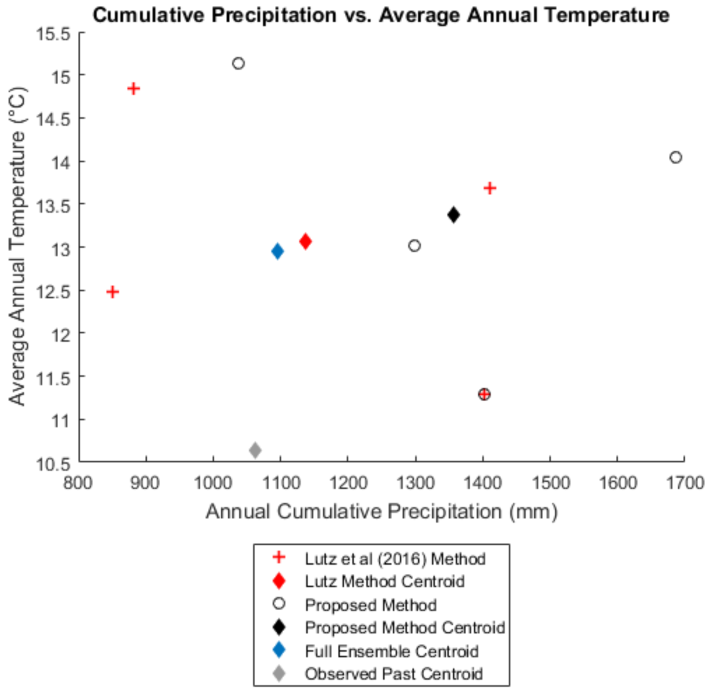

Figure 5: Average Temperature and Cumulative Precipitation Centroids for All Models and the Ensembles

Figures 6 and 7 plot annual cumulative peak streamflow as a function of annual cumulative precipitation, which were the two metrics that the ordinal method strove to maintain variability in. As mentioned previously, the ordinal method favored high precipitation years, and figures 6 and 7 reflect this fact. This method also apparently favored years with high cumulative peak streamflows, suggesting that high peak flow events occur in wet years with heavy flows. The behavior of the baseline method in figure 6 is inconsistent. Generally, it appeared to select years with low to moderate cumulative stream flows, but the ensemble also included the single year with the highest cumulative peakflow. Figure 6 nevertheless indicates the ordinal method selected ensemble realizations on average with higher cumulative flows than the full ensemble or the baseline method.

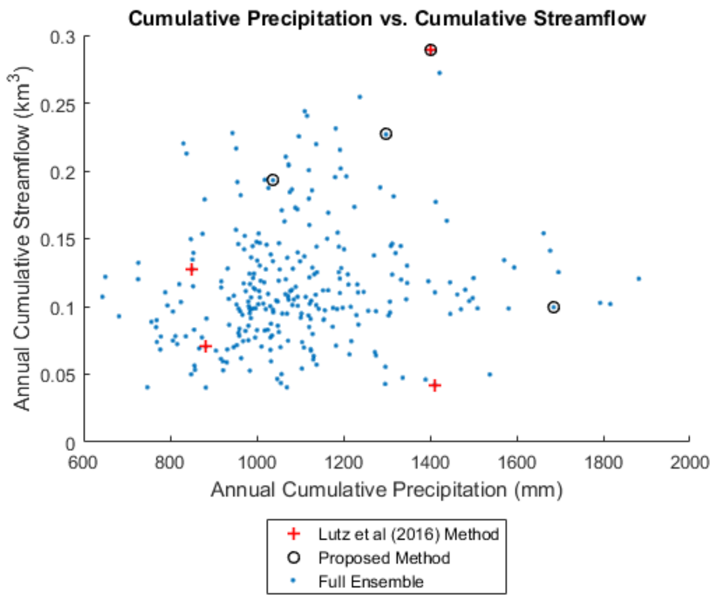

Figure 6: Scatterplot of Annual Cumulative Precipitation vs. Annual Cumulative Streamflow for All Models and the Ensembles

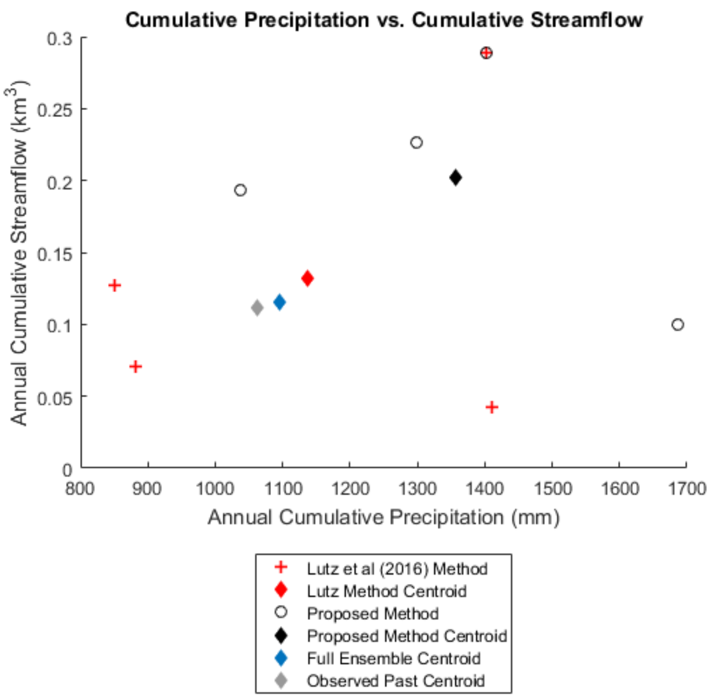

Figure 7: Cumulative Streamflow and Cumulative Precipitation Centroids for All Models and the Ensembles

In both the observed past and the full projected ensemble, the correlation between any two of these factors was typically minimal. There was a modest correlation between peak flows and cumulative streamflows in both cases ($r^2 = 0.26$ and 0.47 for the observed past and full ensemble, respectively), but all other r^2 values were less than 0.1. The ordinal method mimicked this behavior, with a weak correlation between cumulative precipitation and peak flow, and no other meaningful correlations. However, the baseline method created an ensemble with several strong correlations. Peak streamflows and cumulative flows became extremely strongly correlated, with $r^2 = 0.967$, and temperature became strongly correlated with both peak streamflows ($r^2 = 0.844$) and cumulative flows ($r^2 = 0.757$). This suggests that peak streamflow, cumulative streamflow, and temperature are strongly redundant for the projected future, and either one could be used to approximate the behavior of the other two, contrary to what the observed data and full projections suggest.

3.5 Discussion

As expected, the baseline method centroid with respect to temperature and cumulative annual precipitation closely matched the centroid of the full set of climate model years. However, this similarity persisted into traits that the baseline method did not account for in its selection process. This behavior is slightly unexpected, as the correlation between the flow characteristics and climate characteristics were virtually non-existent. As such, there was no reason to assume that average climatic behavior would result in average flow behavior. An explanation for this may lay in the strong correlations between flow and temperature that emerged in the baseline ensemble. It is possible that years characterized by climate extremes – the four corners that the baseline method selects around – exhibit a strong dependence of flow on the temperature value. The sign of the correlation coefficients between temperature and both flow factors are negative, suggesting an inverse relationship. Extremely hot years tend to have low stream flow, and extremely cool years tend to have higher flows. However, it is worth noting that the ensemble generated by the baseline method almost exclusively exhibits lower peak and cumulative flows than the ensemble generated by the proposed method, so it may be more accurate to say that extremely hot years tend to produce low flows, while cool years result in approximately average flows. Also, the emergent relationship between temperature and flow in this baseline ensemble may be entirely coincidental – a larger or smaller ensemble may not produce the same relationship.

The changes the authors proposed to the baseline Lutz et al (2016) method appear to have slightly reduced the diversity of the selected ensemble, but the stated goal of selecting an ensemble with high peak flows was also accomplished. This new method, termed the “ordinal method” in this article, consistently selected model years with higher flows, and resulted in a drastic upward motion in the average maximum peak flow for the entire ensemble, as shown in figure 3. These high flow years also tended to be wetter and have higher cumulative stream flows, as suggested by figure 7. These characteristics, when combined, seem to suggest that high peak flows in the projected future will likely come in wetter years with high average stream flows. The lack of correlation between precipitation and any of the flow metrics suggests that incremental changes within the high cumulative flow and precipitation years do not necessarily result in any consistent change to peak flow values. So, while the high peak flow years in this ensemble all come from the high cumulative flow and precipitation portion of the full ensemble, no internal relationship between these two factors and peak flow exist within this section. Temperature doesn’t seem to be an important predictor for years with high peak flows, as the average temperatures for the ordinally-selected ensembles did not differ much from those of the full ensemble or baseline ensemble, so it appears that consideration of temperature within the selection process would be entirely unnecessary, and excluding it from the proposed method is well justified.

A possible issue with applying the baseline method to single year realizations from a multi-year climate model appears in the final validation step. Predictive skill in the observed past must be assessed for an entire climate model. Consequently, all single year realizations from the same climate model receive the same skill score, which essentially acts as a model filter. Table 4 contains the skill scores for all eleven climate models. All models in the baseline ensemble were from the MM5I_CCSM_LOCI and WRFG_CGCM3_LOCI parent models, which happen to be the 3rd and 1st most skilled models, respectively. This skill rating cannot differentiate between different model years from the same parent model, so any single years from the same parent model are treated equally, regardless of their performance earlier in the process. The specific algorithm that the authors utilized sorted progressing models according to their parent model ID, then their year in ascending order. So, if any corner contained two single model year realizations from the same parent model, the earlier model would be selected arbitrarily.

Table 4: Validation Skill Scores for all 11 Parent Climate Models

Model (RCM_AOGCM_Bias)	Skill
CRCM_CCSM_LOCI	0.462
CRCM_CGCM3_LOCI	0.475
HRM3_GFDL_LOCI	0.551
HRM3_GFDL_LS	0.553
MM5I_CCSM_DM	0.672
MM5I_CCSM_LOCI	0.643
MM5I_CCSM_LS	0.635
MM5I_CCSM_RAW	0.578
RCM3_CGCM3_LOCI	0.507
WRFG_CGCM3_LOCI	0.697
WRFG_CGCM3_LS	0.587

The ordinal approach overcomes this shortcoming by considering ordinal predictive skill. When considering extreme events, the specific year is not important. It is not overly important whether the highest peak flow occurs in 2048 or 2060, all that matters is the magnitude of that high peak flow. The magnitude of the extreme is the primary design focus, and is what will drive eventual action or inaction. So, as opposed to looking at how well the climate models do at predicting the year to year behavior of the past climate, it may be more informative to assess their ability to characterize extreme years. The question changes from “how well does this model do at predicting every year of the observed climate?” to “how well does this model do at predicting the extreme behavior of the observed climate?” Again, if timing is not the focus for this research, it makes more sense to assess how similar the year with the highest peak flow from a particular model is to the observed year from the validation period with the highest peak flow. By the same logic, it is also advantageous to compare second ranked years with each other, third ranked years with each other, and so on.

3.6 Conclusions

Most methods for selecting climate ensembles focus on climatic, not streamflow, characteristics. In general, years with high precipitation do not necessarily result in high peak flows in streams, so any ensemble selection methods must directly incorporate streamflow characteristics into their methodology. The authors chose to modify a method developed by Lutz et al (2016) to incorporate flow data into the selection process, and also altered it to select individual years from a larger climate model as opposed to entire models. This method then is specially developed for any application where minimizing computing time is a central consideration.

An initial peak flow filter was added to the beginning of the selection process to screen out any years that did not have high peak flows. From here, the method applies an ordinal approach to model selecting, comparing equally ranked model years to each other. The goal was to provide an ensemble of high flow model years that result from diverse projected climatic conditions. The method appears to have accomplished this goal. While the diversity of the resulting ensemble was reduced somewhat – the selection process favored years with high precipitation and cumulative stream flows – diversity in average temperatures was maintained, and the selected climate model years all occupy different corners of the high precipitation and cumulative flow space.

The method developed by Lutz et al (2016) exhibited some unexpected behavior, suggesting a strong relationship between flow and temperature within the selected ensemble that was not present in any other data set. The relationship indicated that years with high average temperature tend to exhibit low stream flows, but no clear relationship between temperature and flow seems to exist for high peak flows. Further research may be warranted to explore this possible relationship further, and assess whether this was a random emergent behavior specific to this ensemble, or persists as ensembles of varying sizes are created.

The method proposed in this article appears well suited to studies of extreme flow events in streams, and is also flexible. While this research specified one year lengths for the ensemble methods, multi-year realizations can also be considered. The ordinal approach grants the modeler the freedom to select the length of the modelling period that is appropriate for their specific application.

3.7 References

Arpita, M. and P. P. Mujumdar, 2016. Detection of Change in Flood Return Levels under Global Warming. *Journal of Hydrologic Engineering* **21**.

Babbar-Sebens, M., R. C. Barr, L. P. Tedesco and M. Anderson, 2013. Spatial identification and optimization of upland wetlands in agricultural watersheds. *Ecological Engineering* **52**:130-142.

Babbar-Sebens, M. and B. S. Minsker, 2012. Interactive Genetic Algorithm with Mixed Initiative Interaction for multi-criteria ground water monitoring design. *Applied Soft Computing* **12**:182-195.

Babbar-Sebens, M., S. Mukhopadhyay, V. B. Singh and A. D. Piemonti, 2015. A web-based software tool for participatory optimization of conservation practices in watersheds. *Environmental Modelling & Software* **69**:111-127.

Block, P. J., F. A. Souza, L. Q. Sun and H. H. Kwon, 2009. A Streamflow Forecasting Framework using Multiple Climate and Hydrological Models(1). *Journal of the American Water Resources Association* **45**:828-843.

Clark (ed), 1980. *The Indiana water resource ; availability , uses , and needs*, American Geosciences Institute.

Dankers, R., N. W. Arnell, D. B. Clark, P. D. Falloon, B. M. Fekete, S. N. Gosling, J. Heinke, H. Kim, Y. Masaki, Y. Satoh, T. Stacke, Y. Wada and D. Wisser, 2014. First look at changes in flood hazard in the Inter-Sectoral Impact Model

Intercomparison Project ensemble. *Proceedings of the National Academy of Sciences* **111**:3257-3261.

IPCC, 2012. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaption: A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change. *In,* Editor (Editor)^(Editors). Intergovernmental Panel on Climate Change, Cambridge, UK, and New York, New York.

IPCC, 2014. Climate Change 2014: Synthesis Report: Contribution of Working Groups I, II, and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *In,* Editor (Editor)^(Editors). IPCC, Geneva, Switzerland.

Javaheri, A. and M. Babbar-Sebens, 2014. On comparison of peak flow reductions, flood inundation maps, and velocity maps in evaluating effects of restored wetlands on channel flooding. **73**:132-145.

Jena, P. P., C. Chatterjee, G. Pradhan and A. Mishra, 2014. Are recent frequent high floods in Mahanadi basin in eastern India due to increase in extreme rainfalls? **517**:847-862.

Knutti, R., R. Furrer, C. Tebaldi, J. Cermak and G. A. Meehl, 2009. Challenges in Combining Projections from Multiple Climate Models. *Journal of Climate* **23**:2739-2758.

Luo, L. F., E. F. Wood and M. Pan, 2007. Bayesian merging of multiple climate model forecasts for seasonal hydrological predictions. *Journal of Geophysical Research-Atmospheres* **112**:13.

Lutz, A. F., H. W. ter Maat, H. Biemans, A. B. Shrestha, P. Wester and W. W. Immerzeel, 2016. Selecting representative climate models for climate change impact studies: an advanced envelope-based selection approach. **36**:3988-4005.

Madsen, H., D. Lawrence, M. Lang, M. Martinkova and T. R. Kjeldsen, 2014. Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. **519, Part D**:3634-3650.

Mearns, L. O., W. Gutowski, R. Jones, R. Leung, S. McGinnis, A. Nunes and Y. Qian, 2009. A Regional Climate Change Assessment Program for North America. **90**:311-311.

Mote, P., L. Brekke, P. B. Duffy and E. Maurer, 2011. Guidelines for constructing climate scenarios. **92**:257-258.

Novotny, E. V. and H. G. Stefan, 2007. Stream flow in Minnesota: Indicator of climate change. **334**:319-333.

Peterson, T. C., 2005. Climate Change Indices. pp. 83-86.

Piemonti, A. D., M. Babbar-Sebens and E. J. Luzar, 2013. Optimizing conservation practices in watersheds: Do community preferences matter? *Water Resources Research* **49**:6425-6449.

Seidou, O., A. Ramsay and I. Nistor, 2012. Climate change impacts on extreme floods II: improving flood future peaks simulation using non-stationary frequency analysis. **60**:715-726.

Tebaldi, C. and R. Knutti, 2007. The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society a-*

Mathematical Physical and Engineering Sciences **365**:2053-2075.

Thober, S. and L. Samaniego, 2014. Robust ensemble selection by multivariate evaluation of extreme precipitation and temperature characteristics. **119**:594-613.

Walters, K. M. and M. Babbar-Sebens, 2016. Using climate change scenarios to evaluate future effectiveness of potential wetlands in mitigating high flows in a Midwestern US watershed. *Ecological Engineering* **89**:80-102.

Wu, C. and G. Huang, 2015. Changes in heavy precipitation and floods in the upstream of the Beijiang River basin, South China. **35**:2978-2992.

Zhang, X. S., R. Srinivasan and D. Bosch, 2009. Calibration and uncertainty analysis of the SWAT model using Genetic Algorithms and Bayesian Model Averaging. *Journal of Hydrology* **374**:307-317.

Chapter 4. Are Watershed Management Plans Selected and Preferred by Stakeholders Considering Current Climate Conditions Robust against Climate Change Scenarios? A Sensitivity Study of Stakeholders Spatially-Explicit Preferences.

Kenneth R. Cannady-Shultz[1], Meghna Babbar-Sebens[2]

[1]Department of Civil and Construction Engineering, Oregon State University, Corvallis, OR USA (shultzke@oregonstate.edu)

[2]Department of Civil and Construction Engineering, Oregon State University, Corvallis, OR USA (meghna@oregonstate.edu)

Corresponding Author: Meghna Babbar-Sebens.

Journal TBD

4.1 Abstract

Designing watershed management plans is a complicated process that requires consideration of technical standards and social impacts. Effective design optimization requires consideration both of the quantifiable technical and un-quantifiable social constraints defining the watershed system. This article proposes an addition to the current stakeholder inclusion framework in place for WRESTORE – consensus. This should promote solutions that stakeholders agree are suitable according to their in-tangible criteria, and also perform well. This research created surrogate users based upon stakeholder data from the Eagle Creek Watershed to evaluate different designs. When compared to optimized solutions that did not incorporate stakeholders, this research's solutions performed at least as well, exceeding the non-interactive solutions in some cases. Unfortunately, none of the solutions fared well when subjected to future climate conditions, so any attempt to address future conditions must do so explicitly.

4.2 Introduction

Designing an efficient and effective watershed management plan often involves optimizing across several variables simultaneously, many of which are defined by deep uncertainty (Haimes & Hall 1974). As the climate changes, these problems will grow ever more complicated, and the uncertainty will compound (IPCC 2012 & 2014). These problems are difficult from a purely technical perspective, but watershed management is not performed and implemented in a purely technical context. People live in these watersheds, and they have opinions and intangible evaluation criteria that need to be considered as well.

True optimization of watershed management plans not only requires consideration of the technical aspects and performance metrics of the solutions, but also consideration of the solutions' social impacts. The problem is thus re-defined to include a social dimension to optimize over (Babbar-Sebens et al 2015).

Integrated Water Resources Management (IWRM) provides a framework for including stakeholders in the planning process. IWRM is stakeholder driven, and utilizes a holistic problem-solving approach to address the dual technical and social nature of water resources problems (Castelletti & Soncini-Sessa 2006, Tortajada 2003). Applying IWRM successfully is no simple matter, however. River basins frequently cross political boundaries, so creating a successful plan requires cooperation between the planning team, local stakeholders, and several governing bodies. Often, the process will stall, as individual stakeholders reject plan after plan that meet the stated goals, but negatively impact their own interests (Tortajada 2003). An additional complication arises if there is any perceived disconnect between the modeling tool and the user. If a stakeholder believes that their experiential knowledge and judgement is being bypassed and replaced with a "black box," they are not likely to implement the resulting plan. Stakeholder education can partially mitigate this issue, but it is not a solution unto itself (Cox 1996, McCown 2002).

To maximize stakeholder investment in the design process and resulting plan of action, Castelletti and Soncini-Sessa (2006) proposed a planning procedure, Participatory and Integrated Planning (PIP). This framework is designed to allow scientists, stakeholders, and policy makers to work together to define problems, determine acceptable design and modeling methods that all involved parties can accept, and create solutions. Stakeholder involvement becomes central to the planning process, as opposed to acting as a reviewer only after all the design and modeling work has been completed. To avoid bias and increase the credibility of the final product, the stakeholders and design team must consider all possible consequences of the plan, both positive and negative. These consequences must be both articulated and quantified, if possible. This is not a simple task, as these consequences may not all be quantifiable, or well understood. The design team and stakeholders must work together to determine if these unquantifiable consequences should be simplified for analysis, or ignored in the modeling (Castelletti & Soncini-Sessa 2007).

This research is a continuation of a long-standing web based PIP project WRESTORE (Watershed REstoration using Spatio-Temporal Optimization of REsources). WRESTORE expands upon PIP procedures, and includes the biases, needs, and preferences of stakeholders into the optimization process as additional variables to optimize over (Babbar-Sebens 2014, Babbar-Sebens 2015, Piemonti 2013). The NSGA-II that WRESTORE utilizes for its optimization includes user feedback as an additional fitness function. In this way, the objective space is expanded to incorporate those unquantifiable, subjective factors that stakeholders use to judge the quality of a particular solution (Babbar-Sebens 2014).

Piemonti et al (2013) showed that a significant alteration of the Pareto front of optimal solutions occurs when user evaluations are included in the optimization process utilized by WRESTORE. Performance metrics had lower values when user input was added to the optimization process than when the algorithm only considered solution performance. Nitrate reduction performance fell between 2% and 50%, peak flow reduction was set back by 11% to 98%, and sediment reduction suffered a setback of 20% to 77%. However, the authors represented these performance losses as a tradeoff between the traditional metrics and the unquantifiable indicators that stakeholders use to evaluate solutions.

A natural extension of Piemonti's research is to add several stakeholders to the optimization process simultaneously, and subject the solutions to projected climate conditions. The optimization process should use current climate conditions, as it is easiest to evaluate the effectiveness of any proposed change when the baseline is known behavior (Castelletti & Soncini-Sessa 2006). Additionally, farmers' belief in climate change is far from universal. In the Corn Belt, which Indiana is a part of, only 66% of farmers believe in climate change. The proportion of farmers who believe that climate change is due wholly or in part to human actions is even smaller, sitting at 41% (Arbuckle et al 2012). Thus, any stakeholder involvement scheme that centers on climate change will be viewed with extreme skepticism by over half of the participants.

Skeptical stakeholders will largely be unwilling to implement climate mitigation practices (i.e. practices that reduce the emission of greenhouse gases), though they may be more willing to consider adaptive practices to mitigate the possible effects of extreme weather effects. Most common best management practices (BMPs) are considered adaptive practices under these definitions (Arbuckle et al 2012, Chatrchyan et al 2017). This arises from a slight perceived difference between the terms "climate change" and "increased extreme weather events." Many farmers have experienced a change in extreme weather events during their career, so "extreme weather events" has personal meaning to them. By contrast, many consider "climate change" an abstract concept, and do not identify with it (Chatrchyan et al 2017). Utilizing (BMPs) to mitigate the adverse effects of extreme weather effects is more appealing to agricultural land owners than acting to limit greenhouse gas emissions for these reasons.

The sole BMP utilized in this research, wetlands, can be considered an "adaptive practice" using the framework described previously, though Lal et al (2011) suggest that wetlands may also serve some mitigation functions as well, serving as carbon sinks. Floodplain wetlands have a proven record of reducing peak flows and flooding. However, wetlands in general actually have a much more nuanced effect. Upland wetlands are often flood generating features (Acreman & Holden 2013, Bullock & Acreman 2003). Acreman & Holden (2013) hypothesized that constant saturation, a common feature of many upland wetlands in headwater catchments, reduces the ability of wetlands to effectively store water during heavy precipitation, which results in saturation-excess runoff. This type of runoff often results in high peak flows and low times to peak in the receiving waterbody. However, certain topological features, such as hollows, increase the storage capacity of upland wetlands, and allow for peak flow and flood reduction even during heavy rains. Thus, the effect that wetlands have on flooding and peak flows is highly variable and site specific.

A study by Javaheri and Babbar-Sebens (2014) found that wetlands are highly effective at reducing peak flows and flood areas in School Branch, one of the main river branches in the Eagle Creek Watershed. Eagle Creek Watershed is the focus area for this research. Peak flow reductions anywhere between 20 and 41% were reported. The value of peak flow reduction increased as the baseline peak flow increased, but the percent reduction decreased. Flood inundation area was also reduced, with reductions of up to 55% reported. Flood inundation area reduction values were highest upstream, and decreased near the basin's outlet. The velocity of the flood waters was also generally decreased, with a maximum reduction of 13% reported.

According to the USDA (2011), flooding is highly detrimental to agriculture. Flooding can remove soil and destroy crops, or prevent crop harvesting. In addition, valuable equipment, livestock, and buildings may be damaged and destroyed, which can result in substantial monetary loss. Between 2005 and 2008, the value of direct agricultural premiums written globally increased from USD 8 billion to USD 18.5 billion, which is an increase of 131% (Iturrioz 2009). During this same time, the economic value of global agriculture grew by 49% (World Bank 2016). Clearly, flooding is an increasing global problem that agriculture must address, so any BMPs that can reduce flooding will benefit farmers.

4.2.1 Watershed Management as Participatory Modelling.

In a recent study, Basco-Carrera et al (2017) proposed a method for classifying Decision Support Systems (DSS) as participatory or collaborative modelling. The major distinction they proposed between the two is the extent to which key stakeholders are incorporated into the modelling process. Collaborative modelling can be considered a subset of participatory modelling. To transition from participation to collaboration involves fully incorporating key stakeholders into the model design process. In participatory modelling, these stakeholders are consulted by the main design team, and may be partially involved in the modelling process. However, the design team still performs the majority of the modelling work – the model design is not a joint effort. Collaborative modelling, by contrast, includes extensive stakeholder involvement in the model design and execution, and possibly even in the decision-making process. The design and implementation are a joint effort, with stakeholders and the design team working as equal partners.

While this research does not utilize actual stakeholders, the surrogate stakeholders utilized were developed based upon characteristics that real Eagle Creek watershed stakeholders would exhibit. Eagle Creek watershed's land use is primarily agrarian, with most of the existing and potential wetlands occurring in those areas (Babbar-Sebens et al 2015). As such, the surrogate users represent agrarian users with two main concerns modelled after actual Eagle Creek farmers as studied by Reimer et al (2012). Thus, while it is not possible to talk about actual stakeholder characteristics of this research, it is possible to describe the characteristics of the stakeholders that the surrogate users represent. In this way, the DSS this research utilizes is classified using the methods developed by Basco-Carrera et al to determine what mode of participatory or collaborative modelling is appropriate for a specific problem (2017). The classification parameters are shown in Table 5.

By nature, watershed management plans must be a joint action type of cooperation. This requirement arises from a basic truth of watershed management plans. Implementation of these plans will require land owners to install BMPs on their own property. In this research, many of the potential wetland sites specified are actually former wetlands. The wetlands were drained and piped to improve agricultural activity. These former wetlands are now productive farmland under private ownership and operation (Babbar-Sebens et al 2015). Conversion of these sites back to their former state will require cooperation from the local land owners. Thus, these individual stakeholders are each decision makers for their particular portion of the overall management plan. This corresponds to key stakeholder involvement in the decision-making stage, which is the top rung of the ladder of participation represented in Figure 8. The figure identifies this cooperation type as "Joint Action," where key stakeholders are included in a collaborative modelling process, and other stakeholders are included in a participatory fashion. According to this modelling framework, the key-stakeholders – the ones who will decide whether to implement the management plan – should be involved in the development of the model and execution of the plan. Other interested stakeholders are part of the discussion process, but do not necessarily need to participate in the model's development and implementation (Basco-Carrera 2017).

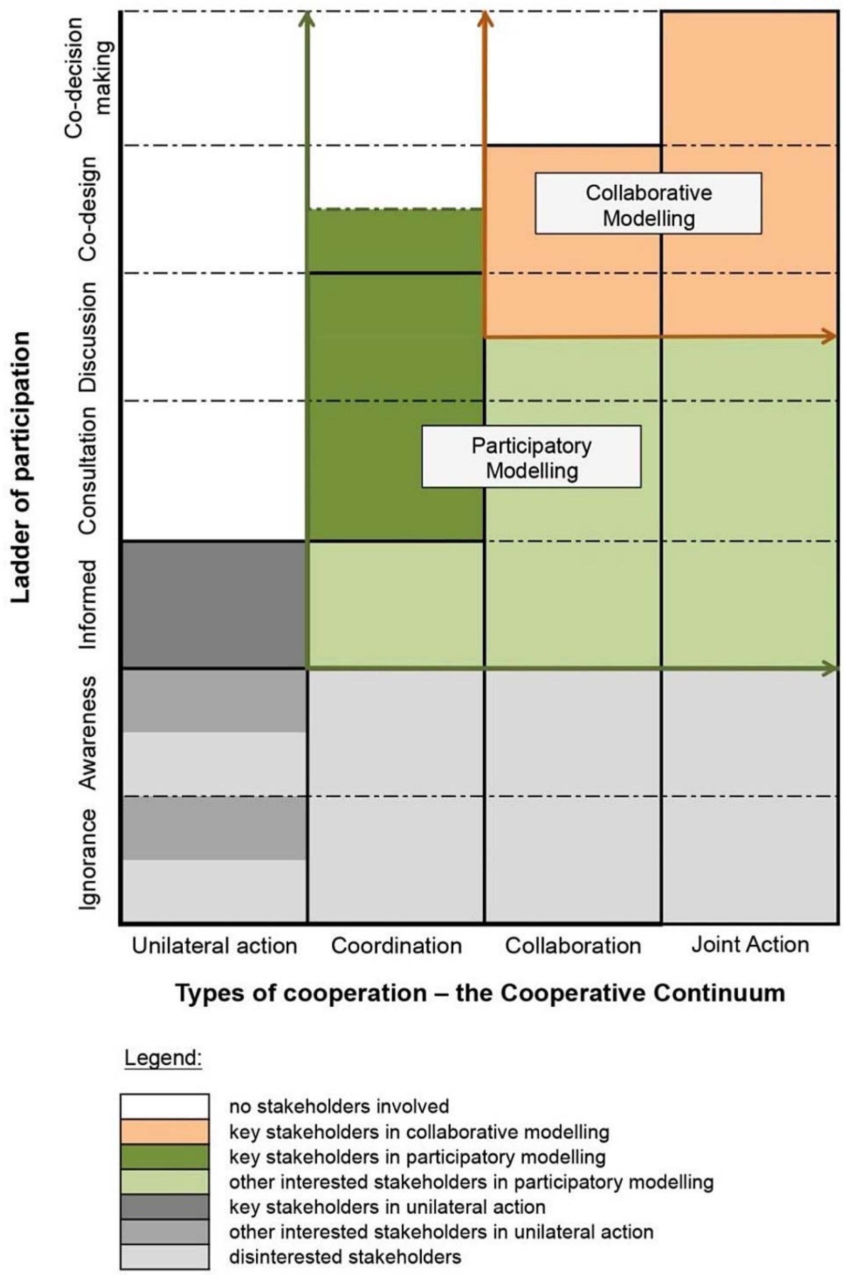

Figure 8: Types of Participatory and Collaborative Modelling (source: Basco-Carrera et al, 2017).

Basco-Carrera et al (2017) offer descriptions of typical parameter values for cooperative and participatory modelling. In general, the parameters described in Table 5 match the expected descriptions for cooperative modeling, but there are some key differences:

1. The modelling tool utilized is not related to the technical skills of the stakeholders. Cox (1996) notes that the majority of agrarian DSS users do not have extensive technical skills related to modelling, so selecting a modelling tool based upon their expertise is likely not possible.
2. The method of stakeholder involvement is not direct, but rather expert facilitated. Just as in point 1, the technical skills of the stakeholders do not allow for them to directly interact with the SWAT model (specify inputs). They do, however, interact directly with the WRESTORE interface, and influence the nature of the inputs by evaluating potential solutions that the optimization algorithm generates.
3. Model selection and construction was performed entirely by the design team. An eventual goal for WRESTORE is to have stakeholders select a hydrologic model specify important input factors for optimization (Babbar-Sebens et al 2015). This tool is still in the early phases of development, however, so the design team selected a hydrologic model they are familiar with.

Overall, this research utilizes a collaborative modeling framework. There are minor deviations from the framework that Basco-Carrera et al (2017) developed, but this is likely due to the difference between the stakeholders they used in their research versus the stakeholders of this research. Basco-Carrera et al applied their framework to a regional groundwater model development project, where the major stakeholders were required to have a baseline level of hydrologic knowledge to contribute. The explicit goal of the project was to develop a unified groundwater model, implying that effective stakeholder inclusion in the project required extensive involvement in the model development process.

By contrast, the stakeholders in this research are purely laypeople, and the end goal is a management policy as opposed to a usable regional model. The difference in goals and stakeholders require slight deviations from the framework as laid out by Basco-Carrera et al, but the overall framework is still useful in defining this problem.

Table 5: Case Study Classification

Factor	Parameter	Description	References
Context and Application	Scale of Action	Predominantly rural Eagle Creek watershed in Indiana, with urban areas near the watershed outlet	(Babbar-Sebens et al 2013)
	Domain	Optimization of watershed management plans using interactive genetic algorithms	
	Interaction Context	Collaborative approach, with a focus on cooperative interaction	
	Problem Structure	Semi-structured to unstructured	
Specific Use	Collaborative Modelling Purpose	Generate best compromise solutions based upon user ratings by interactive optimization that maximize user satisfaction	
Information Handling	Model Type and Software Platform	Genetic algorithm (java language) incorporating a SWAT model. Eventual integration into WRESTORE architecture (web-based interactive optimization tool)	
	Information Delivery Medium	Currently tabular. WRESTORE utilizes GIS and graphical visualization methods	(Babbar-Sebens et al 2015)
Stakeholder Involvement Structure	Type of Cooperation	Joint action	
	Stakeholders Involved	Participation is required for implementation of a management plan. Minimal technical knowledge is required.	
	Participatory Methods	Collaborative	(Babbar-Sebens et al 2015)
	Model Users	Direct interaction with optimization algorithm. Indirect selection and setup of hydrology model	
	Timing of Participation	Model development and selection was performed by research team. Eventual WRESTORE goal is to allow watershed groups to select model and optimization parameters	(Babbar-Sebens et al 2015)
	Participation Mode	Direct interaction with genetic algorithm. WRESTORE utilizes workshops with real stakeholders	(Babbar-Sebens et al 2015)
	Level of Participation	Users direct optimization process, and have final authority to approve or deny a proposed plan	
Modelling and Organizing Team	Team	One principal investigator and one graduate student. WRESTORE project includes IUPUI computer science department and OSU civil engineering department	
	Skills	Knowledge of evolutionary algorithms and hydrologic modelling	
Means	Financial Resources	NOAA Funded	
	Timing	Approximately one year	

4.2.2 Purpose of Study

The research presented here seeks to answer three main questions:

1. How do the solutions that stakeholders prefer differ from solutions that are not preferred?
2. How does including multiple stakeholders in the optimization process alter the performance of the generated solutions? What do the new preferred solutions look like?
3. How do the solutions preferred by stakeholders perform when subjected to the projected climate?

To answer these questions, the research presents methodology to represent stakeholders algorithmically, and incorporate these surrogate users into the Genetic Algorithm (GA) framework utilized by Piemonti (2013). This framework is not currently configured to accommodate multiple users at once, so this research will provide a framework within the GA to accomplish this. Babbar-Sebens et al (2015) did allow for synchronous multi-user operation of WRESTORE by implementing a democratic user that decided the consensus rating of a design, but this process was not included directly in the GA.

The GA used by this research is an Interactive Genetic Algorithm with Mixed Initiative Interaction (IGAMII) that utilizes an NSGA-II optimization process coupled to a SWAT hydrologic model (Babbar-Sebens 2012, Babbar-Sebens et al 2015, Piemonti et al 2013). The run time of this algorithm is not insubstantial, and scales directly with the number of years modelled in SWAT. To minimize the run time for the model, the authors utilized a method for selecting single model year realizations with high peak flows developed by Cannady-Shultz and Babbar-Sebens (2017).

4.3 Methodology

4.3.1 Study Area

Eagle Creek watershed is located northwest of Indianapolis, and features agricultural, urban, and undeveloped land uses. This watershed has a drainage area of approximately 419 km^2 and is part of the larger Upper White River Watershed. Eagle Creek drains to Eagle Creek reservoir, which is a major source of drinking water and recreation for Indianapolis and the surrounding areas. Over 60% of the watershed is agricultural, with the dominant crop being corn. The majority of the possible sites for wetlands in this watershed are on agricultural land, so many of the individuals who would participate in any watershed management plans utilizing only this BMP are farmers (Babbar-Sebens et al 2013, Javaheri & Babbar-Sebens 2014, Piemonti 2013).

The surrogate users utilized in this research represent agrarian users with two main concerns: environmental and financial. Reimer et al (2012) identified these two concerns primary motivators when farmers in Eagle Creek consider implementing BMPs. In addition to these two primary motivators, the study found that farmers also consider on and off farm benefits when making their conservation decision. Typically, individuals who self-identify as stewards of the land will concern themselves with both on and off farm benefits, while financially driven individuals will only consider benefits on their own property. However, no quantitative link between the two primary motivators and the on and off farm benefit considerations was found in the study, so the surrogate users this study will employ treated the motivators and location focus as independent factors.

4.3.2 Climate Model Selection

The climate model year selection methodology discussed in this section is predicated upon work completed by Walters (2016). She explored the performance of the potential wetlands specified by Babbar-Sebens et al (2013) in the Eagle Creek Watershed when subjected to the potential mid-century climate as predicted by six different North American Regional Climate Change Assessment Program (NARCCAP) regional climate model (RCM) atmosphere-ocean general circulation model (AOFCM) pairings, referred to from here on as the "climate models," developed for North America, detailed in Table 6 (Mearns et al 2009, Walters 2016). Each climate model was used to simulate precipitation and temperatures for two time periods: 1970-2000, and 2040-2070 ("mid-century"). Walters (2016) applied four different bias correction techniques to each climate model to calibrate their behavior to observed data. The methods were linear scaling (LS), local intensity scaling (LOCI), power transformation (PT), and distribution mapping (DM). The un-adjusted models (RAW) were also tested. Based on the predictive skill of the different bias corrected or un-adjusted models, 11 models (listed in appendix A) were selected to comprise a final climate model ensemble for the Eagle Creek watershed.

Table 6: Summary of Climate Models

RCM		AOGCM	
Model	**Supporting Agency**	**Model**	**Supporting Agency**
CRCM	Canadian Centre for Climate Modelling and Analysis	**CCSM**	National Center for Atmospheric Research
		CGCM3	Canadian Climate Centre for Climate Modelling and Analysis
HRM3	Met Office Hadley Centre for Climate Science and Services	**GFDL**	Geophysical Fluid Dynamics Laboratory
MM5I	National Center for Atmospheric Research/Pennsylvania State University	**CCSM**	National Center for Atmospheric Research
RCM3	Abdus Salam International Center for Theoretical Physics	**CGCM3**	Canadian Climate Centre for Climate Modelling and Analysis
WRFG	National Center for Atmospheric Research	**CGCM3**	Canadian Climate Centre for Climate Modelling and Analysis

The link between extreme precipitation and flooding is tenuous and inconsistent. Madsen et al (2014) performed an extensive review of papers analyzing extreme precipitation and hydrological floods in Europe. These studies reported findings based upon climate projections and historical trends. The review reported two general trends from the literature: extreme precipitation events are becoming more prevalent, but there are no indications of a similar trend in severe flooding. Some regions have observed increased flooding, but others have observed apparent decreases. These results were also observed in a global study that Dankers et al (2014) performed. Nine global hydrology and land surface models (seven that are part of the Water Model Intercomparison Project, as well as the PCRaster Global Water Balance Model and the Water Balance Model) were used in this research, each operating within a global 0.5° grid. Each of these hydrologic models were driven by five different Coupled Model Intercomparison Project Phase 5 (CIMP5) climate model realizations, and the 30-y return period stream flows were analyzed. Approximately half of the grid points showed an increase in extreme stream flows, but approximately one-third showed a decrease. Again, the areas that showed decreased extreme flows were largely in areas where streamflows are dominated by snowmelt, but this trend was not exclusive.

Several more localized studies have shown similar behaviors. Other locations with snowmelt dominated streamflow regimes, such a Minnesota (Novotny & Heinz 2007) and the Columbia River at the Dalles, Oregon (Arpita & Mujumdar 2016), exhibited similar flat to decreasing trends in extreme river flows. In Arpita & Mujumdar's (2016) study, however, some possible worst-case scenario simulations did show a possible detectible increase in extreme river flows by 2100. In one case, they noted that a detectable increase in the magnitude of the 75-y flow event could happen as early as 2027. Several studies also found an increase in flood risk. A study in the Kemptville Creek watershed in Ontario, Canada found that the exceedance probability for floods ranging from a 2-y to 100-y event increased 34.3% and 17.8% (respectively) for a dam with a 20-year service life (Seidou et al 2012). Jena et al (2014) found that recent increases in high floods in the Mahandi basin in India are linked to increased extreme precipitation in the middle of the basin. Wu and Huang (2015) observed a more nuanced relationship between extreme precipitation and high peak flows. Accounting for changes in peak flows induced by human activity, increasing precipitation lead to increased flooding in June and July, but September increases in precipitation did not. These studies, when viewed in conjunction with the larger studies by Dankers et al (2014) and Madsen et al (2014), all suggest that extreme precipitation events do not necessarily lead to flooding. Thus, if the goal is to select climate models that produce high peak flows, selecting a climate model ensemble based upon precipitation characteristics will not be sufficient. Stream flow characteristics must also be included in the selection process.

Chapter 3 detailed a process for selecting climate model year realizations that have diverse forcing climate factors. The selection process involves four basic steps:

1. The full set of possible climate year realizations is reduced to the set containing the five highest peak flows from each parent model.
2. Four extreme corners for cumulative streamflow and precipitation are selected based upon the 90th and 10th percentiles for these factors. The ensemble size is reduced here. The 8 models nearest each corner are selected to progress to the next step as four different sets (4 sets of 8 progress).
3. Five metrics, maximum five day peak flow and cumulative precipitation, maximum number of consecutive dry days, cumulative rainfall from all days with precipitation above the 95th percentile, and cumulative precipitation are calculated for each model, and the ordinal difference between the past and projected values of these metrics are evaluated. Each model is then assigned five ranks based upon the difference observed in each metric relative to all models in its corner, and awarded five scores equal to their rankings. These scores are averaged, and the four from each corner with the highest score progress to the final step.
4. The skill of each ranked model to predict flow and precipitation characteristics are assessed and averaged, and the five with the highest score become the final ensemble. In their research, Cannady-Shultz and Babbar-Sebens (2017) selected four models. However, Knutti et al (2010) note that there are benefits of including

up to five models in a single ensemble, so this research will follow their recommendation.

4.3.3 Simulated Stakeholders

4.3.3.1 Simulating Stakeholder Preferences with a Scoring Equation

In the absence of actual stakeholder data specific to Eagle Creek, this research developed a method to simulate these stakeholders according to the typical characteristics explored by Reimer et al (2012). As mentioned in section 2.1, Eagle Creek farming stakeholders are defined by two sets of motivators. The first motivator is whether financial gain or environmental considerations drive their conservation behaviour, and the second is whether they primarily consider on or off-farm benefits when making their decisions.

To model stakeholders with these two different motivators, this study utilized a simple scoring equation (equation 12) that represents their relative preference for a specific design alternative as a number between 0 and 1, where a higher score corresponds to a higher preference.

$$\boldsymbol{Score_{ij}} = \boldsymbol{w}\left(\boldsymbol{\gamma}\frac{\boldsymbol{PFR_{ij}}}{\boldsymbol{IdealPFR_j}} + (\mathbf{1}-\boldsymbol{\gamma})\frac{\boldsymbol{IdealArea_j}}{\boldsymbol{Area_{ij}}}\right) + (\mathbf{1}-\boldsymbol{w})\left(\boldsymbol{\gamma}\frac{\boldsymbol{PFR_{it}}}{\boldsymbol{IdealPFR_t}} + (\mathbf{1}-\boldsymbol{\gamma})\frac{\boldsymbol{IdealArea_t}}{\boldsymbol{Area_{it}}}\right) \qquad (8)$$

$Score_{ij}$ is calculated for every stakeholder j for all i design alternatives. The stakeholder score is determined as a weighted average of peak flow reduction (*PFR*) and wetland area (*Area*) fitness function values calculated as part of a study performed by Garrison (2016), where the goal is to maximize *PFR* and minimize *Area*. A subscript j attached to *PFR* or *Area* means that the fitness function value corresponds to the specific subbasin that stakeholder j resides in, while a subscript of t means that the fitness function value was calculated for the entire watershed. The word "*Ideal*" preceding *PFR* or *Area* defines the ideal fitness function values corresponding to the best-case scenarios of all wetlands built or no wetlands built, respectively. A stakeholder's preferences are defined by values of w and γ, which define their preference for local over watershed benefits and environmental benefit over financial gain, respectively. Each preference variable can take any value between 0 and 1, and is assigned randomly for each stakeholder from a uniform distribution. The study by Reimer et al (2012) indicated that there may be some correlation between the values of w and γ, but as the nature of this interdependence is not specified in their study, this research will assume that w and γ are independent. As better quantitative data for Eagle Creek Watershed stakeholder conservation behavior becomes available, the relationship between w and γ can be determined in a fuzzy or deterministic fashion, as appropriate.

WRESTORE uses a likert scale of 1 to 3 to quantify user preference (Piemonti et al 2013). In order to convert the score calculated by equation 7 to a likert scale rating, the following relations were defined:

- $Score_{ij} < 1/3$ corresponds to a likert scale rating of 1
- $Score_{ij} > 2/3$ corresponds to a likert scale rating of 3
- Any other value of $Score_{ij}$ corresponds to a likert scale rating of 2

This research aims to identify design alternatives that are robust when subjected to several different extents of stakeholder participation in implementing watershed management plans. At one extreme, all stakeholders could participate, and allow BMPs to be built on their property. Alternatively, all stakeholders could oppose the management plan, and refuse to construct BMPs on their land. Varying intermediate degrees of stakeholder participation are possible, and the distribution of participants and opponents within the watershed can also vary.

To simulate the varying degrees of stakeholder participation in a Monte Carlo fashion, each simulated stakeholder is randomly designated as a participant or opponent of the plan. If they are designated an opponent of the plan, the simulated stakeholder will automatically award any design alternative that requires them to build a wetland on their property a rating of 1. Nine different extents of user participation are modelled, meaning that each simulated stakeholder receives a set of nine participation designations. The nine different designations are defined as follows:

1. Participation rate = 0%. Complete opposition from the stakeholders.
2. Participation rate = 12.5%. Almost complete opposition from stakeholders.
3. Participation rate = 25%. Opposition from stakeholders.
4. Participation rate = 37.5%. Weak opposition from stakeholders.
5. Participation rate = 50%. Equal opposition and support from stakeholders.
6. Participation rate = 62.5%. Weak support from stakeholders.
7. Participation rate = 75%. Support from stakeholders.
8. Participation rate = 87.5%. Almost complete support from stakeholders.
9. Participation rate = 100%. Complete support from stakeholders.

To address the many different possible distributions of supporters and opponents that can arise from scenarios 2 to 8, 100 different stakeholder distributions of these nine different support scenarios are considered. Thus, each stakeholder rates a single design alternative 90 times.

4.3.3.2 Combining Individual Stakeholder Scores to Determine the Best Design Alternatives

Work by Babbar-Sebens et al (2015) included a procedure for determining the overall ranking of a design alternative that was ranked by several individuals by utilizing a democratic user. Each user preference rating becomes a "vote," and the rating value that is awarded the most often becomes the overall rating for the design alternative. This system successfully integrates the opinions of several users into a single rating, but it also fails to fully represent the varied opinions of each individual. Consider two hypothetical design alternatives being evaluated by 9 people. The first alternative is rated as a 3 by four people, and as a 1 by five people, while the second receives five 3 ratings and four 1 ratings. According to the democratic user, the first design receives an overall rating of 1, and the second receives an overall rating of 3, indicating that the first design is actively liked by the group, while the second design is disliked. The truth, however, is that each alternative is highly controversial, with nearly equal opposition and preference, which the democratic user does not represent.

To remedy this issue, this study instead uses the number of three rankings awarded to each design alternative to represent the consensus opinion. The design alternative that receives the most three ratings is designated as the most preferred alternative, while the design that receives the least three ratings is designated the least preferred alternatives. All remaining designs are ranked intermediately according to the number of three ratings they received, resulting in a set $P = \{V_1, V_2 \dots V_n\}$ for a set of n design alternatives, where V_i is the ranking corresponding the i^{th} alternatives number of three ratings, as well as the raw set of three rating counts *C3*. This approach offers two advantages over the method employed by Babbar-Sebens et al (2015). First, this measure more accurately represents the extent of user agreement by quantifying the extent of user preference for a specific design. Second, the measure offers a higher-resolution picture of user preferences. With a simple democratic user, a design that receives one more three rating than two's or ones and a design that is universally rated as a three receive the same final rating of three. However, the measure proposed by this research would reflect the difference between these two designs, with the second alternative being ranked higher than the first.

Because of the Monte Carlo nature of the simulated stakeholder ratings, each individual simulation is considered a single rating, and thus each three rating an individual vote. So, each user can award up to 900 three rating to a design alternatives, and the total number of possible three votes that any design alternative can receive is equal to the product of 900 and the number of simulated stakeholders. In this research, 108 stakeholders are simulated (one for each subbasin in the Eagle Creek Watershed with a potential wetland), meaning that a design alternative can receive a maximum of 97,200 three ratings.

The method this study uses to represent the consensus opinion of multiple users for a specific design alternative can provide a highly-detailed preference rating for the design, but it does not account for the extent of disagreement among users. In their research on multiperson decision making problems (MPDM), Herrera-Viedma et al (2002) developed a consensus model based upon two criteria: consensus and proximity. Consensus measures the extent of agreement on the rating of a specific alternative, and proximity measures how close an individual user's rating is to the consensus rating. This research adapts their measure of consensus, which is calculated for every alternative, as a supplementary measure to the count of threes for evaluating a design alternative for user preference.

The consensus measure proposed by Herrera-Viedma et al (2002) begins by defining *m* sets $P_i = \{V_1^i,\ V_2^i \dots V_n^i\}$ for each stakeholder i, which will use the same method used to generate the set P. Each user awards different alternatives a different number of three ratings, which will be used to generate the set P_i. The proximity of each expert to the consensus rating for every alternative x_j is calculated using equation 13.

$$p_i(x_j) = \left(\frac{\left|V_j^i - V_j\right|}{n-1}\right) \tag{9}$$

The consensus for the design alternative x_j is then calculated using equation 14.

$$C(x_j) = 1 - \sum_{i=1}^{m} \frac{p_i(x_j)}{m} \tag{10}$$

$C3\{\ j\ \}$ and $C(x_j)$ are each treated as an objective function, and used to perform a Pareto search of the set of solutions generated by Garrison (2016) to find the front of preferred solutions. These will be used in addition to the peak flow reduction and wetland area reduction objective functions used in that same research. Garrison's (2016) optimization also utilized sediment reduction and nitrate reduction objective functions, but computing limitations required this research to remove those functions from the interactive process. The assumption underlying this decision is that total nitrate and sediment reduction are correlated strongly to peak flows: this assumption will be tested both on the data this research produces, as well as on Garrison's (2016) data.

4.4 Results

4.4.1 Climate Models

The climate models selected using the selection method suggested by Cannady-Shultz and Babbar-Sebens (2017) recommended the following three models to best represent the extreme projected climate: MM5I CCSM DM 2057, MM5I CCSM Raw 2055, MM5I CCSM Raw 2062, MM5I CCSM Raw 2065, and WRFG CGCM3 LOCI 2058. The temperature and precipitation predictions from these models will be used in conjunction with a SWAT model to evaluate the response of the different watershed management solutions to the projected climate. These models have an average annual cumulative precipitation of 1271 mm, and an average maximum daily peak flow of 172 cms.

Figure 9 presents the cumulative annual precipitation of the five models and the present climate, and figure 10 presents this same data normalized. The normalized data in figure 10 suggests that the selected climate projections differ from the present climate when considering precipitation timing. During the range of dates representing summer through fall (approximately June to November), two climate models, WRFG CGCM3 LOCI 2058 and MM5I CCSM RAW 2055, depart from the present data. The former model has a drier summer, delaying precipitation until late September into October, and the latter has a wetter summer but drier fall. The MM5I CCSM DM 2057 model exhibits similar behavior to MM5I CCSM LOCI 2055 model, but the rainfall is delayed approximately sixty days, to the late summer. The remaining models generally do not deviate from the observed precipitation pattern. As figure 9 suggests, the present cumulative precipitation value is within the range of cumulative precipitation values predicted by the climate models.

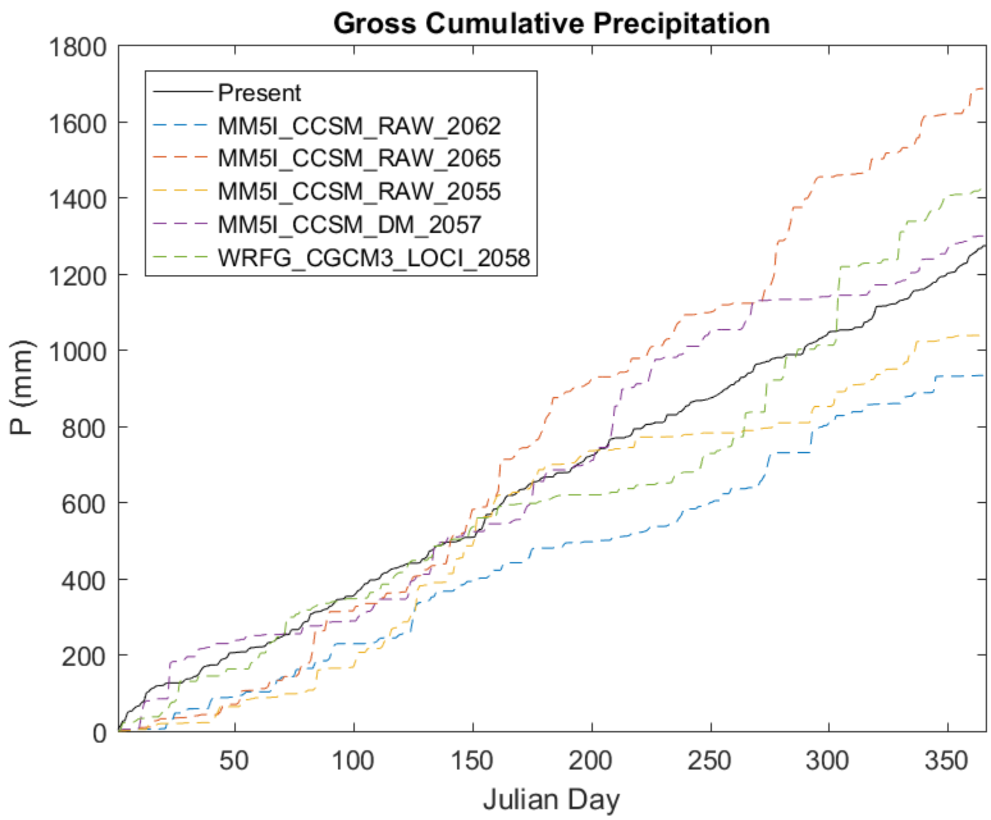

Figure 9: Annual Cumulative Precipitation for Present Climate and Selected Climate Models

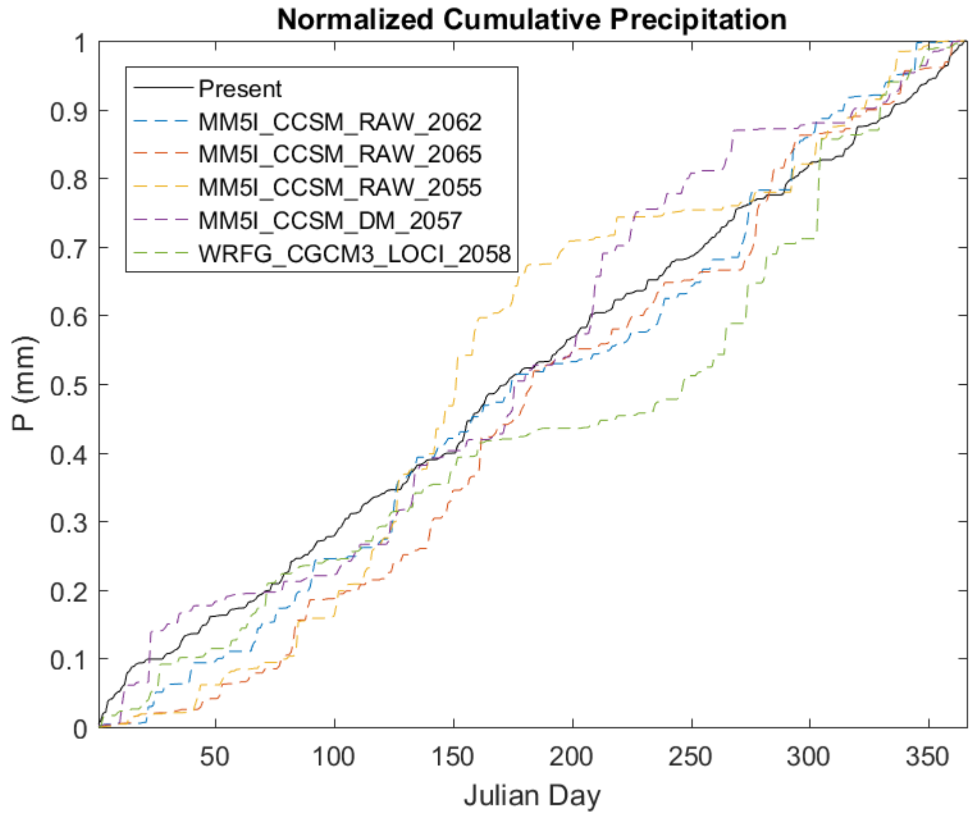

Figure 10: Normalized Cumulative Precipitation for Present Climate and Selected Climate Models

4.4.2 Non-Interactive Optimization

The optimization completed by Garrison (2016) produced solutions that began with higher consensus in earlier generations, then converged towards solutions with lower consensus in the final generations. The proportion of solutions that received ratings of three ("I like it") from the users followed an opposite pattern. Initially, the solutions had a low proportion of solutions that were favored by the simulated users, but this proportion increased through the generations of optimization, converging to a higher value in the final generations. These patterns are shown in figure 11a. The initial values for consensus and proportions of three ratings were 0.9016 and 0.1973, respectively, and the generation values for these same metrics were 0.5359 and 0.3299, respectively. Figure 11b shows the full array of user proximities by generation. Initially, user proximities were relatively uniform and high. As the generations progress, the individual user proximities to the consensus rating for solutions became more varied, and the average value decreased. Initially, the user proximities were characterized by a mean of 0.9016 and a standard deviation of 0.0088; by the final generation, these values were 0.5359 and 0.0992, respectively. The combination of the proximity, consensus, and percent of three ratings patterns in this data suggests that the initial solutions were universally unpopular, with most users in agreement with the overall quality of the offered solutions, and also rating all of the individuals similarly. Optimization introduced dissent into the simulated users, resulting in lower consensus and proximity values. Overall, the users were giving more solutions three ratings, but they could not agree on which solutions were the best. The final product of this optimization then are solutions defined by contention. The final solutions are viewed as better than the initial solutions, but there is very little agreement on which solutions are actually preferable, so implementation of the "optimal" solutions may not be possible.

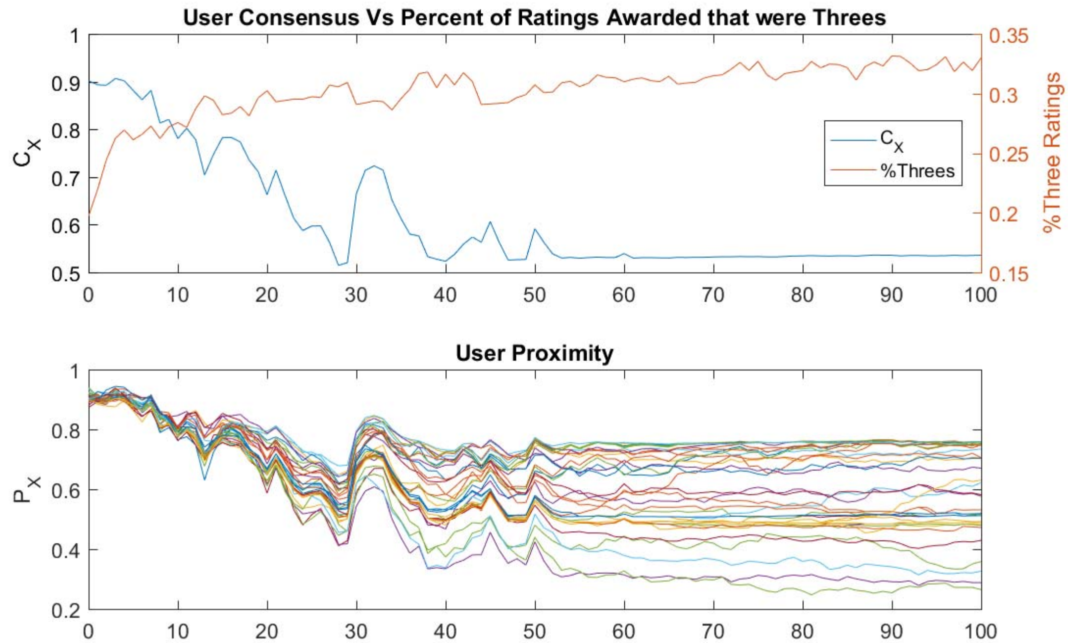

Figure 11: User Rating Characteristics by Non-Interactive Optimization Generation. a. (top): Overall User Ratings and Percent of Three Ratings. b. (bottom): Individual User Proximities to the Consensus Opinion.

Garrison's (2016) optimization solutions converge towards solutions with higher reduction of peak flows and lower wetland areas, as indicated by figures 12a and 13a, respectively. However, the downward trend in wetland area is muted, and much smaller than the overall variability of the wetland area fitness function shown in figure 13a. Figure 12b indicate that most subbasins converge towards higher peak flow reductions, though there is a plurality of these solutions that converge towards less optimal values for this fitness function. Figure 13b tells a more complicated story. Many subbasin wetland area fitness functions actually converged to less optimal values than they began with or reached at some point during the optimization, but a small number of subbasins converged towards values that were much smaller (in some cases as low as 40%) than the largest value attained during optimization. This can be thought of as the algorithm identifying the subbasins containing wetlands that are the most effective at reducing peak flows. Those subbasins that have the greatest impact on peak flow retain higher values of wetland area, while those that are not contributing as strongly to reducing peak flows are assigned lower wetland area values. This dynamic may explain why watershed scale wetland area fitness functions do not get reduced that drastically, and remain so variable. Certain subbasin wetlands are essential for reducing peak flows, so their areas cannot be reduced substantially without disproportionately sacrificing peak flow reduction.

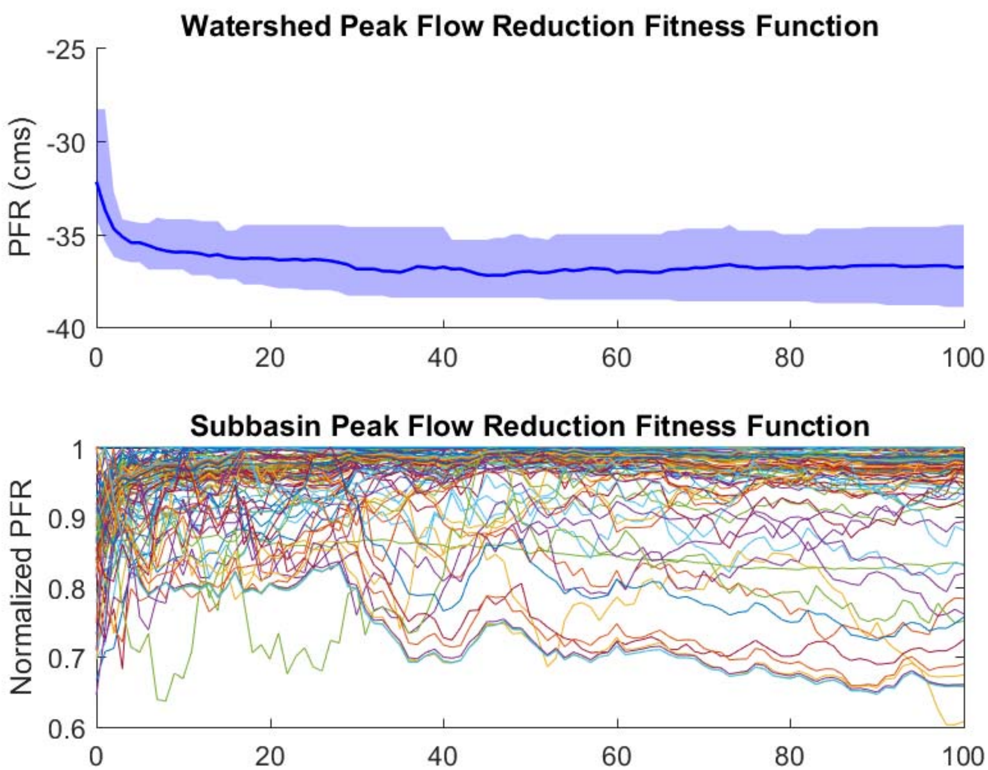

Figure 12: Peak Flow Reduction Fitness Function by Non-Interactive Optimization Generation. a. (Top): Fitness Function with Maximum and Minimum Shown. b. (Bottom): Subbasin Peak Flow Reduction Fitness Function Normalized to the Minimum Value of each Individual Solution.

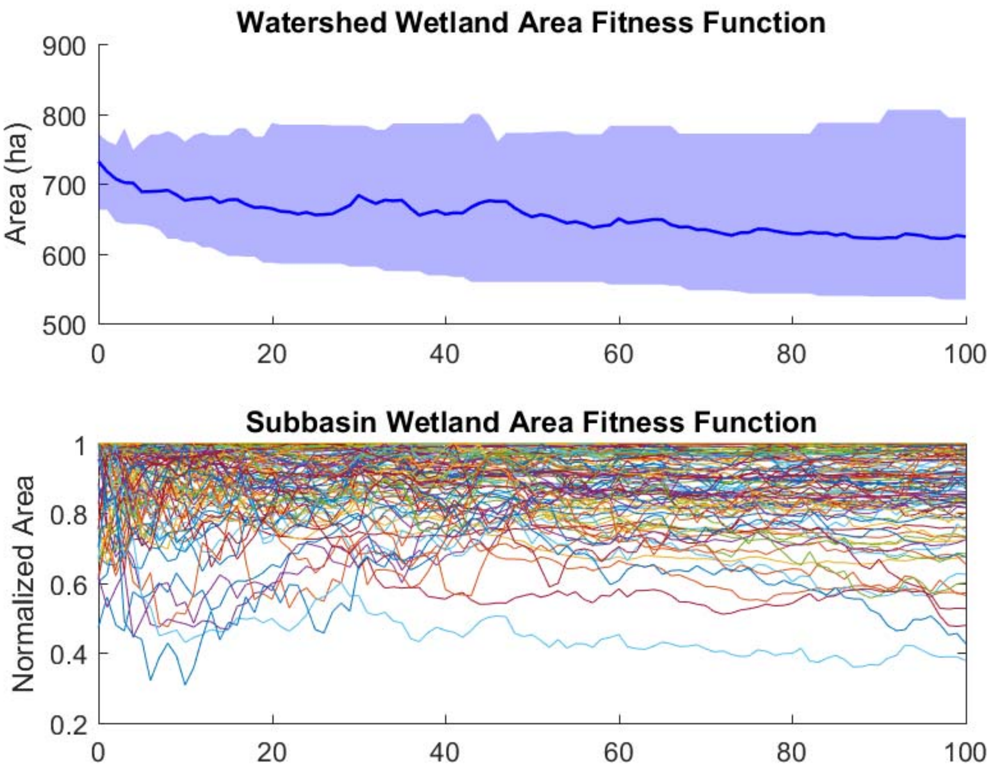

Figure 13: Wetland Area Fitness Function by Non-Interactive Optimization Generation. a. (Top): Fitness Function with Maximum and Minimum Range Shown. b. (Bottom): Subbasin Wetland Area Fitness Function Normalized to the Maximum Value of each Individual Solution.

4.4.3 Interactive Optimization

Incorporating user consensus and ratings into the optimization process altered the behavior of the solutions as the generations progressed, as represented by figure 14a. The random seed of this round of optimization started with a lower consensus than Garrison's (2016) initial population, but an approximately equal proportion of three-rated solutions. Unlike the non-interactive optimization, however, user consensus generally increased as the generations progressed, and the proportion of solutions awarded a user rating of three ("I like it") stayed relatively constant. User consensus had an initial value of 0.5167, and increased to 0.6105 by the final generation. The percentage of three-rated solutions initially was 0.1746, and had only increased to 0.1988 by the final generation. There is a sudden change in consensus and three ratings around generation 30, but this is likely due to an irregularity with the optimization process that occurred within this generation. The computer cluster that was performing the optimization was accidentally shut off, and the process had to be re-started. The recovery data from generation 30 was used to re-initialize the genetic algorithm, and the remaining 70 generations were completed without interruption. This interruption, however, did interrupt the algorithmic search routine, and several factors, such as random number seeds, were re-set, which could explain this irregularity. Judging from figures 14a and 14b, the effect of this interruption had largely dissipated by generation 55.

User proximity increased in average value and variation from the initial generation until approximately generation 15, then remained largely unchanged, as suggested by figure 14b. Initially, user proximity was characterized by a mean of 0.5167 and standard deviation of 0.0322. By the final generation, the mean increased to 0.6105, but the standard deviation had also increased to 0.0961. Similar to the behavior shown in figure 11b, the overall ordering of user proximity values does not vary substantially from generation to generation after the initial generations. Once the evolutionary process has become well established after the somewhat chaotic initial generations, it appears that the users who contribute the most and least to consensus do not change substantially, apart from a substantial perturbation near generation 30 due to the interruption in computing.

The combination of proximity, consensus, and percent of solutions awarded a user rating of three trends suggest that the optimization process incorporating user opinions and consensus does produce designs that have broader consensus. While the proportion of three ratings remains largely unchanged by the optimization process, user consensus steadily increases from generation to generation. While the individuals that were initially the primary sources of contention (lack of consensus) remain the same through the end of optimization, overall proximity is generally increased. The simulated users agree more strongly on the relative quality of the solutions. This represents an improvement over the non-interactive consensus of 13.9%.

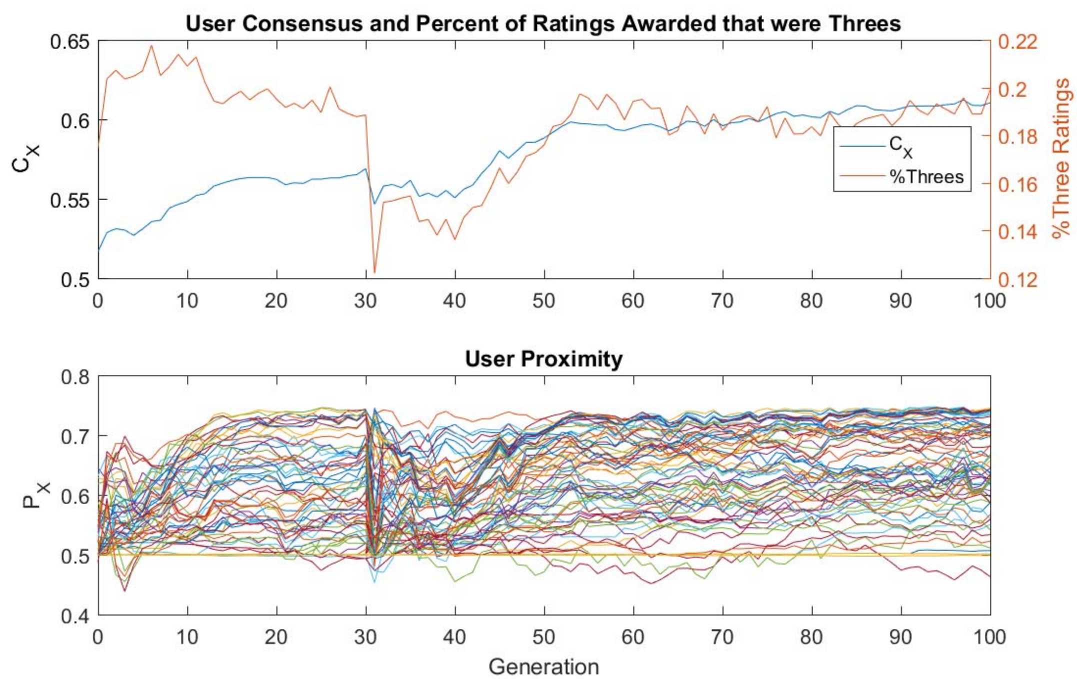

Figure 14: User Rating Characteristics by Interactive Optimization Generation. a. (Top): Overall User Ratings and Percent of Three Ratings. b. (Bottom): Individual User Proximities to the Consensus Opinion.

Peak flow reduction fitness function values decreased slightly from the initial generation to the final generation, as indicated by figure 15a, though the variation from generation to generation is minimal, with the exception to the sudden spike near the generation 30 interruption. The initial peak flow reduction value was -31.33 cms, and only slightly increased to -31.19 cms, though this represents an improvement from a sudden spike of -29.137 cms that occurred at generation 32 (shortly after the interruption). Shortly after the interruption at generation 30, figure 15b indicates that the variation in subbasin peak flow reduction suddenly increased, and only recovered after generation 55. As occurred during the non-interactive run, individual subbasin peak flow reduction fitness function values typically converged to near their optimal value by the final generation, though several individual subbasin values remained below their maximum value attained at some point during the optimization process. Wetland area fitness function values decreased from a high initial value of 729.9 ha to a final value of 643.4 ha. Variation between solutions within each generation remained large, but the final area value is smaller than any of the solutions initially generated. Again, like 14 the non-interactive optimized solutions, several subbasins had near maximum wetland areas, while others were substantially lower than the maximum value attained previously during optimization.

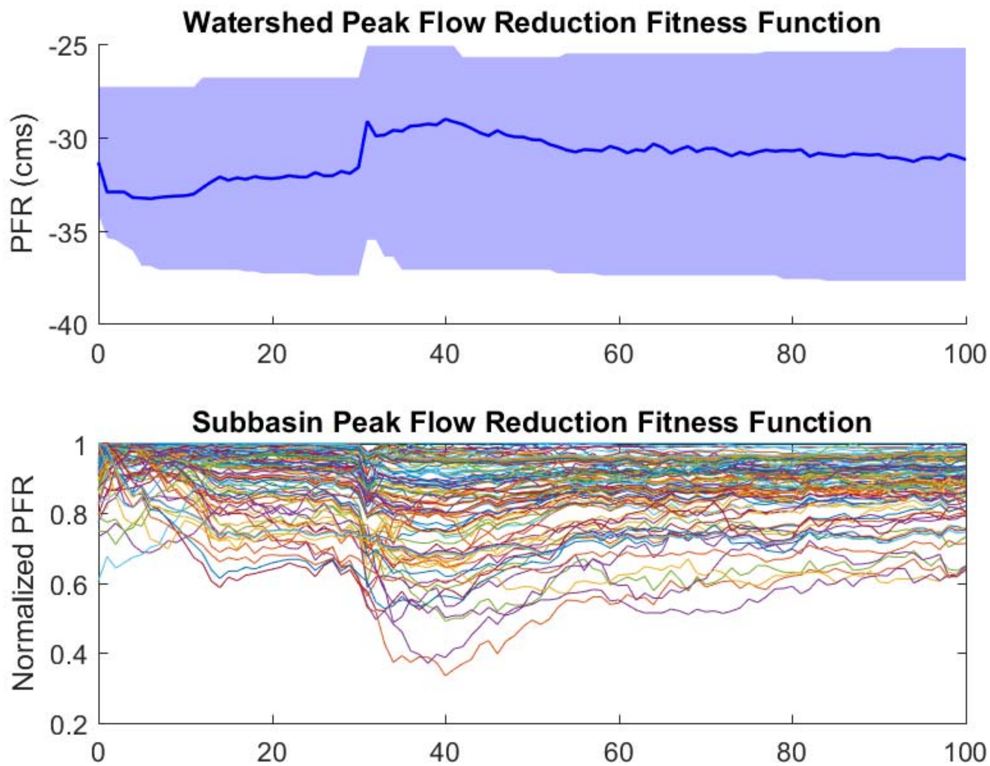

Figure 15: Peak Flow Reduction Fitness Function by Interactive Optimization Generation. a. (Top): Fitness Function with Maximum and Minimum Shown. b. (Bottom): Subbasin Peak Flow Reduction Fitness Function Normalized to the Minimum Value of each Individual Solution.

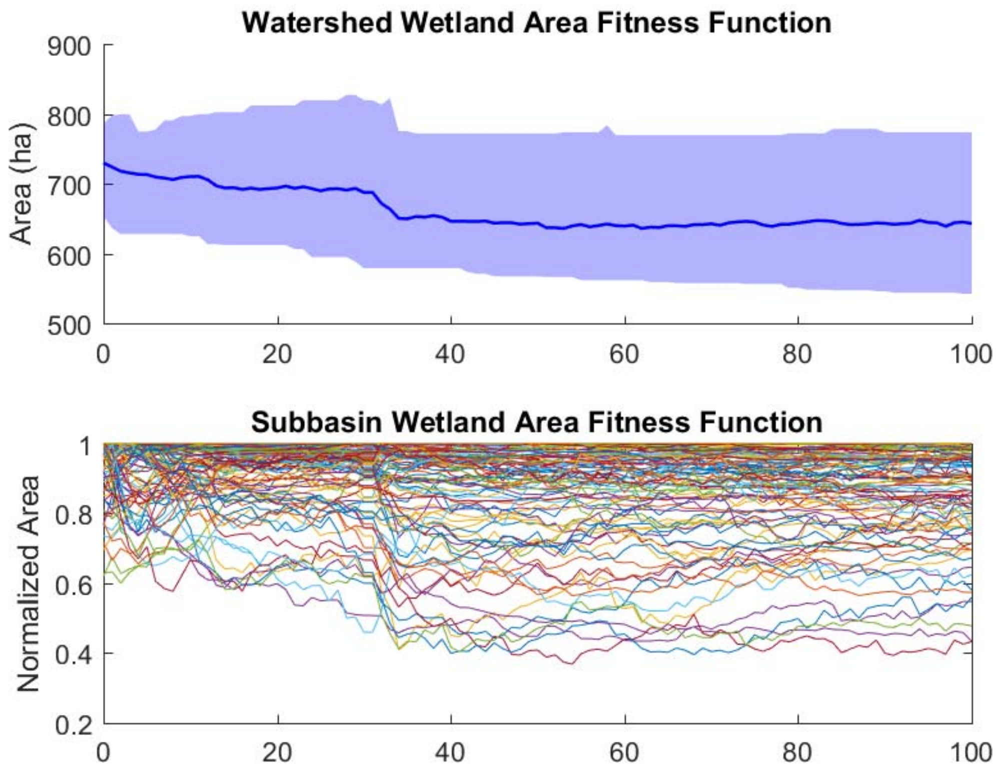

Figure 16: Wetland Area Fitness Function by Interactive Optimization Generation. a. (Top): Fitness Function with Maximum and Minimum Range Shown. b. (Bottom): Subbasin Wetland Area Fitness Function Normalized to the Maximum Value of each Individual Solution.

4.4.4 Effects of Climate Change on Solutions

The final solutions were subjected to the five climate models specified in section 3.1. The results of these runs are summarized by figures 17 and 18. The reduction in peak flow reductions between the present climate and projected climates is drastic. The baseline peak flows (maximum flow in any subbasin on any date modelled without any BMPs implemented) are listed below:

- Present: 215.1 cms
- MM5I CCSM DM 2057: 133.0 cms

- MM5I CCSM RAW 2055: 48.26 cms
- MM5I CCSM RAW 2062: 19.27 cms
- MM5I CCSM RAW 2065: 85.01 cms
- WRFG CGCM3 LOCI 2058: 145.9 cms

While some of the models, such as the 2062 and 2055 realizations of the MM5I CCSM RAW model have peak flows that are between 10% and 25% of the present maximum peak flow, the remaining model flows are within 50% of the present value. If the "optimal" solutions were as effective at reducing peak flow in the projected future as in the present, the model peak flow reduction values should be between 50% and 10% of the present peak flow reduction fitness function values, with their specific value dictated by the relative magnitude of the model's peak flow to the present peak flow. This is not the case in the solutions generated via either optimization method. The model with the highest peak flow relative to the present, WRFG CGCM3 LOCI 2058, exhibits the best gross value of peak flow reduction of any of the models. However, this gross value is only 5% to 10% of the peak flow reduction values from the present climate realization. This model's maximum peak flow is 67.8% of the present peak flow, meaning that this peak flow reduction is not proportionate to the peak flow. The remaining models follow this same pattern, with peak flow reductions being up to 90% less than would be expected if the solutions are equally as good at reducing peak flow in the projected future as in the past.

Even when all of the possible wetlands in the watershed are utilized as a best case scenario for design performance, there is a sharp reduction in peak flow reduction observed from the present to the projected future. Projected maximum peak flow reductions range from

1.08 to 3.70 cms, compared to 41.60 cms in the present. This indicates that the wetland designs used in the management plans are not appropriate for the projected future. The design depths are likely not sufficient to accommodate the more sudden flow increases that are likely to happen in the future, as indicated by the precipitation patterns shown in figure 10. Resultantly, it appears that the only way to ensure adequate performance of a management plan in the projected future is to design it explicitly for the projected future. It is not reasonable to assume that a plan optimized for performance in the present climate will perform well in the projected future as well.

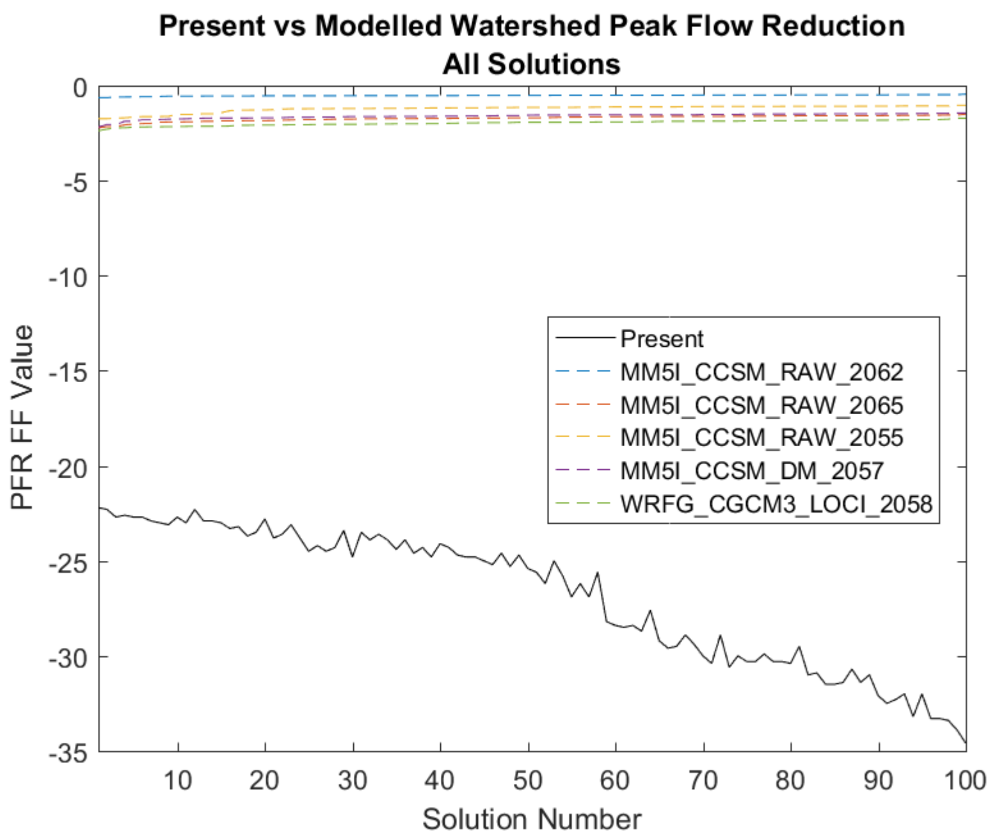

Figure 17: Peak Flow Reduction Fitness Function Values for Non-Interactive Optimization Solutions.

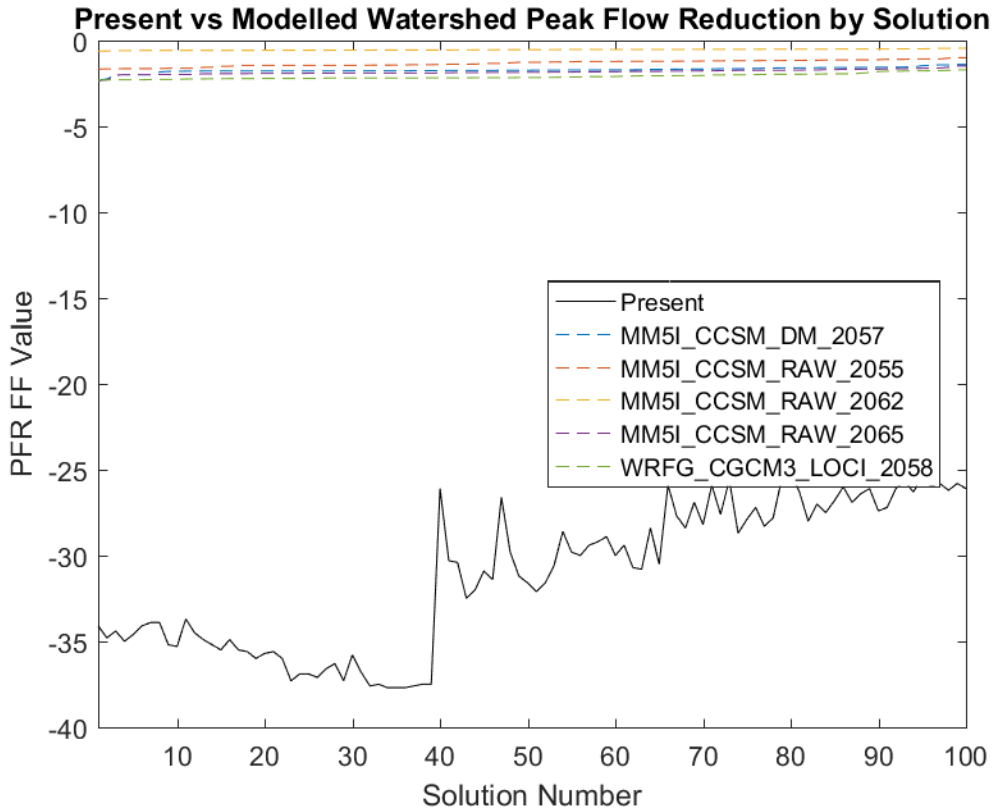

Figure 18: Peak Flow Reduction Fitness Function Values for the Interactive Optimization Solutions.

4.5 Discussion

As discussed previously, computing limitations required the interactive optimization process to neglect nitrate and sediment fitness functions, which were considered in the work completed by Garrison (2016). Removing these two fitness functions was assumed to have a minimal effect on the overall optimization process, as sediment and nitrate reduction functions are likely strongly related to peak flow reductions. The interactively optimized data was post-processed to determine the value of the sediment and nitrate reduction fitness functions, and validate this assumption. In Garrison's (2016) data, sediment and nitrate reduction were strongly correlated to peak flow reduction, with r^2

values of 0.9091 and 0.9265, respectively. Post processing of the interactive solutions revealed similarly strong correlations, with r^2 values of 0.9882 and 0.9972 for sediment and nitrate reduction fitness functions, respectively. In the light of these strong correlations, the assumption that sediment and nitrate reduction fitness functions would be adequately accounted for in the peak flow reduction fitness function appears justified, so the work presented in this paper can safely be compared to work completed by Garrison (2016).

Previous discussion concerning incorporating stakeholders into the WRESTORE process have focused on the idea of a "tradeoff": solutions generated using stakeholder-guided optimization would have reduced performance (sub-optimal fitness function values) compared to purely algorithmically optimized solutions as a tradeoff for the intangible and unquantifiable fitness criteria used by the stakeholders. Piemonti et al noted this effect in their 2013 article. However, this research's results were not consistent with this prediction. Figure 19 plots the Pareto front of the two different sets of optimized solutions, and shows an interesting phenomenon. The interactive solutions occur in two bands, one of which strictly dominates the other, and is defined by a higher proportion of three ratings. The dominated band of interactive solutions overlaps the front of non-interactive solutions. This dimorphic behavior may indicate that the process of convergence was not yet complete for the interactive optimization run, or that there are two distinct types of solutions that users preferred roughly interchangeably. If this is the case, then it is entirely possible that a fully convergent set of interactively optimized solutions may strictly dominate the non-interactive set of solutions, suggesting that the users guide the optimization process into a portion of the decision space not usually explored by purely algorithmic optimization.

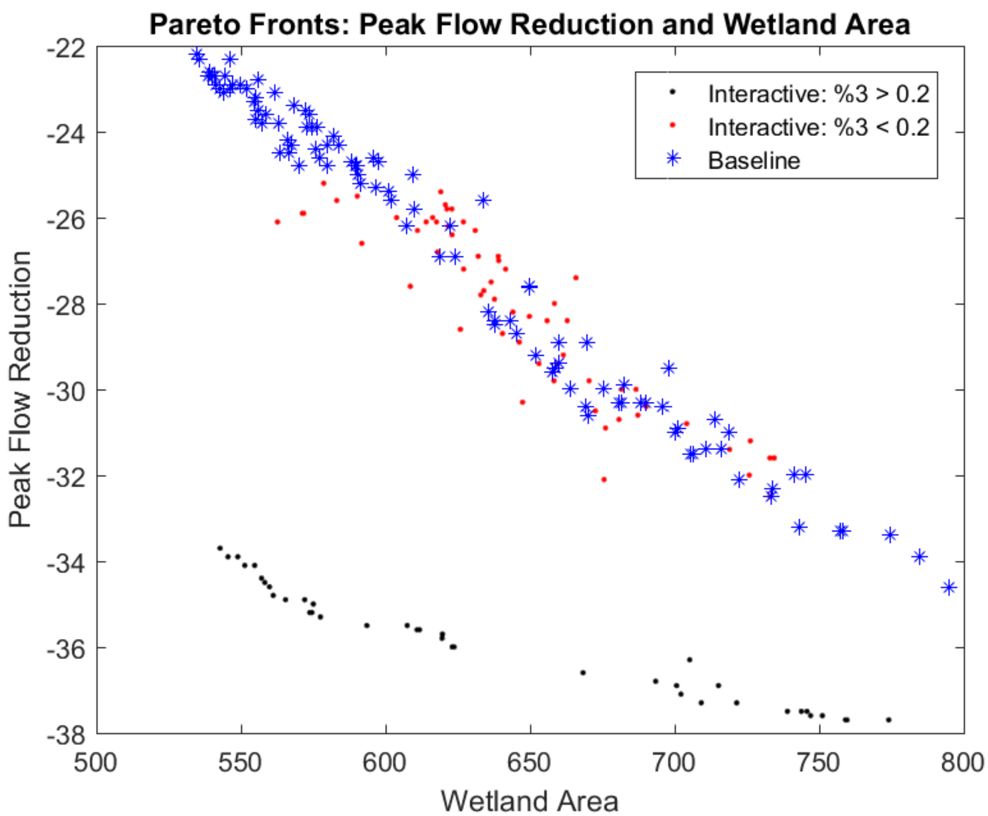

Figure 19: Wetland Area and Peak Flow Reductions Fitness Function Pareto Fronts for Interactive and Non-Interactive Optimization.

The modifications made to incorporate user opinions into the algorithm used by Garrison (2016) for her research placed a higher priority on finding solutions with high user consensus and preference over performance metrics. This represents a slight variation from previous work completed by Babbar-Sebens et al (2015) and Piemonti et al (2013), where only user preferences were considered. Figure 20 is a Pareto plot with user consensus and percent of three ratings, which is a more appropriate plot for examining the interactive solutions. Again, there are two distinct Pareto fronts in the interactively optimized solutions, suggesting incomplete convergence. Interestingly, there are two distinct clusters of solutions within the non-interactive dataset as well, one of which is defined by a relatively lower number of three ratings, and one of which is defined by a larger number of three ratings. The higher number of three ratings cluster that also includes the interactive solutions dominates the other cluster, so there is a subset of the non-interactive solutions that appear to be in the same solution space as some of the interactively-selected solutions. This indicates that post-processing the non-interactive solutions to identify solutions with high consensus and percent of three ratings may yield solutions that are similar to some solutions generated by explicitly incorporating users into the optimization process.

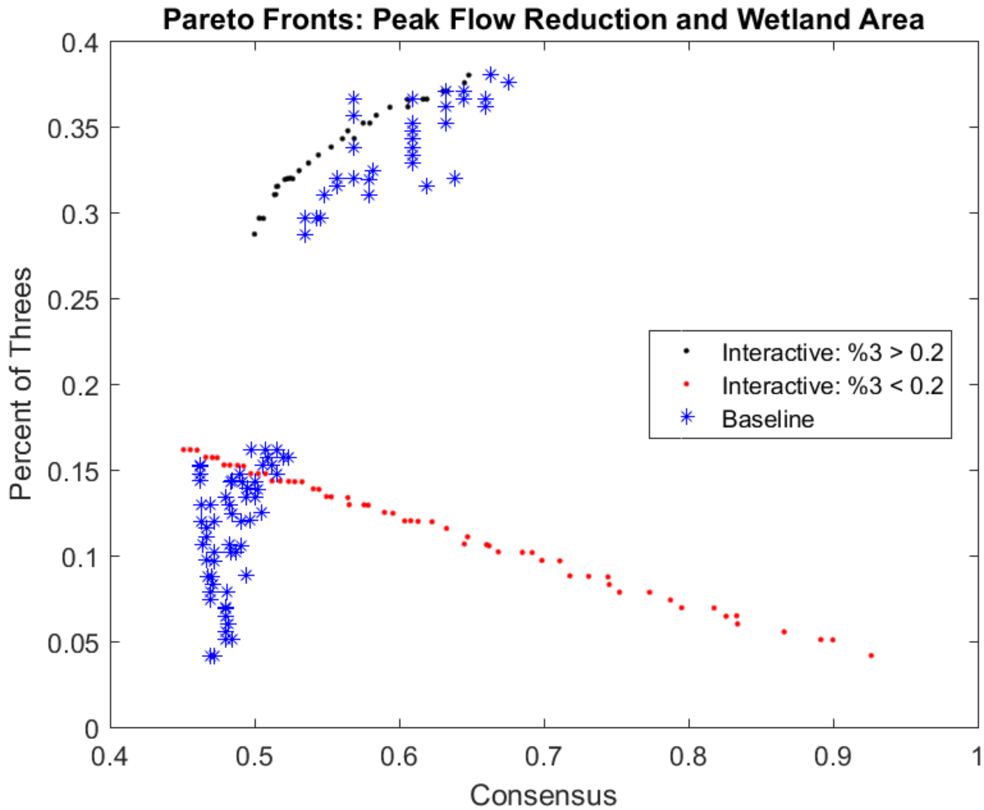

Figure 20: User Consensus and Percent of Solutions Awarded a User Rank Three Fitness Functions Pareto Fronts for Interactive and Non-Interactive Optimization.

The interactive solutions may not be completely reliable. As noted previously, these fronts don't seem to represent a set of fully-converged solutions. Several factors may contribute to this lack of convergence. One possibility is that the genetic algorithm used for this interactive optimization did not use enough generations to achieve convergence. This seems unlikely, considering that previous work by Garrison (2016) used 100 generations, the same number utilized by this research, to achieve convergence. Another possibility is that the new fitness function added by this research, user consensus, is highly volatile. In figure 20, the solutions with the highest consensus also have the lowest number of positive rankings, indicating a fundamental conflict between user ratings and user consensus. It is possible that the consensus measure is highly sensitive to small modifications in input parameters, and minor perturbations in solution characteristics (the number and location of wetlands) result in substantial changes in user consensus. A final possible explanation may be the interruption in optimization that occurred in generation 30. While fitness function values appear to have recovered by generation 55, it is possible that remnants of that interruption persisted at the subbasin scale throughout the remaining generations. Given that the consensus and rating functions used in this research consider both volatile subbasin and more stable watershed fitness function values, it is possible that user consensus could remain sensitive to interruptions of this kind for a longer period than other fitness functions, which do not include the complicated interplay between subbasin and watershed scale values present in the consensus and rating functions.

4.6 Conclusion

The complicated nature of watershed management plan optimization requires the continual development of novel methods to incorporate stakeholders into the design process. Stakeholder investment in the process and the resulting management plan is essential, and is strongly tied to how included in the process these stakeholders are and feel. Work completed by the many authors and researched associated with the WRESTORE project have explored ways to incorporate decision makers and stakeholders into the optimization process for designing watershed management plans, and this research continues their work. This research adds a new mathematical framework for expressing stakeholder preferences: consensus. Due to time constraints, the research could not host a workshop necessary for incorporating actual stakeholder input, so this paper also proposed a method to simulate stakeholders based upon basic information available concerning their opinions and sensibilities. Finally, this research sought to assess how solutions designed considering only current climate conditions would fare when exposed to the extreme conditions of the projected future.

When compared to work performed by Garrison (2016) that did not incorporate user opinions into the optimization process, the method proposed by this article produced solutions that performed as well or better than Garrison's with respect to minimizing wetland area and maximizing peak flow reduction. Garrison's solutions were also optimized to maximize sediment and nitrate loading reduction, but sediment and nitrate removal are strongly correlated to peak flow reduction, so the latter metric can be considered a good surrogate for sediment and nitrate removal. However, neither the interactively nor the non-interactively optimized solutions perform well in the projected future. This suggests that optimizing for present climate conditions will not necessarily yield solutions that fare well in the projected future: the future climate must be explicitly accounted for.

There were apparent irregularities in the interactively optimized data, however, so these findings must be verified before any further conclusions can be drawn. After this research is repeated, and the data issues resolved, the next step would be creating a more sophisticated user rating model to use in the optimization process. More detailed research in the Eagle Creek Watershed focused on its stakeholders will be required to create user models more strongly based upon their actual conservation behaviors.

4.7 References

Acreman, M. and J. Holden, 2013. How Wetlands Affect Floods. **33**:773-786.

Arbuckle, J. G., L. S. Prokopy, T. Haigh, J. Hobbs, T. Knoot, C. Knutson, A. Loy, A. S. Mase, J. McGuire, L. W. Morton, J. Tyndall and M. Widhalm, 2013. Climate change beliefs, concerns, and attitudes toward adaptation and mitigation among farmers in the Midwestern United States. **117**:943-950.

Arpita, M. and P. P. Mujumdar, 2016. Detection of Change in Flood Return Levels under Global Warming. *Journal of Hydrologic Engineering* **21**.

Babbar-Sebens, M., R. C. Barr, L. P. Tedesco and M. Anderson, 2013. Spatial identification and optimization of upland wetlands in agricultural watersheds. *Ecological Engineering* **52**:130-142.

Babbar-Sebens, M. and B. S. Minsker, 2012. Interactive Genetic Algorithm with Mixed Initiative Interaction for multi-criteria ground water monitoring design. *Applied Soft Computing* **12**:182-195.

Babbar-Sebens, M., S. Mukhopadhyay, V. B. Singh and A. D. Piemonti, 2015. A web-based software tool for participatory optimization of conservation practices in watersheds. *Environmental Modelling & Software* **69**:111-127.

Babbar-Sebens, M., A. D. Piemonti, S. Mukhopadhyay and V. B. Singh, 2014. User Modeling And Personalized Optimization For Stakeholder-Driven Watershed Design. *In*: *International Conference on Hydroinformatics.* CUNY Academic

Works, New York, New York.

Bank, W., 2016. Agriculture, value added (current US$). The World Bank, pp. Line chart with year on x-axis and value of global agriculture in trillions of dollars on y-axis.

Basco-Carrera, L., A. Warren, E. van Beek, A. Jonoski and A. Giardino, 2017. Collaborative modelling or participatory modelling? A framework for water resources management. **91**:95-110.

Bullock, A. and M. Acreman, 2003. The role of wetlands in the hydrological cycle. **7**:358-389.

Cannady-Shultz, K. and M. Babbar-Sebens, 2018.

Castelletti, A. and R. Soncini-Sessa, 2006. A procedural approach to strengthening integration and participation in water resource planning. *Environmental Modelling & Software* **21**:1455-1470.

Castelletti, A. and R. Soncini-Sessa, 2007. Coupling real-time control and socio-economic issues in participatory river basin planning. *Environmental Modelling & Software* **22**:1114-1128.

Chatrchyan, A. M., R. C. Erlebacher, N. T. Chaopricha, J. Chan, D. Tobin and S. B. C. e. Allred, United States agricultural stakeholder views and decisions on climate change.e467-n/a.

Cox, P. G., 1996. Some issues in the design of agricultural decision support systems. *Agricultural Systems* **52**:355-381.

Dankers, R., N. W. Arnell, D. B. Clark, P. D. Falloon, B. M. Fekete, S. N. Gosling, J. Heinke, H. Kim, Y. Masaki, Y. Satoh, T. Stacke, Y. Wada and D. Wisser, 2014.

First look at changes in flood hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble. *Proceedings of the National Academy of Sciences* **111**:3257-3261.

Garrison, S., 2016. Determination of Trade-offs between Wetland Ecosystem Services in an Agricultural Landscape *In*: *Civil Engineering.* Oregon State University, Corvallis, OR.

Haimes, Y. Y. and W. A. Hall, 1974. MULTIOBJECTIVES IN WATER-RESOURCE SYSTEMS-ANALYSIS - SURROGATE WORTH TRADE OFF METHOD. *Water Resources Research* **10**:615-624.

Herrera-Viedma, E., F. Herrera and F. Chiclana, 2002. A consensus model for multiperson decision making with different preference structures. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* **32**:394-402.

IPCC, 2012. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaption: A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change. *In,* Editor (Editor)^(Editors). Intergovernmental Panel on Climate Change, Cambridge, UK, and New York, New York.

IPCC, 2014. Climate Change 2014: Synthesis Report: Contribution of Working Groups I, II, and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *In,* Editor (Editor)^(Editors). IPCC, Geneva, Switzerland.

Iturrioz, R., 2009. Agricultural Insurance. *In: Primer Series on Insurance,* Editor (Editor)^(Editors). The World Bank, Washington, DC.

Javaheri, A. and M. Babbar-Sebens, 2014. On comparison of peak flow reductions, flood inundation maps, and velocity maps in evaluating effects of restored wetlands on channel flooding. **73**:132-145.

Jena, P. P., C. Chatterjee, G. Pradhan and A. Mishra, 2014. Are recent frequent high floods in Mahanadi basin in eastern India due to increase in extreme rainfalls? **517**:847-862.

Knutti, R., R. Furrer, C. Tebaldi, J. Cermak and G. A. Meehl, 2009. Challenges in Combining Projections from Multiple Climate Models. *Journal of Climate* **23**:2739-2758.

Lal, R., J. A. Delgado, P. M. Groffman, N. Millar, C. Dell and A. Rotz, 2011. Management to mitigate and adapt to climate change. *Journal of Soil and Water Conservation* **66**:276-285.

Madsen, H., D. Lawrence, M. Lang, M. Martinkova and T. R. Kjeldsen, 2014. Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. **519, Part D**:3634-3650.

McCown, R. L., 2002. Changing systems for supporting farmers' decisions: problems, paradigms, and prospects. *Agricultural Systems* **74**:179-220.

Mearns, L. O., W. Gutowski, R. Jones, R. Leung, S. McGinnis, A. Nunes and Y. Qian, 2009. A Regional Climate Change Assessment Program for North America. **90**:311-311.

Novotny, E. V. and H. G. Stefan, 2007. Stream flow in Minnesota: Indicator of climate change. **334**:319-333.

Piemonti, A. D., M. Babbar-Sebens and E. J. Luzar, 2013. Optimizing conservation practices in watersheds: Do community preferences matter? *Water Resources Research* **49**:6425-6449.

Reimer, A. P., A. W. Thompson and L. S. Prokopy, 2012. The multi-dimensional nature of environmental attitudes among farmers in Indiana: implications for conservation adoption. *Agriculture and Human Values* **29**:29-40.

Seidou, O., A. Ramsay and I. Nistor, 2012. Climate change impacts on extreme floods II: improving flood future peaks simulation using non-stationary frequency analysis. **60**:715-726.

Tortajada, C., 2003. Workshop on Integrated Water Resources Management for South and Southeast Asia. *Water International* **28**:130-131.

USDA, 2011. Help in the midst of FLOOD, USDA (USDA)USDAs). USDA, Washington, DC.

Walters, K. M. and M. Babbar-Sebens, 2016. Using climate change scenarios to evaluate future effectiveness of potential wetlands in mitigating high flows in a Midwestern US watershed. *Ecological Engineering* **89**:80-102.

Wu, C. and G. Huang, 2015. Changes in heavy precipitation and floods in the upstream of the Beijiang River basin, South China. **35**:2978-2992.

Chapter 5. Final Discussion and Conclusions

This research sought to identify how to best incorporate multiple users into the WRESTORE framework simultaneously. Previous work by Babbar-Sebens et al (2015) included methods for allowing multiple stakeholders to collaborate simultaneously during optimization. Their method accumulates the ratings from each stakeholder, and assigns the design being reviewed a rating equal to the mode. However, as noted in chapter 4, this method does have a shortcoming. The mode rating does represent the consensus opinion well, but it provides no information about how unified that opinion is. The algorithm treats a design that has a narrow majority of three ratings the same as one that has unanimous preference, even though the latter is more likely to be implemented. This research adds a second measurement of user preference to provide a measure of user rating unity – consensus.

The proposed measure was adapted from Herrera-Viedma et al (2002), and added to the existing WRESTORE framework. Due to time constraints, this research was unable to involve actual stakeholders or human subjects in the interactive optimization process. Instead, surrogate stakeholders were created, and incorporated into the genetic algorithm. Work completed previously by Piemonti et al (2013) theorized that incorporating stakeholders into the optimization process would result in tradeoffs, wherein some performance would be sacrificed for some gain in the ineffable, unquantifiable criteria they use to judge the solutions. For this research, this expectation corresponds to decreased peak flow reaction and increased wetland area when compared to non-interactively optimized solutions.

The results of this research did not match the expectation. As previously observed in chapter, the solutions resulting from the interactive optimization perform as well as or better than the non-interactively optimized solutions, often attaining superior peak flow reductions for similar total wetland areas. This type of behaviour seems more reminiscent of observations Klau et al (2009) made. An efficient partnership between software and people can increase the convergence speed of certain optimization routines for specific tasks. Humans guide the optimization process, and help the routine explore various parts of the solution space that it would normally not focus on. It is possible that this type of user-software interaction was simulated in this research, and led to the behaviour shown in figure 19. However, this process does not apparently lead to results with high consensus, indicating that there is still contention about what the best results actually look like. The group guides the optimization process, but not with a unified opinion. This discord may actually be the source of this guidance, as conventional wisdom states that healthy discourse is often the basis for innovation.

Another idea this research addressed was whether designing management plans for the current climate produces designs that work well in the projected future as well. Previous work by Walters and Babbar-Sebens (2016) generated a list of eleven thirty-year long climate model realizations that well-represent the projected climate of the Eagle Creek

Watershed through the mid-century. However, utilizing thirty-year long climate realizations in the genetic algorithm utilized by WRESTORE is extremely time consuming, so a prerequisite to using this climate data was developing a method to drastically reduce the length of the model realization while maintaining its accurate representation of the future climate.

The method developed to reduce the length of the climate model realizations used for this research is a modification on the method developed by Lutz et al (2016). The proposed method seeks to maintain diversity while accurately representing future extreme climate in single year realizations. When compared to the baseline method developed by Lutz, the proposed climate model selection method performs well.

Overall, this research seems to indicate that incorporating users into the optimization process produces better solutions than purely non-interactive optimization. Not only does the interactive process foster solutions that have higher consensus, it also apparently improves the performance of those solutions as well. However, these solutions, which only consider the present climate in their design, do not perform well in the projected future. The peak flow reduction is muted relative to the change in maximum peak flow. It appears clear, therefore, that any consideration of climate change in watershed management plan design must be explicit: designs that work well in the present climate may not perform well in the future.

5.1 Engineering Implications and Applications

As the climate changes, the effectiveness of current watershed management plans and practices may change as well. Concerted effort from communities and stakeholders will be necessary moving forward to create plans that will be effective into the future. This research explored how incorporating stakeholders into the management plan design process will change the quality of the plans, and their resilience to future climate flooding scenarios. Management plans optimized under current climate conditions that incorporate stakeholder input within the actual optimization process actually out-perform designs that did not incorporate their input. Generally, increasing stakeholder inclusion in the design process will increase their willingness to implement the plan suggested. Thus, this research suggests that actively including stakeholders into all parts of the design process not only results in better designs, but also increases the likelihood of implementation. As practitioners, the transition from paper designs to built infrastructure is essential. Using the process advocated for in this research increases the likelihood of this all-important transition.

However, this research also indicates that creating management plans that function well in the present do not necessarily perform well in the future. In particular, the wetland-only management plans developed by this research were almost entirely ineffectual at reducing flooding in the future. Even using all of the potential wetland sites available in the watershed, only minimal flood reduction benefits were noted. In this case, the engineering specifications for flood reduction wetlands would have to be altered to account for fundamentally different climate patterns, or different best management practices considered altogether. Simply scaling up the implementation of current designs will not work.

Chapter 6. References

1980. The Indiana water resource ; availability , uses , and needs, American Geosciences Institute.

Acreman, M. and J. Holden, 2013. How Wetlands Affect Floods. 33:773-786.

Ahnstrom, J., J. Hockert, H. L. Bergea, C. A. Francis, P. Skelton and L. Hallgren, 2009. Farmers and nature conservation: What is known about attitudes, context factors and actions affecting conservation? Renewable Agriculture and Food Systems 24:38-47.

Arabi, M., R. S. Govindaraju and M. M. Hantush, 2006. Cost-effective allocation of watershed management practices using a genetic algorithm. Water Resources Research 42:14.

Arbuckle, J. G., L. S. Prokopy, T. Haigh, J. Hobbs, T. Knoot, C. Knutson, A. Loy, A. S. Mase, J. McGuire, L. W. Morton, J. Tyndall and M. Widhalm, 2013. Climate change beliefs, concerns, and attitudes toward adaptation and mitigation among farmers in the Midwestern United States. 117:943-950.

Arpita, M. and P. P. Mujumdar, 2016. Detection of Change in Flood Return Levels under Global Warming. Journal of Hydrologic Engineering 21.

Artita, K. S., P. Kaini and J. W. Nicklow, 2013. Examining the Possibilities: Generating Alternative Watershed-Scale BMP Designs with Evolutionary Algorithms. Water Resources Management 27:3849-3863.

Babbar, M. and B. S. Minsker, 2006. Groundwater remediation design using multiscale genetic algorithms. Journal of Water Resources Planning and Management-Asce 132:341-350.

Babbar-Sebens, M., R. C. Barr, L. P. Tedesco and M. Anderson, 2013. Spatial identification and optimization of upland wetlands in agricultural watersheds. Ecological Engineering 52:130-142.

Babbar-Sebens, M. and B. S. Minsker, 2012. Interactive Genetic Algorithm with Mixed Initiative Interaction for multi-criteria ground water monitoring design. Applied Soft Computing 12:182-195.

Babbar-Sebens, M., S. Mukhopadhyay, V. B. Singh and A. D. Piemonti, 2015. A web-based software tool for participatory optimization of conservation practices in watersheds. Environmental Modelling & Software 69:111-127.

Babbar-Sebens, M., A. D. Piemonti, S. Mukhopadhyay and V. B. Singh, 2014. User Modeling And Personalized Optimization For Stakeholder-Driven Watershed Design. In: International Conference on Hydroinformatics. CUNY Academic Works, New York, New York.

Bank, W., 2016. Agriculture, value added (current US$). The World Bank, pp. Line chart with year on x-axis and value of global agriculture in trillions of dollars on y-axis.

Basco-Carrera, L., A. Warren, E. van Beek, A. Jonoski and A. Giardino, 2017. Collaborative modelling or participatory modelling? A framework for water resources management. 91:95-110.

Block, P. J., F. A. Souza, L. Q. Sun and H. H. Kwon, 2009. A Streamflow Forecasting Framework using Multiple Climate and Hydrological Models(1). Journal of the American Water Resources Association 45:828-843.

Bullock, A. and M. Acreman, 2003. The role of wetlands in the hydrological cycle. 7:358-389.

Castelletti, A. and R. Soncini-Sessa, 2006. A procedural approach to strengthening integration and participation in water resource planning. Environmental Modelling & Software 21:1455-1470.

Castelletti, A. and R. Soncini-Sessa, 2007. Coupling real-time control and socio-economic issues in participatory river basin planning. Environmental Modelling & Software 22:1114-1128.

Chatrchyan, A. M., R. C. Erlebacher, N. T. Chaopricha, J. Chan, D. Tobin and S. B. C. e. Allred, United States agricultural stakeholder views and decisions on climate change.e467-n/a.

Chen, H., C. Y. Xu and S. L. Guo, 2012. Comparison and evaluation of multiple GCMs, statistical downscaling and hydrological models in the study of climate change impacts on runoff. Journal of Hydrology 434:36-45.

Cox, P. G., 1996. Some issues in the design of agricultural decision support systems. Agricultural Systems 52:355-381.

Dankers, R., N. W. Arnell, D. B. Clark, P. D. Falloon, B. M. Fekete, S. N. Gosling, J. Heinke, H. Kim, Y. Masaki, Y. Satoh, T. Stacke, Y. Wada and D. Wisser, 2014. First look at changes in flood hazard in the Inter-Sectoral Impact Model

Intercomparison Project ensemble. Proceedings of the National Academy of Sciences 111:3257-3261.

Dinnes, D. L., D. L. Karlen, D. B. Jaynes, T. C. Kaspar, J. L. Hatfield, T. S. Colvin and C. A. Cambardella, 2002. Nitrogen management strategies to reduce nitrate leaching in tile-drained midwestern soils. Agronomy Journal 94:153-171.

Duff, S. N., D. P. Stonehouse, S. G. Hilts and D. J. Blackburn, 1991. SOIL CONSERVATION BEHAVIOR AND ATTITUDES AMONG ONTARIO FARMERS TOWARD ALTERNATIVE GOVERNMENT POLICY RESPONSES. Journal of Soil and Water Conservation 46:215-219.

Espinoza Felipe, P., S. Minsker Barbara and E. Goldberg David, 2005. Adaptive Hybrid Genetic Algorithm for Groundwater Remediation Design. Journal of Water Resources Planning and Management 131:14-24.

Fraternali, P., A. Castelletti, R. Soncini-Sessa, C. V. Ruiz and A. E. Rizzoli, 2012. Putting humans in the loop: Social computing for Water Resources Management. Environmental Modelling & Software 37:68-77.

Garrison, S., 2016. Determination of Trade-offs between Wetland Ecosystem Services in an Agricultural Landscape In: Civil Engineering. Oregon State University, Corvallis, OR.

Goharian, E., S. J. Burian, T. Bardsley and C. Strong, 2016. Incorporating Potential Severity into Vulnerability Assessment of Water Supply Systems under Climate Change Conditions. Journal of Water Resources Planning and Management 142:12.

Goldberg, D. E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley Longman Publishing Co., Inc.

Haimes, Y. Y. and W. A. Hall, 1974. MULTIOBJECTIVES IN WATER-RESOURCE SYSTEMS-ANALYSIS - SURROGATE WORTH TRADE OFF METHOD. Water Resources Research 10:615-624.

Herrera-Viedma, E., F. Herrera and F. Chiclana, 2002. A consensus model for multiperson decision making with different preference structures. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 32:394-402.

IPCC, 2012. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaption: A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change. In, Editor (Editor)^(Editors). Intergovernmental Panel on Climate Change, Cambridge, UK, and New York, New York.

IPCC, 2014. Climate Change 2014: Synthesis Report: Contribution of Working Groups I, II, and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In, Editor (Editor)^(Editors). IPCC, Geneva, Switzerland.

Iturrioz, R., 2009. Agricultural Insurance. In: Primer Series on Insurance, Editor (Editor)^(Editors). The World Bank, Washington, DC.

Javaheri, A. and M. Babbar-Sebens, 2014. On comparison of peak flow reductions, flood inundation maps, and velocity maps in evaluating effects of restored wetlands on channel flooding. 73:132-145.

Jena, P. P., C. Chatterjee, G. Pradhan and A. Mishra, 2014. Are recent frequent high floods in Mahanadi basin in eastern India due to increase in extreme rainfalls? 517:847-862.

Kaini, P., K. Artita and J. W. Nicklow, 2012. Optimizing Structural Best Management Practices Using SWAT and Genetic Algorithm to Improve Water Quality Goals. Water Resources Management 26:1827-1845.

Klau, G. W., N. Lesh, J. Marks and M. Mitzenmacher, 2010. Human-guided search. Journal of Heuristics 16:289-310.

Knight, R. L., V. W. E. Payne, R. E. Borer, R. A. Clarke and J. H. Pries, 2000. Constructed wetlands for livestock wastewater management. Ecological Engineering 15:41-55.

Knowler, D. and B. Bradshaw, 2007. Farmers' adoption of conservation agriculture: A review and synthesis of recent research. 32:25-48.

Knutti, R., R. Furrer, C. Tebaldi, J. Cermak and G. A. Meehl, 2009. Challenges in Combining Projections from Multiple Climate Models. Journal of Climate 23:2739-2758.

Lal, R., J. A. Delgado, P. M. Groffman, N. Millar, C. Dell and A. Rotz, 2011. Management to mitigate and adapt to climate change. Journal of Soil and Water Conservation 66:276-285.

Luo, L. F., E. F. Wood and M. Pan, 2007. Bayesian merging of multiple climate model forecasts for seasonal hydrological predictions. Journal of Geophysical Research-Atmospheres 112:13.

Lutz, A. F., H. W. ter Maat, H. Biemans, A. B. Shrestha, P. Wester and W. W. Immerzeel, 2016. Selecting representative climate models for climate change impact studies: an advanced envelope-based selection approach. 36:3988-4005.

Luzar, E. J. and A. Diagne, 1999. Participation in the next generation of agriculture conservation programs: the role of environmental attitudes. The Journal of Socio-Economics 28:335-349.

Lynne, G. D., J. S. Shonkwiler and L. R. Rola, 1988. ATTITUDES AND FARMER CONSERVATION BEHAVIOR. American Journal of Agricultural Economics 70:12-19.

Madsen, H., D. Lawrence, M. Lang, M. Martinkova and T. R. Kjeldsen, 2014. Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. 519, Part D:3634-3650.

Maxted, J., M. Diebel and M. Vander Zanden, 2009. Landscape Planning for Agricultural Non-Point Source Pollution Reduction. II. Balancing Watershed Size, Number of Watersheds, and Implementation Effort. Environmental Management 43:60-68.

McCown, R. L., 2002. Changing systems for supporting farmers' decisions: problems, paradigms, and prospects. Agricultural Systems 74:179-220.

Mearns, L. O., W. Gutowski, R. Jones, R. Leung, S. McGinnis, A. Nunes and Y. Qian, 2009. A Regional Climate Change Assessment Program for North America. 90:311-311.

Milly, P. C. D., R. T. Wetherald, K. A. Dunne and T. L. Delworth, 2002. Increasing risk of great floods in a changing climate. Nature 415:514-517.

Mote, P., L. Brekke, P. B. Duffy and E. Maurer, 2011. Guidelines for constructing climate scenarios. 92:257-258.

Novotny, E. V. and H. G. Stefan, 2007. Stream flow in Minnesota: Indicator of climate change. 334:319-333.

Peterson, T. C., 2005. Climate Change Indices. pp. 83-86.

Piemonti, A. D., 2014. Correlating Human Preferences and Optimized Watershed Management Plans. World Environmental and Water Resources Congress 2014.

Piemonti, A. D., 2015. Interactive Genetic Algorithms for Watershed Planning: An Investigation of Usability and Human-centered Design. In: Civil Engineering. Oregon State University, Corvallis, OR.

Piemonti, A. D., M. Babbar-Sebens and E. J. Luzar, 2013. Optimizing conservation practices in watersheds: Do community preferences matter? Water Resources Research 49:6425-6449.

Reimer, A. P., A. W. Thompson and L. S. Prokopy, 2012. The multi-dimensional nature of environmental attitudes among farmers in Indiana: implications for conservation adoption. Agriculture and Human Values 29:29-40.

Seidou, O., A. Ramsay and I. Nistor, 2012. Climate change impacts on extreme floods II: improving flood future peaks simulation using non-stationary frequency analysis. 60:715-726.

Simonovic, S. P. and L. H. Li, 2003. Methodology for assessment of climate change impacts on large-scale flood protection system. Journal of Water Resources Planning and Management-Asce 129:361-371.

Soderqvist, T., 2003. Are farmers prosocial? Determinants of the willingness to participate in a Swedish catchment-based wetland creation programme. Ecological Economics 47:105-120.

Stillman, D. and D. Miller, 2015. What are Climate and Climate Change?, S. May (S. May)S. Mays). NASA.

Tebaldi, C. and R. Knutti, 2007. The use of the multi-model ensemble in probabilistic climate projections. Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences 365:2053-2075.

Thober, S. and L. Samaniego, 2014. Robust ensemble selection by multivariate evaluation of extreme precipitation and temperature characteristics. 119:594-613.

Tortajada, C., 2003. Workshop on Integrated Water Resources Management for South and Southeast Asia. Water International 28:130-131.

Tosakana, N. S. P., L. W. Van Tassell, J. D. Wulfhorst, J. Boll, R. Mahler, E. S. Brooks and S. Kane, 2010. Determinants of the adoption of conservation practices by farmers in the Northwest Wheat and Range Region. Journal of Soil and Water Conservation 65:404-412.

USDA, 2011. Help in the midst of FLOOD, USDA (USDA)USDAs). USDA, Washington, DC.

Walters, K. M. and M. Babbar-Sebens, 2016. Using climate change scenarios to evaluate future effectiveness of potential wetlands in mitigating high flows in a Midwestern US watershed. Ecological Engineering 89:80-102.

Winkler, 2012. Climate Projections for the Midwest: Availability, Interpretation and Synthesis

Wobus, C., M. Lawson, R. Jones, J. Smith and J. Martinich, 2014. Estimating monetary damages from flooding in the United States under a changing climate. Journal of Flood Risk Management 7:217-229.

WRESTORE, 2017. WRESTORE | Watershed REstoration using Spatio-Temporal Optimization of REsources. IUPUI, OSU, and NOAA.

Wu, C. and G. Huang, 2015. Changes in heavy precipitation and floods in the upstream of the Beijiang River basin, South China. 35:2978-2992.

Zedler, J. B., 2003. Wetlands at your service: reducing impacts of agriculture at the watershed scale. Frontiers in Ecology and the Environment 1:65-72.

Zhang, X. S., R. Srinivasan and D. Bosch, 2009. Calibration and uncertainty analysis of the SWAT model using Genetic Algorithms and Bayesian Model Averaging. Journal of Hydrology 374:307-317.

Chapter 7. Appendix A: List of Climate Models

CRCM_CCSM_LOCI

CRCM_CGCM3_LOCI

HRM3_GFDL_LOCI

HRM3_GFDL_LS

MM5I_CCSM_DM

MM5I_CCSM_LOCI

MM5I_CCSM_LS

MM5I_CCSM_RAW

RCM3_CGCM3_LOCI

WRFG_CGCM3_LOCI

WRFG_CGCM3_LS

Chapter 8. Appendix B: Sediment and Nitrate Loading Charts

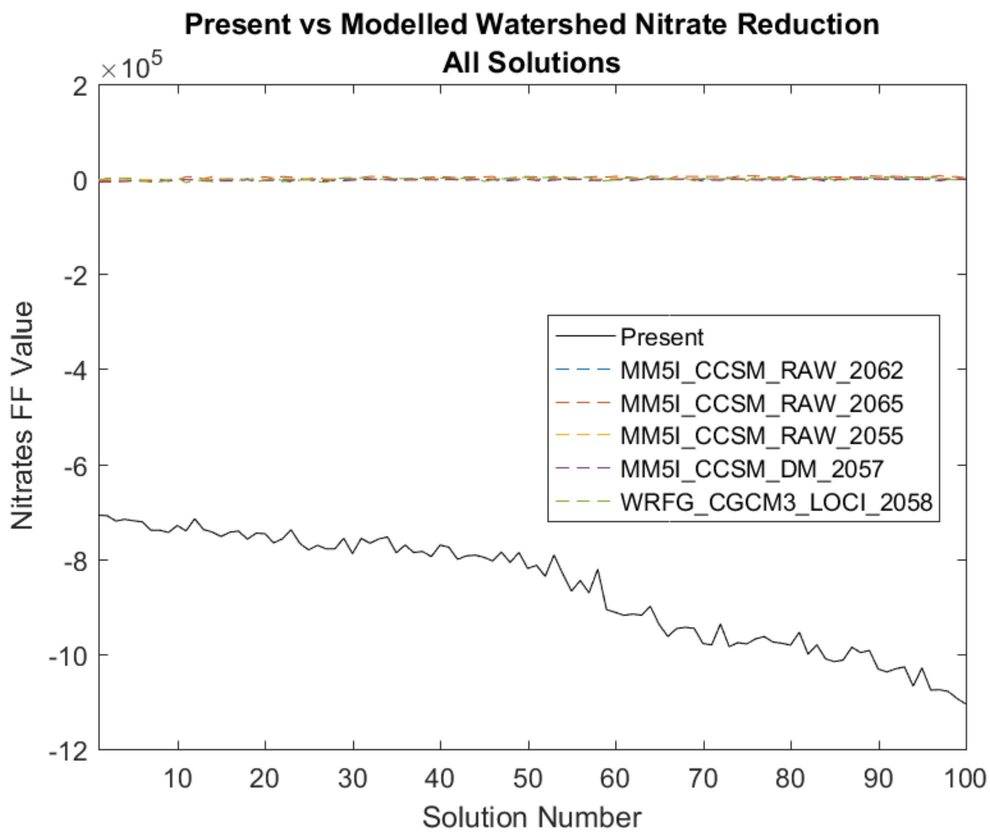

Figure 21: Nitrate Reduction Fitness Function Plot for Non-Interactive Solutions

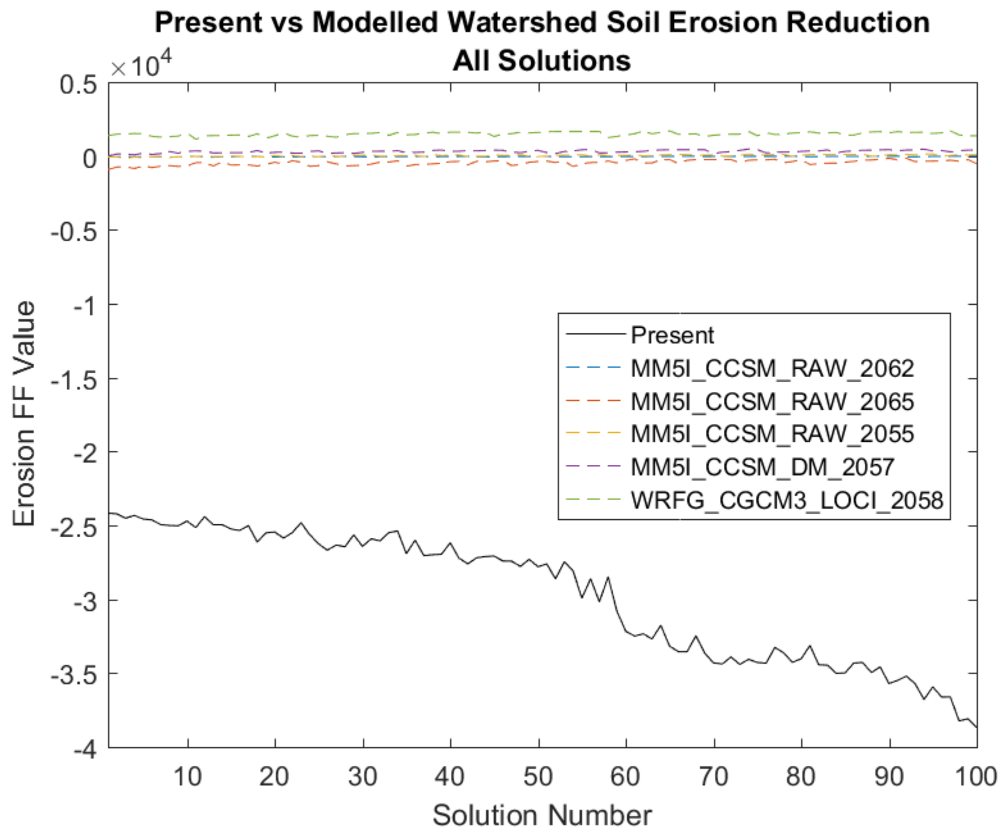

Figure 22: Sediment Reduction Fitness Function Plot for Non-Interactive Solutions

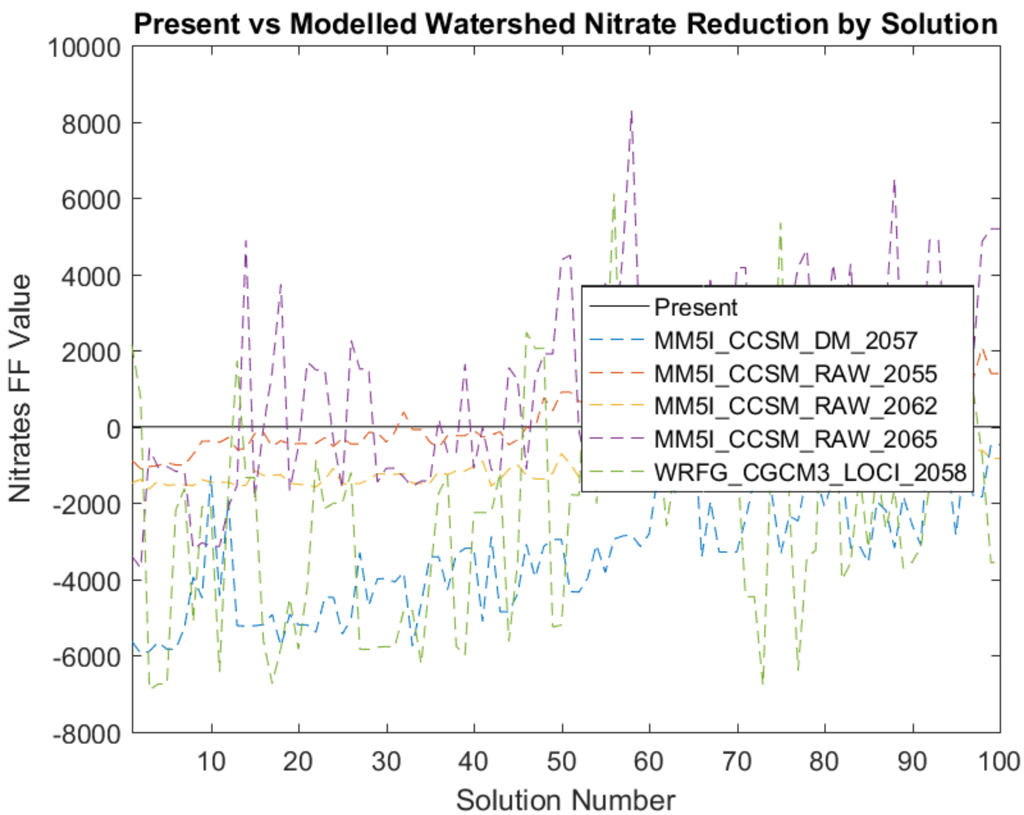

Figure 23: Nitrate Reduction Fitness Function Plot for Interactive Optimization - Projected Climate Only

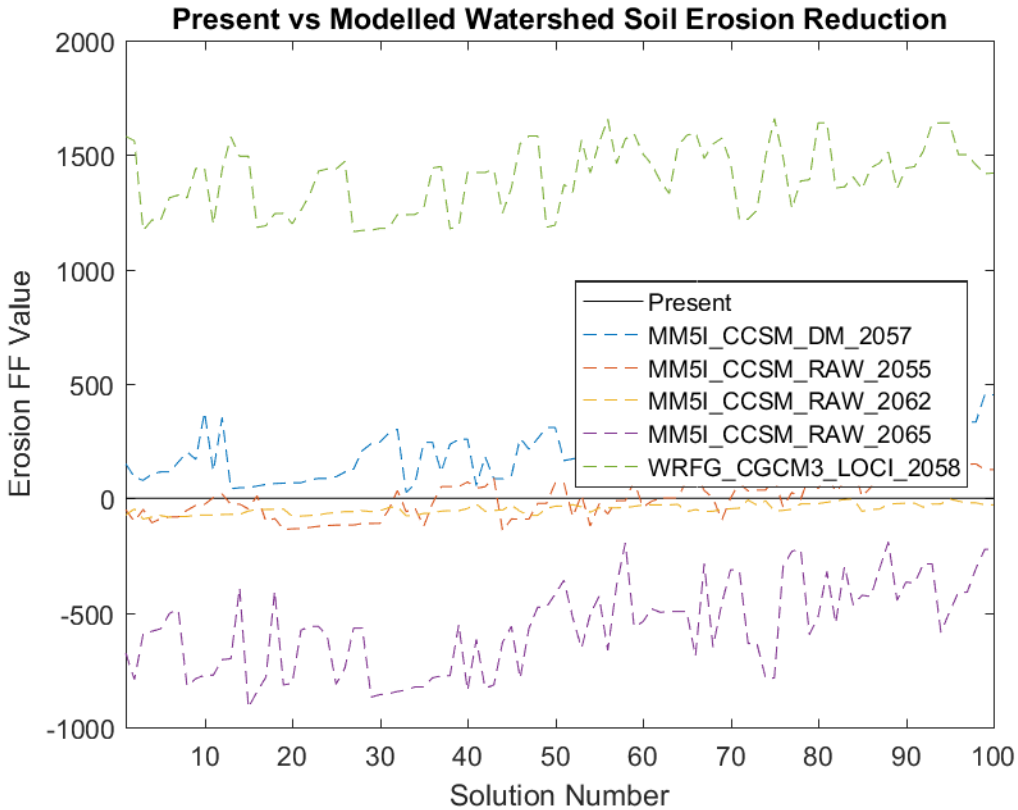

Figure 24: Erosion Reduction Fitness Function Plot for Interactive Solutions - Projected Climate Only

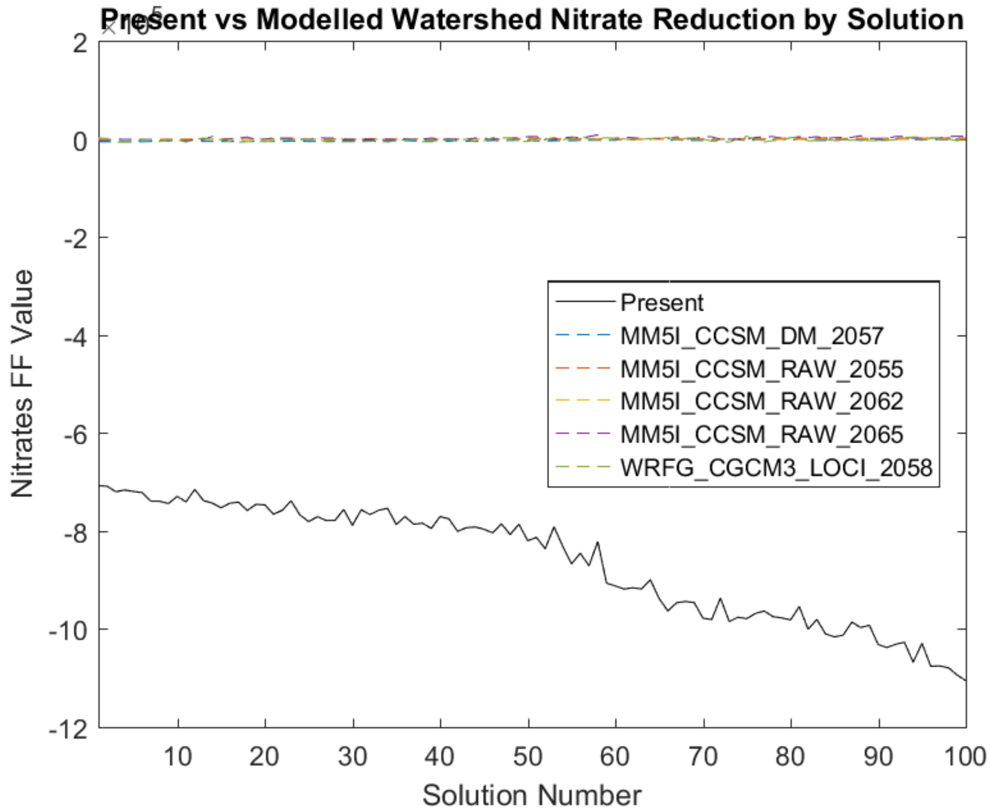

Figure 25: Nitrates Reduction Fitness Function Plot for Interactive Solutions

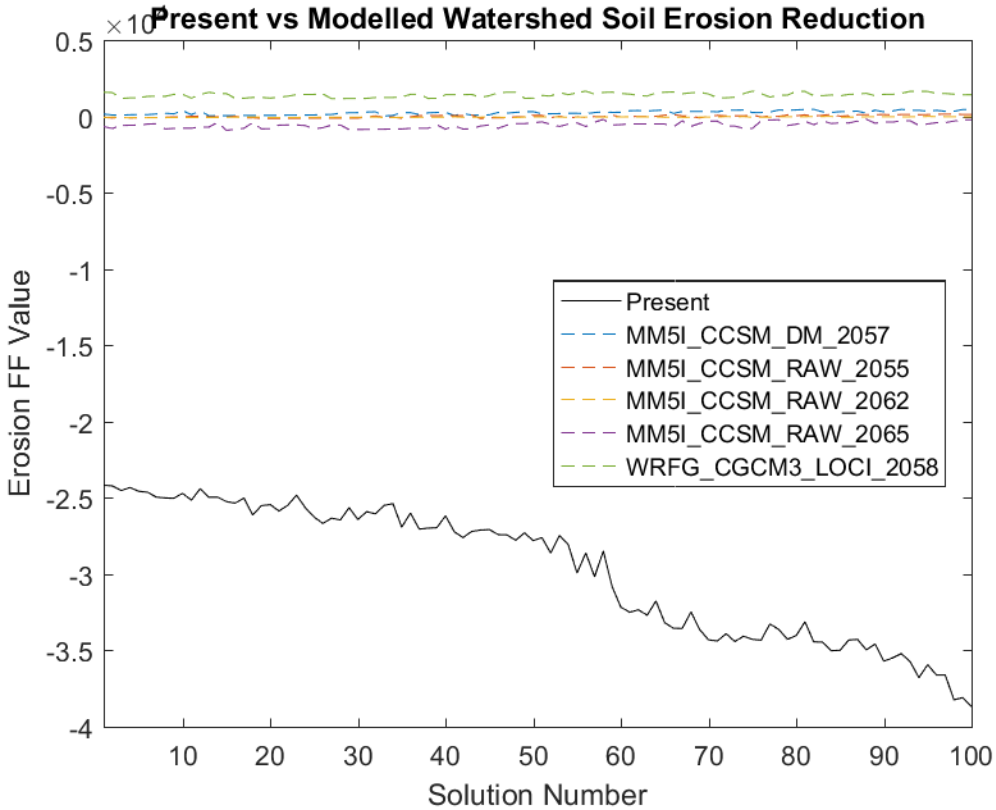

Figure 26: Erosion Reduction Fitness Function Plot for Interactive Solutions.

Chapter 9. Appendix C: Box Plots Comparing Interactive and Non-Interactive Optimization Solutions.

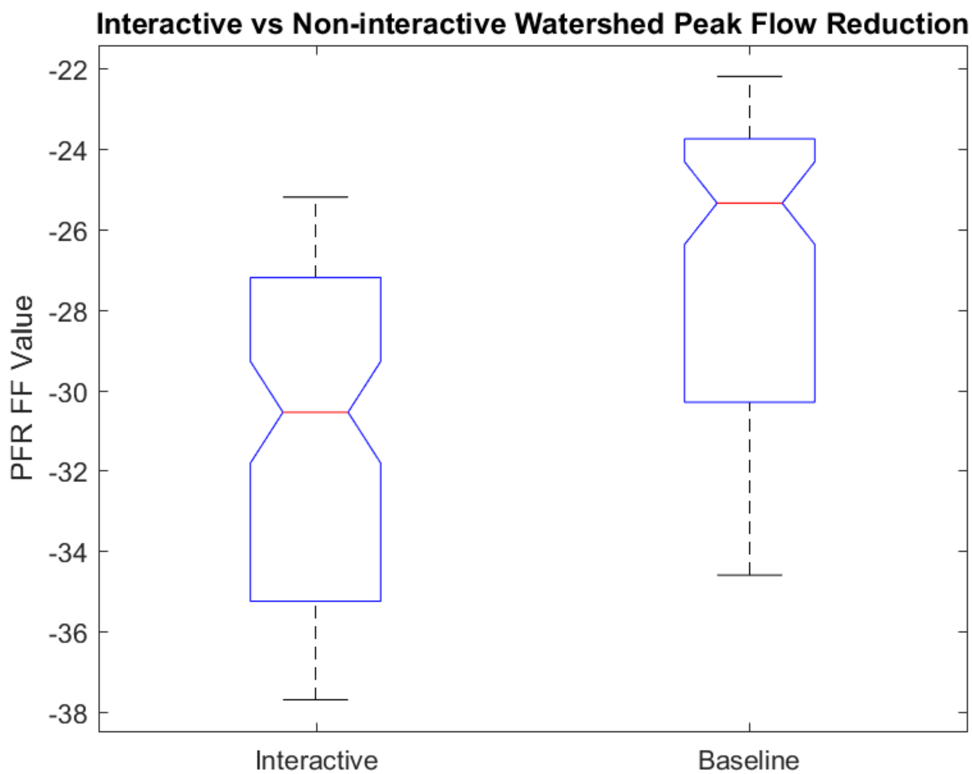

Figure 27: Comparing Interactive and Non-Interactive Peak Flow Reduction Fitness Functions

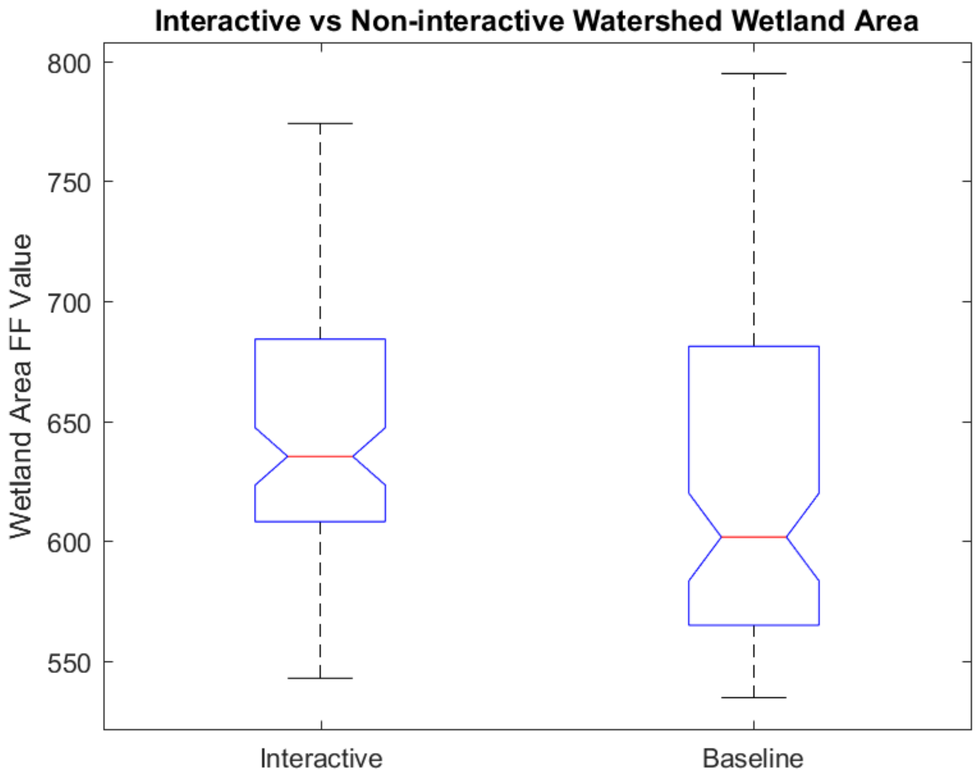

Figure 28: Comparing Interactive and Non-Interactive Wetland Area Fitness Functions

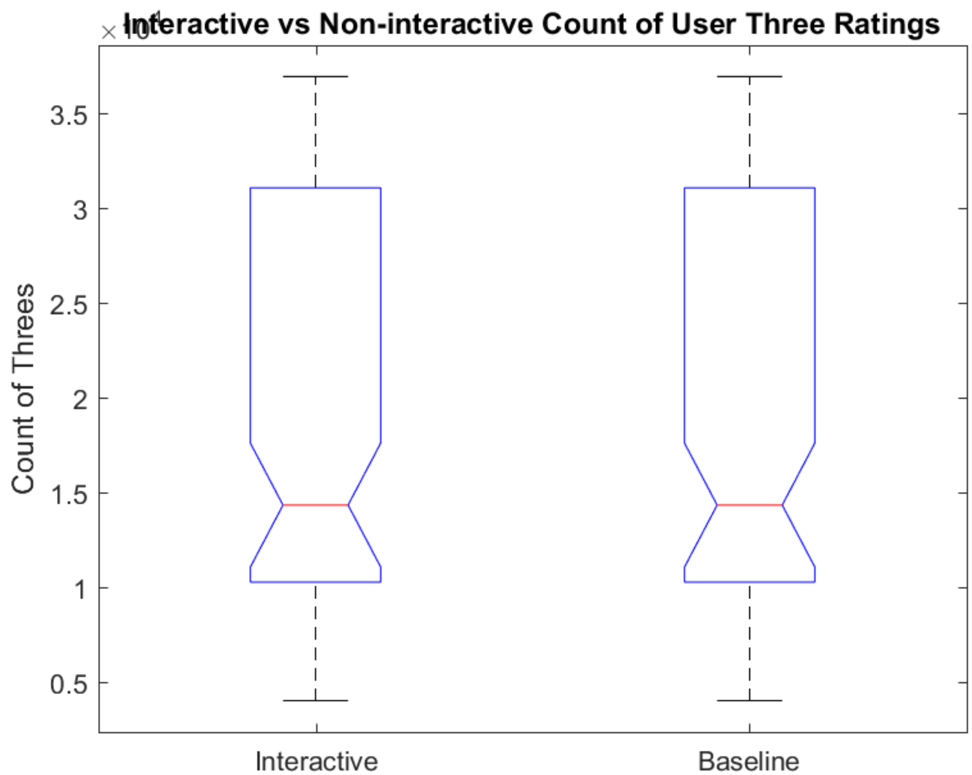

Figure 29: Comparing Interactive and Non-Interactive Three Ratings Counts

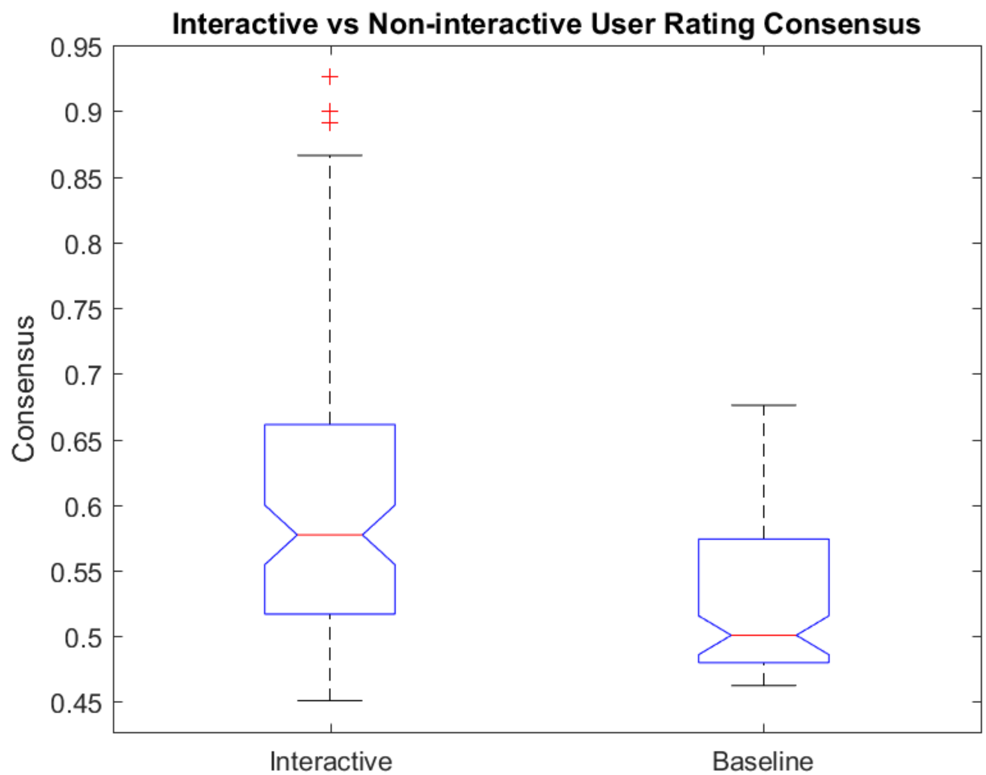

Figure 30: Comparing Interactive and Non-Interactive Consensus Fitness Functions

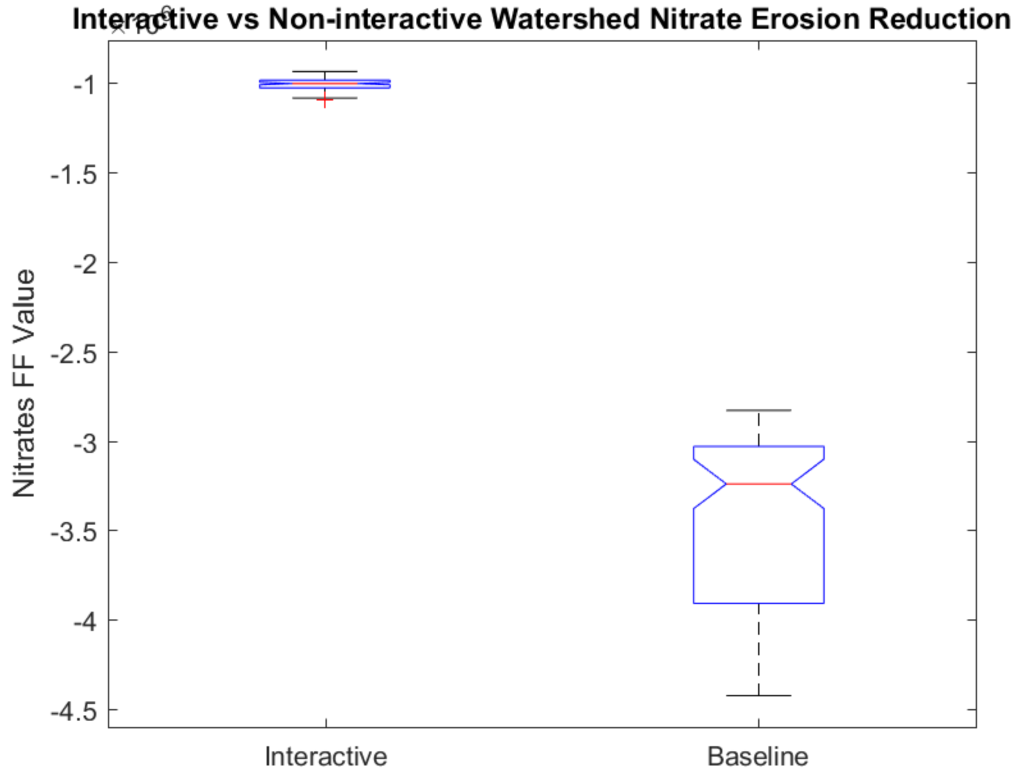

Figure 31: Comparing Interactive and Non-Interactive Nitrate Reduction Fitness Functions

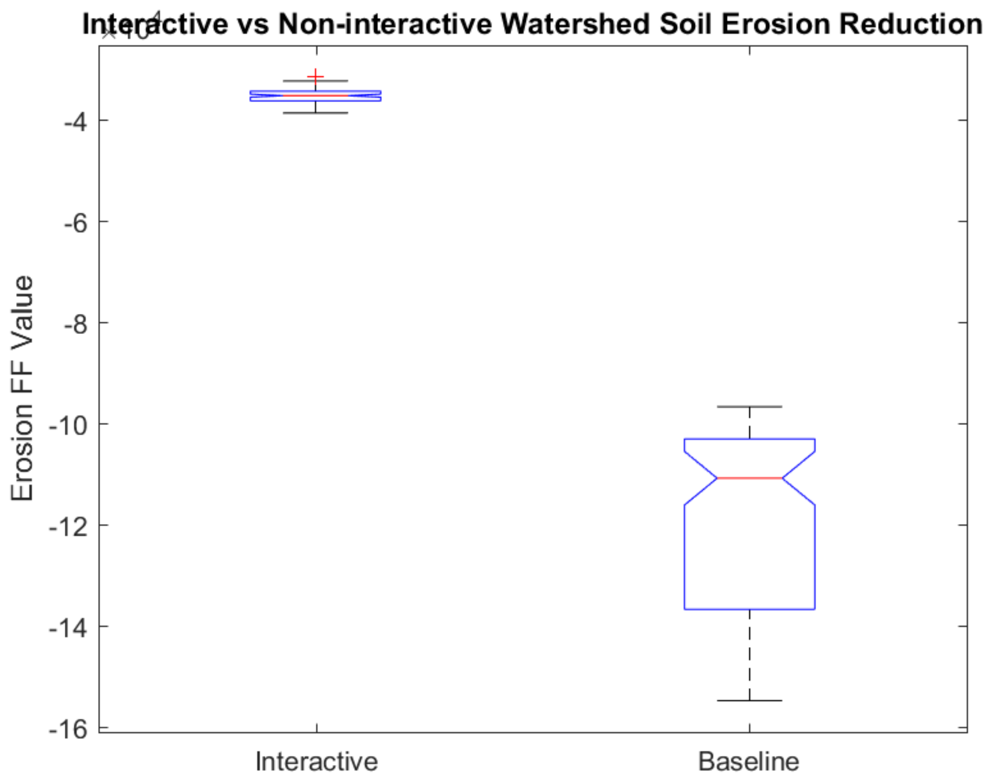

Figure 32: Comparing Interactive and Non-Interactive Erosion Reduction Fitness Functions