

## AN ABSTRACT OF THE DISSERTATION OF

Adriana Debora Piemonti for the degree of Doctor of Philosophy in Civil Engineering presented on December 10, 2015.

Title: Interactive Genetic Algorithms for Watershed Planning: An Investigation of Usability and Human-centered Design

Abstract approved:

---

Meghna Babbar-Sebens

Degradation of watersheds is a major concern in areas where adverse climate effects and unsustainable use of the natural resources have caused extensive stresses to watershed systems (e.g., increased floods, increased droughts, worsened in-stream water quality) through the years. While considerable efforts are being made to generate technical solutions that focus on plans of spatially-distributed conservation practices (e.g., Wetlands, Filter Strips, Grassed Waterways, Crop Management practices, etc.) for restoration of existing conditions in the watersheds, adoption and implementation of these solutions require a better understanding of constraints faced by affected stakeholders and decision makers. Participatory modeling and design approaches have, as a result, become popular in the recent past to support a community's engagement during the modeling process and during development of potential scenarios of plans (or, design alternatives). And now, with new and ongoing developments in Web 2.0 technologies, there is an even greater need for research that examines how large number of stakeholders can be engaged in the development of design alternatives via the internet-based, decision support environments.

The overarching goal of this research is to investigate how stakeholder participation ("humans") and Interactive Genetic Algorithms ("computer") can be coupled in a web-based watershed decision support system (DSS) called WRESTORE (Watershed REstoration using Spatio Temporal Optimization of REsources-<http://wrestore.iupui.edu/>), in order to generate user-preferred design alternatives of

distributed conservation practices on a watershed landscape. An important component of this goal is to also improve the understanding of how human behavior on the graphical user interface (GUI) of the DSS can be observed and evaluated in real-time, and then learned from to further improve the performance of the underlying search algorithm. Four specific objectives were addressed in this work to accomplish the overall goal:

- Objective 1: Observe interactions of multiple users with the GUI of a web-based watershed DSS (WRESTORE, <http://wrestore.iupui.edu/>) during interactive search experiments, and then use Usability metrics (response times, clicking events and confidence levels) to evaluate the differences and similarities in user behaviors and interactions.
- Objective 2: Examine relationships between the type of users (e.g., stakeholders versus surrogates), the Usability metrics, and patterns in the watershed-scale plans of conservation practices generated by the multi-objective Interactive Genetic Algorithm embedded in WRESTORE.
- Objective 3: Examine relationships between the type of users, the Usability metrics, and patterns in the user-preferred, sub-basin-scale plans of conservation practices generated by the multi-objective Interactive Genetic Algorithm embedded in WRESTORE.
- Objective 4: Develop and test novel human-guided search operators that adaptively learn for patterns in user-preferred alternatives generated by the multi-objective Interactive Genetic Algorithm, and, as a result, improve the convergence rate of the search algorithm for generating design alternatives that conserve these learned patterns.

Results show that there is a clear difference on how different types of users interact with the Interactive Optimization system. The observed relationship between confidence levels, time spent on a task, and number of mouse clicking events, indicated that participants who were able to use the WRESTORE GUI to gather more information and had a higher rate of time per number of clicks, tended to increase their levels of self-confidence in their own feedback. Also, when engaging with watershed

stakeholders versus non-stakeholders (or, surrogates), 67% of the stakeholder participants steadily increased their average self-confidence levels as they continued to interact with the tool, in contrast to only 29% of surrogate participants who also showed an increase in their self-confidence levels through time. Such usability and confidence level evaluations provide assessments on which participant was potentially generating reliable feedback data for the search algorithm to use. An analysis of design alternatives generated by the individuals in both stakeholder and non-stakeholder groups showed that a majority (67%) of the stakeholder participants found a higher percentage (on and average 52%) of preferred design alternatives via the interactive search process. Also, users who were focused on assessing the suitability of design alternatives for the entire watershed trended to demonstrate a bias for one of the watershed-scale objective functions. In contrast, users, who were focused on assessing the suitability of design alternatives at only a few local sub-basins in the watershed, did not demonstrate any clear bias for any one of the watershed-scale objective functions. Additionally, patterns were observed in the design of decision alternatives generated by the human-centered search process, which further divulged potential user preferences related to the decision space for example, whether a specific participant preferred a certain practice over another, or a certain location over another for a specific practice. Finally, to improve the convergence rates of the Interactive Genetic Algorithm in WRESTORE, we investigated whether observed patterns in decisions (especially, when users were focused on local sub-regions of the watershed) can be used to improve the search for user-desire designs. A novel Interactive Genetic Algorithm with adaptive, human-guided, selection, crossover and mutation operators was proposed. The new algorithm was tested with six types of simulated participants (three deterministic and three probabilistic users) developed from the feedback data of three real participants. Results of search experiments with the novel adaptive IGA operators indicated a faster convergence than the default IGA, for two out of three deterministic simulated users. However, none of the probabilistic user showed a convergence different than the default values. This indicates that while current results indicate promise, there is need for additional research on adaptive, human-guided IGA operators, especially when noisy/stochastic users participate in the search. Additionally, adaptation of search

operators have the potential to improve convergence rates when participatory design is done via Interactive Genetic Algorithms.

©Copyright by Adriana Debora Piemonti  
December 10, 2015  
All Rights Reserved

Interactive Genetic Algorithms for Watershed Planning: An Investigation of Usability  
and Human-centered Design

by  
Adriana Debora Piemonti

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of  
the requirements for the  
degree of

Doctor of Philosophy

Presented December 10, 2015  
Commencement June 2016

Doctor of Philosophy dissertation of Adriana Debora Piemonti presented on  
December 10, 2015

APPROVED:

---

Major Professor, representing Civil Engineering

---

Head of the School of Civil and Construction Engineering

---

Dean of the Graduate School

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

---

Adriana Debora Piemonti, Author

## ACKNOWLEDGEMENTS

First and foremost I would like to thank God, because He provided me direction and lite my path during this journey. Even during my worst moments, I would turn to my faith and realize that all the events in my life happen for a good and valuable reason.

Next, I would like to thank my advisor, Dr. Meghna Babbar-Sebens, who has been a wonderful person. She gave me the opportunity to work with her and has patiently helped me throughout graduate school. I would have never achieved this goal if it was not for her wise advice, kind words, and, when necessary, the very active brainstorm discussions. I also would like to thank my committee members, especially Dr. Kristen Macuga, for all her support and knowledge during the development of this research. I would like to extend another special acknowledgment to the Geotechnical professors and students. As an honorary member of GIGSO, I enjoy all the activities I was invited to, as well as the great support during my time in OWEN 337.

Special thanks to my family, in particular my mom Emma and my uncle Gonzalo, who has been there to encourage me to continue in every step, and have been always happy to help me when I needed it. I miss you the most, and I am looking forward to your visit next year! My family acknowledgement is not completed without mentioning my little piece of Venezuela, Nash. He is just the most special guy in my life.

Finally, to my friends, new (among them: Maggie, Stephe, Estefania) and old (among them: Jordyn, Daniel, James, Deanna, Monique), getting to meet you was a wonderful part of this incredible experience. I am pretty sure you will never forget me, because I am just that amazing ;). There are three guys that have to live with the idea of having me around for life: Trevor Carey, Andrew Strahler, and Seth Reddy. I do not think you know how much you three mean to me, and wherever I end up, you are always welcomed (as long as you bring spiritual beverages) with open arms (Mi casa es su casa), as you are already part of my most important extended family. I cannot say you are like my brothers, because you cannot be compared (you need to buy a set of “the pants” first, which I hope never happens). But I see you as part of my own cool version of uncles :P. Love u guys, and THANKS for all the memories!



## CONTRIBUTION OF AUTHORS

Dr. Meghna Babbar-Sebens (Assistant Professor at Oregon State University) and Dr. Snehasis Mukopadhyay (Professor at Indiana University-Purdue University Indianapolis) are the lead researchers in the development of the web-based tool that has been tested and analyzed in this research project. Their collaboration is reflected through the development of this work.

Dr. Kristen Macuga (Assistant Professor at Oregon State University), whose expertise lies in Engineering Psychology, Action Observation and Tool Use, is the main collaborator in the usability and cognitive analysis conducted for the web-based WRESTORE tool. Her collaboration has been valuable in the understanding of human-computer interactions, and in interpretation of the different datasets collected via the graphical user interface (GUI).

Mr. Vidya Bhushan Singh and Mr. Mahesh Yerram are Computer Science graduate students at Indiana University-Purdue University Indianapolis, who collaborated in the computer programming tasks and assisted with the execution of user experiments.

Finally, Mr. Austin Kleinberg is a Civil Engineering undergraduate student at Oregon State University, who assisted with data analyses, and generation of plots and images for facilitating an improved understanding of the large set of results.

# TABLE OF CONTENTS

	<u>Page</u>
CHAPTER 1. Introduction.....	1
1.1 Problem Statement.....	1
1.2 Overarching Goal and Objectives .....	4
1.3 Outline.....	5
CHAPTER 2. Literature review .....	7
2.1 Brief overview of decision theory literature .....	8
2.2 Engagement of stakeholders during watershed restoration.....	10
2.3 Hydroinformatics approaches to support participation of stakeholders during design .....	12
2.4 Interaction of humans with search and optimization algorithms .....	14
2.5 WRESTORE DSS for interactive optimization .....	18
2.6 Interactions between Participants and the WRESTORE-Tool.....	26
CHAPTER 3. Case Study .....	30
3.1 Decision Support System .....	30
3.2 Experimental Watershed .....	31
3.2.1. Hydrologic and Water Quality model .....	33
3.3 Participant selection and model modification .....	37
CHAPTER 4. Usability evaluation of an interactive decision support system for planning of conservation practices in a watershed: A case study .....	41
4.1 Abstract .....	42
4.2 Introduction .....	42
4.3 Objectives.....	45
4.4 Methodology .....	46
4.4.1. Case study site.....	46
4.4.2. Web-tool WRESTORE evaluation .....	46
4.4.3. Measures of user interaction with the interface .....	48
4.4.4. Confidence levels .....	50
4.4.5. Relationships between confidence levels, time spent and mouse clicking events .....	51
4.5. Results.....	52
4.5.1. Overall task times.....	52
4.5.2. Mean percentage of time spent in different areas of interest.....	54
4.5.3. Mean percentage of clicking events in different areas of interest.....	55

## TABLE OF CONTENTS (Continued)

	<u>Page</u>
4.5.4. Confidence Levels.....	56
4.5.5. Relationships between <i>confidence levels</i> , time spent and mouse clicking events. ....	57
4.6. Discussion .....	60
4.6.1. Overall task times.....	60
4.6.2. Mean percentage of time spent in different areas of interest .....	61
4.6.3. Mean percentage of clicking events in different areas of interest.....	62
4.6.4. Confidence levels .....	62
4.6.5. Relationships between <i>confidence levels</i> , time spent and mouse clicking events. ....	62
4.7 Conclusions and future work.....	63
4.8 Acknowledgement .....	65
4.9 References .....	65
CHAPTER 5. Participatory design of distributed conservation practices in a watershed: An examination of relationship between user behavior and watershed-scale plans generated via Interactive Optimization .....	67
5.1 Abstract .....	68
5.2 Introduction .....	69
5.3 Research Objective .....	70
5.4 Methodology .....	71
5.4.1. Setup of Participatory Design Experiments.....	71
5.4.2. Evaluation Metrics .....	74
5.5 Results and discussion .....	81
5.5.1. EoS assessment of user searches experiments .....	82
5.5.2. DS assessment of user search experiments .....	99
5.6 Conclusions and future work.....	105
5.7 Acknowledgements.....	107
5.8 References .....	107
CHAPTER 6. Participatory design of distributed conservation practices in a watershed: Effect of spatially-explicit user preferences on Interactive Optimization .....	108
6.1 Abstract .....	109
6.2 Introduction .....	110
6.3 Objectives.....	111
6.4 Methodology .....	112
6.4.1. Spatially-explicit Subbasins of interest .....	112

## TABLE OF CONTENTS (Continued)

	<u>Page</u>
6.4.2. Objective Space Analysis .....	113
6.4.3. Decision space patterns .....	114
6.5 Results and discussion .....	116
6.5.1. Objective Space Analysis .....	116
6.5.2. Decision Space Analysis .....	119
6.6 Conclusions and future work.....	122
6.7 Acknowledgment.....	124
6.8 References .....	124
CHAPTER 7. Human-centered design of spatial allocation of conservation practices in watersheds: Mining user responses to adapt search operators of an Interactive Genetic Algorithm .....	126
7.1 Abstract .....	127
7.2 Introduction .....	128
7.3 Objectives.....	129
7.4 Methodology .....	130
7.4.1. Interactive Optimization Algorithm .....	130
7.4.2. Proposed Adaptive Human-Guided Search (AHS) Operators.....	131
7.4.3. Simulated Participants (SP) .....	133
7.4.4. Operators Update. ....	135
7.4.5. Metrics for Evaluating Convergence and Performance.....	137
7.5 Results and discussion .....	138
7.5.1. Percentage of user rating .....	138
7.5.2 Operators: Crossover and mutation changes .....	139
7.6 Conclusions .....	141
7.7 Acknowledgements.....	143
7.8 References .....	143
CHAPTER 8. Final discussion .....	144
8.1 Conclusions .....	144
8.1.1. Usability Metrics for IGA.....	145
8.1.2. Watershed Scale Plans Generated from IGA.....	147
8.1.3. Local Scale Plans Generated from IGA .....	148
8.1.4. New Adaptive Operators for IGA.....	150
8.2 Future Research .....	151
REFERENCES .....	153

## TABLE OF CONTENTS (Continued)

	<u>Page</u>
APPENDIX.....	165

# LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
Figure 2.1 WRESTORE website home page .....	19
Figure 2.2 <i>Interaction sessions</i> in IGAMII (Babbar-Senbens et al., 2015) .....	20
Figure 2.3 Visualization and Feedback Interface in WRESTORE (Babbar-Sebens et al., 2015) .....	23
Figure 2.4 Development of the WRESTORE tool for an asynchronous experiment Modified from Babbar-Sebens et al., 2015.....	24
Figure 3.1 Location of Eagle Creek Watershed. (Taken from Tedesco et. al, 2005) .	32
Figure 3.2 EWC SB and reach (Taken from Piemonti et al., 2013) .....	34
Figure 3.3 SBs assigned to different participants .....	39
Figure 3.4 Interface for starting a new search experiment for the user’s watershed of interest.....	40
Figure 4.1 Components of WRESTORE tool’s main design and feedback interface and areas of interest (AOI) (Modified from Babbar-Sebens et al., 2015) .....	48
Figure 4.2 Power function representing the learning curves of the mean task times for the surrogates and stakeholders groups across <i>I</i> sessions. Error bars represent the standard error of the mean. ....	53
Figure 4.3 Power function representing the learning curves of the mean task times for the surrogates and stakeholders groups across HS sessions. Error bars represent the standard error of the mean. ....	54
Figure 4.4 Mean percentage of time spent in each AOI for surrogates and stakeholders. Refer to section 4.2.2 and Figure 4.1 for descriptions of each AOI. ....	54
Figure 4.5 Mean percentage of mouse clicking events for surrogates and stakeholders within each AOI. Refer to section 4.2.2 and Figure 4.1 for descriptions of each AOI. ....	55
Figure 4.6 Time spent vs. number of clicks per participant per trend, for <i>Info</i> (upper) and <i>Eval</i> (lower) AOIs. Sections a) and b) show the results for sessions <i>I1</i> to <i>I3</i> . Sections c) and d) show the results for all completed sessions. The empty circles show the change in trends for Participant .....	59
Figure 5.1 Percentage of designs classified by the participants at each level of the user rating <i>Model A-Surrogates</i> .....	84

## LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
Figure 5.2 Percentage of designs classified by the participants at each level of the user rating for <i>Model B-Surrogates</i> .....	85
Figure 5.3 Percentage of designs classified by the participants at each level of the user rating for <i>Model B-Stakeholders</i> .....	86
Figure 5.4 Pareto front representation of watershed performance for Participant 1 (a) and Participant 6 (b)) in <i>Model A-Surrogates</i> group. Participant 1 was asked to provide the <i>user rating</i> for the design alternative based on the watershed performance, while Participant 6 was asked to provide the user rating for the design alternative based on the group of SBs: 103, 105, 106, 121, and 122.....	89
Figure 5.5 Part Pareto front representation of watershed performance for Participant 8 (a) and Participant 20 (b)) in <i>Model B-Surrogates</i> group. Participant 8 was asked to provide the <i>user rating</i> for the design alternative based on the watershed performance, while Participant 20 was asked to provide the user rating for the design alternative based on the group of SBs: 58, 59, 61 and 63. Similarly, the Pareto Fronts for Participants 15 and 11 (c) and d) respectively) in <i>Model B-Stakeholders</i> group are shown. Participant 15 was asked to provide the <i>user rating</i> for the design alternative based on the watershed performance, while Participant 11 was asked to provide the user rating for the design alternative based on the group of SBs: 41,90,92, and 93...	91
Figure 5.6. Examples of the distribution of $ProR3 > 0.5$ (a and b) and $POM > 50\%$ (c and d) for Participants 6 and 7 .....	95
Figure 5.7. Maps of Average $ProbgR3$ (upper row) and Standard Deviation (lower row) for each of the Model groups. From left to right we have <i>Model A-Surrogates</i> (a and d), <i>Model B-Surrogates</i> (b and e) and <i>Model B-Stakeholders</i> (c and f) .....	98
Figure 5.8. Maps of Average $ProbgR3$ (upper row) and Standard Deviation (lower row) for each of the Model groups. From left to right we have <i>Model A-Surrogates</i> (a and d), <i>Model B-Surrogates</i> (b and e) and <i>Model B-Stakeholders</i> (c and f) .....	99
Figure 5.9. Participants 1, 3 and 6 percentage of solutions in <i>user rating</i> $R_i$ by Epochs .....	100
Figure 5.10 Average of percentage of solutions per participant per Epoch.....	102
Figure 6.1 Histogram and correlations of $SB_{int}$ for Participant 2 .....	118
Figure 6.2 Group analysis of the Objective Space in for the participants in the $SB_{int}_1$ .....	119

## LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
Figure 6.3 Individual (bar plots) and average (black lines) values of probability of CoverCrops by SBint for Group 1 .....	121
Figure 6.4 Example of the mode and percentage of the filter strip width for each participant, in SBint1 .....	122
Figure 7.1 AHS flowchart.....	132
Figure 7.2 Flow chart for the activator of AHS .....	136
Figure 7.3 Internal process for the AHS after the activator have been set. ....	137
Figure 7.4 comparisons between the deterministic user and the Stochastic User for Participant 3 .....	139
Figure 7.5 Probability of crossover for Cover Crops (up) and mutation (down) for SB 12 (User 3) for deterministic (left) and stochastic (right) SP through generations...	140
Figure 7.6 Probability of crossover for Filter Strip (up) and mutation (down) for SB 12 (User 3) for deterministic (left) and stochastic (right) SP through generations...	141



## LIST OF TABLES

<u>Table</u>	<u>Page</u>
Table 3.1 Parameters used in the calibration of SWAT .....	36
Table 3.2 Parameters modified for the sediments and nitrates calibration .....	37
Table 4.1 Confidence intervals (CI) for the mean percentages of time spent in each AOI across all sessions by group .....	55
Table 4.2 Confidence intervals (CI) for the mean percentages of mouse clicking events within each AOI across all sessions by group. ....	56
Table 4.3 Classification of confidence levels, mean percentage of clicking events and mean percentage of time spent across participants with the same trend.....	58
Table 4.4 Coefficient of determination and AIC values for the different tested models .....	60
Table 5.1 Objective function formulas used to optimize the design alternatives used by the WRESTORE .....	72
Table 5.2 Changes made on the SWAT model to simulate conservation practices ...	73
Table 5. 3 Equations used in the evaluation of the decision space .....	81
Table 5.4 Distance from design alternatives performance between non-interactive and interactive Pareto Fronts for Peakflow Reductions (PFR), Sediments Reduction (SR) and Nitrates Reduction (NR) for <i>Model A-Surrogates</i> .....	88
Table 5.5 Distance from design alternatives performance between non-interactive and interactive Pareto Fronts for Peakflow Reductions (PFR), Sediments Reduction (SR) and Nitrates Reduction (NR) .....	90
Table 5.6 Group average for distance between non-interactive optimal Pareto Front	92
Table 5.7 Number of SBs with a design alternative pattern for the Probabilities of Cover Crops and the Filter Strip width. ....	94
Table 5.8 Number of SBs with a design alternative pattern for the Probabilities of Cover Crops and the Filter Strip width. ....	96
Table 6.1 ID numbers of SBs inside the SBint groups and participants focused on this areas .....	113
Table 6. 2 Table with the objective function and the type of Objective Function....	114

## LIST OF TABLES (Continued)

<u>Table</u>	<u>Page</u>
Table 7.1 Assignment of user rating based on scaled values.....	134
Table 7.2 Summary table of the selected rates in each SB based on the DCi .....	135

## LIST OF APPENDIX FIGURES

<u>Figure</u>	<u>Page</u>
Figure A.1 Usability metrics Participant 1 .....	165
Figure A.2 Usability metrics Participant 2 .....	165
Figure A.3 Usability metrics Participant 3 .....	165
Figure A.4 Usability metrics Participant 4 .....	165
Figure A.5 Usability metrics Participant 5 .....	166
Figure A.6 Usability metrics Participant 6 .....	166
Figure A.7 Usability metrics Participant 7 .....	166
Figure A.8 Usability metrics Participant 8 .....	167
Figure A.9 Usability metrics Participant 9 .....	167
Figure A.10 Usability metrics Participant 20 .....	167
Figure A.11 Usability metrics Participant 21 .....	167
Figure A.12 Usability metrics Participant 22 .....	168
Figure A.13 Usability metrics Participant 24 .....	168
Figure A.14 Usability metrics Participant 25 .....	168
Figure A.15 Usability metrics Participant 11 .....	169
Figure A.16 Usability metrics Participant 13 .....	169
Figure A.17 Usability metrics Participant 14 .....	169
Figure A.18 Usability metrics Participant 15 .....	169
Figure A.19 Usability metrics Participant 16 .....	170
Figure A.20 Usability metrics Participant 18 .....	170
Figure A.21 EoS Percentage of design alternatives per user rating Participant 1 .....	171
Figure A.22 EoS Percentage of design alternatives per user rating Participant 2 .....	171
Figure A.23 EoS Percentage of design alternatives per user rating Participant 3 .....	171
Figure A.24 EoS Percentage of design alternatives per user rating Participant 4 .....	171

LIST OF APPENDIX FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
Figure A.25 EoS Percentage of design alternatives per user rating Participant 5 .....	172
Figure A.26 EoS Percentage of design alternatives per user rating Participant 6 .....	172
Figure A.27 EoS Percentage of design alternatives per user rating Participant 7 .....	172
Figure A.28 EoS Percentage of design alternatives per user rating Participant 8 .....	173
Figure A.29 EoS Percentage of design alternatives per user rating Participant 9 .....	173
Figure A.30 EoS Percentage of design alternatives per user rating Participant 20 ...	173
Figure A.31 EoS Percentage of design alternatives per user rating Participant 21 ...	173
Figure A.32 EoS Percentage of design alternatives per user rating Participant 22 ...	174
Figure A.33 EoS Percentage of design alternatives per user rating Participant 24 ...	174
Figure A.34 EoS Percentage of design alternatives per user rating Participant 25 ...	174
Figure A.35 EoS Percentage of design alternatives per user rating Participant 11 ...	175
Figure A.36 EoS Percentage of design alternatives per user rating Participant 13 ...	175
Figure A.37 EoS Percentage of design alternatives per user rating Participant 14 ...	175
Figure A.38 EoS Percentage of design alternatives per user rating Participant 15 ...	175
Figure A.39 EoS Percentage of design alternatives per user rating Participant 16 ...	176
Figure A 40 EoS Percentage of design alternatives per user rating Participant 18 ...	176
Figure A.41 DS Percentage of design alternatives per user rating Participant 2 .....	176
Figure A.42 DS Percentage of design alternatives per user rating Participant 4 .....	176
Figure A.43 DS Percentage of design alternatives per user rating Participant 5 .....	176
Figure A 44 DS Percentage of design alternatives per user rating Participant 7 .....	176
Figure A.45 DS Percentage of design alternatives per user rating Participant 8 .....	177
Figure A.46 DS Percentage of design alternatives per user rating Participant 9 .....	177
Figure A.47 DS Percentage of design alternatives per user rating Participant 20 .....	177

LIST OF APPENDIX FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
Figure A.48 DS Percentage of design alternatives per user rating Participant 21 ....	177
Figure A.49 DS Percentage of design alternatives per user rating Participant 22 ....	177
Figure A.50 DS Percentage of design alternatives per user rating Participant 24 ....	177
Figure A.51 DS Percentage of design alternatives per user rating Participant 25 ....	178
Figure A.52 DS Percentage of design alternatives per user rating Participant 11 ....	178
Figure A.53 DS Percentage of design alternatives per user rating Participant 13 ....	178
Figure A.54 DS Percentage of design alternatives per user rating Participant 14 ....	178
Figure A. 55 DS Percentage of design alternatives per user rating Participant 15 ....	178
Figure A.56 DS Percentage of design alternatives per user rating Participant 16 .....	179
Figure A.57 DS Percentage of design alternatives per user rating Participant 18 .....	179
Figure A. 58 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 2 .....	179
Figure A. 59 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 3 .....	179
Figure A.60 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 4 .....	180
Figure A.61 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 5 .....	180
Figure A. 62 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 7 .....	180
Figure A.63 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 9 .....	181

LIST OF APPENDIX FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
Figure A.64 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 21 .....	181
Figure A.65 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 22 .....	181
Figure A.66 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 24 .....	181
Figure A. 67 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 25 .....	182
Figure A.68 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 13 .....	182
Figure A.69 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 14 .....	182
Figure A.70 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 16 .....	183
Figure A.71 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 18 .....	183
Figure A.72 Matrix plot for performance of Peak flow reduction (PFR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 2 .....	184
Figure A.73 Matrix plot for performance of Cost of the design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 2.....	184
Figure A.74 Matrix plot for performance of Sediment reduction (SR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 2 .....	185
Figure A.75 Matrix plot for performance of Nitrates reduction (NR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 2 .....	185

LIST OF APPENDIX FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
Figure A.76 Matrix plot for performance of Peak flow reduction (PFR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 9 .....	186
Figure A. 77 Matrix plot for performance of Cost of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 9.....	186
Figure A.78 Matrix plot for performance of Sediment reduction (SR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 9 .....	187
Figure A.79 Matrix plot for performance of Nitrate reduction (NR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 9 .....	187
Figure A. 80 Matrix plot for performance of Peak flow reduction (PFR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 16 .....	188
Figure A.81 Matrix plot for performance of Cost of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 16.....	188
Figure A.82 Matrix plot for performance of Sediment Reduction (SR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 16 .....	189
Figure A.83 Matrix plot for performance of Sediment Reduction (SR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 16 .....	189
Figure A.84 Ratio group analysis for SB of interest in group 2 .....	190
Figure A.85 Ratio group analysis for SB of interest in group 3 .....	190
Figure A.86 Ratio group analysis for SB of interest in group 4 .....	190
Figure A.87 Ratio group analysis for SB of interest in group 5 .....	190
Figure A.88 Ratio group analysis for SB of interest in group 6 .....	191
Figure A.89 Bar plots for Probabilities of Cover Crop in SB of interest in group 2 .	191
Figure A.90 Bar plots for Probabilities of Cover Crop in SB of interest in group 3 .	192
Figure A.91 Bar plots for Probabilities of Cover Crop in SB of interest in group 4 .	192
Figure A.92 Bar plots for Probabilities of Cover Crop in SB of interest in group 5 .	193
Figure A.93 Bar plots for Probabilities of Cover Crop in SB of interest in group 6 .	193
Figure A.94 Bar plots for Filter Strip mode and Percentage of design alternatives equal to the mode in SB of interest for group 2.....	194

LIST OF APPENDIX FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
Figure A.95 Bar plots for Filter Strip mode and Percentage of design alternatives equal to the mode in SB of interest for group 3.....	195
Figure A. 96 Bar plots for Filter Strip mode and Percentage of design alternatives equal to the mode in SB of interest for group 4.....	195
Figure A.97 Bar plots for Filter Strip mode and Percentage of design alternatives equal to the mode in SB of interest for group 5.....	196
Figure A. 98 Bar plots for Filter Strip mode and Percentage of design alternatives equal to the mode in SB of interest for group 6.....	196
Figure A.99 Percentage of design alternatives in different generations for: Deterministic Simulated user 1 (left) and Probabilistic Simulated user 1 (right).....	197
Figure A. 100 Percentage of design alternatives in different generations for: Deterministic Simulated user 5 (left) and Probabilistic Simulated user 5 (right).....	197



## **CHAPTER 1. Introduction**

### **1.1 Problem Statement**

The Environmental Protection Agency (EPA) defines “Watersheds” as the *area of land where all the water that is under it, or drains off of it goes into the same place* (<http://water.epa.gov/type/watersheds/whatis.cfm>). Watersheds provide multiple resources (such as water supplies, food supplies, wood supplies, minerals, etc.) and *ecosystem services* (defined by MEA (2005) as the benefits obtain from the ecosystems, such as, hydrologic services – water storage and availability-, nutrient cycling, biological regulations, etc.) that are necessary for create sustainable living environments for communities (Strange et al., 1999; Smith et al., 2006, Brauman et al., 2007). Therefore, effective management of watersheds is critical for supporting the activities of settled communities.

However, alteration of hydrologic cycle and impairment of streams has become a reality in a large number of watersheds in the United States and across the world, because of changes in land use that were necessary to facilitate the development and growth of settled communities. In the late 1800’s, the United States Federal Government became concerned about the growing impacts of ineffective watershed management on the general well-being of human and ecological communities living in the watersheds. The rapid increases in the population had begun to stress natural resources, in particular, water resources, which were necessary for ensuring water supply and waterway navigations (Cole et al., 2002). However, even though Federal entities were working towards finding solutions to overcome the impacts of developments (e.g., urban and agriculture) and population growth on the natural resources (water, forestry and soil), it was soon realized that successful adoption of prescribed actions by the stakeholders of the watershed was critical to ensuring effective restoration and remediation of degraded watershed conditions (Gregersen et al., 2007). Over time, it became even more important for watershed planners to effectively communicate with stakeholders the multiple consequences of different scenarios of restoration plans that could potentially be adopted by them.

According to the United States Department of Agriculture's (USDA) "Budget Summary and Annual Performance Plan," the conservation programs and have received approximately \$66 billion in the last five years to assist with the implementation of plans that seek to diminish environmental damage (<http://www.obpa.usda.gov/budsum/FY15budsum.pdf>). Currently, the NRCS (National Resources Conservation Service) is the leading federal agency to assist in the restoration of watershed services providing technical and financial assistance to those private landowners committed to the re-naturalization of ecosystem services.

The complex relationships and tradeoffs between the available watershed resources, available restoration budgets, and a community's needs and aspirations have presented significant challenges in managing and restoring the quality of watershed resources (Johnson et al., 2002; Poch et al., 2004; Bonnell and Koontz, 2007; Jakeman et al., 2008; Cohen and Davidson, 2011; Margerum and Robinson, 2015), and the ecosystem services provided by these resources. Pahl-Wostl (2005), argues that, because of communities' dissatisfaction with prescribed expensive and non-sustainable solutions, there is a need to find decentralized technological alternatives that include individual stakeholders in the problem-solving process and help them find acceptable and sustainable solutions. Several other studies have also recommended participation of different professionals and public in watershed planning and management projects as means for improving the adoption of management plans (Moote et al. 1997; Hamalainen et al., 2001; Lubell, 2004; Mostert et al., 2008; Kronaveter and Shamir, 2009; McFadden et al., 2011; Evers et al., 2012).

One of the major tasks of planners and managers, who are responsible for managing a community's resources, is to generate design alternatives for watershed management that satisfy community interests and criteria without overriding individual rights. Gregersen et al., (2007) indicate that successful watershed management should consider design alternatives for watershed's built-in complex systems based on both technical and institutional perspectives. These perspectives may be related to rules, norms and laws established and accepted by the community, and may consider the range of technological solutions available for meeting management goals. Therefore, new plans and designs must

be developed under a legal framework that protects the individual stakeholders' rights, while also enabling sustainability of the watershed's resources.

However, it is almost impossible to create design alternatives where different management objectives and community interests do not conflict. For example, a community constrained by budget limitations may end up adopting restoration alternatives that cost less but do not necessarily meet all water quality goals. Therefore, multi-objective methods (e.g., multi-objective optimization) have become essential for assisting a community to search for multiple scenarios of non-dominated alternatives that satisfy their distinct goals and criteria to varying degrees of satisfaction and under a range of possible conditions (Krettek et al., 2009, Porzecanski et al., 2012; Kelly et al., 2012). To incorporate the diverse goals and criteria, integrated watershed management that relies on a well-developed plan for public participation has become critical. Participation must include stakeholders in the decision-making process where they can: 1) get intimately involved from the conception of the project through the implementation and maintenance phases, and 2) become better informed about their problems and available choices for actions through a collective learning process in which the stakeholders can exchange information and accept responsibilities for the decisions made by the community (Moote et al. 1997).

However, incorporating participation of stakeholders in the development of design alternatives via multi-objective optimization can pose significant challenges, due to the existence of qualitative, unknown, unquantifiable, and unrepresented preferences, goals, and biases of a large number of stakeholders who are affected by watershed decisions. Hence, there is a critical need to develop stakeholder-centric decision support systems (DSS) that enable groups of stakeholders and communities to develop design alternatives in the participatory design environment, where their preferences and biases can be learned about, expressed, shared, and evaluated. Engagement of the public in using such DSS can help them learn about potential solutions that could be used to sustain or improve their livelihood and awareness of risks (Gregersen et al., 2007; Evers et al., 2012). Hence, a growing body of researcher (Abbott, 1991; Gregersen et al., 2007; Evers et al., 2012; Delipetrev et al, 2013; Babbar-Sebens et al., 2015) has suggested that information and communication technology (ICT) should be used to assist with engagement of stakeholders

in participatory design efforts in watershed planning, in addition to facilitating effective outreach in a community.

Development of the DSS that use ICT to support participatory processes requires consideration of multiple factors, including human. Theoretical and practical concepts from interdisciplinary fields such as civil engineering, engineering psychology, economics, and computer sciences, should be used towards rigorously evaluating the virtual design and visualization environments, behaviors of expected DSS users, and interactions between users and the different elements of the DSS. McIntosh et al. (2011) recently reported that most DSS used for supporting environmental decisions are not evaluated for their usability and face multiple hurdles related to interactions of the end users with the DSS. They further emphasized that there is a critical need to increase such evaluations as part of the DSS development and testing process so that standards and common issues of implementation can be identified.

This dissertation contributes to the development and evaluation of a web-based DSS for participatory design, along with a novel human – computer collaborative design approach to incorporating stakeholder participation in the development of design alternatives for watershed restoration.

## **1.2 Overarching Goal and Objectives**

The overarching goal of this research is to investigate how stakeholder participation (“humans”) and interactive search algorithms (“computers”) can be coupled in a web-based watershed DSS called WRESTORE (Watershed REstoration using Spatio Temporal Optimization of REsources- <http://wrestore.iupui.edu/>), in order to generate user-preferred design alternatives of distributed conservation practices on a watershed landscape. An important component of this goal is also to improve the understanding of how human behavior on the GUI of the DSS can be observed and evaluated in real-time and then learned from to improve further the performance of the underlying search algorithm. Four specific objectives were addressed in this work to accomplish the overall goal:

- **Objective 1:** Observe interactions of multiple users with the GUI of a web-based watershed DSS (WRESTORE) during interactive search experiments, and then use

- Usability metrics (response times, clicking events and confidence levels) to evaluate the differences and similarities in user behaviors and interactions.
- **Objective 2:** Examine relationships between the type of users (e.g., stakeholders versus surrogates), the Usability metrics, and patterns in the watershed-scale plans of conservation practices generated by the multi-objective Interactive Genetic Algorithm embedded in WRESTORE.
  - **Objective 3:** Examine relationships between the type of users, the Usability metrics, and patterns in the user-preferred, subbasin-scale plans of conservation practices generated by the multi-objective Interactive Genetic Algorithm embedded in WRESTORE.
  - **Objective 4:** Develop and test novel human-guided search operators that adaptively learn for patterns in user-preferred alternatives generated by the human-guided search algorithm, and, as a result, improve the convergence rate of multi-objective Interactive Genetic Algorithm for generating design alternatives that conserve these learned patterns.

### 1.3 Outline

In this research, the Web-based DSS WRESTORE was used to engage the multiple human participants in the interactive design of “user-preferred” alternatives for allocating conservation practices in a watershed. The test watershed chosen for this study was Eagle Creek Watershed (HUC-05120201120), which is located in Central Indiana near the city of Indianapolis. Chapter 2 reviews the essential background literature necessary to understand this body of work, followed by a brief description of the WRESTORE DSS and the multi-objective human-guided search algorithm IGAMII (Interactive Genetic Algorithm with Mixed-Initiative Interaction) used in WRESTORE to accomplish the workflow. Chapter 3 describes the test site Eagle Creek Watershed, followed by a brief overview of the simulation models used to simulate two conservation practices – cover crops and filter strips – in this watershed for flood and water quality benefits, and finally the formulation of the optimization problem for a distributed allocation of conservation practices on the landscape. Chapter 4 presents the methodology for the interactive search experiments that were conducted with a group of participants and the WRESTORE DSS

to accomplish the specific Objective 1, followed by results on the usability of the interactive design and visualization environment in WRESTORE. In Chapter 5, a methodology for examining patterns in the decision space and objective space at watershed scale is presented and then used to examine relationships between user-preferred alternatives generated by participants, the participant types, and the Usability data (collected for Objective 1). Chapter 6 conducts a study similar to that achieved in Chapter 5 but addressed Objective 3 by examining for patterns in the decision space and objective space at local sub-basins that were of interest to the participants. Chapter 7 uses the findings from Chapter 4, 5, and 6, to propose, develop, and test new search operators for the IGAMII algorithm in WRESTORE. These new search operators are adaptive in the sense that they learn from the patterns in the generated alternatives during the ongoing search process and steer the search process to converge quickly to alternatives that conserve the features found desirable by the user. Chapter 8 concludes the main findings from this research and provides recommendations for future investigations.

## CHAPTER 2. Literature review

This chapter gives an overview of the different subjects that have played a significant role in this research focused on interactive decision support systems (DSS) for watershed planning and management. DSS are useful information and communication technology (ICT) tools that can be used to assist stakeholders and decision makers (DMs) in building scenarios of watershed plans. These plans generally include a range of decisions – such as, strategies for on-the-ground, measures, and/or policies – that optimize a set of pre-determined, performance-based goals. Those goal may include quantitative bio-physical-economic achievements in the watershed (e.g., flooding impacts, water quality impacts, ecological impacts, and construction and maintenance costs), as well as qualitative individual personal criteria (e.g., landowner subjective preferences, local knowledge, and socio-cultural constraints). This research aims to evaluate and improve interaction of real humans (e.g., stakeholders) with a web-based DSS that has been previously developed to assist with generation of scenarios of spatial allocation of best management or conservation practices in a watershed.

The work investigated in this dissertation is also related to the concept of *human-centered design*. Hence, the discussion on background literature is also focused on issues pertinent to the problem of engaging individuals in optimization-based, human-centered design of watershed solutions. ISO 9241-210 (2008) provides guidelines to developers of computer software and hardware on how to generate usable and useful interactive technologies via a *human-centered design* process, so that they are able to assess user needs and expectations throughout the life cycle of the technology. These guidelines recommend use of multiple human factors/ergonomics, and usability knowledge and techniques to evaluate and improve such interactive systems. Nemeth (2004) has defined *user-centered design* as the consideration of users (or humans) and the technical subsystems where “*users are consulted through the design process*”. The definition of *optimization-based, human-centered design* adopted in this research is related to not only the ability of users to understand and effectively use the DSS technology, but to also assist the underlying optimization algorithm within the DSS in finding design alternatives that improve on

quantitative as well as qualitative stakeholder preferences. This work especially investigates how the graphical user interface elements of a DSS, the users' (or stakeholders') unique perspectives, and an interactive search/optimization algorithm should be combined together to provide an effective participatory design process for the allocation of best management practices in a watershed.

The sub-sections below describe the current understanding in related fields of decision theory, optimization, watershed planning and management, hydroinformatics, and participatory decision support systems, which provide the theoretical underpinnings for the proposed work. Section 2.1 provides an overview of the treatment of decision makers in the general Decision Theory literature, followed by discussion on how decision makers and stakeholders play an important role in watershed planning and management in Section 2.2. Section 2.3 introduces the concept of hydroinformatics and its role in facilitating engagement of stakeholders, followed by a discussion (in Section 2.4) on multiple types of methods and algorithms that have been previously used to accomplish the different levels of stakeholder participation in optimization of water resources and watershed optimization problems. Finally in Section 2.5, the WRESTORE (Watershed REstoration using Spatio Temporal Optimization of Resources) DSS, which is used to support the *optimization-based, human-centered design* process investigated in this research, is described.

## **2.1 Brief overview of decision theory literature**

In Rational Choice Theory (von Neumann and Morgenstern, 1944), decision-making processes must follow a rational and logical procedure, where the decision maker identifies a problem, defines causes and relationships, thinks about and identifies a possible set of solutions, and finally, selects the “best” or “optimal” option from a proposed set of solutions or alternatives. However, decision making is not an easy task and may include cognitive tasks, criteria, and actions performed by humans that may not seemingly be rational. Multiple researchers in Decision Theory and Behavioral Economics (Tversky, 1969; Kahneman and Tversky, 1979; Tversky and Kahneman, 1981; Kahneman and Tversky 1984; Baron, 2000) now claim that humans often do not follow principles of rational choice, and limitations in humans' cognition play an important role in determining their choices (Simon, 1955, 1956 and 1957). Simon also suggested that when people make



choices, many of them will “satisfice” instead of “maximizing” their goals. An individual who is a “satisficer” will tend to evaluate choices based on some degree of satisfaction, until a threshold of acceptability is achieved. When that occurs, the individual will select the choice that is both satisfactory and acceptable. Simon noted that when multiple choices are presented to a satisficer, he/she might choose a new choice if it is ranked higher than the previous one; therefore, a satisficer may sometimes move in the direction of maximization without realizing it. Therefore, algorithms that are able to generate discrete scenarios of choices can be useful for all types of human decision makers, irrespective of whether they are maximizers or satisficers, because it can provide them with a set of candidate alternatives/options that could potentially meet their individual criteria and/or constraints for decision making.

Designing or generating alternatives is an integral part of problem-solving and decision making processes. In commonly used models (and their adaptations) of decision-making processes, such as those proposed by Simon (1977) and Mintzberg et al. (1976), the design of alternatives usually occurs in the second phase of a three phase process that includes – (1) problem identification and definition phase, (2) problem development and alternatives generation phase, and (3) negotiation and selection phase. The first phase involves interaction with stakeholders and experts to identify, structure, and define the problem at hand. The second phase involves use of various computational tools, such as simulation models and search/optimization techniques along with the parameters of the search/optimization algorithm and quantitative representations of the problem objectives and constraints defined in phase 1, to generate optimal/feasible sets of alternatives that would satisfy or outperform the problem objectives. Once, the search has ended in phase 2, the alternatives are then presented to the stakeholders in phase 3 for decision making and selecting a final alternative for implementation.

Decision makers generally undergo through an exhaustive analytical process to identify ways to formulate an optimization problem, so that the solutions identified by the optimization algorithm represents their interests and knowledge (Tsitsiklis, 1984; Fonseca and Flemming, 1995; Peng et al., 2011). Therefore, it is crucial to recognize that a simplified representation of a decision maker’s interests and local knowledge of

stakeholders in the formulation of an optimization problem could lead to identification of seemingly “optimal” solutions that may actually be unacceptable to a decision maker, especially if he/she is not consistently consulted with and represented in the design process. Hence, participation of decision makers and other stakeholders throughout the entire process of decision making, including during the design process where alternatives are generated, is of paramount importance. This is particularly true for problems involving natural resources (such as, watersheds), where, beyond the typical quantifiable objectives and constraints, representation of subjective criteria in the design process is equally, if not more, crucial in order to ensure acceptance of alternatives by multiple stakeholders. However, developing problem formulations that represent the range of criteria relevant to an entire community, and where the required local actions agree with the interest of local stakeholders in the community, is a challenging process. These challenges are further complicated by the fact that the number of individuals affected by the decisions may vary through time and from one location to other, and also by the fact that individuals may differ in their perceptions of proposed choices based on their unique set of beliefs and values, social concerns, economic interests, awareness of the consequences of their decision, and personal cognitive bias and learning. As a result, a need has arisen for the development of heuristic and human-guided optimization methods that integrate knowledge from different fields, with assistance from social computing technologies for effective communication, to enable a more direct inclusion of subjective criteria and expression of preferences by multiple stakeholders during the decision making processes (Fraternali et al., 2012).

## **2.2 Engagement of stakeholders during watershed restoration**

Many watershed planning and management problems deal with allocation, use, and regulation of resources that are directly related to spatio-temporal necessities of humans co-existing in a shared area of the watershed. Federal concerns about careful planning and management of watersheds in the United States started in late 1800's when the increasing stresses on existing natural resources due to a rapidly growing population necessitated actions and policies to protect the resources, in particular, protection of water resources to assure water supply and waterway navigations (Cole et al., 2002). However, even though Federal entities aimed to find solutions to overcome the negative impacts from rapid

watershed developments and population growth, it was soon realized that only solutions that were generated in collaboration with the local stakeholders of the watershed could lead to successful adoption of restoration solutions (Gregersen et al., 2007).

Therefore, it has now become a priority to effectively communicate, instruct, and include stakeholders in the generation of designs of watershed restoration plans. These plans are expected to effectively target and achieve physical, biological, chemical and socio-economical goals determined in collaboration with the community stakeholders. Opportunities that enable stakeholders to contribute to the design process may also give them a sense of ownership, which may further increase the probability of adoption and acceptance of recommended plans (Moote et al., 1997, Johnson et al., 2001, Williams et al., 2012). Close interaction with stakeholders can also facilitate successful negotiations and conflict resolution, during the implementation of prescribed plans.

Engagement of stakeholders becomes especially relevant when restoration is achieved via spatial landscape management practices where stakeholders live and work. These landscape practices are called as conservation practices (CP) or best management practices (BMPs), and include wetlands, filter strips, grassed waterways, crop management practices (e.g., cover crops, no-till practices), etc.. These practices have been proposed as potential strategies for preventing or reducing pollutant loads in water bodies, and for mitigation of flood events through runoff control and peak flow reduction (Ice, 2004; Arabi et al., 2007; Artita et al., 2008; Kelly and Merritt 2010). Achieving an optimal selection and spatial allocation of BMPs that is also acceptable to the stakeholder community is an inherently complex process. Multiple researchers (Arabi et al., 2006; Kelly and Merrit, 2010; Lethbridge et al., 2010; Tilak et al., 2011; Kaini et al., 2012) have investigated coupled simulation models and optimization algorithms to identify the optimal distribution of BMPs in a watershed. In these studies, simulation models were used to estimate the landscape responses and evaluate related design objectives, whereas the optimization algorithms were used to search through large decision spaces for design alternatives that enhance one or more objectives in the watershed. While these analytical techniques have the benefits of generating optimized scenarios of design alternatives with respect to quantifiable goals (many of which maybe developed via stakeholder engagement), they are

limited in their ability to incorporate diverse subjective, unquantifiable or unrepresented criteria (such as personal or social values, beliefs, interests, biases, local knowledge, etc.) and preferences of stakeholders. For example, a stakeholder's preference might be constrained by her/his lifestyle and family situation, which might be difficult to learn about and/or quantify in the problem formulation. Additionally, Clearfield and Osgood (1986) reported that many times stakeholders may be more driven by subjective personal and social constraints than by economic considerations, when they are in the process of considering specific conservation practices for their land. Hence, unique stakeholder circumstances should be taken into consideration (Greiner et al., 2009) when developing plans for implementing practices on land. Current inability to incorporate these stakeholder-specific subjective issues and circumstances in the design process has likely contributed to the unsuccessful adoption of "optimal" design alternatives, and because of this the use of optimization algorithms in watershed planning has been criticized by some (Mendoza and Martins, 2006).

### **2.3 Hydroinformatics approaches to support participation of stakeholders during design**

In 1977, Heidegger reported that individuals in a community will perceive a resource as being vital to their lives if the consequences of poor management of resources are revealed to them, and/or if they are directly affected by these consequences. Hence, when communities are engaged in watershed planning activities, education of the various costs and benefits of decisions, and the consequences of a decision on the ecosystem services provided by the watershed become necessary. Ecosystem services are based on four types of functions served by watershed resources (de Groot et al., 2002):

- 1) Functions that regulate the biogeochemical cycles (*regulation functions*),
- 2) Functions that provide refuge and reproduction habitat to native plants and animals (*habitat functions*),
- 3) Functions that build a large variety of carbohydrate structures to provide goods ranging from food to raw materials and energy resources (*production functions*),  
and

- 4) Functions that provide spiritual enrichment, cognitive development, and recreational and aesthetic experiences necessary to maintain human health (*information functions*).

Quantifying and representing these types of multiple functions as co-benefits during the design process can require integration of data and knowledge from multiple fields. Abbot (1991) proposed the development of a new field of *hydroinformatics* as a field that could harness advances in the information and communication technologies (ICTs) to deliver complex information related to water to communities, and thereby, enable an increased use of and integration of heterogeneous types of data to educate and instruct communities about watershed resources, environmental processes, and ecosystem services. According to the Institute for Water Education from UNESCO-IHE (<https://www.unesco-ihe.org/msc-programmes/specialization/hydroinformatics-modelling-and-information-systems-water-management-2>), *hydroinformatics* is a field that employs a wide range of simulation modeling techniques, software tools, and information technologies to solve complex problems related to hydraulics, hydrology, and environmental engineering, for the better management of water. One of the unique strengths of *Hydroinformatics* approaches is their ability to fuse different types of data, information, and knowledge via mechanistic as well as heuristic methods, in order to appropriately cope with the complexity of natural and socio-economical systems in a watershed (Abbot, 1999; Evers et al., 2012; Delipetrev et al., 2013).

In an attempt to find more effective ways to include stakeholders in the design process and potentially increase adoption of BMPs, researchers (Rabotyagov et al., 2010; Evers et al., 2015; Smit et al., 2015) have explored different hydroinformatics methods. Many of these methods also employ current information technologies (such as social networking technologies) to develop platforms via which the affected stakeholders can express their unique socio-economic and subjective constraints during the design process (Feather and Amacher, 1994; Rahelizatovo and Gillespie, 2004; Prokopy et al., 2008). For example, analytical approaches such as Systems Dynamics models (Metcalf et al., 2010), Bayesian Networks (Castelletti and Soncini-Sessa, 2007; Zorrilla et al., 2010), Fuzzy sets and cognitive maps (de Kok et al., 2000), Agent-based models (Barreteau and Abrami, 2007),

and Human-guided/Interactive Optimization (Babbar-Sebens and Minsker, 2011) have all been used to create participatory decision support platforms aimed at improving the engagement of stakeholders. These semi-structured and interactive DSS elicit data, knowledge, and feedback from stakeholders, and then use that information to generate design alternatives.

Recent reviews by Jakeman et al. (2008) and McIntosh et al. (2011) expose relevant user interaction issues that should be considered when developing Environmental DSS (EDSS) to engage stakeholders. McIntosh et al. (2011) also emphasized the importance of the evaluation of human factors such as needs, goals, fatigue, learning, etc. in the design of interfaces used by EDSS. Similarly, a growing number of research studies in water management have indicated interest in improving EDSS usability to enhance the users' understanding of the underlying problem and of the design alternatives proposed for solving the problem (Johnson, 1986; Power and Sharda, 2007; Kirchoff et al., 2013; Babbar-Sebens et al., 2015). For example, Mysiak et al. (2005) emphasized the importance of end users feedback in the development of a decision support tool, while McIntosh et al. (2011) described the need for establishing a process that measures the evaluation and usability of this systems, in order to reach standards, create parameters, and identify common user-interaction issues that affect most of the decision support systems. Many authors in the computational sciences, industrial engineering, and web development field have also suggested that testing the usability of a product is crucial to determine its adoption among the community. It has been demonstrated by several studies in the biomedical field (Kushniruk et al., 1996; Patel et al., 2000) that the quality of the data provided by users will depend on the reasoning of the individuals motivated by their experience with the tool, and the information collected while interacting with the tool. Therefore, the usability of a tool or product is critical when the goals of a tool includes delivery of precise and clear information to its users via effective visualization of data in the DSS's GUI.

#### **2.4. Interaction of humans with search and optimization algorithms**

Heuristic search and optimization algorithms, such as Genetic Algorithms, Simulated Annealing, Ant Colony Optimization, and Strength Pareto evolutionary algorithm, have

been widely applied in the water resources planning and management research to solve problems that have system uncertainties, multiple criteria and constraints, ill-defined search spaces, and are NP-complete<sup>1</sup> in complexity. Applications of these methods include reservoir operations (Wurbs, 1993; Haddad et al., 2006; Mathur and Nikam, 2009; Adeyemo, 2011), groundwater modeling and management (McKinney and Lin, 1994; Sidiropoulos and Tolikas, 2008; Siegfried et al, 2009; Babbar-Sebens and Minsker, 2011), surface water management (Savic and Walters, 1997; Bekele and Nicklow, 2005; Nicklow et al., 2010; Babbar-Sebens et al. 2011), water distribution (Lansey and Mays, 1989; Cunha and Sousa, 1999; Eusuff and Lansey, 2003; Maier et al., 2003) and pipeline optimization (Golberg and Kuo, 1987; Simpson and Goldberg, 1994; Geem et al., 2001).

For successful design of watershed plans, heuristic search and optimization algorithms must employ decision makers (or stakeholders) as key agents that affect the search process, as also discussed earlier in Section 2.2. One way to include stakeholder participation within heuristic search is via advancements in the communication and participatory media, such as the Social Web that has supported social interactions among people via the internet. These types of participatory media also encourage stakeholder learning, and providing a more informed environment for humans to understand the consequences of implementing a solution (Hare et al., 2001; Clark and Aufderheide, 2009; Vervoort et al., 2010). However, stakeholders are also one of the major sources of uncertainty, constraints, and criteria during decision making, because of the existent complexities and uncertainties embedded in the human behavior (Osman, 2010); hence investigation of the effect of human behavior on heuristic search is necessary. For example, there is still a need to understand, compare, and analyze how the uncertainty in human agents (stakeholders) and their interactions with the interface or algorithm per se affect the performance of search and optimization algorithms. Additionally, human behavior will affect the search process in different ways depending upon how participation of stakeholders is incorporated directly

---

<sup>1</sup> In computational complexity, the nondeterministic polynomial time (NP) is a fundamental complexity class where a problem's solution can be verified quickly in polynomial (P) time, but there is no efficient way to determine the solution. These kinds of problems cannot be solved through a fast algorithm (Eppstein, 1996), hence heuristic and approximation algorithms are often used to solve such problems.

or indirectly in the heuristic optimization process. Some of the direct and indirect approaches investigated by researchers include:

- Wang et al. (2001) explored a reference point method to support multi-objective decision making through interactive optimization algorithms. Theoretically, this method provides a changing reference point to satisfy decision maker (DM) preferences with just a single interaction. The authors did not provide experiments that involved real users (or DMs), but claimed that the technique guarantees the feasibility of the desired reference point. Their illustration example is limited by a two variable minimization problem that does not reflect the complexity that real user interaction scenarios may bring to the system. Also, they examined the reference point in the objective space, while many humans may be more interested in the values given in the decision space.
- Avigad and Moshaiov (2009) proposed an Interactive Concept-based, Multi-Objective Evolutionary Algorithm (IC-MOEA) that provides extensive search in the decision space giving users the opportunity to build a set of solutions with similar conceptual criteria. The conceptual criteria are identified in the early stage, where an intensive use of the human creativity is required. Their approach seeks to include the users preferences based on a weighted vector that interacts (in a ranking process system) with the set of solutions found by traditional non-interactive optimization process. Their application studied the consequences of different scenarios in a more realistic application environment. However, their study also lacked results with real DMs for evaluation of the search optimization process, and instead used simulated users to introduce a human-based function.
- Maringanti et al. (2009) tried to incorporate some aspects of landowner preferences in their optimization approach using a combination of NSGA-II and the simulation model SWAT, to optimize several best management practices in a watershed. The preferences were inferred using records from stakeholders and county agents of most current and commonly adopted BMPs, in their watershed sites. Those BMPs (i.e., nutrient management, buffers, conservation till and no-till) were then used within the optimization approach. Their approach did not investigate the potential effects of underlying spatial distributions of attitudes and preferences of stakeholders in the



selection of these BMPs. Due to the complexity associated with the preferences and behaviors of spatially-explicit individuals and communities, development most watershed plans are based on averaged data on farmer preferences (similar to the approach used by Maringanti et al., 2009). Commonly, the data is obtained from surveys, workshops or interviews conducted at a particular point in time in the watershed. At the time of implementation, many farmers may not perceive conservation practices in the same way as they did in the survey. Additionally, they may have emerging attitudes, biases, and preferences that could be associated with their own cognitive learning affected by new information and new experiences. This can lead to deterioration in the performance of original optimized watershed plan/design when stakeholders/landowners end up transforming the design into a sub-optimal modified design to suit their constraints, defeating the original purpose and effort of spatially optimizing conservation practices in a large watershed (Piemonti et al., 2013; Babbar-Sebens and Minsker, 2008).

- Babbar-Sebens and Minsker (2011) developed an Interactive Genetic Algorithm with Mixed Initiative Interaction (IGAMII) algorithm that aids with human-centered search/optimization of alternatives. Its main goal is to assist users in the identification of designs that will satisfy user's preferences, and at the same time enabling a learning process in participating humans. The algorithm was tested on a groundwater monitoring problem, where the optimization problem was based on two conflicting objectives: 1) minimize the number of active monitoring wells, and 2) minimize the error between measured benzene concentrations and the concentrations estimated from active monitoring wells via interpolation models. The researchers explored the use of IGAMII in the selection of sets of active monitoring wells, which based on the judgement of human participants (novice and experts) generated the most representative plume map. The researchers were also able to determine temporal changes in mean confidence ratings (i.e., a participant's level of self-confidence in her/his own evaluation of design alternatives), as well as the differences between design alternatives generated from non-interactive and interactive optimization. Their results support the theory of learning process (even for expert users), and suggest the effectiveness of human-algorithm

collaboration during the optimization algorithm's search process in generating viable solutions that also satisfy user preference.

A new system that uses interactive optimization techniques have been developed in order to accommodate for the new advances in participatory media that relies in the Web 2.0 framework. The next section explains this novel participatory system and how it allows the human-centered design.

## **2.5 WRESTORE DSS for interactive optimization**

Currently there are a wide variety of watershed modeling programs (such as SWAT, HEC-RAS, WMOST, etc.) and decision support systems (such as NSDSS, MODOLUS, CLAM, WEAP, etc.) that have been developed to support structured and unstructured watershed planning and management problems. These decision support systems use embedded simulation model to evaluate the multiple performance-based objective functions (Cox, 1996; Mysiak et al., 2005; Jakeman et al., 2008; Elmahdi and McFarlane, 2012), which may be relevant to the water management problem. Additionally, since numerous decision variables may play an important role in the modeling and decision making process, the size of the decision space can explode and overwhelm decision makers. Hence, DSS researchers have resolved this issue by incorporating automated search/optimization algorithms that provide either a single optimal solution, or a set of non-dominated optimal solutions (Pareto fronts) that lie on a tradeoff curve of objective functions (Arabi et al., 2006; Artita et al., 2008; Maxted et al., 2009; Kaini et al., 2012).

In this research, the WRESTORE decision support system has been used for human-centered optimization of BMPs in a watershed. Readers are recommended to review Babbar-Sebens et al (2015) for detail description of this tool. WRESTORE (or Watershed REstoration using Spatio Temporal Optimization of Resources – Figure 2.1) is a web-based environmental decision support system (DSS) specifically developed to include the participation of individuals and communities in the design process and selection of optimal allocation of BMPs in agricultural watersheds in order to restore natural processes within agricultural watersheds. The web-based architecture of WRESTORE allows the DSS to be accessible for a large community of stakeholders, and facilitate the participatory watershed planning concept.

WRESTORE's underlying search algorithm is a modification of the IGAMII algorithm (Interactive Genetic Algorithm with Mixed Initiative Interaction, Babbar-Sebens and Minsker, 2011) that aids with human-centered search/optimization of alternatives. Its main goal is to generate optimal design alternatives that satisfy not only the physical objectives and constraints, but also individual stakeholders and community preferences. Within the system the users can 1) simulate single or multiple types of BMPs in a watershed plan alternative, 2) gathering information on the performance of the design alternative at a local or regional scale, 3) compare the performance of different design alternatives that differ in the spatial allocation of the conservation practices, 4) select their preferred designs based on their personal criteria, and 5) propose those designs to the community and agencies that provide assistance in implementation of BMPs and watershed plans.

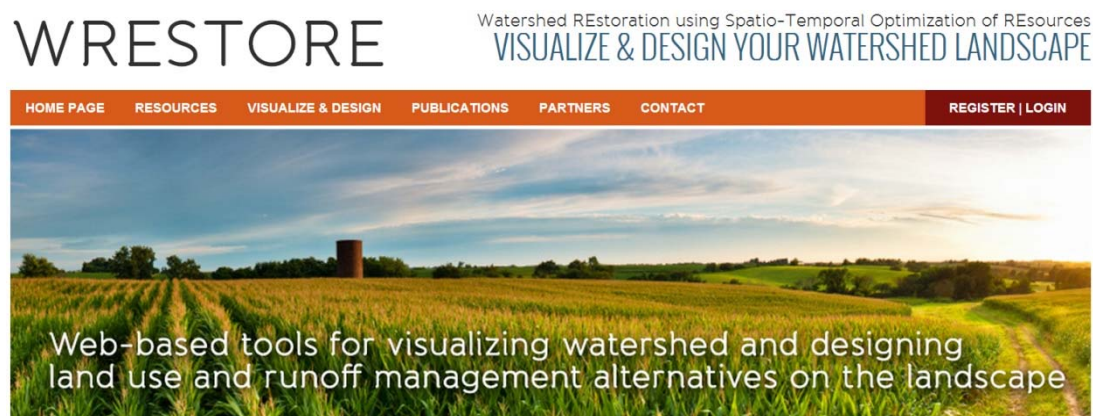
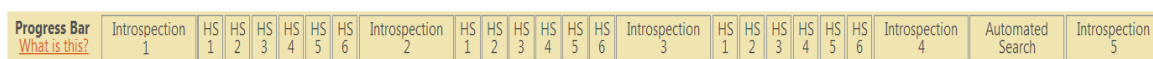


Figure 2.1. WRESTORE website home page

The IGAMII algorithm is a human-guided (or, *human-centered*) optimization algorithm that engages with users in an iterative manner via visualization interfaces. In every iteration, which is called an *interaction session*, both the decision space of the alternatives (via maps) and the objective space of the alternatives (via graphs) are displayed to the user. Once the user has evaluated the alternatives, he/she can provide his/her feedback on the quality of the alternative to the IGAMII's underlying optimization algorithm via a *user rating* determined on a Likert-type scale (e.g. "I like it", "Neutral", "I do not like it"). The IGAMII's optimization algorithm uses the *user rating* as an additional objective function (in addition to four quantitative objective functions) to identify the user's preferences. The underlying optimization algorithm is critical to enabling search of new alternatives, and

though the IGAMII uses a multi-objective Genetic Algorithm called NSGA-II (Deb et al., 2002), WRESTORE is not restricted by the type of multi-objective optimization technique and has the capabilities to select from a variety of other search approaches (e.g., Decentralized Pursuit Learning Automata (Singh, 2013)).

*Interaction sessions* in WRESTORE can be of three types (Figure 2.2 shows the sequence of sessions): *introspection* sessions, *human-guided search* (HS) sessions, and *automated* search sessions. An *introspection session* is used for improving the learning efficiency of the human user by enabling the user to re-examine previously viewed and rated alternatives that are stored in a *case-based memory* (Craw, 2003; Shi and Zhang, 2005), and re-assess his/her own thoughts, reasoning process, emotions, biases, consciousness, and *user ratings* of these previously assessed alternatives.



**Figure 2.2. Interaction sessions in IGAMII (Babbar-Senbens et al., 2015)**

For example, Figure 2.2. illustrates an IGAMII experiment in which five introspection sessions occurred at different times during the progress of the experiment. Each of the *human-guided search* (HS) sessions is an iteration of the underlying optimization technique (or, *generation* in the case when a Genetic Algorithm is used as the search method in IGAMII) where new alternatives created by the underlying optimization operators are shown to the user. In IGAMII, when a *human-guided search* is conducted, a small population micro-genetic algorithm is used. Hence, the number of alternatives shown in a typical HS session is typically equal to the population size of this micro-genetic algorithm. Every alternative (or, the genetic algorithm chromosome) is evaluated in its performance using a suite of mathematical objective functions and process simulation models (e.g., the SWAT model of a watershed); and then the values of these performance-based objective functions are displayed to the user, in addition to the alternative decision variables using maps and graphs. The user provides the feedback via the Likert scale-based *user rating* and then this *user rating* is used by the micro-genetic algorithm operators to create the next generation of new alternatives (or, new chromosomes in the case of Genetic Algorithm).

Hence, HS sessions are always presented successively and are equal to the number of generation of the micro-genetic algorithm.

The use of such web-systems provides a safe environment to the stakeholder allowing a true introspection of their needs, concerns, and criteria (Yigitcanlar and Okabe, 2002; Kelly et al., 2012). In the same context, this tool enables researchers to understand, how stakeholders are exploring and interacting with such DSS, allowing the observation of preferences' changes under a dynamic behavior. This is particularly relevant for agricultural stakeholders, whose mental maps, perceptions, behaviors and attitudes influence in their understanding of their environment and the interactions between systems, affecting the intrinsic motivation to adapt or accept changes. McCown (2002) emphasizes that the implementation of DSS need to face a change where the emphasis relies more on the 'learning' of what DMs (in this case stakeholders) are learning, without diminishing the design aspect. For the DMs accomplish this education there is a need for iterative learning and practice change process. These two features are embedded in WRESTORE through an iterative search process that supports not just the user's growing knowledge, but also give researchers a window on how to explore the learning and selection process.

The *automated search* session (as seen in Figure 2.2. between *introspection sessions* 4 and 5) is the third type of session, which is a more computationally intensive optimization run and is performed by replacing the human user with a heuristic model of *user ratings* (or, a *simulated decision maker model*). The main purpose of automated search is to minimize user fatigue by replacing the human user with the simulated user, and hence no visual interfaces are shown to the user when automated search is running. Data on *user ratings* collected in earlier *introspection* and HS sessions are generally used to create the personalized and heuristic *simulated decision maker* models for every user. For example, Babbar-Sebens and Minsker (2011) used fuzzy logic models that related design parameters to *user ratings*, whereas in WRESTORE we have included multiple linear and non-linear classification models, neural networks, fuzzy logic models, and deep learning models (Singh, 2013) to create *simulated decision maker* models.

In IGAMII, the sequence of *interaction sessions* ( similarly to Figure 2.2.) is decided via a flexible *mixed initiative interaction* (Hearst, 1999) (managed in the system by the

mixed initiative manager -MIM-) strategy that monitors the individual user learning and *simulated decision maker model's* accuracy to identify when *human-guided search* should be conducted and when *automated search* should be conducted. Monitoring and tracking user learning is an active topic of research in Human-Computer Interaction and Cognitive Psychology. While additional research investigations will enable advanced tracking techniques to inform the mixed initiative interaction strategies, WRESTORE currently uses the technique proposed by Babbar-Sebens and Minsker (2011). This technique monitors the trends in users' self-reported confidence in their *user ratings* to identify how fast human users are learning by interacting with the tool. In this manner, it is possible to use the human user and the simulated user models for search/optimization when they are most suitable for evaluation of alternatives. After every optimization run, irrespective of whether it is *human-guided search* or *automated search*, an *introspection* session is invoked to facilitate a user's re-reflection of previously generated alternative and improve his/her own cognitive learning.

In the first *introspection session*, the MIM will access the case-based memory to select potential design alternatives found earlier in a different search or by an offline optimization run that did not involve any *user ratings* (e.g. a preliminary non-interactive optimization run proposed by Babbar-Sebens and Minsker 2011). The MIM then calls the IM (Individual Manager), which sends these alternatives to the web server to show the alternatives to the user by means of a web-based interface (Figure 2.3). This same interface is also currently used for all *human-guided search* sessions and is being further improved for better engagement with users. The User Program will then trigger the EmailM (Email Manager) to send an email to the user whenever a session is available for viewing on the web server.

After the user logs into the website, he/she is able to visualize and compare the previously evaluated alternatives, which have now been made available to him/her for viewing in the first introspection session. The user evaluates all the alternatives shown by the interface (Figure 2.3.) based on his/her assessment of how BMPs are sited and sized in the entire watershed or in their local sub-basins of interest (viewed in the map space). The bar graphs allow the user to evaluate the performance of the design alternative with respect to quantitative goals in the entire watershed or in their local subbasins of interest (SBint).

The user provides feedback on his/her assessment of the quality of the alternative via *user ratings*, and this data along with other usability data (e.g., confidence ratings, mouse clicks, clock time spent on interface features, etc.), are collected and sent back from the web server to the database for archiving and use by WRESTORE's software managers.

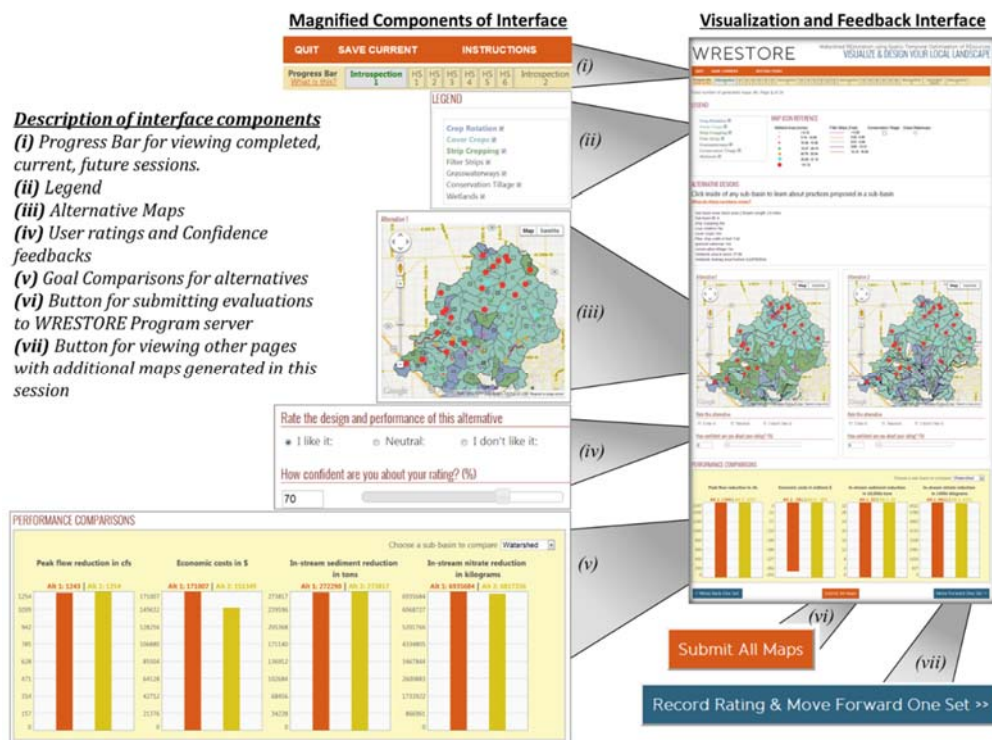


Figure 2.3. Visualization and Feedback Interface in WRESTORE (Babbar-Sebens et al., 2015)

Implementing WRESTORE in a watershed involves three phases: pre-processing, real-time participatory design experiments, and post-processing. Currently, WRESTORE has been implemented, and tested for user learning, and multi-users engagement issues, and overall tool improvements at the test site of Eagle Creek Watershed, Indiana. But the flexible architecture of WRESTORE allows other watershed groups, in the future, to include their own simulation models, design parameters, and data related to their region. Figure 2.4. provides a synopsis of the three phases.

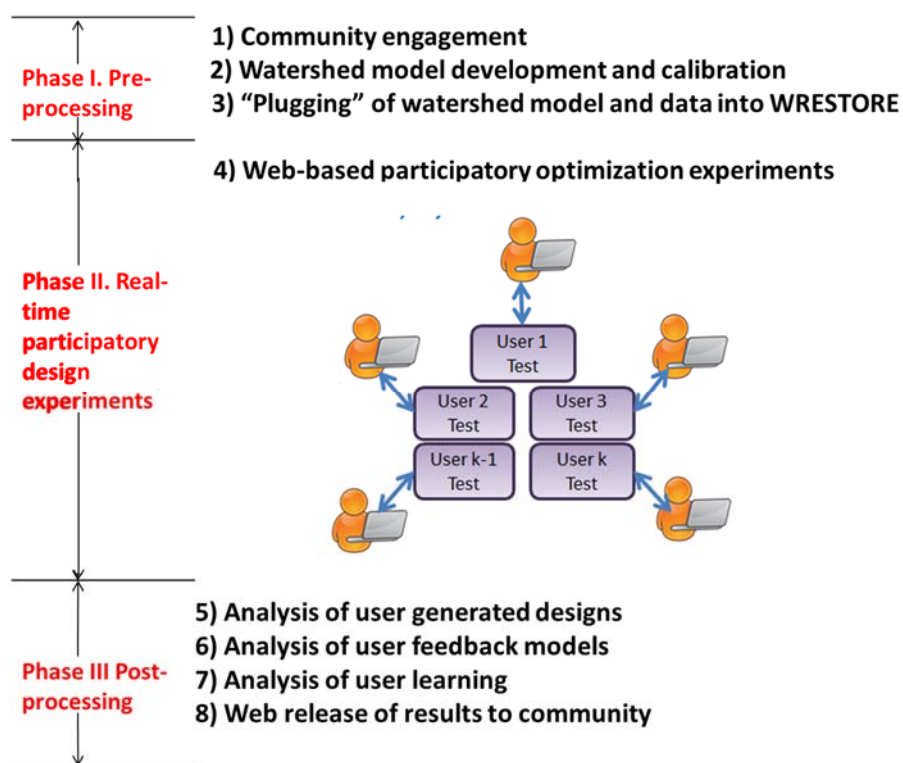


Figure 2.4. Development of the WRESTORE tool for an asynchronous experiment  
Modified from Babbar-Sebens et al., 2015.

Phase I. Pre-processing phase: In this phase, a watershed community's agency personnel or stakeholder council group/alliance is expected to first engage with the various parties of interest to identify conservation practices of interest and specific areas where potential sites for these practices could exist. The watershed community is expected to then develop an appropriate simulation model of their study area, preferably via participatory modeling approaches (e.g. Palmer, 1996). We have currently used the SWAT model to simulate effectiveness of new conservation practices in the test site, but WRESTORE's software architecture is not constrained by a specific hydrology or water quality model. Once a community-supported process simulation model to evaluate effect of conservation practices on the watershed has been developed and calibrated, the watershed group leaders can then submit the model files to the WRESTORE administrative team for setting up a WRESTORE project for their watershed. Copies of the folders of the simulation model input/output/executable files are saved on the WRESTORE program server, from where the program makes copies and saves in different nodes to allow the simulation whenever is required. Besides the simulation models, various GIS files identifying the watershed



boundaries, subbasins (SB), and stream network are also required for the interface. This GIS data are stored into Google Fusion Tables so that Google Maps API can be used in the interface. We are, currently, in the process of developing a separate interface that will enable watershed group leaders to automate this setup process of site data and models for any watershed via the web, in the future.

Phase II. Real-time participatory design experiments: Once the WRESTORE project for the application watershed has been setup, it is then available for release to the general community and “crowdsource” feedbacks in the design of alternatives. There are multiple synchronous and asynchronous approaches (Babbar-Sebens et al., 2015) via which watershed groups could engage their stakeholders in conducting web-based, multi-user participatory optimization experiments in WRESTORE. Here, we present the asynchronous approach that was used in this research.

- i. *Asynchronous multi-user experiments:* In this type of experiment, every user can initiate his/her own human-computer collaborative search for exploring spatial implementation of conservation practices that are of interest to her/him. Hence, multiple instances of User Program will be generated in this experiment type. When a user logs in and begins the WRESTORE workflow (discussed earlier in Section 2.4), he/she can choose from a set of available BMPs and goals for his/her watershed site. Multiple users can begin their experiments independent of others, and hence can asynchronously explore the effect of different types and combinations of conservation practices in the watershed. Since these experiments are conducted asynchronously (in a parallel fashion), WRESTORE currently does not assume preferred SBs of interest in advance, and, therefore, presumes that BMPs chosen (as shown in the maps for Figure 2.3) are applicable to all SBs in the watershed specified the potentially places for chosen BMPs. Additionally, because of this assumption WRESTORE uses the values of the quantitative goals at the watershed scale (as shown in the bar graph for Figure 2.3) as the objective functions for the underlying optimization algorithm. The future interface of WRESTORE will enable more detail settings for individual users, where user will be able to declare a narrower sub-region of interest. The user-feedback-driven search and the

learning process in the WRESTORE's underlying algorithms are, however, customized to individual participating users. One advantage of this kind of asynchronous engagement with multiple users is that it provides users the flexibility to explore alternatives at a time that suits them the most, without being dependent on the feedback of others.

**Phase II. Post-processing:** Once user experiments are finished, alternatives generated by the multiple users can then be post-processed for similarities and dissimilarities in spatial plans of practices (i.e. alternatives) liked or disliked by the users. Additionally, simulated decision maker models generated by the WRESTORE program can be processed for identifying underlying parameters and variables that best explain the *user ratings*. Data collected via the interface on users' transient confidence ratings, numbers and types of clicks, time spent in sessions, etc. can also be post-processed to understand how each participant engaged with the interface and whether any detectable learning or changes in opinions were observed. Once this post-processing is completed, the analyses can be released to the user community for decision making and for identifying how individual user's behavioral factors affected identification of promising alternatives.

#### **2.4 Interactions between Participants and the WRESTORE-Tool**

Twenty common selected designs were shown, in sets of two at the time, providing information regarding the spatial distribution of the BMPs from a pool of non-dominated design alternatives found by a multi-objective genetic algorithm that used only the four physical objective functions (Babbar-Sebens et al., 2015). In this session, all participants saw the same twenty designs. After *I1*, participants transitioned into human-guided search sessions (i.e., six *HS* sessions - *HS1* to *HS6*) that had a one-to-one correspondence with the six iterations (or *generations*) of a micro-Interactive Genetic Algorithm (micro-IGA; Babbar-Sebens and Minsker, 2011; Babbar-Sebens et al., 2015). In each of these *HS* sessions twenty new designs (*population* of micro-IGA), generated by the micro-IGA, were presented to the user. After the micro-IGA was over (i.e. at the end of *HS6*), designs that had high *user ratings* and/or were on the best non-dominated front were saved into a Case-Based Memory (CBM). Next, the participant revisited twenty random designs from the CBM in *Introspection* session 2 (*I2*), where he/she was provided the chance to modify

his/her previous *user ratings* and *confidence levels* based on any newly formed preferences and perceptions. This revisiting technique aimed to improve the participants' skills in evaluating the performance of the design alternatives using the information gathered in the *HS* sessions. If the user is modified his/her evaluation of the revisited design in the *Introspection* session, the changes were updated in the CBM. At the end of *I2*, a sub-set of designs from the CBM were again inserted in the initial *population* of the next micro-IGA, and a new set of six *HS* sessions ensued. It is worth noting here that the WRESTORE interactive search tool was originally developed for an additional set of alternating *I* and *HS* sessions and an *Automated Search* session, all occurring after *I3*. The *Automated Search* session uses a Machine Learning model to learn from the participant's *user ratings* collected in the earlier *HS* sessions and then conducts an exhaustive search on behalf of the participant by using the Machine Learning model as a simulated user. Interested readers are encouraged to refer to Babbar-Sebens et al. (2015) for details on these sessions. In order to limit human fatigue and keep the workload the same for all users, researchers requested that participants complete all alternating *HS* and *I* sessions, at least until the end of *I3*. Several participants in the *surrogates* group, however, did choose to complete sessions beyond *I3*. Nevertheless, these additional sessions were not considered as part of the first analyzes conducted to examine the hypotheses above.

After the participant had submitted his/her *user ratings* for design alternatives, ten to fifteen minutes of waiting time was needed for the underlying optimization algorithm and hydrologic model to generate and evaluate new design alternatives. To allow participants to complete the experiment at their own pace, an email notification system was used to send them an automated reminder whenever the next session was available for user interaction. The automated e-mail included a log-in link to the WRESTORE web page, and when users clicked on the link to log into their account, a set of instructions on how to conduct the ongoing experiment was redisplayed on the screen (Babbar-Sebens et al., 2015).

When users progress through the workflow of the IGAMII algorithm in WRESTORE, they undergo two types of sessions: *Introspection (I)* sessions and *Human-guided Search (HS)* sessions. While the human-guided search involves the use of a micro-interactive

Genetic Algorithm to create new design alternatives, the introspection enables participants to reflect on and reassess their satisfaction with previously found design alternatives. In this study, each participant was requested to progress through at least fifteen sessions (3I and 12 HS). Each session displayed twenty design alternatives to the users (in sets of two), who were then expected to evaluate the suitability of the designs (based on individual considerations of decision variables, goals, and their local constraints and knowledge) via a Likert-type scale. The scale included three *user rating* classes - “I like it”, “Neutral” or “I do not like it”. *Introspection session 1 (I1)* was the starting point in the WRESTORE workflow, for all participants. In this session, twenty designs, randomly selected from a pool of 219 Non-Dominated design alternatives were displayed to the users. The pool of non-dominated design alternatives had been found by a prior simulation-optimization experiment that used a multi-objective genetic algorithm along with the four physical objective functions (i.e., Costs, Peak Flow Reduction, Nitrates Reduction, and Sediment Reduction) (Babbar-Sebens et al., 2015). To maintain same initial starting conditions for the participants, this same set of randomly selected design alternatives was used for all the user experiments. After I1, participants transitioned into six human-guided search sessions (i.e., six HS sessions - HS1 to HS6), in which each session had one-to-one correspondence with the iterations (or *generations*) of the micro-Interactive Genetic Algorithm (micro-IGA; Babbar-Sebens and Minsker, 2011, Babbar-Sebens et al., 2015) within the IGAMII. In each of these HS sessions, twenty new designs (*population* of micro-IGA), generated by the micro-IGA, were presented to the user for obtaining *user ratings*. After the micro-IGA was over (i.e. at the end of HS6), designs that had “I like it” *user ratings* and/or were on the best non-dominated front were saved into a Case-Based Memory (CBM). Next, the participant’s workflow transitioned into the *Introspection session 2 (I2)* where they re-examined twenty random designs from the CBM. The introspection session provided the participant a chance to modify his/her previous *user ratings* and *confidence levels* based on any newly formed preferences and perceptions at the end of recently completed HS sessions. If the participant is modified his/her evaluation of the re-examined design in the *Introspection session*, the changes were updated in the CBM. At the end of I2, a sub-set of designs from the CBM were again inserted in the initial *population* of the next micro-IGA,

and a new set of six *HS* sessions ensued. It is worth noting here that the WRESTORE interactive search tool was originally developed for an additional set of alternating *I* and *HS* sessions and an *Automated Search* session, all occurring after *I3*.

After the participant had submitted his/her *user ratings* for design alternatives, ten to fifteen minutes of waiting time was needed for the underlying optimization algorithm and hydrologic model to generate and evaluate new design alternatives. An email notification system was used to send participants an automated reminder whenever the next session was available for user interaction. The automated e-mail included a log-in link to the WRESTORE web page, and when users clicked this link to log into their account, a set of instructions on how to conduct the ongoing experiment was redisplayed on the screen (Babbar-Sebens et al. 2015). The email notification allowed participants to complete the experiment at their own pace.

## **CHAPTER 3. Case Study**

Chapter 3 was divided into four different sections where each of the components to be studied is explained. Section 3.1 briefly describe the decision support system (DSS) used in this investigation (as it was explained in detailed in Chapter 2, section 2.6). Section 3.2 describe the location and site use as experimental watershed to test the capabilities of the DSS and where a select group of stakeholders and decision makers was willing to address comments and run experiments using the developed tool. This section also describes the details of the construction and calibration of the selected hydrologic model (Soil and Water Assessment Tool -SWAT) use as a baseline by WRESTORE, as well as how the conservation practices were simulated by the hydrologic model, i.e., how the decision variables are represented in the decision space. Finally, section 3.3 introduces the participants and their association with different simulated models for this research.

### **3.1 Decision Support System**

As described in Chapter 2, WRESTORE seeks to optimize the spatial distribution of conservation practices (or Best Management Practices-BMPs) in an agricultural watershed. The system was developed in a web-based environment to allow easy access and outreach larger stakeholder communities that wished to be included in the participatory planning efforts of their watershed.

WRESTORE uses an innovative technique as the underlying multiobjective search algorithm, centered in adding the user's subjective criteria in the search optimization process through an additional objective function. The search algorithm is a modification of the Interactive Genetic Algorithm with Initiative Interaction (IGAMII), developed by Babbar-Sebens and Minsker (2011). IGAMII seeks to find optimal solutions for a design problem, using a collaborative strategy between human-computer interactions. The advantage of this strategy relies on teaching the search algorithm about the desirable objective space and decision space, developing solutions that have a better agreement with the human (users) constraints (specifically subjective and qualitative criteria, such as preferences in design, priorities of physical objectives function, and/or aesthetics).

WRESTORE uses an iterative process that requires providing a classification for each of the shown designs. To keep a control of the variables, two conservation practices widely accepted in the watershed study (cover crops and filter strips) and four physical goals (maximize peak flow reduction, minimize economic cost, maximize sediment reduction, and maximize nitrates reduction) were asked to be chosen by each of the participants.

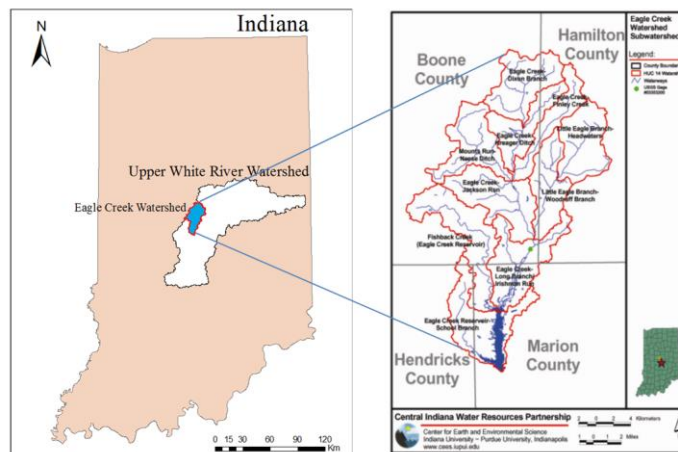
The Graphical User Interface (GUI) had seven different components (as shown in Figure 2.3 in chapter 2) that might be used in the experiment execution: 1) a progress bar that inform the participant the stage of the experiment, 2) a set of two maps with the spatial allocation of the conservation practices, 3) a legend indicating the meaning of symbols and colors on the map, 4) a set of two bar graphs that indicate the performance of the system (at watershed and local levels), 5) a rating section, where the agreement with the design alternative is reflected and the confidence level on this rate is provided, 6) two back and forward buttons to facilitate navigation across all the designs in a session, and 7) the submit all button that will allow the search algorithm to gather the information from the user.

### **3.2 Experimental Watershed**

Eagle Creek Watershed (ECW) is an HUC-11 watershed (05120201120) located in central Indiana across four different counties: Marion, Hamilton, Hendricks and Boone (Figure 3.1), about 16 km northwest of Indianapolis, Indiana. Its drainage area is approximately 419.26 Km<sup>2</sup> into the Eagle Creek Reservoir (ECR), one of the major recreational and water drinking supplies for Indianapolis. The reservoir was developed as a flood control method of the seasonal inundated northwest area of Indianapolis and Speedway. This reservoir has been impaired mainly by sediments, pesticides, herbicides and fertilizers from the agricultural land in the upstream areas (Tedesco et al., 2005) transported in the streams. The watershed topography is relatively flat to undulating, with some dissection near Eagle Creek reservoir.

Agriculture is the dominant land-use in the upstream area of the watershed (approximately 60%), with predominantly corn and soy-bean based crops (*Census of Agriculture, 2007*). Urban development has occurred mostly in the southeast region of the watershed, due to growth in the Indianapolis population (U.S. Department of Agriculture's

(USDA) crop data layers database). According to USA.com reports, Indianapolis growth population rate since 2000 is 4.93 (<http://www.usa.com/indianapolis-in.htm>)



**Figure 3. 1 Location of Eagle Creek Watershed. Taken from Tedesco et. al, 2005**

Dominant soils association in the area consists of the Crosby-Treaty-Miami association in the headwaters (USDA's Soil Survey Geographic Data Base). These soils are generally deep, poorly drained, and nearly level to gently sloping soils formed in a thin silty layer overlying glacial till. Whereas downstream areas are dominated by Miami-Crosby-Treaty association, generally deep well drained to somewhat poorly drained, and nearly level to moderately steep soils formed in a thin silty layer and the underlying glacial till. The Eagle Creek valley has a minor soils association that consists of Sawmill-Lawson-Genesee. In the northwestern boundary, two minor associations exist Fincastle-Brookston-Miamian association and Mahalassville-Starks-Camden association. The minor soils also vary in their drainage characteristics based on the composition.

The climate in this area is predominantly temperate continental and humid (Clark, 1980; Newman, 1997), with an average annual temperature of approximately 11° C. The average annual precipitation varies from 97 to 102 cm, with late spring being the wettest seasonal period and February being the driest. Most of this average annual precipitation occurs during the 5-6 months of the frost-free growing season.

All four counties have farmland owners with similar race, age, principal operator's gender and principal crop. The agriculture community population consists mainly of Caucasian males in their mid-fifties. The community mostly produces corn and soybeans row crops

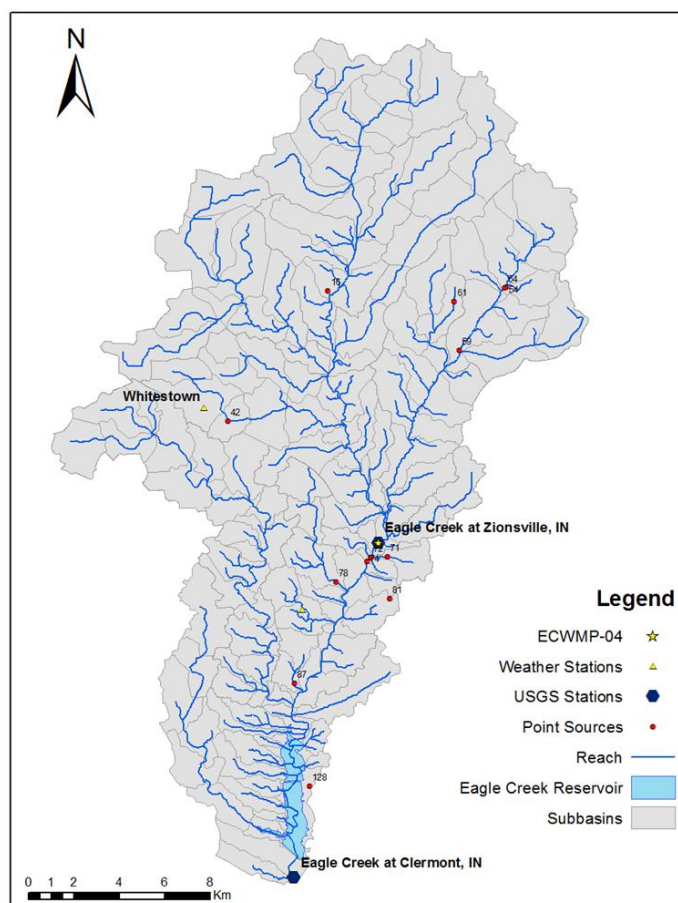


(Dobbins, C.L., and Cook, K., (2002-2012b), Purdue Agricultural Economics Report, Purdue Agricultural Economics Extension. <http://www.agecon.purdue.edu/extension/pubs/paer/archive.asp>).

### **3.2.1. Hydrologic and Water Quality model**

The hydrology and water quality were simulated using the Soil and Water Assessment Tool 2005 (SWAT 2005) model. This modeler was developed by Dr. Jeff Arnold for the USDA Agricultural Research Service (Arnold et al., 1998; Neitsch et al., 2005). SWAT is a physically based, time model that can be operated from ArcGIS interface (via ArcSWAT). It simulates and predict the impact of management practices at subbasin (SB) and watershed scales. The spatial factors such as topography, land use, soil type, and climate are necessary inputs for the development of the model.

Eagle Creek Watershed, the SWAT model, was built on a daily time step for a short time period of five years (i.e. 2004-2008). The watershed was divided into 130 SBs (average area of 327.41 hectares), and 130 reaches for modeling purposes. For each SB, the program calculated the SB outlet on the stream network based on the digital elevation model (10 meter DEM) and pre-defined boundaries. Once outlets were fixed, the point sources (National Pollutant Discharge Elimination System (NPDES) located in SBs: 16, 42, 54, 59, 61, 71, 72, 74, 81, 87, and 128) and the reservoir (located in SB 128) were added (Figure 3.2).



**Figure 3.2 EWC SB and reach (Taken from Piemonti et al., 2013)**

The land use (USDA crop data layer, 2008) and soil type (USDA SSURGO) maps were then added to the model. These two maps were then combined with the land slope map (classified into three classes of 0-1%, 1-2%, and >2%) to divide the SBs into hydrologic response units (HRUs). The HRUs are disconnected, unique combinations of land use, soil type, and slope, in the SWAT model, and are used as a basic spatial unit for the mass balance in the watershed processes.

Daily climate data for precipitation and temperature were obtained from the National Oceanic and Atmospheric Administration (NOAA) stations at Whitestown, IN (Station ID GHCND: USC00129557, latitude 39.996°, longitude -86.354°) and Indianapolis Eagle Creek, IN (Station ID GHCND: USC00124249, latitude 39.920°, longitude -86.313°).

Various model input parameters were modified using specific values for the Eagle Creek Watershed. Model parameter values for tile drains are listed in Table 1 based on typical

values found for tile drains in Central Indiana. For estimating the runoff routing, the curve number method was chosen. While the Muskingum routing method was chosen for channel routing.

Daily flow measurements at the USGS station at Clermont (# 03353460) were used to represent dam releases. For flow calibration, daily data from 2005-2008 (2004 year was let as a warming period for the model) of the USGS gage stations Zionsville gage station and Clermont gage station were compared with the outflows of SBs 70 and 128, respectively. To estimate the efficiency of the model calibration, Nash-Sutcliffe efficiency ( $E_{NS}$ ) (Nash and Sutcliffe, 1970), given by equation (1), was used.

$$E_{NS} = 1 - \frac{\sum_{i=1}^n (O_i - M_i)^2}{\sum_{i=1}^n (O_i - O_{avg})^2} \quad (3.1)$$

Where  $O_i$  is the observed data on day  $i$ ,  $M_i$  is the model data in day  $i$  and  $O_{avg}$  is the average value of the observed data. Pearson's product-moment correlation coefficient ( $R^2$ ) (Legates and McCabe, 1999), given by equation (3.2), was also used to estimate the model performance.

$$R^2 = \frac{\left(\sum_{i=1}^n (M_i - M_{avg})(O_i - O_{avg})\right)^2}{\left(\sum_{i=1}^n (M_i - M_{avg})^2\right)\left(\sum_{i=1}^n (O_i - O_{avg})^2\right)} \quad (3.1)$$

where  $M_{avg}$  is the model data average. For both efficiency estimations equations, a value close to 1 indicated superior model performance. Table 3.1 presents the parameters that were adjusted in order to improve the efficiency of the model for prediction of stream flows.

**Table 3.1 Parameters used in the calibration of SWAT**

Parameter	Description	File	Parameter range	Calibrated value
ALPHA_BF	Base flow alpha factor (days)	.gw	0-1	0.048
CH_K2	Effective hydraulic conductivity in main channel (mm/h)	.rte	0-150	10
CH_N2	Manning's n value for main channel	.rte	0-1	0.01
CN_FROZ	Frozen soil adjustment on infiltration/runoff	.bsn	0 or 1	1 (Active)
CN2	Initial SCS runoff curve number for moisture condition II	.mgt	Specific to land use	<b>For Land Use:</b> AGRR, CORN, SOYB: 0.8075 * CN2 <sub>default</sub> <b>For Land Use:</b> HAY: 1.045 * CN2 <sub>default</sub> <b>Other land-use:</b> 0.95 * CN2 <sub>default</sub>
ESCO	Soil evaporation compensation factor	.hru, .bsn	0-1	0.95
GW_DELAY	Groundwater delay time (days)	.gw	0-50	31
GW_REVAP	Groundwater "revap" coefficient/transfer of water from the shallow aquifer to unsaturated zone	.gw	0.02-0.2	0.02
GWQMN	Threshold depth of water in the shallow aquifer required for return flow to occur (mmH <sub>2</sub> O)	.gw	0-5000	0
HRU_SLP	Average slope steepness (m/m)	.hru	Specific to HRU	2 * HRU_SLP <sub>default</sub>
LAT_TTIME	Lateral flow travel time (days)	.hru		4
SLSUBBSN	Average slope length (m)	.hru	10-150	2 * SLSUBBSN <sub>default</sub>
SMFMN	Melt factor for snow on December 21 (mmH <sub>2</sub> O/°C-day)	.bsn	0-10	1.4
SMFMX	Melt factor for snow on June 21 (mmH <sub>2</sub> O/°C-day)	.bsn	0-10	6.9
SOL_AWC	Available water capacity of the soil layer (mmH <sub>2</sub> O/mm soil)	.sol	0-1	1.5 * SOL_AWC <sub>default</sub>
SURLAG	Surface runoff lag coefficient	.bsn	0-10	6

Water quality observed data collected by the Center of Environmental and Earth Sciences (CEES) of IUPUI (Station ID: ECWMP-04, latitude 39.946°, longitude -86.260°) was used for water quality calibration (Figure 3.2). Since only monthly data from March 2007 to December 2008 were available for sediments and nitrates, we decided to expand the calibration dataset for sediments and nitrates by using LOADEST (Runkel et al., 2004), and then compare the interpolated daily data with the SWAT model daily predictions (White and Chaubey, 2005). Table 3.2 shows the variables that were modified for the sediments and nitrates.

**Table 3.2 Parameters modified for the sediments and nitrates calibration**

<b>Sediments</b>				
<b>Parameter</b>	<b>Description</b>	<b>File</b>	<b>Parameter range</b>	<b>Calibrated value</b>
SPCON	Linear parameter to calculate the maximum amount of sediment re-entrained during channel	.bsn	0.0001-0.1	0.001
SPEXP	Exponential parameter to calculate sediment re-entrained in channel sediment routing	.bsn	0.0-2.0	0.65
PRF	Peak rate adjustment factor for sediment routing in the main channel	.bsn	0.0-2.0	0.01
CH_COV	Channel cover factor	.rte	0.001-1.0	0.12
CH_EROD	Channel erodibility factor	.rte	0.05-0.08	0.08
ADJ_PKR	Peak rate adjustment factor for sediment routing in the sub basin (Tributary channels)	.bsn		0.01
<b>Nitrates</b>				
NPERCO	Nitrate percolation coefficient	.bsn	0.0-1.0	0.7
SDNCO	Denitrification threshold water content	.bsn		0.8
CDN	Denitrification exponential rate coefficient	.bsn	0.0-3.0	0.7
RSDCO	Residue decomposition coefficient	.bsn	0.02-0.2	0.2
IPND2	Ending month of mid-year nutrient settling “season”	.pnd	0-12	12
RCN	Concentration of nitrogen in rainfall (mg N/L)	.bsn	0.0-15.0	3
RS4	Rate coefficient for organic N settling in the reach at 20°C (day <sup>-1</sup> )	.swq	0.001-0.1	0.001
RS3	Benthic source rate for NH <sub>4</sub> -N in the reach at 20°C (mg NH <sub>4</sub> -N/(m <sup>2</sup> *day)	.swq	0-1	1
N_UPDIS	Nitrogen uptake distribution parameter	.bsn		15
SOL_NO3	Initial NO <sub>3</sub> concentration in the soil layer (mg N/Kg soil or ppm)	.chm	0.0-100.0	100
AI1	Fraction of algal biomass that is nitrogen (mg N/mg alg)	.wwq	0.07-0.09	0.071
RHOQ	Algal respiration rate at 20°C/day	.wwq	0.05-0.5	0.5
NSETLW1	Nitrogen settling rate in wetlands for months IPND1 through IPND2 (m/year)	.pnd	0.0-20.0	0.8
NSETLW2	Nitrogen settling rate in wetlands for months other than IPND1-IPND2 (m/year)	.pnd	0.0-20.0	0.8

### 3.3 Participant selection and model modification

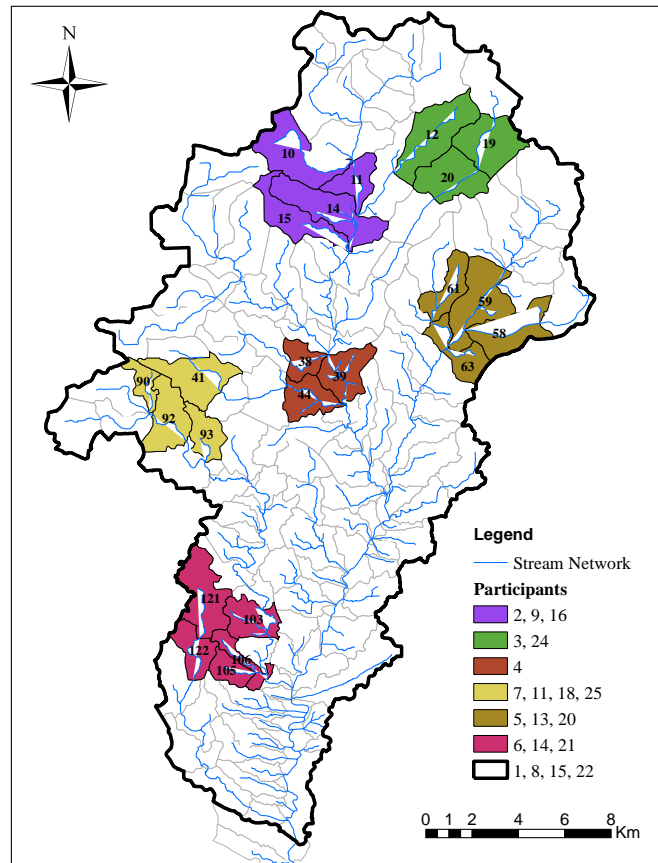
Twenty-three participants volunteered for the study. We divided them into two groups based on their affiliation with the watershed: *stakeholders* and *surrogates*. Each group was treated as a predictor variable. The *stakeholders* group contained eight *stakeholders* from the Eagle Creek watershed who work at federal and state agencies and non-governmental organizations in programs that support the implementation of conservation practices in the watershed. The *surrogates* group contained fifteen non-stakeholder volunteers with science and engineering backgrounds, who were not directly involved in the watershed.

Three participants' data sets were excluded from the analysis. One participant (a *stakeholder*) quit the study before finishing, leaving an incomplete set of answers. The other two participants (a *stakeholder* and a *surrogate* user) failed to follow the instructions. Therefore, twenty participants' data sets were analyzed (Six *stakeholders* – five males, one female; fourteen *surrogates* – six males, eight females). We also want to note that for seven participants in the *surrogates* group, the tool used a different underlying hydrology model (used by the optimization algorithm to calculate the four physical objective functions). This was to create an enhancement of the BMPs performances and observe if there exist significant changes in the responses due to a better performance of the practice

### ***3.3.1. Design and procedure of experiment***

In this research, we investigated the asynchronous aspect of WRESTORE. The experiment was set up to be performed at the participant own convenience time. An initial workshop with the explanation of the capabilities of WRESTORE and the general instruction on how the optimization search process worked was provided to all the participants before they start the test.

For each participant, a group of SBs were assigned. Four out of twenty participants were asked to classify the design alternatives based on their perception of design at a watershed scale. The other sixteen were requested to observe carefully and provide a judgement of the design, based on a selected region (as shown in figure 3.3). We call these sets the SB of interest (*SB<sub>int</sub>*). They will be evaluated and examined later on as part of the analysis at a local scale.



**Figure 3.3 SBs assigned to different participants**

The first page of the registration process and before the experiment is shown in figure 3.4. In this page, all the participants were asked to select the four objective functions (Peak flow reduction, Sediment Reduction, Nitrates Reduction and Economic Cost) and two BMPs (Cover Crops and Filter Strip width). This was set to have an understanding of the representation of two common conservation practices that were widely accepted in the watershed community and that represent one binary variable and one real variable.

**WRESTORE** Watershed REStoration using Spatio-Temporal Optimization of REsources  
VISUALIZE & DESIGN YOUR WATERSHED LANDSCAPE

RETURN TO MAINPAGE BACK TO LOGIN

## GETTING STARTED

You have now entered the design workspace for identifying new types of conservation practices, along with their locations and sizes, in the Eagle Creek Watershed, Indiana. Multiple alternatives will be proposed to you, each of which will have a unique layout of the practices you want to explore.

You will also be able to test the costs and benefits of alternatives for the baseline years 2005-2008. In other words, you will be able to test the impacts of the new practices if they had existed on the farmland in the years 2005-2008.

Please select the options for practices and goals you want to explore, and then click on "Submit" button below to initiate the design process.

>> Which conservation practices or best management practices (BMPs) do you want to work with? (select one or more)

Strip Cropping

Crop Rotation

Cover Crops

Filter Strips

Grassed Waterway

Conservation Tillage

Wetlands

>> Which of these goals are important to you? (select one or more)

Decrease net costs

Decrease flooding

Decrease fertilizer losses

Decrease erosion losses

Submit

## GET INVOLVED

We provide web-based tools for visualizing watershed and designing land use and runoff management alternatives on the landscape.

How does Wrestore work?

Meet our partners

Learn about our tools

Questions? Contact us

**Figure 3.4** Interface for starting a new search experiment for the user’s watershed of interest.

Then, users will start the experiment as described in Chapter 2, sections 2.4 and 2.5. This is an iterative process where the user classifies the design alternatives based on his/her own subjective criteria, and provides information on his/her confidence level. We were also able to track the interaction of the user, such as time spent in areas of interests and clicks events, allowing to gather the appropriated data to test the usability of the tool based on the ISO Standard 9241 that defines usability as the “extent to which a product can be used by the specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use.” The gathered data will provide information about the effectiveness and efficiency of the tool.



**CHAPTER 4. Usability evaluation of an interactive decision support system for planning of conservation practices in a watershed: A case study**

**Adriana D. Piemonti<sup>1</sup>, Kristen L. Macuga<sup>2</sup>, and Meghna Babbar-Sebens<sup>1</sup>**

<sup>1</sup> School of Civil and Construction Engineering, Oregon State University, Corvallis, Oregon, USA.

<sup>2</sup> School of Psychological Science, Oregon State University, Corvallis, Oregon, USA.

#### 4.6 Abstract

This paper evaluates the usability of a web-based watershed design tool called WRESTORE (<http://wrestore.iupui.edu/>). WRESTORE is an Environmental Decision Support System (EDSS) based on an interactive, multi-objective optimization framework. This framework uses participant's ratings of design alternatives, collected via graphical user interface (GUI), as an additional objective function, guiding the search for user-desired designs. Usability of WRESTORE's GUI was evaluated for two groups of participants (*surrogates* and *stakeholders*) via 1) task times across sequential sessions, 2) percentage of time spent and percentage of mouse clicking events in different regions of the GUI (e.g., *Info* vs. *Eval* areas of interest), and 3) trends in self-reported user confidence levels. Results for task times showed that participants followed theoretical models of learning curves across *Introspection* sessions, with coefficients of determination of  $R^2 > 0.9$  for both groups. *Stakeholders*, however, spent 15% more time than *surrogates* on *information gathering* regions. Similarly, *stakeholders* made 14% more mouse clicks than *surrogates* in *information gathering* areas. Confidence level trends increased over time in 67% of the *stakeholder* participants, while only 29% of the *surrogate* participants showed this increase. Analysis of the relationship between time spent, mouse clicking events, and trends in average confidence levels indicated that most participants with higher number of mouse clicks per unit time in *information gathering* areas, increased their self-reported confidence levels. The approach presented provides a useful methodology that can be applied by other EDSS developers to evaluate the usability of their tools.

#### 4.7 Introduction

ISO Standard 9241 defines usability as the “extent to which a product can be used by the specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use.” Estimating usability attributes for different EDSS can potentially help developers to identify users' interactions with their system, and gain a better understanding of their behaviors, preferences and needs. Some researchers, such as Haklay and Tobon (2003), Slocum et al. (2003), Jankowski et al. (2006), Nyerges et al. (2006) have examined the factors that influence the development and usability of EDSS. However, appropriate measurements and evaluations for EDSS usability, as well as

standardized techniques, have not been extensively studied, applied, or developed for these systems.

For example, in Slocum et al. (2003), the authors used a combination of cognitive walkthrough methods, think aloud protocols, and pluralistic inspection while developing software intended to visualize the uncertainties of global water balance. The cognitive walkthrough methods allow developers and human factors specialists to evaluate the steps of tasks-scenarios. The think aloud protocol permits users to verbalize their thought processes as they perform the tasks, while pluralistic inspection allows users with different expertise and backgrounds to contribute towards the improvement of the system. However, such approaches do not present quantitative results, but rather qualitative comments centered on the expertise of each of the tested groups.

Other studies by Jankowski et al. (2006) and Nyerges et al. (2006) evaluated usability of a collaborative spatial EDSS that sought to create a consensus for solutions/options for surface water allocation and groundwater pumping rates, using qualitative and quantitative methods. The qualitative methods involved nonstandard usability questionnaire data, while quantitative methods provided information on task times for different activities, but they did not consider any measurement of the GUI's usability. Jankowski et al. (2006) and Nyerges et al. (2006) provided communication results of two different groups (of ten stakeholders each) that collaborated together to propose scenarios of regulation laws for water allocation and groundwater pumping rates. However, their study examined the analysis and evaluation of collaborative interaction rather than the evaluation of EDSS based on individual stakeholders's interaction and user's behavior. Despite these advances, there are still gaps in the use of quantitative evaluation methods to determine the usability of such participatory EDSS.

Tullis and Albert (2013) proposed quantitative measures such as task time, number of mouse clicks, and percentage of time spent in specific areas of interest of a web interface, as potential approaches that researchers can use for quantifying the usability of software and webpages. In this study, we investigated the use of these quantitative measures for evaluating the usability of a new, web-based participatory EDSS called WRESTORE (Watershed Restoration using Spatio-Temporal Optimization of Resources;

<http://wrestore.iupui.edu/>). WRESTORE has been recently developed with the goal of allowing stakeholders, policy makers, and planners to participate in the design of a spatially distributed system of BMPs in a watershed. WRESTORE uses a modification of the IGAMII (Interactive Genetic Algorithm with Mixed Initiative Interactions – Babbar-Sebens and Minsker, 2011) algorithm to engage users in a human-guided search process that is targeted towards identification of user-preferred alternatives. In WRESTORE, users are shown multiple design alternatives and are asked to provide a qualitative rating of the candidate designs based on their preferences and subjective criteria. This *user rating* is then used as an additional objective function in the search algorithm to search for similar or better alternatives that would have design features liked by the user. In summary, WRESTORE optimizes five objective functions: four objective functions are related to the physical performance of the BMPs within the watershed, and one objective function (*user ratings*) is related to the user's qualitative rating of the newly found design alternatives. Babbar-Sebens et al. (2015) have extensively explained how the *user ratings* for design alternatives are obtained via the GUI. Because these *user ratings* are what is used by the underlying optimization algorithm to search for new solutions/design alternatives that agree with user's qualitative preferences (such as land management, biases towards practices and their implementation, individual local constraints and needs, etc.), evaluating the usability of the GUI is critical. In addition to the *user ratings*, the interface is also used to collect *confidence levels* from the users. The self-reported *confidence levels* indicate how confident a participant was regarding his/her own rating of a candidate design alternative (Babbar-Sebens and Minsker, 2011). The *confidence levels*, along with the user's interaction, offer potential insights into the quality of user's input. This information is also valuable for assessing the reliability of a user-guided search.

Usability testing was conducted for WRESTORE, developed by Babbar-Sebens et al. (2015), and this article summarizes the findings on the nature of user behavior observed for two different types of users (*stakeholders* and *surrogates*) who interacted with the search tool's GUI. In Section 4.3, we present the objectives and research hypotheses investigated in this work. Section 4.4 has the description of the methodology and the experimental design, followed by Sections 4.5 and 4.6 that contain the analysis and

discussion of the results. Finally, in Section 4.7, we present some concluding thoughts, future work and recommendations that may address the usability testing of WRESTORE and similar EDSS.

## 4.8 Objectives

The aim of the study was to use an observational approach to determine how participants used the GUI of the WRESTORE EDSS, as they gathered information and made decisions in order to rate different design alternatives. This kind of system, that includes direct participation of stakeholders in the optimization process, is relatively new in the field of watershed planning and management, and also in the field of EDSS. As mentioned earlier, since the GUI is the primary mechanism to collect the *user ratings* from end users, it is important to determine if it can support the necessary functions for a user to be able to easily (a) gather information about the candidate design alternatives, (b) conduct comparisons between candidate design alternatives, and (c) provide meaningful feedback with improved confidence in his/her own evaluation of candidate design alternatives. Analyses of the user's data will help to better characterize his/her interaction with the tool, enabling improvements in the efficiency of the underlying search algorithm used by WRESTORE (i.e, IGAMII). We tested the tool's performance with two groups: *stakeholders* (watershed end users) and *surrogates* (non-stakeholder volunteers). Because *surrogate* users' data is often used for prototype development of interfaces due to time and cost constraints, it is critical for us to know and understand the extent to which volunteers can be used as proxies for future end users.

The study examined the following research hypotheses:

**H1.** Over time, users should become more efficient in using WRESTORE's GUI. As users learn how to navigate and use the GUI's features, overall task times should decrease across repeated sessions. This decrease should follow the theoretical learning curves (Yelle, 1979; Newell and Rosenbloom, 1981; Estes, 1994), specifically, DeJong's learning formula (Jaber, 2011).

**H2.** Because *stakeholders* are directly affected by the issues and implementation decisions related to the watershed, we expect that the percentage of time they spend in *information gathering* areas of the interface will be greater than for *surrogate* users.

**H3.** Similarly, we expect that *stakeholders* will have a higher percentage of mouse clicks than *surrogates* in the *information gathering* areas of the web-interface.

**H4.** As users gain experience by interacting with the tool and develop a better understanding of the performances of different designs alternatives, we expect that their overall self-reported confidence levels will increase over time, resulting in a positive trend.

**H5.** A comparison among confidence level trends, time spent, and mouse clicking events should show that when users spend more time and make more mouse clicks in *information gathering* areas, their confidence levels increase over time.

A final goal was to set a basis for evaluating the interactions between human factors and GUIs in participatory decision support tools. Such protocols will allow tool developers to test the usability of these systems for their user community, determine what improvements should be made based on specific user populations, and learn how to facilitate the participation of stakeholders in similar watershed design tools.

## **4.9 Methodology**

### **4.4.1. Case study site**

The web-based WRESTORE tool (<http://wrestore.iupui.edu/>) was developed by Babbar-Sebens et al. (2015) to enable watershed communities to engage in participatory design efforts and influence the spatial design of conservation practices on their landscape. WRESTORE software is currently being tested at the study site of Eagle Creek Watershed, Indiana, where multiple researchers (Tedesco et al., 2005; Babbar-Sebens et al., 2013; Piemonti et al., 2013) have conducted watershed investigations.

### **4.4.2. Web-tool WRESTORE evaluation**

Chapter 3 presents a detailed description of the selected participants and how the testing was performed. In summary, twenty participants are selected and divided into three different groups, depending on their background and the hydrologic model that was used. The first group (referred as *Model A-Surrogates*), was integrated by seven participants, with a majority of students related to the research group. The second group (*Model B-Surrogates*) was also integrated by seven different participants, associated with engineer and sciences, but not directly related to the research group. The third and final group

(*Model B-Stakeholders*) was integrated for actual stakeholders concerned about the solutions and design plans in the watershed.

In general, the main goal is to rate 20 design alternatives in each session, with the aim to generate new design alternatives that will satisfy not just physical criteria and conditions, but also the participant's subjective perception. To understand and evaluate the results, we provided a semi-structure design, where Cover Crops and Filters Strips must be optimized at a watershed scale (for five different users) and a local Subbasin (SB) scale. Figure 4.1 shows the components of the interface for each pair of designs, where the participant is allowed to see the decision and objective space, and at the same time provide the classification on a rating scale, addressing the level of agreement (or disagreement) and the confidence level of that classification that will serve to generate future new design alternatives under a human-guided search.

To measure interface interaction, we tracked the time spent and number of mouse clicks in three main areas of interest (AOI). The first AOI was associated with *information gathering* ( $AOI_1 = Info$ ) and included user interactions with the legend (*component (ii)*), the Alternative maps (*component (iii)*) and/or the drop-down (*sub-drop* in *component (v)*) menu. The drop-down menu allowed users to compare the performance of two design alternatives at watershed scale or at a local sub-basin scale (*component (v)*). The second AOI was associated with the evaluation process ( $AOI_2 = Eval$ ), and included user interactions with the Likert-scale *user ratings* and the confidence sliders (*component (iv)*). The final AOI ( $AOI_3 = Other$ ) included all other user activity outside the areas covered by  $AOI_1$  and  $AOI_2$ , i.e., if the participant clicked in an area outside of the purple (*Info*) and green (*Eval*) boxes shown in Figure 4.1.

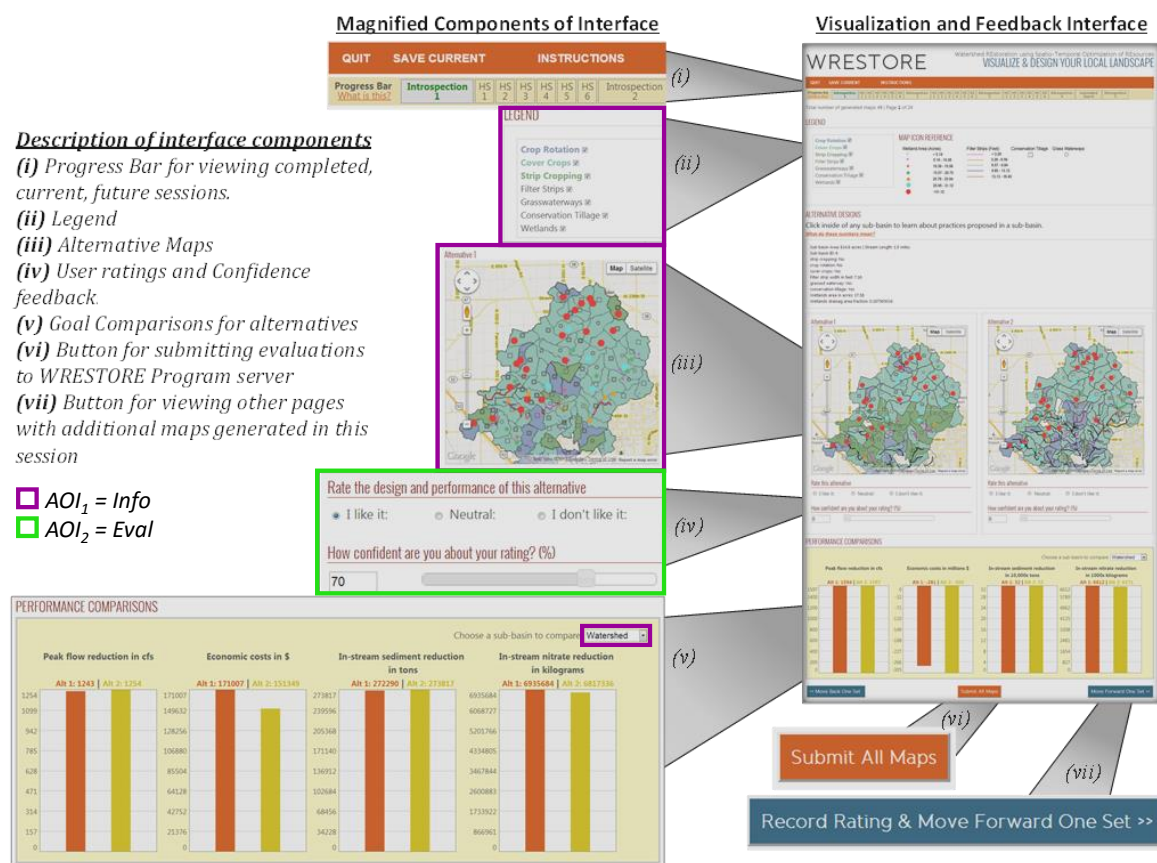


Figure 4.1. Components of WRESTORE tool's main design and feedback interface and areas of interest (AOI) (Modified from Babbar-Sebens et al., 2015)

#### 4.4.3. Measures of user interaction with the interface

This section introduces the different usability metrics or response variables that were employed to evaluate the user's interaction and usability of the WRESTORE tool.

##### 4.4.3.1. Overall task times

We assessed the participant's ability to navigate and learn the tool by evaluating the mean task times across successive *I* sessions and also across successive *HS* sessions, during which design alternatives were presented to the user via the GUI. Task times for each session were recorded and used to infer how quickly participants learned to use the tool interface. When the participants were not using the tool (e.g. taking a break), they were instructed to press the "Save all" button, so that we could consider these off-task time



intervals as outliers and exclude them from the task time analyzes. These events were saved in a database as “Save all Maps”, and removed from the post-processing analysis, along with “Quit” events. However, there were some occasions when some participants failed to click the “Quit” or “Save all” buttons, resulting in excessively long task times. To remove these outliers, we excluded the task time values that were greater than two standard deviations from the mean task time across all considered sessions, for each participant.

#### 4.4.3.2. Mean percentage of time spent in different areas of interest

In order to compare time spent in each AOI (*Info*, *Eval* and *Other* described in section 4.4.2), the percentage of time spent ( $pts_{ijm}$ ) was calculated using:

$$pts_{ijm} = \frac{\sum_{h=1}^H \sum_{k=1}^L \Delta t_{ijm kh}}{\sum_{h=1}^H ttotal_{jmh}} * 100 \quad (4.1)$$

where  $pts_{ijm}$  is the percentage of time spent,  $i$  is an index that goes from one to three and represents each AOI,  $j$  is a vector that goes from one to five and represents  $I$  or  $HS$  sessions. Data from  $HS1$  to  $HS6$  sessions were grouped together into blocks of  $HS$  sessions in order to conduct a temporal analysis that was based on alternating session types. Therefore, if  $j$  is an even number it represents an *Introspection* session, and the variable  $H$  is equal to one (because *Introspection* sessions are not evaluated in blocks), but if  $j$  is an odd number it represents a *Human-guided Search* session, and  $H$  is equal to six (representing each of the iterations in the micro-IGA). The index for each participant is represented by  $m$ . The variable  $\Delta t_{ijm kh}$  is the interval of time between two events ( $k$ ),  $L$  is the total number of events associated with the AOI $_i$  and  $ttotal_{jmh}$  is the participant's ( $m$ ) total time in each session ( $j$ ).

The mean time spent was calculated by adding the percentage of time spent ( $pts_{ijm}$ ) per AOI $_i$  per participant ( $m$ ). This mean was grouped by session type and by the group. Therefore, to calculate the mean percentage of time spent by AOI we used:

$$MPTS_{ij} = \frac{\sum_{m=1}^N pts_{ijm}}{N} \quad (4.2)$$

where  $MPTS_{ij}$  is the mean percentage of time spent in each AOI $_i$ , and  $N$  is the total number of participants in the group.

#### 4.4.3.3. Mean percentage of mouse clicking events by area of interest

The same AOIs described in section 4.2.2 (*Info*, *Eval* and *Other*) were used to track the mouse clicking events. As the total number of clicking events varied between participants, the percentage of clicking events in each session was calculated for each participant, and for each group (*surrogates* and *stakeholders*).

For each participant, we used the following general formula:

$$PMC_{ijm} = \frac{\sum_{h=1}^H NC_{ijmh}}{\sum_{h=1}^H TC_{jmh}} * 100 \quad (4.3)$$

where,  $PMC_{ijm}$  is the percentage of mouse clicks per participant per AOI,  $NC_{ijm}$  is the number of clicks per  $i^{th}$  AOI, per  $j^{th}$  session, per  $m^{th}$  participant, and  $TC_{jmh}$  is the total number of clicks per  $j^{th}$  session per  $m^{th}$  participant. As in the mean percentage of time spent,  $H$  will vary with the session type. Therefore, for  $j$  equal to an even number (representing  $I$  sessions), the variable  $H$  is equal to one, and for  $j$  equal to an odd number (representing  $HS$  sessions) the variable  $H$  is equal to six. This allowed us to compare the clicking interactions of individuals within each group.

The percentage for each group was calculated using the mean value of all the participants within the group. Therefore:

$$PGMC_{ij} = \frac{\sum_{m=1}^N PMC_{ijm}}{N} \quad (4.4)$$

where  $PGMC_{ij}$  is the percentage per group per  $i^{th}$  AOI, and  $N$  is the total number of participants per group.

#### 4.4.4. Confidence levels

*Confidence level* indicates how confident the participant felt about his/her own *user rating* (*I like it*, *Neutral*, or *I do not like it*). User's confidence in his/her *user ratings* were indicated via the *confidence level* slider bar (*component (iv)* in Figure 4.1 **Error! Reference source not found.**) that ranged in its scale from 0 to 100, and could be modified by the user during the session. However, changes in the *confidence levels* for the same designs that appeared multiple times during the experiment could not be tracked, as new changed values replaced previous values in the archive database.

We classified participants by *confidence level* trends in the following manner. First, we conducted a nonparametric Mann-Kendall hypothesis test (Helsel and Hirsch, 2002) via a Matlab script (Burkey, 2006) to assess whether the trends in average values of *confidence levels* (estimated from data on *confidence levels* in each session) were monotonically increasing or decreasing. This test indicated whether or not participants presented a trend, at a significance level alpha of 0.1, and the Sen' Slope (S value) determined if the trend was positive or negative. Since the main focus of this test was to minimize the risk of not detecting an existing trend (i.e. Type II error), a larger alpha value was chosen. Participants were thus separated into three *confidence level* trend groups (*positive*, *negative* and *no* trend). Finally, we attempted to identify similarities and differences between participants who showed *positive*, *negative* and *no* trend in *confidence levels*, and relate them to their interface behavior (i.e., time spent and mouse clicks) concerned with *information gathering*.

#### 4.4.5. Relationships between confidence levels, time spent and mouse clicking events

We fitted trend curves to usability data (i.e. time spent and number of mouse clicks) in order to determine any underlying patterns and relationships for participants within each *confidence level* trend (*positive*, *negative*, or *no*). We also compared the differences in these trends for *Info* and *Eval* type events. To select the best trend curve model, we used a combination of the coefficient of determination, and three versions of the Akaike Information Criterion (AIC).

The AIC provides the relative quality of a proposed model in a given data set, dealing with the goodness of fit and the complexity of the model. The approaches for calculating AIC were based on the following equations:

$$AIC = -2(\log(L)) + 2K \quad (4.5)$$

$$AIC_{Residual} = n(\log(\sigma^2)) + 2K \quad (4.6)$$

$$AIC_C = -2(\log(L)) + 2K + \frac{2K(K+1)}{(n-K-1)} \quad (4.7)$$

where  $L$  is the likelihood of the model,  $K$  is the number of parameters,  $n$  is the sample size,  $AIC$  is the definition for the Akaike's Information Criterion,  $AIC_{Residual}$  is based on the least

square regression (assuming normal distribution) and  $AIC_C$  is the second order AIC that includes a penalization for small sample sizes. (Mazerolle, 2004)

## 4.5. Results

In this section, we present the results from the data analysis for the task time, percentage of time spent in each AOI, percentage of mouse clicks in each AOI, and confidence level trends, as well as the relationships between these variables.

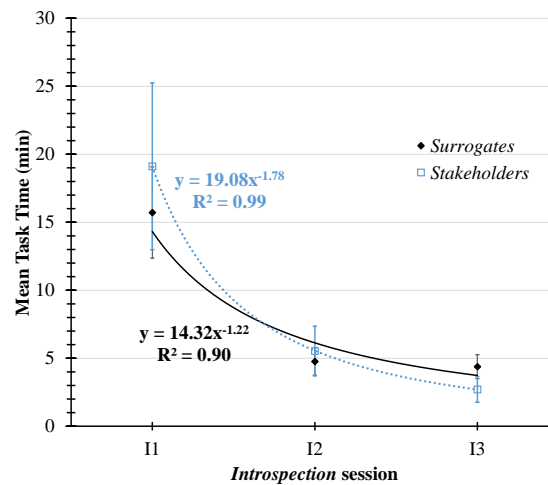
### 4.5.1. Overall task times

We first assessed differences between participant groups based on their mean task times for each session. The mean for each group (*surrogates* and *stakeholders*) was calculated and the results were separated according to the type of sessions (i.e., *I* or *HS*). Results for *surrogates* showed that in earlier sessions (for both *I* and *HS*) the mean task time and the standard errors were greater than in later sessions. Similar results were found for *stakeholders* in *I* sessions. However, *stakeholders*' mean task times for *HS* sessions were somewhat more variable.

Figure 4.2 presents the mean task times for *surrogates* and *stakeholders* for all completed *I* sessions, where the main task of participants was to evaluate initial designs and/or re-evaluate their previous *user ratings*. Despite the limited number of *Introspection* sessions, both groups show a good estimation of the theoretical power learning curve typically used to represent learning processes (Yelle, 1979; Newell and Rosenbloom, 1981; Estes, 1994; Jaber, 2011), with a coefficient of determination  $> 0.9$  for both groups. However, we observed that the *stakeholders*' learning curve started at a higher value of mean task time than that of the *surrogates*. *Stakeholders*' mean task times decreased by 85% from *I1* to *I3*, while *surrogates*' mean task times decreased by 74% from *I1* to *I3*. This effect shows that the task might have seemed more challenging for the *stakeholders* in the first *I* session when they were still getting used to the interface, but as time progressed, their mean task time decreased to a value lower than the *surrogates*', in the last *I* session. Variability also decreased over time for both groups.

We also compared the task times of *HS* sessions for the two groups (Figure 4.3), where the main task of participants was to compare and evaluate newly generated design

alternatives, and provide feedback on these designs to help guide the search algorithm in creating new design alternatives for the next *HS* session. **Error! Reference source not found.**3 shows two different learning curves for *surrogates* and *stakeholders*. *Surrogates* showed continuous learning with a decrease in average time across *HS* sessions. *Surrogates*' behavior fit the power learning curve with a coefficient of determination of  $R^2 = 0.87$ . The mean task times for *stakeholders*, though generally higher than *surrogates*, poorly fit the theoretical power learning curve, resulting in a coefficient of determination of  $R^2 = 0.23$ . This fit is substantially lower than the  $R^2$  value previously obtained for the *stakeholders*' *I* sessions, primarily due to the higher variability of the mean task times within *HS* sessions. On average, the standard error for *stakeholders* in the *HS* sessions was 62% larger than that for *surrogates*.



**Figure 4.2. Power function representing the learning curves of the mean task times for the surrogates and stakeholders groups across *I* sessions. Error bars represent the standard error of the mean.**

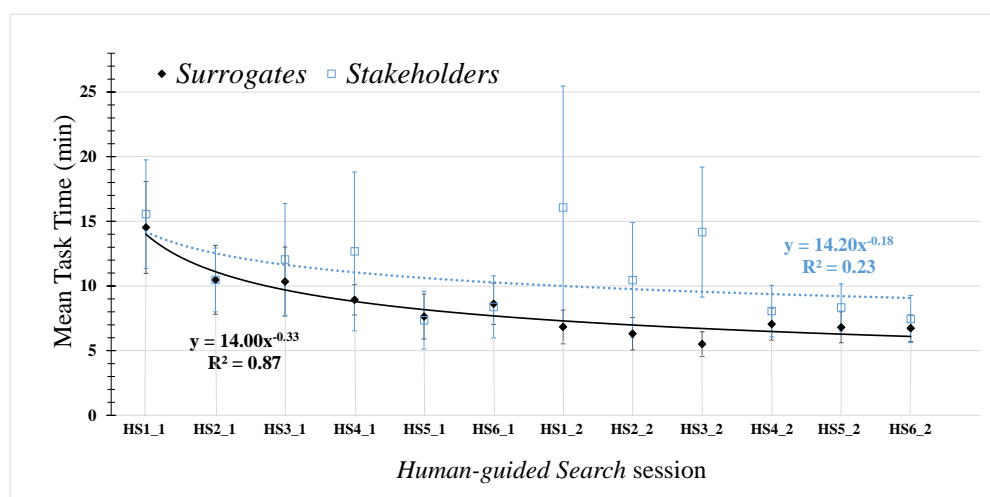


Figure 4.3. Power function representing the learning curves of the mean task times for the surrogates and stakeholders groups across HS sessions. Error bars represent the standard error of the mean.

#### 4.5.2. Mean percentage of time spent in different areas of interest

We also analyzed the mean percentage of time spent within the different AOIs for each group. Figure 4.4 shows a pie chart table that compares *surrogates* and *stakeholders* for all sessions. On average, *stakeholders* spent 15% more time on *information gathering* than *surrogates*, while *surrogates* spent more time in the *Eval* and *Other* AOIs (8% and 7% respectively greater than *stakeholders*).

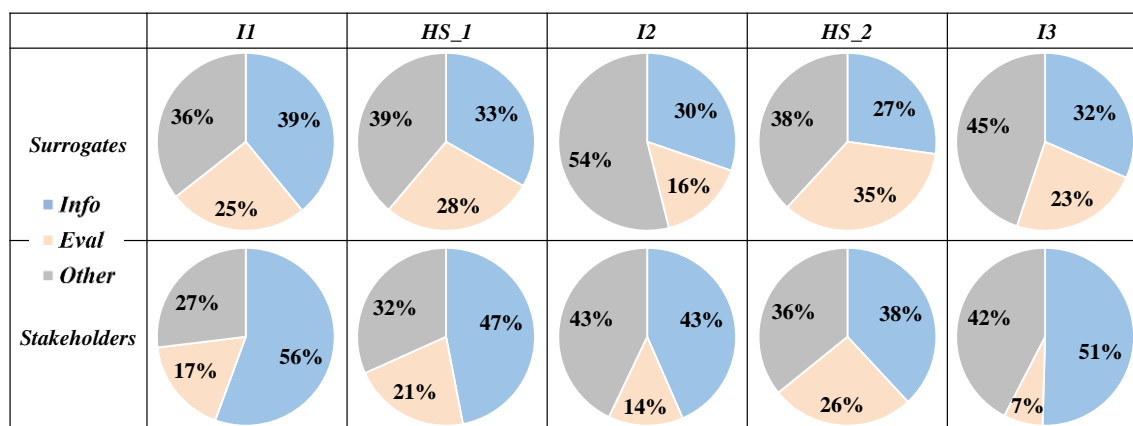


Figure 4.4. Mean percentage of time spent in each AOI for surrogates and stakeholders. Refer to section 4.2.2 and Figure 4.1 for descriptions of each AOI.

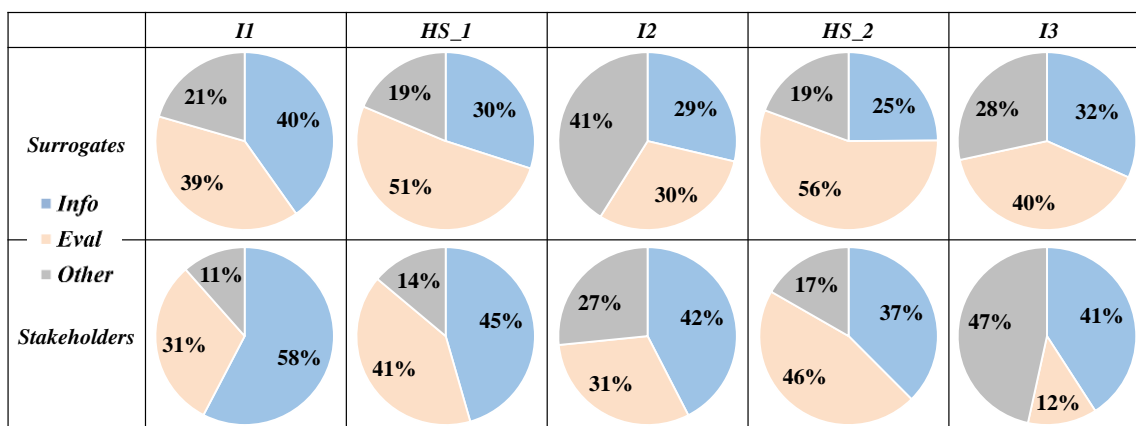
Table 4.1 summarizes the data in Figure 4.4, and presents the overall mean percentages of time spent (averaged across sessions) and 95% CIs for *surrogates* and *stakeholders* in each AOI.

**Table 4.1 Confidence intervals (CI) for the mean percentages of time spent in each AOI across all sessions by group**

AOI	Group	Mean	95% CI
<i>Info</i>	<i>Surrogates</i>	32	[25, 39]
	<i>Stakeholders</i>	47	[36, 57]
<i>Eval</i>	<i>Surrogates</i>	25	[21, 30]
	<i>Stakeholders</i>	17	[13, 22]
<i>Other</i>	<i>Surrogates</i>	42	[36, 48]
	<i>Stakeholders</i>	36	[27, 45]

#### 4.5.3. Mean percentage of clicking events in different areas of interest

Similarly to Figure 4.4, Figure 4.5 displays a summary pie chart table with the mean percentages of mouse clicks in the *Info*, *Eval* and *Other* AOIs for the two groups across the different types of sessions. On average, *stakeholders* clicked 14% more in *Info* areas than *surrogates*, while *surrogates* had a greater percentage of clicks in the *Eval* and *Other* areas of interest in comparison to *stakeholders*. The percentage of clicks in the *Eval* AOI was 11% greater for the *surrogates* group, and the percentage of clicks in the *Other* AOI was 3% greater for *surrogates*.



**Figure 4.5. Mean percentage of mouse clicking events for surrogates and stakeholders within each AOI. Refer to section 4.2.2 and Figure 4.1 for descriptions of each AOI.**

We calculated 95% Confidence Intervals (CI) for overall mean mouse clicking events, similar to the analyses in Table 4.1 for time spent. Table 4.2 **Error! Reference source not found.** presents the overall mean percentages of mouse clicks (averaged across sessions) for *surrogates* and *stakeholders* in each AOI.

**Table 4.2 Confidence intervals (CI) for the mean percentages of mouse clicking events within each AOI across all sessions by group.**

<b>AOI</b>	<b>Group</b>	<b>Mean</b>	<b>95% CI</b>
<i>Info</i>	<i>Surrogates</i>	31	[24, 38]
	<i>Stakeholders</i>	45	[34, 56]
<i>Eval</i>	<i>Surrogates</i>	43	[38, 49]
	<i>Stakeholders</i>	33	[24, 42]
<i>Other</i>	<i>Surrogates</i>	26	[21, 30]
	<i>Stakeholders</i>	22	[14, 30]

#### 4.5.4. Confidence Levels

For each participant, and for each of the rating classes (i.e., *I like it*, *Neutral* and *I do not like it*), a Mann-Kendall trend test was performed to identify if there were monotonic trends in the mean *confidence levels* across consecutive sessions. The results were separated according to *Positive*, *Negative*, or *No* trends, based on the results over time. A *Positive* trend indicated that users were becoming more confident over time about the ratings they provided for the designs. On the other hand, a *Negative* trend indicated that they were becoming less confident over time. In summary, 40% of all of the participants showed a *Positive* trend in at least one of the rating classes. For the *stakeholders* group, 67% of the participants showed a *Positive* trend, while just 29% of the participants in the *surrogates* group showed a *Positive* trend. It is, however, important to mention here that these trends were calculated for the sessions that lasted until the end of *I3*. It is possible that if participants continued beyond *I3* then they could change their trends, especially if



the additional engagement during the interactive search process improved their reasoning process and led to a change in their *confidence levels*.

#### **4.5.5. Relationships between *confidence levels*, time spent and mouse clicking events.**

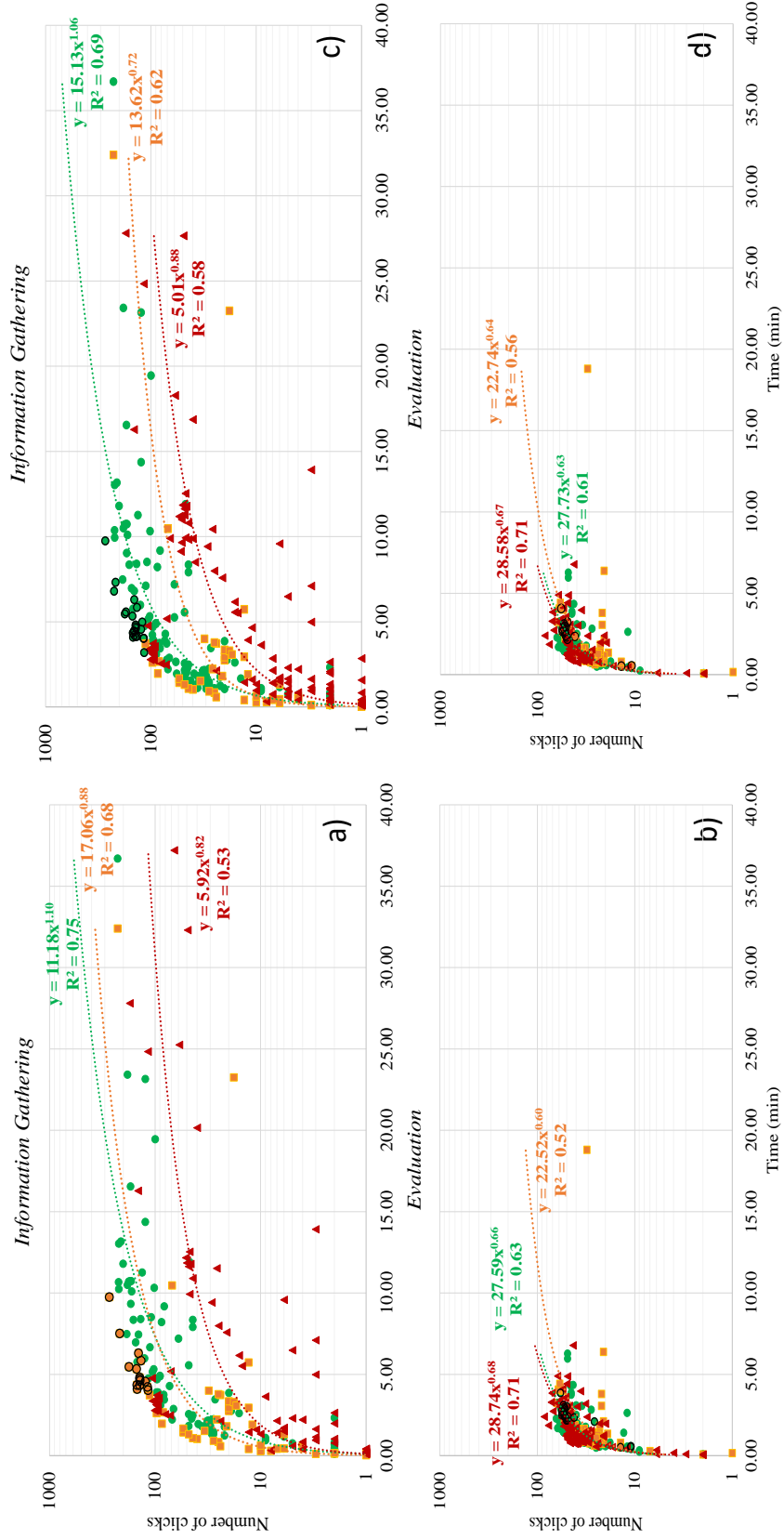
We also associated the mean percentage of mouse clicks for *information gathering* and the mean percentage of time spent for *information gathering* (Table 4.3) with each of the trends in mean *confidence levels* reported in the previous sub-section. Results showed that participants with a *Positive* trend in mean *confidence levels* also had 13% more mouse clicks than participants with a *Negative* trend, and 9% more mouse clicks than participants with *No* trend in areas of interest related to *information gathering*. Participants with a *Positive* trend in mean *confidence levels* spent 1% more time than participants with a *Negative* trend, and 12% more time than participants with *No* trend in areas of interest related to *information gathering*.

As it was suggested earlier, it is possible for participants to experience changes in the trends of the means of their *confidence levels*, during the course of their interaction with the tool. Therefore, we identified participants who had interacted with WRESTORE tool beyond the *I3* session, and re-evaluated the trends in their interaction data by including the data from the additional sessions after *I3*. Figure shows four plots that relate time spent vs. number of mouse clicking events for each participant in the *Info* and *Eval* AOI. Figure a) and 4.6b) show the relationships for the data collected from sessions completed by all participants until the end of *I3*. Figure c) and 4.6d) show the results for all of the available data from each participant, which includes data from additional sessions for participants who progressed beyond *I3*. It can be seen that while the classification of trends in mean *confidence levels* remained the same for the majority of participants (see Figure a) and 4.6c)), the trend for Participant 2, however, changed from *No* trend (for sessions from *I1* to *I3*) to *Positive* trend (for sessions from *I1* to *I5*). This indicates that when Participant 2 engaged actively with the tool longer than *I3*, she/he was eventually able to improve his/her confidence in the *user ratings* provided during the experiment. Results also showed that,

for the *Eval* AOI (Figures 4.6b) and 4.6d)), there is no clear separation between the responses for each trend, irrespective of how long the experiment lasted.

**Table 4.3 Classification of confidence levels, mean percentage of clicking events and mean percentage of time spent across participants with the same trend.**

Trend	% of Total participants	Mean % of clicking events in <i>Info</i> AOI	Mean % of time spent in <i>Info</i> AOI
<i>Positive</i>	40	42	41
<i>Negative</i>	30	29	40
<i>No</i>	30	33	29



● Positive ■ No ▲ Negative ○ Participant 2

Figure 4.6 Time spent vs. number of clicks per participant per trend, for *Info* (upper) and *Eval* (lower) AOIs. Sections a) and b) show the results for sessions *I1* to *I3*. Sections c) and d) show the results for all completed sessions. The empty circles show the change in trends for Participant

Table 4.4 shows the comparison of the different AIC values obtained for four different approximations generated using the results for the sessions completed by all the participants. The AIC values indicate that a power approximation represents the best-fit model for any of the trends, with the exception of a negative trend in “All Available Sessions”, where a linear function seems to have a better fit.

**Table 4.4 Coefficient of determination and AIC values for the different tested models**

Trend	Model	Sessions I1 to I3				All Available Sessions			
		R <sup>2</sup>	AIC	AIC <sub>Residual</sub>	AIC <sub>c</sub>	R <sup>2</sup>	AIC	AIC <sub>Residual</sub>	AIC <sub>c</sub>
Positive	Linear	0.38	-1.64	302.55	-1.59	0.17	-1.65	478.38	-1.62
	Quadratic	0.67	0.18	281.16	0.32	0.57	0.14	442.11	0.23
	Third	0.73	2.11	277.28	2.39	0.67	2.06	430.93	2.24
	Power	0.75	-1.85	321.37	-1.80	0.69	-1.89	502.60	-1.86
No	Linear	0.06	-1.13	219.08	-1.06	0.29	-0.44	145.04	-0.35
	Quadratic	0.45	0.46	207.22	0.67	0.31	1.42	146.48	1.69
	Third	0.60	2.40	205.28	2.83	0.66	2.99	149.38	3.54
	Power	0.68	-1.38	215.91	-1.31	0.62	-0.77	141.38	-0.68
Negative	Linear	2.35E-03	-0.69	242.36	-0.64	0.22	-0.71	325.21	-0.67
	Quadratic	0.14	1.07	239.52	1.24	0.25	1.19	325.49	1.30
	Third	0.17	2.83	250.73	3.17	0.35	3.08	321.65	3.31
	Power	0.53	-0.73	239.70	-0.68	0.58	-0.55	325.90	-0.52

## 4.6. Discussion

In this paper we investigated the usability of a novel web-based tool, WRESTORE, which supports decision makers in the search and design of alternatives for allocating conservation practices in agricultural watersheds.

Four main usability metrics were considered for the analysis: task time evaluation, percentage of time spent in different areas of interest, percentage of mouse clicking events in different areas of interest, and trends in mean *confidence levels*.

### 4.6.1. Overall task times

We evaluated overall task times in order to assess participants' efficiency in using the tool interface. Overall task times decreased across repeated sessions following a power

learning curve. The results are consistent with hypothesis **H1** where it was stated that over time, users should become more efficient in using the interface, as they learn how to navigate and use the different features. *I* sessions followed a power learning curve for both *surrogates* and *stakeholders*, but *stakeholders* showed a greater decrease in overall task time for *I* sessions than *surrogates*.

*HS* sessions also followed a power learning curve for the *surrogates*. However, the mean task times for *stakeholders* were much more variable, resulting in a poor fit to the expected learning curve. The differences in mean task times across *stakeholders* may be due to the following potential reasons: 1) some abnormally large task times that were still within two standard deviations from the mean and hence were not excluded as outliers, possible related with a re-learning process due to the lag time between sessions, 2) small sample size of *stakeholders*, or 3) existence of two learning curves instead of one overall learning curve as seen by an apparent increase in the time spent again at the start of the second block of *HS* sessions (i.e., HS1\_2 in Figure 4.3).

Overall task times can provide estimates of the tool's efficiency to developers. Information on how fast the users learn via the GUI can be expected to assist the tool designers in the re-evaluation of the users' interactions during the sessions, and of the tool's ability to use participation in guiding the search process. Further, an insightful understanding of time needed to complete goals and provide useful feedback is also important for eliminating extra sessions, which could increase user fatigue and introduce noise without meaningful benefit to the search algorithm.

#### **4.6.2. Mean percentage of time spent in different areas of interest**

The analysis of these data provided us with insightful evidence about the percentage of time each group spent in each of the areas of interest (AOIs). The groups showed fairly consistent behavior across sessions. However, *stakeholders* expended more time in the *Info* AOI, while *surrogates* expended more time in *Eval* and *Other* AOIs. This supports hypothesis **H2**, where we predicted that because the *stakeholders* are directly affected by issues and actions in the watershed, their percentage of time spent in *information gathering* areas of the interface would be greater than for the *surrogates*.

#### 4.6.3. Mean percentage of clicking events in different areas of interest

These results help to understand how different groups were using the interface to perform the tasks. The higher percentage of mouse clicking events for *stakeholders* vs. *surrogates* in *Info* areas of interest supports hypothesis **H3** and is consistent with the results reported above for mean percentage of time spent. *Surrogates* and *stakeholders* behave differently regarding *information gathering*. On average, the majority of *stakeholders* tend to make more mouse clicks to gather information from the user interface in order to make their decisions. *Surrogates*, on the other hand, do not explore the *information gathering* areas of the interface as much, making fewer mouse clicks in these regions. This could be a consequence of lack of interest in the task, or a lack of information about the tool's goals.

#### 4.6.4. Confidence levels

Previous work using *confidence levels*, showed a positive trend as experimental time progressed (Babbar-Sebens and Minsker, 2011). A positive trend indicates that mean confidence levels increase over time as users gained experience with the tool. However, we did not find that mean confidence levels increased for all of the users. Therefore hypothesis **H4**, was only partially supported.

Our results showed that just 40% of participants exhibited a *Positive* trend in mean confidence levels over time. Nevertheless, there was a clear distinction between *surrogates* and *stakeholders* in relation to these trends. Approximately 67% of the total *stakeholder* participants presented a positive trend, while just 29% of the total *surrogate* participants presented a positive trend.

#### 4.6.5. Relationships between *confidence levels*, time spent and mouse clicking events.

The section examines the results on the relationships between trends in mean *confidence levels*, time spent and mouse clicking events. The results in Table 4.3 help to support hypothesis **H5** that states “when users spend more time and make more mouse clicks in information gathering areas, their mean *confidence levels* increase over time”. An analysis of relationship between time spent, clicking events and mean *confidence levels* trends indicates that participants with a *Positive* trend had a larger number of clicking events per

unit time spent in *information gathering* areas, compared to participants with a *Negative* trend. This may indicate that if a participant interacts with the interface to gather more information on the design alternatives for the same amount of interaction clock time (i.e., the time spent), then there exists a probability for them to either improve their self-confidence over time or maintain a steady value over time. In addition to the slope, the fitted value of the exponent in power curves was also lower for the *Negative* trend than for the *Positive* trend in mean *confidence levels*. In the *Eval* AOI, no clear difference across *confidence level* trends was observed.

#### **4.7 Conclusions and future work**

This research presented some techniques widely used in usability testing and adapted them to a set of data collected with the WRESTORE tool to evaluate its performance and usability. Overall, this work provided three important contributions: 1) determination and validation of usability and metrics for participatory design tools based on information technologies, 2) evaluation of differences between how *surrogates* (volunteers) and *stakeholders* (end users) use such interactive design tools, and 3) suggestions for possible improvements and considerations for similar web-based watershed design tools.

WRESTORE was developed for the Eagle Creek Watershed, in Indianapolis, IN, with the goal of providing a more democratic venue for stakeholders to engage with the watershed community in the design of alternatives for spatial allocation of conservation practices. The tool was initially tested by *surrogates* who were not intimately involved with the issues and concerns in the watershed. Their feedbacks were recorded and saved for later analysis. Then, the tool was tested with *stakeholders* (i.e., potential end users) to determine if the findings from *surrogates* held true for the actual end users.

As the majority of usability tests are performed by students or volunteers, we wanted to track possible differences in responses between the tested group (*surrogates*) and the end users (*stakeholders*), and analyze to what extent the results from *surrogates* would be reflected in the behaviors of *stakeholders*. From the overall task time analysis, we concluded that the participants of both groups became more efficient as they learned how to navigate and use the tool's features. Overall task times decreased across repeated

sessions. Therefore, *surrogates* can potentially be used as proxies for *stakeholders* for overall task time analysis and improvements.

For time spent on the areas of interest (AOIs), results showed that *stakeholders* expended more time than *surrogates* in gathering information about the performance of the different design alternatives. This could be a result of the motivation of *stakeholders* in creating a designed distribution of Best Management Practices (BMPs) that better suits their interests. *Surrogates* that were not involved with the watershed may have lacked this motivation. Similarly, as we predicted, a higher percentage of mouse clicks were made on the *information gathering* areas by *stakeholders* vs. *surrogates*.

We also noticed that the majority of the *stakeholders* showed an increase in their mean *confidence levels* over time, while the *surrogates* did not. A comparison among trends in mean *confidence levels* showed that *Positive* and *No* trends were associated with more *information gathering* activity. *Surrogates* that did less *information gathering* were more likely to show a decrease in their confidence levels. As observed, for one of the participants (Participant 2), extensive *information gathering* over repeated sessions can also lead to a later positive change in the trend of mean *confidence levels* (see Figure 4.6).

Some suggestions to improve the WRESTORE tool, which may also be of interest to other researchers developing EDSS are:

- 1) decrease the time that user has to expend for giving feedbacks, particularly reducing the number of sessions they need to go through; overall task times could be a good indicator to determine how long user interactions should last,
- 2) motivate the use of the areas of interest related to *information gathering* to increase the confidence levels of the users. This motivation could be achieved through interface development that emphasizes the exploration of areas where users have the opportunity to gather more information via menus, graphs, maps, and improved data visualizations to allow better comparisons among design alternatives, and
- 3) provide a final “summary” session that recapitulates the findings and designs of desirable alternatives found by the users.



## 4.8 Acknowledgement

We would like to acknowledge the funding agency - National Science Foundation (Award ID #1014693 and #1332385). We would also like to thank all of our collaborators from the different agencies and institutions: Empower Results LLC team for facilitating user testing, Dr. Snehasis Mukhopadhyay and Mr. Vidya B. Singh for computational support, and all workshop participants.

## 4.9 References

- Babbar-Sebens, M., and Minsker, B.S. (2011), Interactive Genetic Algorithm with Mixed Initiative Interaction for multi-criteria ground water monitoring design, *Applied Soft Comp.*, 12, 182-195.
- Babbar-Sebens, M., Barr, R.C., Tedesco, L. P., and Anderson, M. (2013), Spatial identification and optimization of upland wetlands in agricultural watersheds, *Ecol. Eng.*, 52, 130-142.
- Babbar-Sebens, M., Mukhopadhyay, S., Singh, V. B., and Piemonti, A. D., (2015), “WRESTORE- A web-based planning software for participatory optimization of conservation practices in watersheds,” *Environ Mod and Soft.*, 69, 111-127.
- Burkey, J. (2006), *A non-parametric monotonic trend test computing Mann-Kendall Tau, Tau-b, and Sens Slope written in Mathworks-MATLAB implemented using matrix rotations.* King County, Department of Natural Resources and Parks, Science and Technical Services section. Seattle, Washington. USA. [Online]. Available: <http://www.mathworks.com/matlabcentral/fileexchange/authors/23983>
- Estes, W.K. (1994), *Classification and Cognition*, New York, NY: Oxford University Press-Clarendon Press.
- Haklay, M., and Tobon, C., (2003), Usability evaluation and PPGIS: towards a user-centred design approach, *Int. J. Geo. Info. Sci.*, 17(6), 577-592.
- Helsel, D.R., and Hirsch, R.M., (2002), Statistical Methods in Water Resources Techniques of Water Resources Investigations, Book 4, chapter A3. U.S. Geological Survey. 522 pages.
- Jaber, M.Y. (2011), *Learning curves: theory, models, and applications*, Boca Raton, FL: CRC Press.
- Jankowski, P., Robischon, S., Tuthill, D., Nyerges, T., and Ramsey, K. (2006), Design considerations and evaluation of a collaborative spatio-temporal decision support system, *Transac. in GIS*, 10(3), 335-354.
- Mazerolle, M.J., (2004), *Mouvements et reproduction des amphibiens en tourbières perturbées*, Ph.D. dissertation, Dept. Forest. And Geomat., Univ. du Quebec, Quebec, Montreal, Canada.
- Newell, A., and Rosenbloom, P.S. (1981), Mechanisms of skill acquisition and the law of practice, in *Cognitive skills and their acquisition*, Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., pp. 1-51.

- Nyerges, T., Jankowski, P., Tuthill, D., and Ramsey, K. (2006), Collaborative water resource decision support: results of a field experiment, *Ann. Assoc. Amer. Geograph.*, 96(4), 699-725.
- Piemonti, A. D., Babbar-Sebens, M., and Luzar, E. J. (2013), Optimizing conservation practices in watersheds: Do community preferences matter? *Water Resour. Res.*, 49(10), 6425-6449.
- Slocum, T.A., Cliburn, D.C., Feddeman, J.J., and Miller, J.R. (2003), Evaluation the usability of a tool for visualizing the uncertainty of the future global water balance, *Cart. And Geo. Info. Sci.*, 30(4), 299-317.
- Tedesco, L.P., Pascual, D.L., Shrake, L.K., Hall, R.E., Casey, L.R., Vidon, P.G.F., Hernly, F.V., Salazar, K.A., Barr, R.C., Ulmer, J., Pershing, D., 2005. Eagle Creek Watershed Management Plan: An Integrated Approach to Improved Water Quality. Eagle Creek Watershed Alliance, CEES Publication 2005-07, IUPUI, Indianapolis, 182p.
- Tullis, T., and Albert, B., *Measuring the User Experience: collecting, analyzing and presenting usability metrics*. Waltham, MA: Elsevier Inc, 2013.
- Yelle, L.E. (1979), The learning curve: historical review and comprehensive survey, *Decision Sciences*, 10(2), 302-328.

**CHAPTER 5. Participatory design of distributed conservation practices in a watershed: An examination of relationship between user behavior and watershed-scale plans generated via Interactive Optimization**

**Adriana Piemonti<sup>1</sup>, Meghna Babbar-Sebens<sup>1</sup>, Snehasis Mukhopadhyay<sup>2</sup>, and Austin Kleinberg<sup>1</sup>**

<sup>1</sup> School of Civil and Construction Engineering, Oregon State University, Corvallis, Oregon, USA.

<sup>2</sup> Department of Computer and Information Sciences, Indiana University-Purdue University Indianapolis, Indianapolis, Indiana, USA.

To be submitted to the Water Resources Research Journal.

## 5.1 Abstract

Participatory decision support systems (DSS) aim to provide a framework for increasing participation of stakeholders in the decision making process, and have become increasingly popular in the recent past. However, there is a critical lack of understanding of how stakeholder participation and interaction with such DSS affect the function and performance of underlying computational techniques and models in participatory DSS. This paper presents the results and analyses of observational experiments where multiple participants interacted with a web-based DSS called WRESTORE, for interactive optimization of scenarios for allocating conservation practices in a watershed. In WRESTORE, user interactions and feedback on the graphical user interfaces are used to guide the search process of underlying optimization algorithms. The main goal of this study is to use different metrics (such as, the percentage of desirable design alternatives found via interactive optimization, measure for relative distances between Pareto fronts in the objective space, and similarity measures for identifying patterns in the decision space of design alternatives) for quantifying and evaluating how the interactive optimization algorithm adapts to individual user's participation. Results for different participants (*surrogates* and *stakeholders*) are presented, with the goal to evaluate the tool's ability to capture user preferences under a variety of semi-controlled experimental conditions. Our conclusions show clear differences in how the surrogate test group interacts with the DSS in comparison to the group of stakeholders. Stakeholders interacted more with multiple features of the DSS GUI, and demonstrated a correlated increase in their self-confidence in which designs they liked (or did not like). This work also proves that the variability across surrogated individuals is clearly higher than that shown by the stakeholders. In conclusion, the presence of patterns on the design alternatives may assist with the evaluation of the decision space preferences to have a better understanding of the decision making process of optimal design alternatives. Also, the identification of possible noise in the user's input data, using metrics such as confidence level trends and interactions with the system is very helpful to disregard those users that may disturbing the performance of the optimal search in a collaborative process.

## 5.2 Introduction

Alteration of land uses in watersheds generate consequences and challenges for the entire watershed community, including agency managers and planners. Alterations that degrade the natural systems in a watershed and alterations that restore the watershed lead to two opposing system-scale transformations driven by the same agents: Stakeholders. Hence, engagement of stakeholders in the planning and management of their watersheds is of paramount importance. Participatory modeling and decision support environments (DSS) are needed to enable stakeholders to (a) reflect on their use of available natural resources, and (b) create multiple scenarios of land use alternatives that can potentially improve impaired watersheds, while also being sustainable and acceptable.

The literature provides with several systems that claim to be participatory in nature. Systems such as, Web-based, Water-Budget, Interactive, Modeling Program (WebWMPI) (Matsuura et. al, 2009), Sierra Nevada Adaptive Management Project (SNAMP) (Fry et al., 2015), Environmental Risk Assessment Management (eRAMS) (<https://erams.com/>), and WRESTORE (Babbar-Sebens et al., 2015), among others, aim to adapt management plans in order to include users criteria as part of the design process. In most of this systems the main task is to inform the users about processes and results, giving the opportunity to evaluate complex systems via visually friendly systems. For example, WebWMPI uses the information of weather input to simulate hydrological effects in a region. The capabilities of WebWMPI are limited to water balance, but it has the advantage and facility of a web-based environment. On the other hand, SNAMP serve as a more collaborative site, where information can be shared through documents or data, and where the public participation in concerns is encouraged. However, SNAMP does not offer the assistance of generating optimal plans that may improve or change the stakeholders concern. Kaunda-Bukenya et al. (2012) have focused their efforts in develop tools that will assist the state and federal agencies to develop a case-by-case scenario, providing municipal officials with a system that combines modeling and geographical information systems (GIS), to assist in the spatial data processing required for land changes, land modification, and consequences of land use decision in a case by case scenario. However, they do not specifically evaluate the different

plans before implementing an alternative, but shows the consequences of the land uses changes in a specific region.

On the other hand, WRESTORE is a novel tool that now just support the optimization of management plan for an agricultural watershed, but also a system that encourage the user to interact with the different available options in order to understand the effects as consequences of the changes in a decision.

### 5.3 Research Objective

This study focuses on participatory DSS WRESTORE, which has been developed for the purpose of engaging individuals in the design of conservation practices in a watershed. Specifically, the objective of this study is to investigate relationships between two types of users (i.e., stakeholders versus non-stakeholders), users' interaction with the interface when they inspect proposed alternatives and give feedback (via usability metrics), and patterns in the watershed-scale plans of conservation practices generated by the multi-objective Interactive Genetic Algorithm embedded in WRESTORE. Four research questions related to the research objective have been examined:

1. How efficient is the interactive optimization algorithm in generating design alternatives that are preferred by a user, when users are included within the simulation-optimization loops of the search algorithm? To investigate this question, the percentage of design alternatives in each of the levels of WRESTORE's Likert-type *user rating* scale were calculated. A high percentage value for "I like it" *user rating* scale reflected the ability of the optimization algorithm to identify participant's preferences and generate new alternatives that reflect those preferences. Since the *user rating* calssifications are used as values of an additional objective function in the interactive optimization algorithm, it is expected that the selection pressure from this objective function will guide the optimization algorithm towards regions in decision space that meet user preferences. Hence, if the percentage of design alternatives classified as "I like it" is high, then we could assume that the algorithm was able to determine "user preferred" features (either in the objective or decision space) related to the spatial design of conservation practices.

2. In objective functions space, how different or similar are the design alternatives found by user-driven interactive search experiments in comparison to the design alternatives found by a typical non-interactive search? To examine this question, the similarity measure proposed by Piemonti et al. (2013) was used to estimate the distances between Pareto Fronts in the objective functions space, since it measures both the spread and the distance between the centers of mass of any two fronts. The measure was calculated using values of physical objective functions (e.g., cost, peak flow reductions, nitrate reductions, and sediment reductions) estimated at the scale of the entire watershed.
3. In decision variables space (or decision space), how different or similar are the design alternatives for different types of users? The similarities and differences between design alternatives based on the spatial allocations of conservation practices were examined for multiple users, and for every *user rating*. This analysis aimed to determine if there were distinctive patterns in the spatial distribution of the conservation practices that could help identify what types of decisions were preferred by specific individuals or user groups (e.g., stakeholder group and non-stakeholder/surrogate group).
4. How do *user ratings* during the search experiment affect the performance of the interactive optimization algorithm over time? One of the critical limitations of inserting humans in the search/optimization loop is the issue of human fatigue. With increased workload, the tired user is also at a risk of providing noisy feedback that can detract the interactive optimization algorithm's convergence rate. However, on the other hand, continuous interaction with a search algorithm can provide opportunities for users to learn about their problem via experimentation and reflection. Hence, the dynamic nature of the user behavior and search process was examined using usability metrics and the temporal patterns in objective and decision spaces of alternatives preferred (or, not preferred) by multiple users.

## **5.4 Methodology**

### **5.4.1. Setup of Participatory Design Experiments**

The WRESTORE tool was developed by Babbar-Sebens et al. (2015) to enable communities to engage in participatory design efforts and guide the search algorithm in the

spatial allocation of conservation practices on stakeholders' landscape. The study site that was used to test WRESTORE in this research is Eagle Creek Watershed (ECW), located 10 miles NW of Indianapolis, IN (Babbar-Sebens et al., 2013; Piemonti et al., 2013).

The hydrology and water quality in ECW for baseline conditions and for conditions when conservation practices are implemented on the landscape were simulated using the Soil and Water Assessment Tool 2005 (SWAT 2005) model (Arnold et al., 1998; Neitsch et al., 2005). SWAT uses the topography, land use, soil type and regional weather information to estimate the water routing and the water quality through the watershed on a daily time step. Piemonti et al. (2013) and Chapter 3 give a detailed description of the model construction, calibration, and how the SWAT model outputs were used to calculate four physically based objective functions (Peak Flow Reduction, Sediment Reduction, , Nitrates Reduction, and Economic Costs,) relevant to this study. These objective functions are shown in Table 5.1

**Table 5.1 Objective function formulas used to optimize the design alternatives used by the WRESTORE**

Objective	Function
Peakflow Reduction (PFR)	$PFR = \min[-\max_{i,t}(PF_{i,t,base} - PF_{i,t,alt})]^*$
Sediments Reduction (SR)	$SR = \min\{-\sum_{i=1}^N[\sum_{t=1st\ day}^{last\ day} \frac{T2}{T1}(Sout_{i,t,base} - Sout_{i,t,alt})]\}^{**}$
Nitrates Reduction (NR)	$NR = \min\{-\sum_{i=1}^N[\sum_{t=1st\ day}^{last\ day} \frac{T2}{T1}(Nout_{i,t,base} - Nout_{i,t,alt})]\}^+$
Economic Cost (EC)	$EC = \min[\sum_{i=1}^N NPV_i]^{++}$

\* $i$  = SB ID,  $t$  = day,  $PF$  = peakflow,  $base$  = result of baseline of calibrated model,  $alt$  = result of alternative' simulation

\*\* $N$  = total number of SBs,  $T1$  = initial year of simulation,  $T2$  = final year of simulation,  $Sout$  = sediment production

+ $Nout$  = nitrate production

++ $NPV$  = Net of Present Value (Piemonti et al., 2013; Babbar-Sebens et al., 2015)

ECW was divided into 130 different sub-basins (SBs) to simulate the local implementation of a set of conservation practices. Currently, WRESTORE is capable of generating design alternatives for seven different BMPs (strip cropping, crop rotation, cover crops, filter strips, grassed waterways, no-till practices, and wetlands) in all the sub-basins considered for allocation. In each of the targeted sub-basins, decisions for implementing the BMPs are modeled as binary variables or as a real variables (Piemonti et



al. 2013). For this study, the researchers focused only on two practices Cover Crops (CC) and Filter Strips (FS), which represented binary and real decision variables respectively. Table 5.2 explains how the search algorithm's decision variables were converted into parameters relevant to the practice.

**Table 5.2 Changes made on the SWAT model to simulate conservation practices**

Practice	SWAT Variable Modified	File	Range	Installation																																
Filter Strips (implemented in a sub-basin with the decision variable value FILTERW)	FILTERW	.mgt	0-5 m	A typical installation requires a 19 ha field and a 37 m length																																
Cover Crops (implemented in a sub-basin when decision variable has a value 1)	Operation Schedule	.mgt		<p>An example of Corn-Winter Wheat in one year.</p> <p>This operation changes at a HRU level for Corn and Soybeans</p> <table border="1"> <thead> <tr> <th>Year</th> <th>HU*</th> <th>Operation</th> <th>Kg/ha</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>0.28</td> <td>Harvest and Killing</td> <td></td> </tr> <tr> <td>1</td> <td>0.1</td> <td>Pesticide application</td> <td>1.12</td> </tr> <tr> <td>1</td> <td>0.12</td> <td>Plant Corn</td> <td></td> </tr> <tr> <td>1</td> <td>0.3</td> <td>Fertilizer application</td> <td>200.00</td> </tr> <tr> <td>1</td> <td>1.5</td> <td>Harvest and Killing</td> <td></td> </tr> <tr> <td>1</td> <td>0.997</td> <td>Generic Fall Tillage</td> <td></td> </tr> <tr> <td>1</td> <td>0.998</td> <td>Plant Winter Wheat</td> <td></td> </tr> </tbody> </table>	Year	HU*	Operation	Kg/ha	1	0.28	Harvest and Killing		1	0.1	Pesticide application	1.12	1	0.12	Plant Corn		1	0.3	Fertilizer application	200.00	1	1.5	Harvest and Killing		1	0.997	Generic Fall Tillage		1	0.998	Plant Winter Wheat	
Year	HU*	Operation	Kg/ha																																	
1	0.28	Harvest and Killing																																		
1	0.1	Pesticide application	1.12																																	
1	0.12	Plant Corn																																		
1	0.3	Fertilizer application	200.00																																	
1	1.5	Harvest and Killing																																		
1	0.997	Generic Fall Tillage																																		
1	0.998	Plant Winter Wheat																																		

In WRESTORE, besides the four cost/benefit objective functions, a *user rating* objective function is also used for the search process of the interactive optimization/search algorithm. The underlying search algorithm used by the web-based participatory decision support system WRESTORE is based on IGAMII (Interactive Genetic Algorithm with Mixed-Initiative Interaction) (Babbar-Sebens and Minsker, 2012). IGAMII includes alternating sessions of search and introspection. The values for the *user rating* function is decided by stakeholders who are engaged in the search process. The values are based on a Likert-type scale – “I like it” ( $R_3$ ), “Neutral” ( $R_2$ ), and “I do not like it” ( $R_1$ ) – that users can utilize to indicate their preference for an alternative. At every iteration of the search process (see Babbar-Sebens et al., 2015 for details), the human visualizes the design alternatives on a graphical user interface, learns about the decision space and the objective space performance, and then provides a value for *user rating* to indicate the degree of his/her satisfaction with the design alternative.

In this study, twenty participants were asked to interact with the tool to test WRESTORE's capability for finding solutions based on their preferences. The participants

were divided into three different groups according to their affiliation with the watershed (i.e. *Surrogates* (non-stakeholders) and *Stakeholders*) and based on the SWAT model parameters (i.e. *Model A* and *Model B*). In the following paragraphs, these groups will be identified as *Model A-Surrogates*, *Model B-Surrogates*, and *Model B-Stakeholders*. The *Model A-Surrogates* and *Model B-Surrogates* groups contained seven participants (four females and three males for both groups). The *Surrogates* participants were volunteers with science and engineering backgrounds, who were not directly involved in the watershed but are useful representatives of potential participants in a community who may be only cursorily interested in the decisions. The group *Model B-Stakeholders* consisted of six participants (five males, one female) associated with the ECW through land ownership or work. Many of them belonged to federal agencies, state agencies, and non-governmental organizations who are promoting the implementation of conservation practices in the area. Please note that because of limited resources and logistical constraints we were not able to conduct a *Model A-Stakeholders* experiment. However, the findings of this study are not expected to be affected significantly by the absence of data from such an experiment.

The *Model B* was a SWAT model of the watershed created by artificially enhancing the flow and water quality benefits predicted by the calibrated SWAT model (i.e. *Model A-Surrogates*). Enhancement in benefits was accomplished by activating a few wetlands in the watershed. This change considerably increased the benefits of peakflow reduction, sediment reduction, and nitrate reduction, when the conservation practices under consideration – i.e. Cover Crops and Filter Strips – were allocated in the SBs by the search optimization algorithm. However, the users were not informed of this enhancement and were asked to evaluate the design alternatives on the basis of the assumption that only Cover Crops and Filter Strips were being implemented in the watershed. The rationale for this artificial enhancement of benefits was to examine if an improvement in benefits would change the participants' preferences for practices, design alternatives, or spatial locations on the interface.

#### **5.4.2. Evaluation Metrics**

In this study, the results from the user experiments were analyzed using two different perspectives – one that examined the overall outcome of the search at the end of the

interactive optimization algorithm (i.e., *End of Search* or EoS), and another that examined the convergence (or, lack of convergence) in search outcomes through time as the interactive optimization algorithm progressed through its iterations (i.e., *During Search* or DS).

For the DS perspective, the search outcomes were assessed at specific time intervals called *Epochs*. Three Epochs were used in this study – the first one that examined the data on initial design alternatives the were shown to user in the first introspection (*I1*) session, the second Epoch that examined all design alternatives from human-guided search sessions (*HS1* to *HS6*) after *I1*, and, finally, the third Epoch that examined all design alternatives from human-guided search sessions (*HS1* to *HS6*) after second introspection (*I2*) session.

For analyzing results relevant to the EoS and DS perspectives, assessment metrics was set up based on 1) participant's *user ratings*, 2) similarities between discovered design alternatives in the objective space, 3) and the similarities between discovered design alternatives in the decision space. Below are descriptions on how these metrics were set up.

#### **5.4.2.1. Metrics based on user ratings**

##### **A. EoS: Percentage of design alternatives per user rating**

To evaluate and compare the final number of design alternatives with a specific *user rating* at the end of the search experiment (EoS), we used a metric based on the percentage of design alternatives in each *user rating* classification. This metric,  $PRate_{ik}$ , was calculated using Equation 5.1.

$$PRate_{ik} = \frac{X_{ik}}{TD} * 100 \quad (5.1)$$

where  $PRate_{ik}$  is the percentage of design alternatives of  $i^{th}$  *user rating* ( $R_i$ ), for the  $k^{th}$  user,  $X_{ik}$  is the total number of designs in *user rating*  $R_i$  for the same  $k^{th}$  user and  $TD$  is the total number of designs presented to every participant. As explained earlier, all the participants were shown at least 260 design alternatives that were included in the first two cycles of the

*Introspection – Human-guided Search* sessions (i.e., I1 → HS1-HS6 → I2, → HS1-HS6). Therefore, TD had a value of 260.

To examine the overall outcome, for users in each of the three groups (i.e., *Model A-Surrogates*, *Model B-Surrogates*, and *Model B-Stakeholders*), a global assessment of this metric ( $GPRate_i$ ) was estimated using the average of the  $PRate_{ik}$  in each group.

$$GPRate_i = \frac{\sum_{k=1}^N PRate_{ik}}{N} \quad (5.2)$$

where  $N$  is the total number of users in the group.

### ***B. DS: Percentage of design alternatives per user rating***

The percentage of design alternatives within a *user rating* was also calculated through time, for every Epoch that occurred as the experiment progressed. This was calculated for each participant in the following manner:

$$PRate_{ikt} = \frac{X_{ikt}}{TD_t} * 100 \quad (5.3)$$

where,  $PRate_{ikt}$  is the percentage of design alternatives of  $i^{th}$  *user rating* ( $R_i$ ), for the  $k^{th}$  user of the group, and at  $t^{th}$  Epoch.  $X_{ikt}$  is the total number of designs in *user rating*  $R_i$  for the same  $k^{th}$  user in the  $t^{th}$  Epoch, and  $TD_t$  is the total number of designs presented to the  $k^{th}$  user in the  $t^{th}$  Epoch.

To examine the overall outcome, for users in each of the three groups (i.e., *Model A-Surrogates*, *Model B-Surrogates*, and *Model B-Stakeholders*), a global assessment ( $GPRate_{it}$ ) was estimated using the average of the  $PRate_{ikt}$  in each group.

$$GPRate_{it} = \frac{\sum_{k=1}^N PRate_{ikt}}{N} \quad (5.4)$$

where  $N$  is the number of users in the group.

#### ***5.4.2.2. Metrics based on Similarities in Objective Space***

The similarities and dissimilarities between design alternatives in objective space were evaluated using a metric proposed by Piemonti et al., (2013) for the overall distance

between Pareto Fronts. This distance metric was first estimated for each of the participants, and then an average of the metric values across the participants was calculated to summarize the results for participants in each groups.

For every participant, in order to assess the impact of interactive optimization on search results, the distance in the objective space between the design alternatives found via the participant's interactive search experiment and the design alternatives found via a non-interactive search was estimated. Note that even though the participant was not involved in the non-interactive search process, he/she had the opportunity to review and rate twenty of the non-interactive search's final non-dominated design alternatives. The twenty designs were shown to the participant in the first introspection session *II*. All design alternatives were first separated into three separate groups based on their user rating  $R_i$  (i.e. "I don't like it", "Neutral", and "I like it"), and then the distance metric was calculated for alternatives in each group. Two physical objectives at a time (e.g., cost and peak flow reduction, or cost and nitrates reduction) were selected to calculate the distance. This metric is based on an average relative euclidean distance (Equation 5.5) and compares the position of every  $j^{th}$  design alternative from the non-interactive Pareto Front, with each  $l^{th}$  design alternative that has  $i^{th}$  user rating and is from the Pareto Front found via interactive search.

$$DR_{ik}(A, B) = \frac{\sum_{j=1}^J \sum_{l=1}^{L_{ik}} \sqrt{(A_j - A_{ilk})^2 + (B_j - B_{ilk})^2}}{J * L_{ik}} \quad (5.5)$$

Then,  $DR_{ik}(A, B)$  represents the distance between two objective functions ( $A$  is the performance of peakflow reduction, sediment reduction or nitrate reduction, and  $B$  is the economic cost);  $J$  is the total number of alternatives in the non-interactive Pareto Front shown to the user (design alternatives in Introspection *II*), and  $L_{ik}$  is the total number of design alternatives with  $i^{th}$  user rating for the  $k^{th}$  participant.

Once the distance metric was calculated for each participant, a representative group distance metric was calculated for the multiple participants belonging to a group (e.g., *stakeholder* group or *surrogates* group). Equation 5.6 shows how this group metric,  $GDR_i$ ,

was calculated as the average of all the participants' distances,  $DR_{ik}$ , in each of the *user ratings*.

$$GDR_i(A, B) = \frac{\sum_{k=1}^N DR_{ik}(A, B)}{N} \quad (5.6)$$

where  $N$  is the total number of participants in the group.

Similarly, distance was also calculated using Equation (5.7) to examine changes in distance between Pareto Fronts through time, for every Epoch that occurred as the experiment progressed. . This metric is based on an average relative euclidean distance and compares the position of every  $j^{th}$  design alternative from the non-interactive Pareto Front, with each  $l^{th}$  design alternative that has  $i^{th}$  *user rating* and is from the Pareto Front found via interactive search

$$DR_{ikt}(A, B) = \frac{\sum_{j=1}^J \sum_{l=1}^{L_{ikt}} \sqrt{(A_j - A_{ilkt})^2 + (B_j - B_{ilkt})^2}}{J * L_{ikt}} \quad (5.7)$$

then,  $DR_{ikt}(A, B)$  represents the distance between two objective functions ( $A$  is the performance of peakflow reduction, sediment reduction or nitrate reduction, and  $B$  is cost);  $J$  is the total number of alternatives in the non-interactive Pareto Front shown to the user (design alternatives in Introspection *II*) and  $L_{ikt}$  is the total number of designs with  $i^{th}$  *user rating* for  $k^{th}$  participant at  $t^{th}$  Epoch.

To examine changes in group distance over time, a group distance metric ( $GDR_{it}(A, B)$ ) at each Epoch ( $t$ ) was also calculated. Equation 5.8 shows how this temporal distance,  $DR_{ikt}$ , was estimated for each *user rating* ( $i$ ) and for each participant ( $k$ )

$$GDR_{it}(A, B) = \frac{\sum_{k=1}^N DR_{ikt}}{N} \quad (5.8)$$

where  $N$  is the total number of participants in the group.

#### 5.4.2.3. Metrics based on similarities in Decision Space

Similarities and dissimilarities between design alternatives in the decision space were estimated in order to identify what types of solutions were preferred (or not preferred) by participants. Metrics for similarities were estimated differently based on whether a decision variable representing the BMP (or, conservation practice) was a Binary variable or a Real variable. The Binary variables represent whether a BMP is implemented (when the variable value is 1) or not implemented (when the variable value is 0) in a sub-basin. Whereas, the Real Variables specify a design parameter value related to the BMP (e.g. decisions for the BMP Filter Strips were represented as Filter Strip Widths, which are real numbers that can lie between a minimum and a maximum real number value). Piemonti et al., (2013) and Table 5.3 give additional information on how the binary and real decision variables are incorporated in to the optimization algorithm, and how they are used to simulate a practice in the watershed simulation model.

To estimate similarities in decision variable values across multiple design alternatives, we first classified all the design alternatives rated by the participants according to their *user rating* ( $i$ ). Next, a set of summary metrics were estimated for each of the genes ( $g$ ) in the chromosome used by the genetic algorithm, IGAMII, to represent the design alternative (Piemonti et al., 2013; Babbar-Sebens et al., 2015). The following general approach was used to estimate these summary metrics:

1. For genes representing binary decision variables, we calculated a “probability of implementation” of the BMP in a SB,
2. For genes representing real decision variables, we calculated the mode (or most repeated design variable value in a gene) and calculate the percentage of design alternatives with this value in a specific gene.

Table 5.3 shows the compilation of the different specific metrics and their equations, which were used to find the patterns in design alternatives with the same *user rating*. Following are the descriptions of these metrics:

- $Prob_{gRik}$  is the probability of BMP implemented in the  $g^{th}$  gene of design alternatives that have the same *user rating*  $R_i$  and have been evaluated by the same participant ( $k$ ).

- $ccimp_{gilk}$  is the decision variable value (1/0) in gene  $g$ , which indicates whether a BMP has been implemented in the SB represented by that gene. Additionally,  $i$  is the specific *user rating* given to the  $l^{th}$  design alternative by the participant  $k$ .
- $L_{ik}$  is the total number of design alternatives that have the same *user rating*  $i$  given by the  $k^{th}$  participant.
- For every unique  $f$  value of  $FSW$  (filter strip width) in a chromosome of  $G_i$  genes we calculated the mode (most repeated  $FSW$ ) using:

$$FSW_f = \sum_{g=1}^{G_i} \left\{ \begin{array}{l} 1 \rightarrow \text{if } FSW_g = FSW \\ 0 \rightarrow \text{otherwise} \end{array} \right\} \quad (5.9)$$

$$Mode = \max(FSW_f) \quad (5.10)$$

- $POM_{gRik}$  is the percentage of design alternatives (that have same *user rating*  $R_i$  given by participant  $k$ ) with decision variable values in gene  $g$  equal to the mode of all values in gene  $g$ .
- $DEM_{gRik}$  is the number of design alternatives (that have same *user rating*  $R_i$  given by participant  $k$ ) with decision variable values in gene  $g$  equal to the mode of all values in gene  $g$ .
- $GProb_{gRi}$  is the average probability of BMP implemented in the  $g^{th}$  gene across all participants in the same group, and for the set of design alternatives with the same *user rating*  $R_i$ .
- As previously stated,  $N$  is the total number of participants in the group.
- $GPOM_{gRi}$  is average value of percentage of design alternatives with design parameter (or, decision variable value) in gene  $g$  equal to the mode across all participants in the same group, and for the set of design alternatives with the same *user rating*  $R_i$ .

For *During the Search* (DS) analysis, the same definition for these variables were considered, but each variable was calculated for a specific  $l^{th}$  Epoch.



**Table 5.3 Equations used in the evaluation of the decision space**  
**Variable type in  $g^{th}$  gene**      **Metric Equation**

		Variable type in $g^{th}$ gene	Metric Equation	
Individual participant	EoS	Binary	$Prob_{gRik} = \frac{\sum_{l=1}^{L_{ik}} ccimp_{gilik}}{L_{ik}}$	(5.11)
		Real	$POM_{gRik} = \frac{DEM_{gRik}}{L_{ik}} * 100$	(5.12)
	DS	Binary	$Prob_{gRikt} = \frac{\sum_{l=1}^{L_{ikt}} ccimp_{gilikt}}{L_{ikt}}$	(5.13)
		Real	$POM_{gRikt} = \frac{DEM_{gRikt}}{L_{ikt}} * 100$	(5.14)
Groups	EoS	Binary	$GProb_{gRi} = \frac{\sum_{k=1}^N Prob_{gRik}}{N}$	(5.15)
			$Stdev_{gRi} = \sqrt{\frac{\sum_{k=1}^N (Prob_{gRik} - GProb_{gRi})^2}{(N - 1)}}$	(5.16)
		Real	$GPOM_{gRi} = \frac{\sum_{k=1}^N POM_{gRik}}{N}$	(5.17)
			$VarPOM_{gRi} = \frac{\sum_{k=1}^N (GPOM_{gRik} - GPOM_{gRi})^2}{(N - 1)}$	(5.18)
	DS	Binary	$GProb_{gRit} = \frac{\sum_{k=1}^N Prob_{gRikt}}{N}$	(5.19)
			$Stdev_{gRit} = \sqrt{\frac{\sum_{k=1}^N (Prob_{gRikt} - GProb_{gRit})^2}{(N - 1)}}$	(5.20)
		Real	$GPOM_{gRit} = \frac{\sum_{k=1}^N POM_{gRikt}}{N}$	(5.21)
			$VarPOM_{gRit} = \frac{\sum_{k=1}^N (GPOM_{gRikt} - GPOM_{gRit})^2}{(N - 1)}$	(5.22)

## 5.5 Results and discussion

As mentioned earlier, the user search experiments conducted in this study included participants from two types of groups (e.g. stakeholders and non-stakeholders (also called *surrogates*)), and included two types of watershed models (*Model A* and *Model B*), in order to examine:

- potential differences in participant biases and preferences, and their user behavior on the GUI, and
- potential differences in users' overall evaluation of how acceptable the proposed design alternatives are to them, when they also have to consider the performance-based physical objective functions estimated by different watershed models into their decision making.

Additionally, as also mentioned earlier, two different perspectives (i.e. EoS and DS) were used to examine the effectiveness of the interactive search process that included the above

scenarios of participants and watershed models. Hence, this section has been divided into multiple sub-sections in order to examine the results relevant to multiple types of experimental setups, participants, and the research objectives (mentioned in Section 5.3). First, results related to the overall search outcomes at the end of every user experiment (i.e. EoS) are discussed in section 5.5.1, followed by a discussion of results that were obtained at various intermittent times of the search process (i.e. DS) in Section 5.5.2. Within each of the sections 5.5.1 and 5.5.2, results are first evaluated using the metric for *User Ratings* (i.e., Sections 5.5.1.1 and 5.5.2.1), followed by discussions on similarities and dissimilarities between results in Objective Space (i.e., Sections 5.5.1.2 and 5.5.2.2), and discussions on similarities and dissimilarities between results in Decision Space (i.e., Sections 5.5.1.3 and 5.5.2.3). In each of the sub-sections (i.e., 5.5.1.1-5.5.2.3), the discussion first focuses on how individual users working with a specific watershed model (i.e. *Model A* or *Model B*) found similar or dissimilar results, followed by summarizing discussion on the group results. Comparison of results found by individual participants, and then those found by the groups of participants, offer the following benefits:

- a) Evaluation of results of individual users has the potential to help us understand how effective interactive optimization can be in generating personalized design alternatives in a human-guided search process for different types of user preferences and behaviors.
- b) Evaluation of users' results that are grouped based on the type of participants or watershed model type can help identify agreements between participants and variability across participants that belong to a certain group. This further provides insight into whether the personalized interactive search can also assist in the generation of desirable alternatives that satisfy the requirements of most of the community members within a group.

### **5.5.1. EoS assessment of user searches experiments**

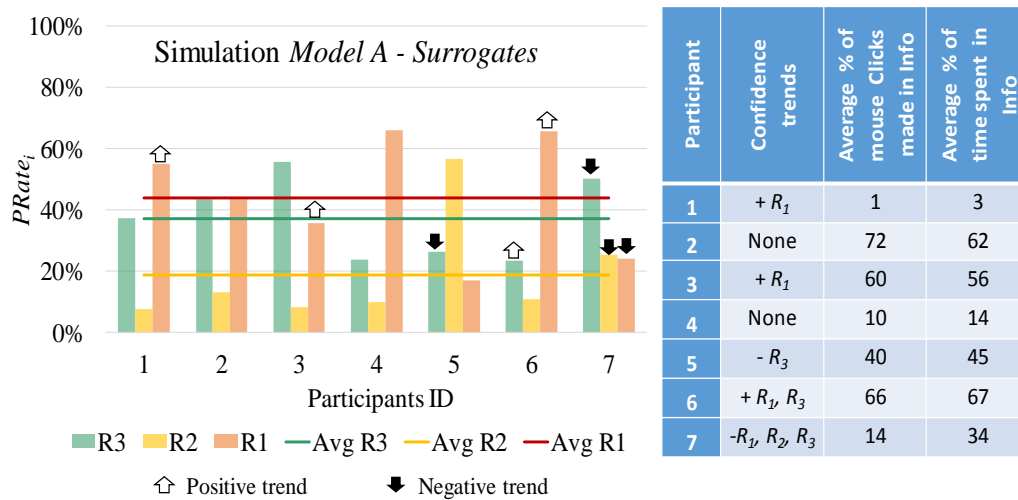
#### ***5.5.1.1. Percentage of design alternatives per user ratings***

##### **A. Assessment of Individual Participants**

The percentage of *user ratings* provide insight into how successful the interactive optimization algorithm was in identifying desirable alternatives (e.g., those with user rating  $R_3$ ) for different participants and when different watershed models were used to evaluate cost-benefits of the design alternatives. The percentages of *user ratings* (i.e.,  $PRate_1$  to  $PRate_3$ ) for each of the participants were also associated with the trends in average confidence level (calculated in previous chapter 4) for each of the user rating classifications. Note that the trends in confidence levels provide an insight into the participant's learning process and a self-assessment of the accuracy of his/her own evaluation of alternatives through time. Figure 5.1 shows the percentage of user ratings ( $PRate_i$ ) at the end of the search experiment, and for each  $i^{th}$  rating ( $R_i$ ) ranked by *surrogate* participants (bars) who were working with watershed *Model A*. The horizontal red, yellow, and green lines are the average values for  $PRate_1$ ,  $PRate_2$ ,  $PRate_3$  bars across all participants, respectively, and are discussed in next Section B on group statistics. The confidence level trends are represented by white arrows (positive trend) and black arrows (negative trend) in this figure.

By the end of the experiments, shown in Figure, we noticed that four (i.e. 57%) out of seven of these *surrogate* participants have a higher value of  $PRate_1$  (for  $R_1$  or “I don't like it” alternatives) than  $PRate_3$ . This indicates that majority of these participants did not like most of the design alternatives produced by the interactive search algorithms. However, only two (i.e. 50%) of these four *surrogate* participants indicated an increase (white arrows) in the average confidence levels of  $R_1$  design alternatives over time (Participant 1 and Participant 6 – Figure 5.1), suggesting that only half of these four individuals by the end of the experiment were increasingly self-confident about the design alternatives that they did not like. Moreover, when data from the participant's interaction with the GUI was considered (Piemonti et al., (in review) and Chapter 3), we observed that the Participant 6, unlike Participant 1, spent a considerable amount of time and made a large number of mouse clicks in tasks involving gathering information on the GUI. Therefore, we could assume that his/her conclusions about giving a *user rating*  $R_1$  to design alternatives and self-confidence in his/her own assessment of designs were well supported by a learning process that considered substantial information on how the design will benefit the

watershed, and meet their personal constraints and interests. Also, notice that for “I like it” alternatives, Participant 3 and Participant 7 have the highest  $PRate_3$  than  $PRate_2$  and  $PRate_1$ . However, *user rating* provided by Participant 3 may be more reliable due to no changes in confidence levels through time for  $PRate_3$ , increasing confidence levels for  $PRate_1$ , and the high percentage of his/her interaction with the tool in the information gathering areas (67% of time and 66% of mouse clicks in gathering information). Participant 7, on the other hand, had significantly lower interaction with the GUI (34% of time and 14% of mouse clicks) in information gathering areas of the GUI, and also experienced a decrease in all of his/her confidence levels through time. This indicates that even though Participant 7 liked most of his/her designs, the user behavior, and self-confidence in his/her feedback do not suggest that the *user ratings* may be a reliable portrayal of her/his assessment of designs alternatives.

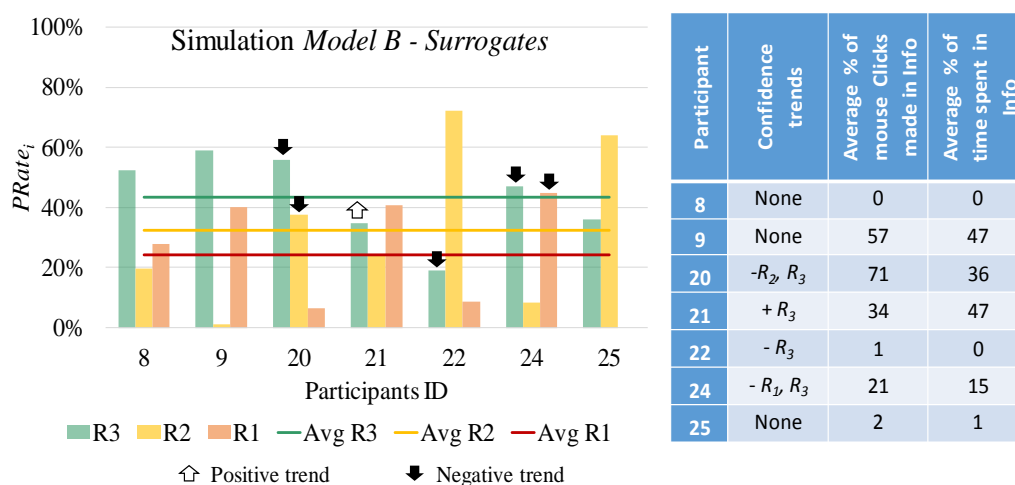


**Figure 5.1** Percentage of designs classified by the participants at each level of the user rating *Model A-Surrogates*

Figures 5.2 and 5.3 show the percentage of *user ratings* for the *Model B-Surrogates* and *Model B-Stakeholder* experiments respectively. Contrary to *Model A-Surrogates* (Figure 5.1), user experiments with *Model B* (i.e. *Surrogates* and *Stakeholder*) generated a higher percentage of participants that found a high proportion of design alternatives that they liked. Specifically, four out of seven (i.e. 57%) of *Model B-Surrogates* participants and four out of six (66%) of the *Model B-Stakeholders* participants had values for  $PRate_3$  higher than  $PRate_1$  and  $PRate_2$ . These results indicate that participants (stakeholders and non-

stakeholders) seemed to be a lot more satisfied with their design alternatives when the watershed simulation model over-predicted the performance of the conservation practices, than participants in *Model A-Surrogates*.

However, on closer observation of the level confidence trends, *Model B-Surrogates* had only one participant (Participant 21) whose confidence level increased over time for design alternatives rated  $R_3$ , even though his/her  $PRate_3$  was smaller than  $PRate_1$ . This participant has a moderate amount of interaction (the average percentage time in information gathering was 34% and the average percentage of mouse click in information gathering was 47%) with the GUI. Participants 20, 22 and 24, on the other hand, had  $PRate_3$  higher than  $PRate_1$  (Figure 5.2) but still presented a decrease in their confidence level trends over time. The interaction of Participants 20, 22, and 24 with the GUI also varied from 0 % on average percentage time spent and 1% in average percentage mouse click, for Participant 22 to 34 % on average percentage time spent and 71% in average percentage mouse click, for Participant 20, suggesting that the activities in the areas of interest are not consistent with the expected tendencies of mouse clicks events with time spent in information gathering for positive trends in confidence levels.



**Figure 5.2** Percentage of designs classified by the participants at each level of the user rating for *Model B-Surrogates*

On closer observation of *Model B-Stakeholders*, the majority of these participants (i.e., four out of six participants or ~66%) were found to have increasing confidence levels over time. In particular, Participant 15 had  $PRate_3$  close to 70 %, suggesting that he/she was

satisfied with most of the design alternatives found by him/her. However, unlike other participants, Participant 15 did not spend a lot of time interacting with the GUI to collect information on the design alternatives. This can be attributed to the fact that Participant 15 was interested in the entire watershed and was not focused on a smaller region of interest. Hence, this user did not need to do additional mouse clicks or use drop down menus in order to procure information for the watershed. Also, only one of the participants in this group presented a decrease in the confidence level trends (Participant 16) for design alternatives with *user rating*  $R_2$ . In summary, even though  $PRate_3$  was found to be highest for the majority of the participants in both the two groups using Model B, most of the participants in *Model B-Stakeholders* were found to be increasingly confident in the accuracy of their user ratings throughout the experiment than the participants in *Model B-Surrogates*.

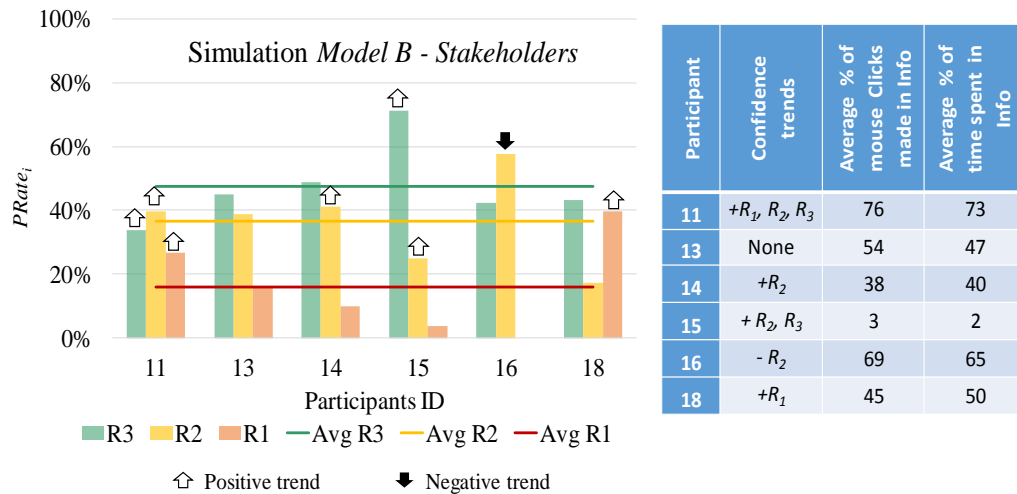


Figure 5.3 Percentage of designs classified by the participants at each level of the user rating for *Model B-Stakeholders*

## B. Overall Group Assessment

As mentioned earlier, the horizontal red, green, and yellow lines in Figures 5.1, 5.2 and 5.3 provide information on the average of the percentage of solutions (for each *user rating*) across all participants in a group. Figure 5.1 shows that on an average  $PRate_1 > PRate_3 > PRate_2$  for the *Model A-Surrogates*, whereas Figures 5.2 and 5.3 show that  $PRate_3 > PRate_2 > PRate_1$  for *Model B-Surrogates* and *Model B-Stakeholders*. This indicates that when participants used the watershed *Model B*, the search algorithm seemed to have a better

performance in capturing their preferences and delivering more design alternatives that the participants were satisfied with.

### 5.5.1.2. Similarities in Objective Space

The physical objective space reflects the range of physical environmental benefits and costs (or, revenues) that different design alternatives would be expected to deliver, as well as the tradeoffs between the non-dominated alternatives. As also mentioned earlier, the similarity between results from multiple user experiments in objective space was assessed by calculating the distance (Equation 5.3) between the set of a non-dominated design alternatives found by every participant and the initial set of optimized non-interactive design alternatives. Design alternatives in each set were separated into three groups based on the user ratings  $R_i$ . Below is a discussion of similarity and dissimilarities in objective space of various design alternatives with *user rating*  $R_i$  in each set, and found by the participants in *Model A-Surrogates*, *Model B-Surrogates*, and *Model B-Stakeholders*.

#### A. Assessment of Individual Participants

Table 5.4 present the results of distance metric calculated for participants in *Model A-Surrogates* group, for each of the user rating  $R_i$ , and using two objectives at a time. It can be observed that for any one of the specific pair of objective functions (e.g., cost vs. *PFR*), the distances (e.g.,  $DR_j$ ) for the same *user rating* (e.g.  $R_j$ ) was similar across all users. Note that this distance metric not only represents how far the center of masses of two sets of design alternatives are from each other, but it also represents the spread of the alternatives around the center of mass. For all the participants in this group, it can be seen that  $DR_1$  was greater than  $DR_3$ , suggesting that designs classified as  $R_3$  (i.e., “I like it”) are closer to the design alternatives in the non-interactive Pareto Front. For some participants - i.e., Participants 1 and 6 -  $DR_3$  was greater than  $DR_2$ . This could be attributed to the fact that for these participants (Participant 1 and Participant 6), the Pareto Front of  $R_2$  design alternatives was less spread than the  $R_3$  design alternatives, leading to a smaller value of  $DR_2$ .

Figure 5.4 shows a common example of the distribution of *user ratings* in the objective space for two of these participants – Participant 1 who was interested in the performance of design alternatives at the scale of the entire watershed (Figure 5.4a), and Participant 6

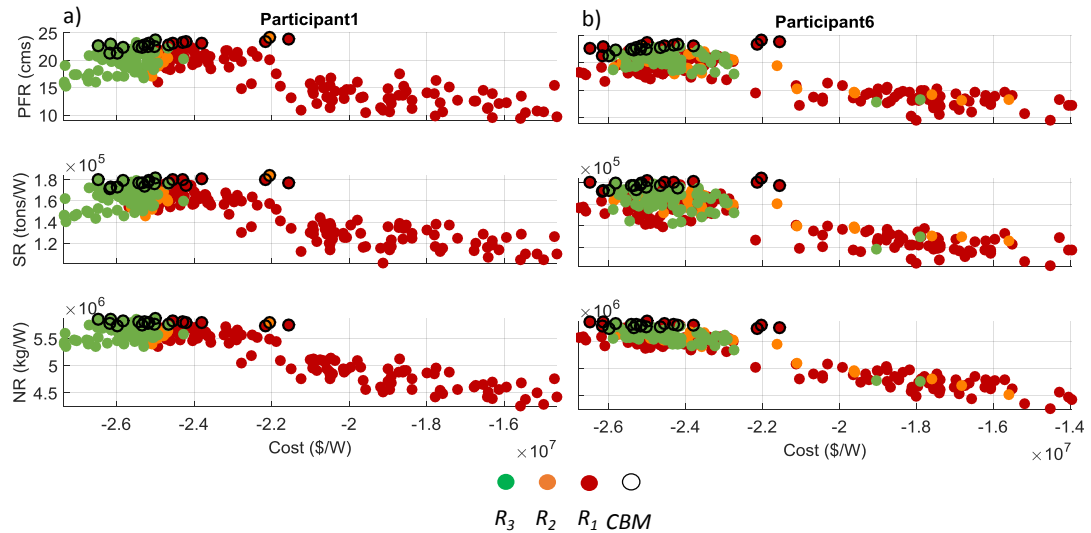
who was interested in the performance in a small set of local sub-basins (SBs) (Figure 5.4b). This figure shows that for Participant 1 there are clear clustered regions where the user found most of her/his preferred and less-preferred design alternatives. In this example, Cost at the watershed scale seems to be the deciding criteria based on which a user decided the acceptability of alternatives. However, this clear distinction in the regions of desirable and less-desirable alternatives does not seem to exist for Participant 6, who was less concerned about watershed scale performance. Even when most of the preferred design alternatives lie on the low Cost (on the left side) region of the objective space, there are multiple design alternatives in the same region that were rated  $R_I$  by the participant.

**Table 5.4 Distance from design alternatives performance between non-interactive and interactive Pareto Fronts for Peakflow Reductions (PFR), Sediments Reduction (SR) and Nitrates Reduction (NR) for *Model A-Surrogates***

PARTICIPANT	Objective 1: COST, Objective 2: PFR			Objective 1: COST, Objective 2: SR			Objective 1: COST, Objective 2: NR		
	$DR_1$	$DR_2$	$DR_3$	$DR_1$	$DR_2$	$DR_3$	$DR_1$	$DR_2$	$DR_3$
1	0.22	0.09	0.12	0.24	0.08	0.11	0.22	0.07	0.09
2	0.25	0.16	0.11	0.27	0.17	0.12	0.23	0.15	0.10
3	0.21	0.14	0.11	0.24	0.15	0.14	0.20	0.14	0.11
4	0.19	1.3E-4	0.11	0.21	1.4E-4	0.11	0.18	1.2E-4	0.10
5	0.24	0.14	0.16	0.27	0.16	0.18	0.22	0.13	0.16
6	0.23	0.08	0.10	0.25	0.10	0.12	0.23	0.08	0.09
7	0.31	0.12	0.12	0.34	0.15	0.15	0.32	0.12	0.12
<b>Average</b>	0.24	0.10	0.12	0.26	0.12	0.13	0.23	0.10	0.11
<b>Standard deviation</b>	0.04	0.05	0.02	0.04	0.06	0.03	0.04	0.05	0.02

Figure 5.4 also shows the set of design alternatives (labeled as CBM or Case Base Memory) that were selected for the initial evaluation in Introspection 1 (I1). These design alternatives were found after an exhaustive non-interactive search was performed using the 4 physical objective functions. The set of the CBM gives the participant a starting point for those solutions known to have an optimal performance.





**Figure 5.4** Pareto front representation of watershed performance for Participant 1 (a) and Participant 6 (b) in *Model A-Surrogates* group. Participant 1 was asked to provide the *user rating* for the design alternative based on the watershed performance, while Participant 6 was asked to provide the *user rating* for the design alternative based on the group of SBs: 103, 105, 106, 121, and 122

Table 5.4 presents the  $DR_i$  for *Model B-Surrogates* and *Model B-Stakeholders*. Notice that the values of these distances are significantly higher than for *Model A-Surrogates*. This is an effect of the artificial enhancement of the peak flow reduction, nitrate reduction, and sediment reduction benefits, as explained in section 5.4.1.

Results of *Model B-Surrogates* showed that 43% of the participants seem to have  $DR_1 > DR_3$ , while 83% of the participants for *Model B-Stakeholders* showed distances values of  $DR_1 > DR_3$ . These results suggest that not all Participants in both groups preferred solutions that have enhanced PFR, SR, and NR values, and more spread out and farthest away from the Pareto Front of non-interactive design alternatives.

Figure 5.5. shows some demonstrative examples of the distribution of *user ratings* in the objective space quantified at the entire watershed scale. Note that two of these participants rated design alternatives based on their performance at the scale of the entire watershed (Figure 5.5a and 5.5c), and two of these rated designs based on the performance of alternatives in a particular subset of local sub-basins (Figure 5.5b and 5.5d). Notice that even when Figure 5.5a and Figure 5.5c are participants who were concerned with optimizing the solutions for the entire watershed, the design alternatives with different *user ratings* are more scattered in the objective space for the Participant 15 in Figure 5.5c. Participant 8 in Figure 5.5a, on the other hand, has well-defined clusters of  $R_1$ ,  $R_2$ , and  $R_3$

design alternatives, similarly to Participant 1 from *Model A-Surrogates* (shown in Figure 5.4). However, on the other hand, note that Participant 8 did not have any clear trend in confidence levels (Figure 5.2), whereas Participant 15 had an increase in self-confidence levels over time for  $R_2$  and  $R_3$  ratings (Figure 5.3). Then, the results of Participant 15's user experiments should be considered more compelling, in spite of the lack of clear clusters in objective space. Two other examples in figure (5.5) Participant 11 and Participant 20 rated the design alternatives based on a particular set of SBs. Their results in the objective space might look similar, however, Participant 11 have an increase of confidence levels for all the trends, while Participant 20 have a negative trend for alternatives with *user ratings*  $R_3$  and  $R_2$ . Even when both participants seem to have a good percentage of mouse click events (76 % and 71% respectively), the discrepancies in the average percentage time spent (73 % and 36% respectively) suggest that the reliability of the classification for Participant 11 may be higher than for the classification of Participant 20.

**Table 5.5 Distance from design alternatives performance between non-interactive and interactive Pareto Fronts for Peakflow Reductions (PFR), Sediments Reduction (SR) and Nitrates Reduction (NR)**

PARTICIPANT	Objective 1: COST, Objective 2: PFR			Objective 1: COST, Objective 2: SR			Objective 1: COST, Objective 2: NR		
	$DR_1$	$DR_2$	$DR_3$	$DR_1$	$DR_2$	$DR_3$	$DR_1$	$DR_2$	$DR_3$
8	0.50	0.41	0.33	0.47	0.41	0.35	0.46	0.41	0.35
9	0.27	0.01	0.21	0.38	0.01	0.32	0.33	0.01	0.28
20	0.63	0.47	0.29	0.58	0.44	0.28	0.57	0.45	0.29
21	0.33	0.37	0.46	0.33	0.37	0.45	0.32	0.36	0.45
22	0.34	0.39	0.41	0.38	0.38	0.39	0.37	0.38	0.40
24	0.40	0.37	0.41	0.39	0.37	0.40	0.39	0.38	0.40
25	0	0.39	0.38	0	0.39	0.36	0	0.38	0.35
<b>Average</b>	0.35	0.34	0.36	0.36	0.34	0.36	0.35	0.34	0.36
<b>Standard deviation</b>	0.20	0.15	0.09	0.18	0.15	0.06	0.18	0.15	0.06
11	0.24	0.13	0.26	0.37	0.19	0.42	0.32	0.17	0.37
13	0.29	0.28	0.18	0.41	0.36	0.26	0.36	0.32	0.23
14	0.24	0.30	0.18	0.36	0.37	0.26	0.31	0.34	0.23
15	0.30	0.11	0.24	0.47	0.16	0.37	0.41	0.13	0.33
16	0	0.23	0.28	0	0.30	0.36	0	0.28	0.33
18	0.45	0.23	0.42	0.44	0.23	0.41	0.45	0.23	0.41
<b>Average</b>	0.25	0.21	0.26	0.34	0.27	0.35	0.31	0.25	0.32
<b>Standard deviation</b>	0.15	0.08	0.09	0.17	0.09	0.07	0.16	0.08	0.07

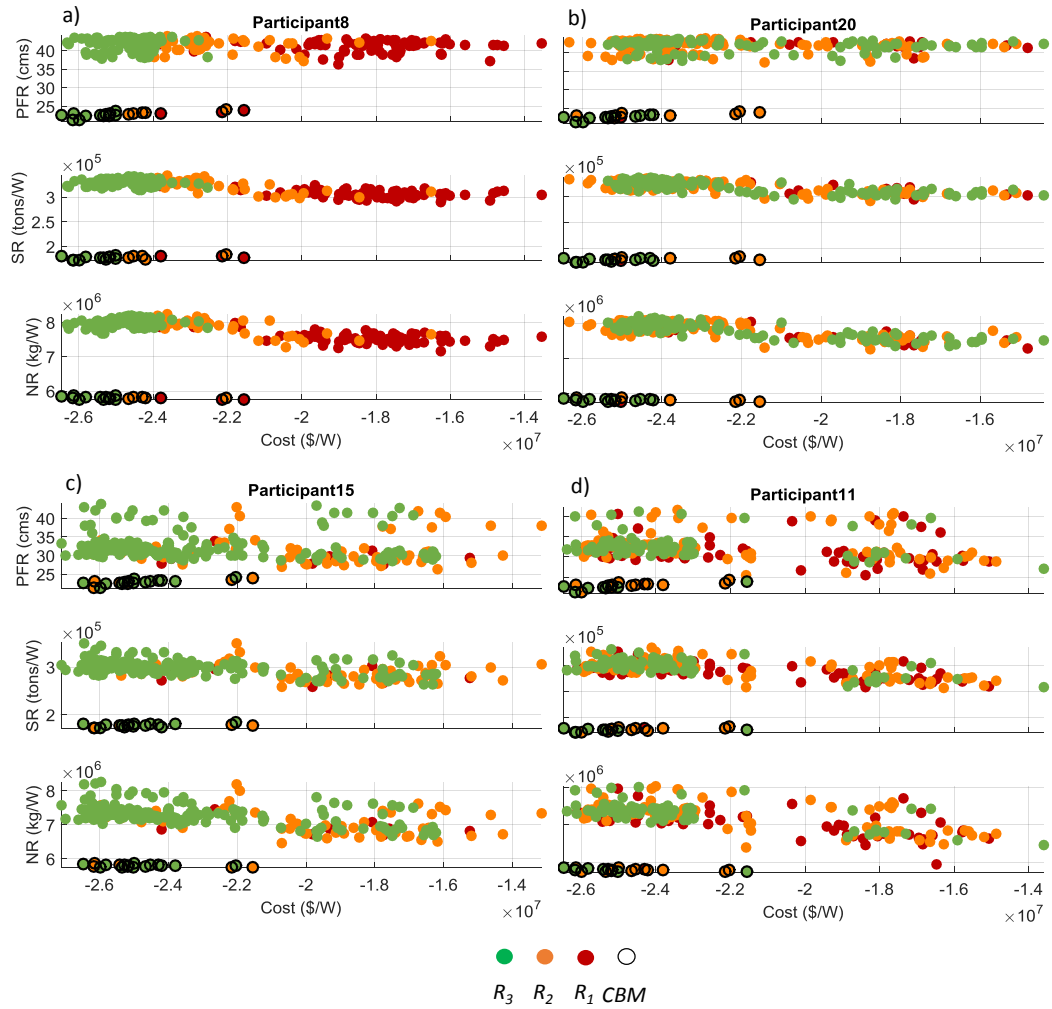


Figure 5.5 Part Pareto front representation of watershed performance for Participant 8 (a) and Participant 20 (b) in *Model B-Surrogates* group. Participant 8 was asked to provide the *user rating* for the design alternative based on the watershed performance, while Participant 20 was asked to provide the user rating for the design alternative based on the group of SBs: 58, 59, 61 and 63. Similarly, the Pareto Fronts for Participants 15 and 11 (c) and (d) respectively) in *Model B-Stakeholders* group are shown. Participant 15 was asked to provide the *user rating* for the design alternative based on the watershed performance, while Participant 11 was asked to provide the user rating for the design alternative based on the group of SBs: 41,90,92, and 93

## B. Overall Group Assessment

Table 4 show the group averages in distance metrics for the different participant groups and models, calculated using Equation 3. Notice that in general  $GDR_1$  was found to be greater than  $GDR_3$  suggesting that the groups, in spite of the model type (i.e. *Model A* or *Model B*), on an average, preferred design alternatives located closer to the design alternatives on the non-interactive Pareto Front. While this outcome seems reasonable for participants working with *Model A-Surrogates*, the finding seems counterintuitive, at first

glance, for participants working with *Model B-Surrogates* or *Model B-Stakeholders*. Note that *Model B* over-estimated the *PFR*, *SR*, and *NR* benefits in comparison to the *Model A*, and *Model A* was used to estimate benefits of alternatives on the non-interactive Pareto front. Hence, a user assessing the quality of the design alternative based on only the physical objective function values estimated by *Model B* would have been expected to prefer designs with higher  $GDR_3$  than  $GDR_1$ . However, this was not always observed, indicating that even *Model B* participants may not have been entirely motivated by the performance of design alternative estimated at the watershed scale in order to decide what design alternatives they liked. Additional factors may have been more important to these participants when they were evaluating design alternatives. For example, some participants may have been more influenced by the value of the design decisions (e.g., certain locations may be more favorable for a BMP from the user’s perspective, in spite of the performance). This issue is examined in the next section.

**Table 5.6 Group average for distance between non-interactive optimal Pareto Front**  
**Objective 1: COST,**  
**Objective 2: PFR**

GROUP	Objective 1: COST, Objective 2: PFR			Objective 1: COST, Objective 2: SR			Objective 1: COST, Objective 2: NR		
	GDR <sub>1</sub>	GDR <sub>2</sub>	GDR <sub>3</sub>	GDR <sub>1</sub>	GDR <sub>2</sub>	GDR <sub>3</sub>	GDR <sub>1</sub>	GDR <sub>2</sub>	GDR <sub>3</sub>
<i>Model A-Surrogates</i>	0.24	0.10	0.12	0.26	0.11	0.13	0.23	0.10	0.11
<i>Model B-Surrogates</i>	0.41	0.34	0.36	0.42	0.34	0.36	0.39	0.30	0.34
<i>Model B-Stakeholders</i>	0.30	0.21	0.26	0.41	0.27	0.35	0.37	0.25	0.32

### 5.5.1.3. Similarities in Decision Space

#### A. Assessment of Individual Participants

As described in the methodology, the decision space was represented by two different BMPs, Cover Crops and Filter Strips. The BMPs were modeled using two different types of variables, binary and real. In this study, the binary variable represents whether cover crops exist in a particular SB, while the real variable represents the design parameter of filter width for the BMP filter strips. Each BMP can be simulated in one of the 108 different SBs identified in an earlier study Babbar-Sebens et al. (2015). Therefore, the chromosome in the genetic algorithm (used for interactive search in WRESTORE) consisted of 216 genes, where each gene value represents the decision for one of the two BMPs. For each of the variables we calculated a pattern metric (as shown in section 5.4.3.3), depending on the variable type.

*Model A-Surrogates* results are shown in Table 5.7. As with the percentage of user ratings, we added user behavior information related to their interaction with the GUI and the trends in their self-reported, average confidence levels. To estimate the level of agreement (or disagreement) within the watershed for cover crops, we first identified those SBs with  $Prob_{gR3k} > 0.5$  and then the set of SBs with  $Prob_{gR1k} > 0.5$ . Since,  $R_1$  and  $R_3$  are the extreme values for a participant's *user rating*, we then identified a sub-set of sub-basins in the entire watershed where (a) the participant clearly prefers cover crops (i.e., the conditions  $(Prob_{gR3k} > 0.5)$  and  $(Prob_{gR3k} > Prob_{gR1k})$  are met for the SB), and (b) the participant clearly does not prefer cover crops (i.e., the condition  $(Prob_{gR1k} > 0.5)$  and  $(Prob_{gR1k} > Prob_{gR3k})$  are met for the SB). Table 5.7 shows these results. For all the participants there are a higher numbers of SBs where cover crops implementation is higher in their  $R_3$  design alternatives. Additionally, the probability of implementing cover crop is greater than 50% (since  $Prob_{gR3k} > 0.5$ ), leaning towards an agreement of the use of this BMP in large majority of their design alternatives and SBs.

Table 5.7 also shows the results for the design alternative patterns of the Filter Strip width. Similarly to the cover crop analysis, we calculated the number of SBs where the  $POM_{gR3k}$  was greater than 50%. This means that majority of the design alternatives in  $R_3$  have their design value for the filter width equal to the mode of all values, indicating a clear preference by the participant. Similar to the results for cover crops, there are a greater number of SBs where Mode value is distinctly the filter width preferred by the user.

These results also allow us to compare designs across participants. We can see that according to Table 5.7, Participants 6 and 7 seem to have a close agreement on the design alternatives (in either Cover Crops or Filter strips width). However, Participant 7's interaction with the tool and trends in her/his confidence levels indicate that she/he did not gather enough information to develop a confident judgment. Whereas, Participant 6 demonstrated an increase in his/hers confidence trends, and has a high level of interaction with the GUI in the information gathering areas. These user behavior parameters seem to indicate that even when similar design alternatives were found by these two participants, the feedback provided by Participant 6 seemed to be more reliable. The maps in Figure 5.6 show examples of the probability distribution within the watershed for Participant 6 and 7.

This example is the pattern for cover crop Probabilities in the designs rated  $R_3$  and  $R_1$  by Participant 6. We also noticed that as all the participants present a lower  $Prob_{gR3k}$  in the Southern region of the watershed (close to the reservoir). This suggest that in optimized solutions, regardless of the participant, the allocated system of Cover Crops and Filter Strips are more desired in the upstream region.

**Table 5.7 Number of SBs with a design alternative pattern for the Probabilities of Cover Crops and the Filter Strip width.**

Participant	Confidence Level trends	Avg. Clicks % Info	Avg. Time % Info	Number of SBs where $Prob_{gR3k} > 0.5$ & $Prob_{gR3k} > Prob_{gR1k}$ (out of 108)	Number of SBs where $Prob_{gR1k} > 0.5$ & $Prob_{gR1k} > Prob_{gR3k}$ (out of 108)	Number of SBs where $POM_{gR3k} > 50\%$ & $POM_{gR3k} > POM_{gR1k}$ (out of 108)	Number of SBs where $POM_{gR1k} > 50\%$ & $POM_{gR1k} > POM_{gR3k}$ (out of 108)
1	+ $R_1$	1	3	81	10	108	0
2	None	72	62	79	13	59	2
3	+ $R_1$	60	56	66	26	45	39
4	None	10	14	82	17	51	2
5	- $R_3$	40	45	53	39	62	29
6	+ $R_1, R_3$	66	67	82	8	73	0
7	- $R_1, R_2, R_3$	14	34	82	7	77	0
<b>Average</b>		38	40	75	17	68	10
<b>Std. deviation</b>		29	24	11	12	21	16

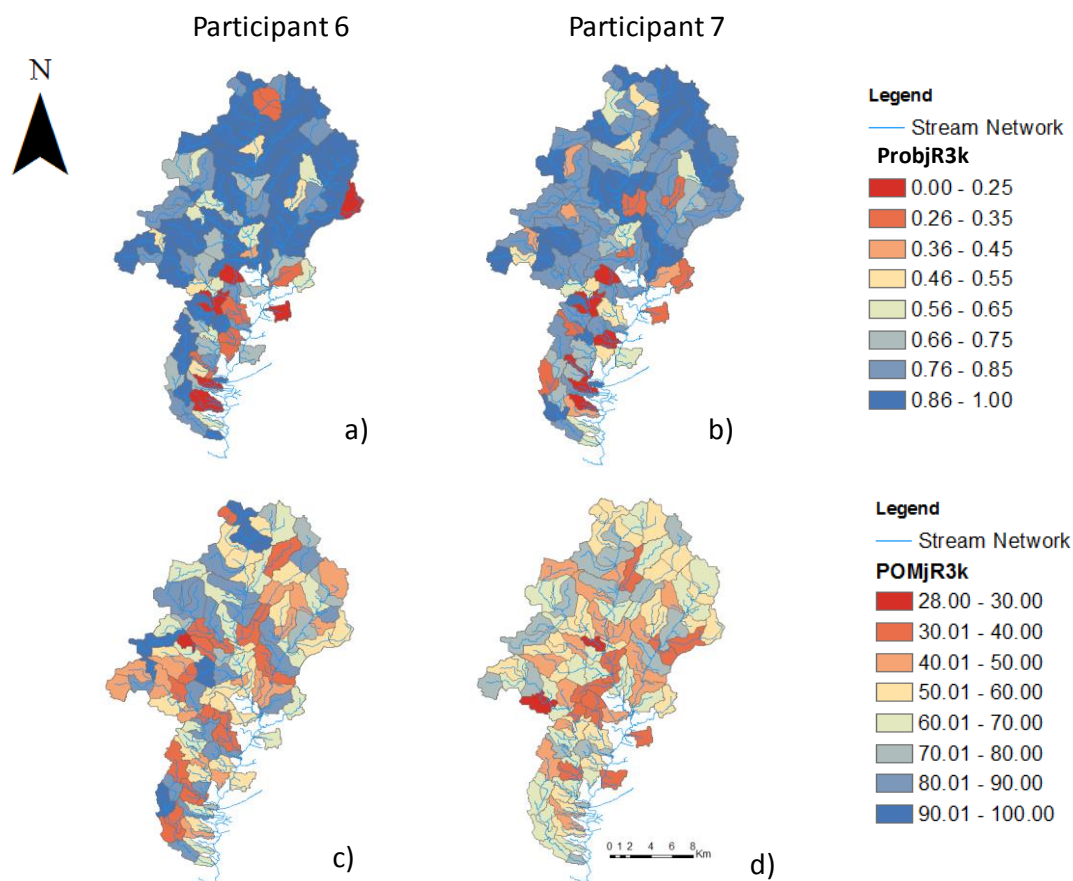


Figure 5.6. Examples of the distribution of  $ProR3 > 0.5$  (a and b) and  $POM > 50\%$  (c and d) for Participants 6 and 7

*Model B-Surrogates* and *Model B-Stakeholders* results are presented in Table 5.8. The average number of SBs where conditions ( $Prob_{gR3k} > 0.5$ ) and ( $Prob_{gR3k} > Prob_{gR1k}$ ) are met are 67 and 72, for *Model B-Surrogates* and *Model B-Stakeholders* respectively. Hence, the number of SBs where cover crops are clearly desired are slightly lower for these participants, than for the participants in *Model A-Surrogates* where cover crops were desired in 75 SBs on an average across all participants. Similarly, for filter strips, the average number of SBs where conditions ( $POM_{gR3k} > 0.5$ ) and ( $POM_{gR3k} > POM_{gR1k}$ ) are met are 36 and 39, for *Model B-Surrogates* and *Model B-Stakeholders* respectively, in comparison to *Model A-Surrogates* where the Modes of filter strip widths were clearly desired in 66 SBs on an average across all participants. This suggests that *Model B-Surrogates* and *Model B-Stakeholders* participants are satisfied with designs where not too many new cover crop and filter strip practices are added into the current baseline watershed landscape. Additionally, between the individual participants there are also unique

differences in the manner they prefer one practice over other. For example, Participant 16 indicates a clear preference for cover crops in majority of the SBs (83) in her/his  $R_3$  design alternatives, in comparison to cover crops in her/his  $R_1$  design alternatives, and in comparison to filter strips in all of his  $R_1$  and  $R_3$  design alternatives. This participant also has high amount of interaction (69% of mouse clicks and 65% of time spent) with the GUI to gather information, and no discernable increasing or decreasing confidence trend in design alternatives rated  $R_3$  or  $R_1$ . The same trends and preferences are also presented by Participant 25, except for the amount of interaction (2% of mouse clicks and 1% of time spent in gathering information) that this participant had with the GUI.

**Table 5.8 Number of SBs with a design alternative pattern for the Probabilities of Cover Crops and the Filter Strip width.**

Participant	Confidence Level trends	Avg. Clicks % Info	Avg. Time % Info	Number of SBs where $Prob_{gR3k} > 0.5$ & $Prob_{gR3k} > Prob_{gR1k}$ (out of 108)	Number of SBs where $Prob_{gR1k} > 0.5$ & $Prob_{gR1k} > Prob_{gR3k}$ (out of 108)	Number of SBs where $POM_{gR3k} > 50\%$ & $POM_{gR3k} > POM_{gR1k}$ (out of 108)	Number of SBs where $POM_{gR1k} > 50\%$ & $POM_{gR1k} > POM_{gR3k}$ (out of 108)
<b>Model B-Surrogates</b>							
8	None	0	0	82	10	90	0
9	None	57	47	73	13	79	0
20	$-R_2, R_3$	36	71	62	34	11	0
21	$+R_3$	47	34	57	37	4	1
22	$-R_3$	0	1	56	38	22	7
24	$-R_1, R_3$	15	21	61	27	45	16
25	None	2	1	79	0	0	0
<b>Average</b>		22	25	67	23	36	3
<b>Std. deviation</b>		24	27	11	15	36	6
<b>Model B-Stakeholders</b>							
11	$+R_1, R_2, R_3$	76	73	77	13	99	0
13	None	54	47	82	15	35	0
14	$+R_2$	38	40	59	27	30	2
15	$+R_2, R_3$	3	2	71	17	24	0
16	$-R_2$	69	65	83	0	0	0
18	$+R_1$	45	50	57	30	46	8
<b>Average</b>		48	46	72	17	39	2
<b>Std. deviation</b>		26	25	11	11	33	3



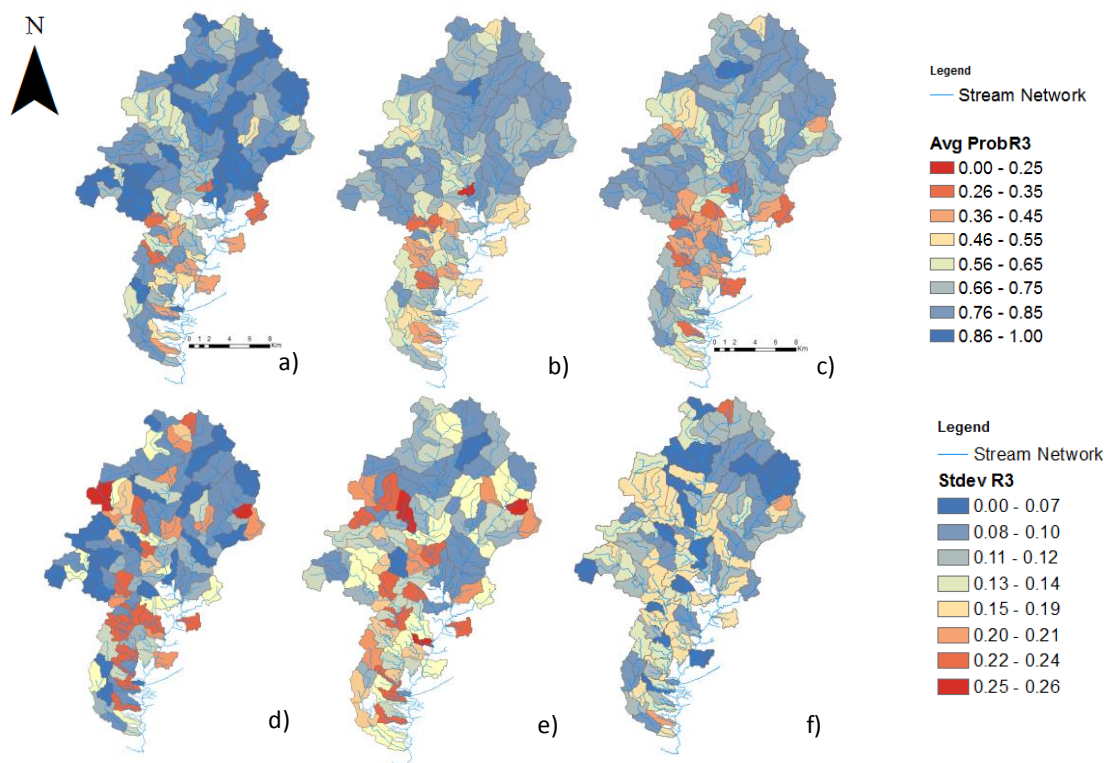
In summary, by identifying patterns across the different design alternatives liked or disliked by the user, it is possible to categorize which and where are practices more favored by different individuals. However, to examine fidelity of the preferences expressed by users for locations and practices, the user-GUI interaction data provides useful supporting information.

## B. Overall Group Assessment

Group statistics for the decision space were calculated using Equations 5.13 and 5.14, in order to examine any patterns in the preferences for locations or practices expressed by a specific group of participants. In this section, results for design alternatives rated  $R_3$  by the participants are discussed, in order to focus on designs that were considered acceptable by the participants.

Figure 5.7. shows three maps (maps (a), (c), and (e)), representing the  $GProb_{gR3}$  (Average Probability of Cover Crop across participants for design alternatives rated  $R_3$ ) for each of the three participant groups. In addition to the  $GProb_{gR3}$ , maps (b), (d), and (f) show the variability across participants in the probability of cover crops at every SB. This variability was estimated by calculating the standard deviation ( $Stdev_{R3}$ ) of the  $GProb_{gR3}$  values across participants. The maximum value for  $Stdev_{R3}$ , across all SBs and all groups, was found to be 0.26. Observe that 59% of the SBs in *Model B-Surrogates* present a  $Stdev_{R3} >$  half of the maximum value (i.e.,  $0.26/2 = 0.13$ ), indicating a high disagreement in these SBs. While for *Model A-Surrogates* and *Model B-Stakeholders*, these percentages are lower (i.e., 41% and 31% of SBs with  $Stdev_{R3} > 0.13$  in *Model A-Surrogates* and *Model B-Stakeholders*, respectively). These results suggest that participants from the *Model A-Surrogates* and *Model B-Stakeholders* groups may have a better chance for negotiating implementation of cover crops in most of their SBs, since participants in these groups strongly disagree (i.e., when  $Stdev_{R3} > 0.13$ ) with each other on fewer number of SBs in the set of design alternatives rated  $R_3$ . Figures 5a), 5b) and 5c) also show that the percentage of SBs with  $GProb_{gR3} < 0.5$  (i.e., SBs where the average probability of cover crop across participants is low) is low for the three groups. For example, for *Model A-Surrogates*, there are less than ~20% (22 SBs) of SBs with  $GProb_{gR3} < 0.5$ . These SBs are located mostly in the downstream area of the Watershed, close to the Eagle Creek Reservoir. When

variability (or, disagreements) across participants are considered in the *Model A-Surrogates example*, there are only ~11% (12 SBs) where cover crop probability is low ( $GProb_{gR3} < 0.5$ ), but not all participants agree with the low value of the probability (i.e.,  $Stdev_{R3} > 0.13$ ).

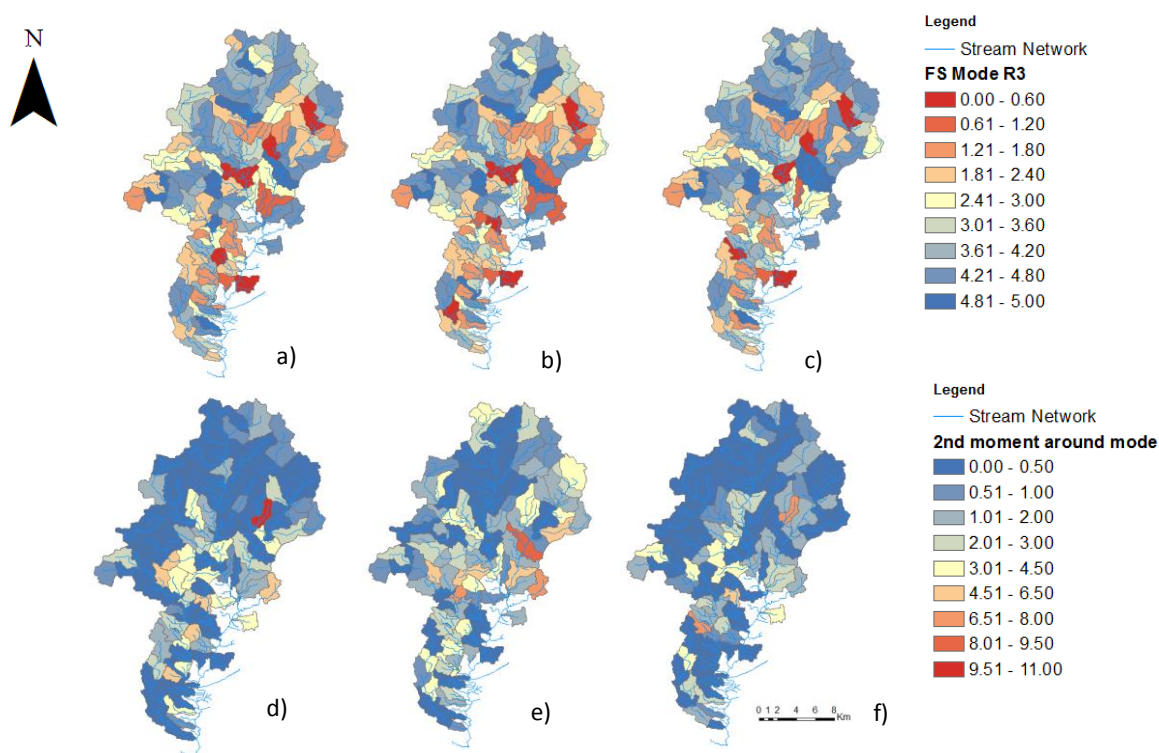


**Figure 5.7. Maps of Average ProbR3 (upper row) and Standard Deviation (lower row) for each of the Model groups. From left to right we have *Model A-Surrogates* (a and d), *Model B-Surrogates* (b and e) and *Model B-Stakeholders* (c and f)**

In an overall analysis, we can say that *Model B-Stakeholders for the  $GProb_{gR3}$*  present a more uniform preference and assessment of the design alternatives across all participants, while the *Model A-Surrogates* and *Model B-Surrogates* seem to develop more variable patterns across participants.

For the analysis of the group assessment of Filter Strip width, we calculated the Mode of Modes across participants ( $GPOM_{gR3}$ ) to estimate the group's central tendency, and also calculated the second moment around the Mode of Modes to estimate the variability across participants ( $VarPOM_{gR3}$ ). Figure 5.8. show that, in this case, there is a more consistent preference for filter strip modes in SBs across the groups and less variabilities across participants (see figure 5.8.b), 5.8.d) and 5.8.f)). Similarly to the process for Cover Crops,

the maximum value for  $VarPOM_{gR3}$ , across all SBs and all groups, was found to be 5.2. Again, there indicates a higher agreement across participants in *Model B-Stakeholders*, with approximately 91% of the SBs (98 SBs) present a  $VarPOM_{gR3} < 3$  value (i.e., second moment around the mode is less than 3 – as indicated in figure 5.8.). While the variability is higher for the *Model A-Surrogates* and *Model B-Surrogates*, the variability is high for fewer than 17% of the total SBs (18 SBs *Model A-Surrogates* and 10 SBs *Model B-Stakeholders*).



**Figure 5.8.** Maps of Average ProbR3 (upper row) and Standard Deviation (lower row) for each of the Model groups. From left to right we have Model A-Surrogates (a and d), Model B-Surrogates (b and e) and Model B-Stakeholders (c and f)

### 5.5.2. DS assessment of user search experiments

In the previous section 5.5.1 (and its sub-sections), we examined the design alternatives generated via interactive optimization, at the end of the users' search experiments. In this section, results are presented that examine the same metrics related to user ratings, objective space, and decision space, but through time for all users. Such a temporal perspective is expected to provide insights into convergence properties of the user-guided search process.

### 5.5.2.1. Percentage of design alternatives per user ratings

#### A. Assessment of Individual Participants

In this section, the changes in the percentage of solutions in every *Epoch* are described. As explained in section 5.4.3., an *Epoch* is a specific time interval related to a session type within the experiment where the user provides design alternatives' rates for the first time, to generate more design alternatives. Therefore, *Epoch 1* is associated with the first introspection (*I1*) session, *Epoch 2* with design alternatives from human-guided search sessions (*HS1* to *HS6*) after *I1*, and, finally, the *Epoch 3* with design alternatives from human-guided search sessions (*HS1* to *HS6*) after the second introspection (*I2*) session.

The three participants in *Model A-Surrogates* who had positive trends in the confidence levels (Participants 1, 3, and 6) demonstrated a decrease in the percentage of design alternatives that were rated  $R_3$  as the search progressed from *Epoch 1* to *Epoch 2*. However, from *Epoch 2* to *Epoch 3*, the change in percentage switched to an increasing trend for two of these participants while the other two continued towards a declining trend in percentage (see Figure 5.9 for details on these three participants)

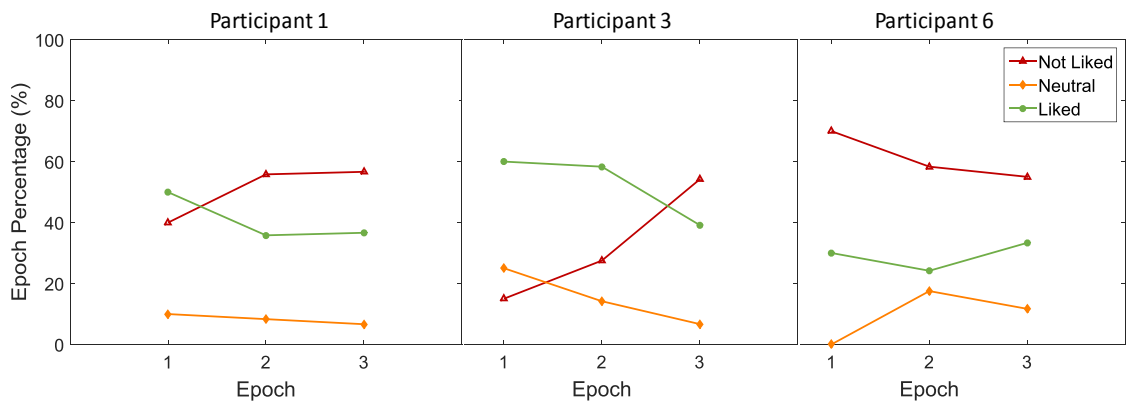


Figure 5.9. Participants 1, 3 and 6 percentage of solutions in user rating  $R_i$  by Epochs

The change in the trend could be an effect of multiple factors that influenced the search algorithm. For example, they could be because of potential changes in the users' reasoning process for rating  $R_3$  in the *Epoch 2 – Epoch 3*, or they could have occurred because of the stochastic nature of the exploration-exploitation operators (in the underlying Genetic Algorithm, IGAMII) that led to a more successful search in the later *Epoch 2 – Epoch 3* period. We also observed that two out of these three participants (Participants 1 and 3) experienced an increase in their percentage of design alternatives rated  $R_1$  as the *Epochs*

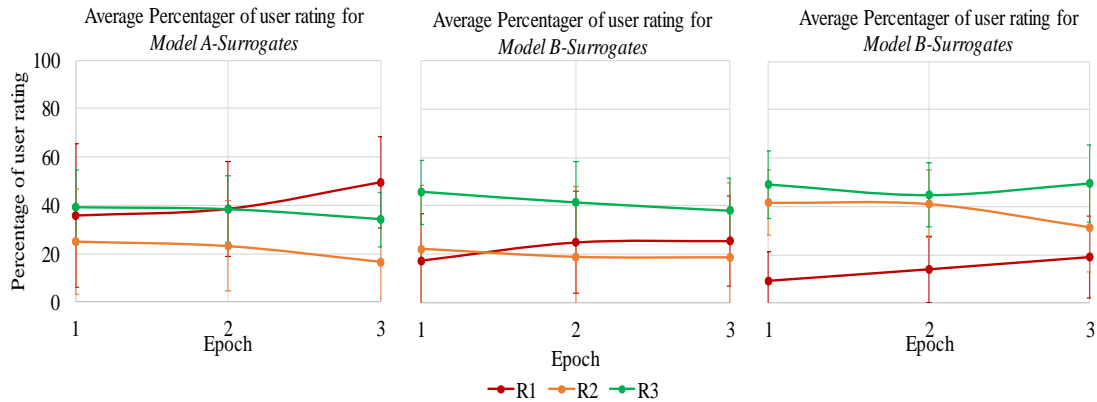
progressed. These same participants, however, also seemed to have the percentage of  $R_3$  design alternatives in *Epoch 1* be greater or equal to 50%.

For the case of *Model B-Surrogates*, just one (Participant 21) demonstrated a positive trend in his/her confidence levels over time. This participant experienced a decrease in the percentage of design alternatives rated  $R_3$  from *Epoch 1* to *Epoch 2*, but then experienced an increase in the percentage of design alternatives rated  $R_3$  from *Epoch 2* to *Epoch 3*. *Model B-Stakeholders* results show that four out of six participants consistently experience an increase in the percent of design alternatives rated  $R_1$ . While only one of them consistently increases design alternatives rated  $R_3$ . However, four out of six participants will present an increase in *Epoch 3*. We also observed that five out of six participants present an increase in design alternatives rated  $R_1$ , and four of these participants will start with 15% or fewer designs rated  $R_1$ . For this group, the participants with positive trends show an increase of  $R_1$  design alternatives in *Epoch 1* to *Epoch 2* and *Epoch 2* to *Epoch 3*. While the participant with none presents a decrease in *Epoch 1* to *Epoch 2* and then an increase from *Epoch 2* to *Epoch 3*. The result indicates that participants with positive trends increase their design alternatives rated  $R_1$ , and in the majority of the cases also those designs rated  $R_3$ .

This analysis shows the dynamic across participants over time. An early identification of patterns in the design may improve the search process and avoid the repetition of designs that were clearly disregarded by the user. A combination of confidence levels values, previous percentages of designs in each rate and the information about the amount of interactions may avoid losing key features of the design that meets with the user's criteria.

## **B. Overall Group Assessment**

Figure 10 shows the average percentage per user rating at each epoch. The results show that *Model A-Surrogates* participants do not show a consistent trend through the *Epochs*. However, we observe that participants with positive confidence level trends increase the number of  $R_1$  design alternatives at each *Epoch*. In fact, from *Epoch 2* to *Epoch 3*, there is just one participant that decreases the percentage of user rating alternatives ( $PRate_i$ ) rated  $R_1$ . This is an indication that the solutions found by the system are in disagreement with the goals of the participants in this model.



**Figure 5.10** Average of percentage of solutions per participant per Epoch

Again, the case of *Model B-Surrogates*, we find a lot of variability regarding the percentage of the design alternatives. However, we noticed that three participants consistently decrease the percentage of design alternative rated *R3*. We also notice that there is not a consistent decrease for the design alternatives rated *R1*, i.e., the percentage of *R1* alternatives is always increasing across participants of this group. *Model B-Stakeholders'* group presents a more uniform patterns across the *Epochs* and participants, especially in *Epoch 3* where there is an increase of designs rated *R1* and *R3* in the majority of the participants of this group. These results then show how the community that is directly affected and involve with the location seems to have a better agreement on the required actions to improve the watershed management.

Figure 5.10 shows the average percent of *user rating* across participants in each group. Notice that the error bars represent the variability (standard deviation) for each group. As we stated before, *Model B- Stakeholders* shows a better agreement in the percentage of solutions in each of the rates, with an average standard deviation (error bars) of 15%, compare to the *Model B-Surrogates* and *Model A-Surrogates* (21% and 18% respectively).

### 5.5.2.2. Similarities in Objective Space

#### A. Assessment of Individual Participants

This information provides an evaluation of how the user rating is affecting the other objective functions, and what is the direction the new user-tool designs are taken. At the same time the distances give us an estimation regarding how the non-interactive Pareto

Front with physically found optimal design alternatives and allow us to quantify the reduce of the performance in the objective space.

For *Model A-Surrogates*, the changes in the distances for the Pareto front are similar to the PFR vs. Cost, SR vs. Cost and NR vs. Cost measurements. Three participants (Participants 3, 4 and 5) presented a higher distance in *Epoch 2*. Meaning that in *Epoch 3* they found *R1* solutions closer to the non-interactive Pareto Front. However, this same three participants also found *R3* solutions farther from the Pareto Front in *Epoch 3*. This could be a consequence of improvement in the particular areas of interests. However, the improvements in local SBs reduce the performance of the entire system. From these three participants, just one of them (Participant 3) have a positive trend of the confidence levels in *R1* and the highest activity in the information gathering area. This indicates that the designs he classified in *R1* and the calculated distances may be more reliable.

For the case of *Model B-Surrogates*, the distances in *Epoch 2* and *Epoch 3* are nearly identical, suggesting that there are not significant changes in the objective space through time. However, this technique is not capturing the dissimilarities in the objective space across participants, because the distances are fairly equal for all the participants, except for one (Participant 9). However, the group's usability metrics are very low, telling that the interaction with the tool was minimal, with no increases in the confidence levels trends (or negative trends) weaken the reliability of the input provided by the participant. The distance metric in the *Model B-Stakeholder* for the designs rated *R1* present some variability from *Epoch 2* to *Epoch 3*. For *Epoch 3*, participants select designs closer to the non-interactive Pareto Front, while also having similar distances for Pareto Fronts rated *R3*.

## **B. Overall Group Assessment**

*Model B-Surrogates* and *Model B-Stakeholders* temporal analysis, suggests that overall there is not a significant variation in the objective space performance for *Epoch 2* to *Epoch 3*. We also noticed that *Model B-Stakeholders* choose and lead the search towards designs that are closer to the non-interactive Pareto Front, this is reflected in the lower average distance metric presented by the stakeholders.

### 5.5.2.3. *Similarities in Decision Space*

#### **A. Assessment of Individual Participants**

Because we would like to understand the characteristic of desirable designs, this analysis is center on those design alternatives rated  $R_3$  by the different participants. The results in the decision space for the temporal analysis showed that for the average probability of implementation in the binary variable (Cover Crop), there is a general decrease of the probability of implementation from *Epoch 1* to *Epoch 2* to *Epoch 3* in all the user ratings. However, two participants present an exception in the trend from *Epoch 2* to *Epoch 3*, where there is an slightly increase in the average of probabilities across SBs. Due to the decrease of the probabilities in all three user ratings, we could not conclude that the decrease is a consequence of the user's rating. Also, the probability found here is associated with the entire watershed. In order to draw significant conclusions we will need a more detailed consideration about the local scale performance.

For the real variable case (Filter Strips width) we found that the averages in the whole watershed ranged from 2.74 m to 3.19 m. Remember that in this case we calculated the percentage of solutions with the filter width value equal to the mode of the design alternative ( $POMR_i$ ). The behavior of the average of  $POMR_3$ , is very erratic across participants. There is a 45% of the participants that decreases  $POMR_3$  from *Epoch 1* to *Epoch 2*, but then increases within *Epoch 2* and *Epoch 3*. Then, another 25% of the participants exclusively increases the  $POMR_3$  values, a 20% that exclusively decreases the  $POMR_3$  and then there is just one participant who increases from *Epoch 1* to *Epoch 2*, but decreases from *Epoch 2* to *Epoch 3*. This analysis does not provide any conclusive results, and therefore there is a need to explore the parameter further (i.e., at local scale) to truly understand if there is a particular design that motivates the participant to select a certain particular rate.

#### **B. Overall Group Assessment**

As a group, the results show an increase of the variability and a common decrease of the probability of implementation in the pattern analysis.  $GProbR_i$  for *Model A-Surrogates*, *Model B-Surrogates* and *Model B-Stakeholders* have a similar behavior, however, the second case presents a lower variability across participants. This indicates that



As section 5.5.1.1, we used the average of the percentage of modes to observe the trend of selections of “I liked” alternatives for each of the participants. In most of the cases by the end of epoch three there is an increase in the average percentage of modes, indicating that more solutions present the same pattern for the real number. The standard deviations seem to decrease. However, the present high uncertainty may be a result of the average of the watershed scale results.

## 5.6 CONCLUSIONS AND FUTURE WORK

This research provided a very useful observational study where users’ answers, and feedbacks are analyzed in order to understand the decision process and validate the user’s input data supplied to a participatory search algorithm. The uniqueness of this research relies in the understanding on how user’s input data (as an additional objective function) is used by the optimization algorithm in order to search for designs that will meet the user’s criteria. Intuitively we can say that if a design alternative is rated  $R_3$  there is a higher chance for it to be implemented by the stakeholders because they agree with either the decision space or the objective space performance.

To gather the *user rating* a Liker-type scale with three classifications was provided: “I liked” ( $R_3$ ), “Neutral” ( $R_2$ ) and “I do not like it” ( $R_1$ ). In general results show that 55% of participants present a higher percentage of design alternatives in  $R_3$ . This percentage suggests that over half of the population agree with the design alternative found for them by the IGA.

The objective space evaluates the distances between the Pareto Front of design alternatives found in a non-interactive exhaustive search (designs presented in *II*) and the Pareto Front generated by the user’s interaction with the tool. The measure calculates not just the accuracy but also the precision of the design alternatives of different classifications with respect to those design alternatives found on the non-interactive search. In general, the design alternatives classified as  $R_1$  have a greater distance and are less precision than those design alternatives classified as  $R_3$ . The classification is more evident for those participants that evaluate the design alternatives based on the watershed level performance (Piemonti, 2014)

The differences between the probabilities and the variability across participants demonstrate the ability of the search algorithm to identify individual preferences, and how they affect the designs in complex systems. Even when all the groups were given similar areas to evaluate, stakeholders have a common goal/interest that must be achieved within the watershed. Therefore, the results are consistent with the expected behavior of participants with a common goal. On the other hand, Surrogate participants are focused in SBs assigns as part of the experimental process. Therefore, there is not a common objective in improving the overall area, but a very deterministic tasks of improving their own areas, regardless of the performance of other SBs. It is possible that the selection of the design alternative based on specific areas leads the design to generate solutions that may underestimate the performance in other SBs. Therefore, future research must include the analyzes of localized design patterns based on those SBs on interest, establishing comparisons across participants, and showing where the possible agreements and disagreements may occur. A general evaluation of the decision space gave detailed information regarding the number of SBs where designs present common features such as implementation or designed. As was mention before, the decision space was divided into two different analysis depending on the BMP in the design alternative. For the *binary BMP* (Cover Crops), the general evaluation suggests that all the participants would agree to the implementation in a higher number of SBs, regardless their interaction with the tool, or their confidence level trends. While, the *real BMP* (Filter Strip width) there is more variability regarding the number of SBs where the practice should be implemented.

The analysis in the decision space presented distinct patterns on the designs in the individual and group analysis. It seems that the search may identify agreements more accurately for the real variable designs, where the results show a lower variability on the mode selected for those designs rated *R3*. However, there is still a need to understand better the selections based on local scale preferences. We suggest and encourage to take a closer look at the results for how the human-guided optimization process affects not just SBs of interest, but also other participants SBs of interest.

## 5.7 ACKNOWLEDGEMENTS

We would like to acknowledge the funding agency - National Science Foundation (Award ID #1014693 and #1332385).

## 5.8 REFERENCES

- M. Babbar-Sebens, and B.S., Minsker, “Interactive Genetic Algorithm with Mixed Initiative Interaction for multi-criteria ground water monitoring design, ” *Applied Soft Comp.*, vol. 12, pp. 182-195, Sept. 2011.
- Arnold, J.G., Srinivasan, R., Muttiah, R.S., and Williams, J.R., (1998), Large Hydrologic Modeling and Assessment: Part I. Model Development, *JAWRA*, 34(1), 73-89.
- M. Babbar-Sebens, Barr, R.C., Tedesco, L. P., and Anderson, M., “Spatial identification and optimization of upland wetlands in agricultural watersheds,” *Ecol. Eng.*, vol. 52, pp. 130-142, Mar. 2013.
- M. Babbar-Sebens, Mukhopadhyay, S., Singh, V. B., and Piemonti, A. D., “WRESTORE- A web-based planning software for participatory optimization of conservation practices in watersheds,” *Environ Mod and Soft.*, 69, 111-127, July 2015.
- Fry, D.L., Battles, J.J., Collins, B.M., and Stephens, (2015), Sierra Nevada Adaptive Management Project. Fire and Forest Ecosystem Health report.
- Kaunda-Bukenya, N., Tadesse, W., Fu, Y., Tsegaye, T., and Wagaw, M., (2012), Spatial Decision Support System (SDSS) for Stormwater Management and Water Quality Assessment, in Voudouris, K., and Voutsas, D., Published: April 5<sup>th</sup>, 2012, ISBN: 978-953-51-0486-5. InTech
- Matsuura, K., Willmott, C., Legates, D., 2009. WebWIMP: the web-based water-budget interactive modeling program. <<http://climate.geog.udel.edu/~wimp/>> (last accessed 28.12.15).
- Piemonti, A. D., Babbar-Sebens, M., and Luzar, E. J. (2013), Optimizing conservation practices in watersheds: Do community preferences matter? *WRR*, 49(10), 6425-6449.
- Piemonti, A.D., (2014), Correlating Human Preferences and Optimized Watershed Management Plans. Proceedings of the World Environmental and Water Resources Congress 2014: pp. 1961-1969

**CHAPTER 6. Participatory design of distributed conservation practices in a watershed: Effect of spatially-explicit user preferences on Interactive Optimization**

**Adriana Piemonti<sup>1</sup> and Meghna Babbar-Sebens<sup>1</sup>**

<sup>1</sup> School of Civil and Construction Engineering, Oregon State University, Corvallis, Oregon, USA.

To be submitted to the Water Resources Research Journal.

## 6.1 Abstract

Participatory modeling and design have been recommended by multiple researchers as effective means for supporting a more active role of stakeholders towards Integrated Management of their watersheds (Johnson et al, 2002; Gregersen et al., 2007; Lubell, 2014; Kelly et al., 2012; Evers et al., 2012). Interactive Optimization (IO) is a human-in-the-loop type participatory technique that has the potential to include different types of community stakeholders in the design process. IO aims to use stakeholder participation towards generation of scenarios of *satisficing* (*satisfy* + *suffice*) watershed plans that fulfil people's subjective constraints and criteria, in addition to achieving quantified physical cost-benefit goals (Piemonti et al. 2013; Piemonti 2014; Babbar-Sebens et al., 2015). However, little research had been done that improves our current understanding of how participation of humans, who may be interested in only a few sub-areas of a watershed, can affect the overall search process of an interactive optimization algorithm that is set up to generate plans for the entire watershed. Such investigations have the potential to generate insightful evaluation of links between IO-based participatory design algorithms and the preferences and criteria expressed by stakeholders willing to participate in the planning and management process. In a previous study (Chapter 5), we examined IO-generated design alternatives and analyzed them for the presence of spatial patterns in locations and practices at the entire watershed scale. In this study, we have examined how stakeholder feedback that is more driven by the local decisions in sub-areas of a watershed affects the search process of the IO algorithm centered on the entire watershed. Results show that, in 100% of the cases, there is a correlation of performance of the SBs in the objective space for those SBs connected through a stream. In the decision space, we observed that, in general, there is an acceptance towards an implementation of Cover Crops in 95% of the SBs for the majority of the participants (even if they are not focused on the region). For Filter Strips width, there is a better agreement for similar designs in those SBs upstream. However, the average percentage of designs equal to the mode is less than 60% for all the *SBint*. This analysis enables researchers, designers, and decision makers to evaluate discrepancies in preferred designs that are considered optimal by different individuals, and at the same time

provides valuable information about the crowdsourced solutions to generate democratic and participatory watershed management plans.

## **6.2 Introduction**

Generation of scenarios of watershed plans (or, designs) based on physical goals and constraints have been extensively researched before (Seppelt and Voinov, 2002; Arabi et al, 2006; Artita et al, 2008; Babbar-Sebens et al., 2013). Multiple researchers have recommended that multi-objective optimization-simulation approaches can be notably effective in incorporating multiple conflicting criteria towards the development of restoration plans for degraded watersheds (Randhir et al., 2000; Perez-Pedini et al., 2005; Lethbridget et al., 2010; Kaini et al, 2012). On the other hand, optimization based on subjective criteria still remains a challenge, because of the diversity in socio-economic and cultural conditions that may exist in a watershed community. Complexities from diverse conditions influence how people determine their priorities, beliefs, interests, and biases via their underlying cognitive processes, and express individual subjective preferences during decision making (Mintzberg and Westley, 2001; Cash et al, 2002; Prokopy et al., 2008; Hoag et al., 2012; Ahmadi et al., 2014). Hence, there is a dire need for studies that examine approaches for incorporating the complex human preferences in to the design of watershed plans, while also evaluating the effect of such stakeholder integration.

In this paper we examine the problem of interactive optimization of conservation practices in agricultural watersheds, based on not only goals that target land services, water quality, and flooding problems, but also stakeholders' diverse subjective criteria and preferences pertinent to local sub-areas of the watershed. Many location-specific stakeholder factors, e.g. a land owner's personal constraint in using a specific patch of her/his land for implementing a conservation practice, past unsuccessful experience with a specific practice or a specific location, or limited available personal financial resources to implement and maintain a practice, etc., may influence the stakeholder's decision to select a prescribed conservation practice at a recommended location. However, when developing watershed-scale plans for a community, planners and managers may overlook or not be able to pre-determine these stakeholder factors (Cole, et al., 2002; Greiner et al; 2009; Hoag et al., 2012). Exclusion of spatially-explicit stakeholder subjective factors may

weaken the acceptability of proposed practices in individual sub-areas, and create hurdles in the adoption of the larger management plan.

Analyzing the design alternatives generated by Interactive Optimization at the local sub-basic scales relevant to individual stakeholders, provides an opportunity to understand how the stakeholders' interaction with the tool and their spatially-explicit preferences may influence the IO's assignment of decisions at the sub-basin scale. Especially decisions in sub-basins, which lie in a stakeholder's region of interest, can not only influence watershed processes in her/his region, but also other surrounding regions that may be ecologically and/or hydrologically connected to her/his region. Additionally, for multiple stakeholders focused on the same local sub-basins, the disagreements between stakeholders at sub-basin scales can be assessed based on what alternatives they like or do not like.

### 6.3 Objectives

In this study, we focus on the web-based tool WRESTORE, a participatory framework for Interactive Optimization of conservation practices in watersheds (Babbar-Sebens et al., 2015). This study examines the relationships between the types of users, their usability behavior, and patterns in alternatives generated by users interacting with WRESTORE. A previous study identified the presence of patterns in the decision space of design alternatives classified as desirable or undesirable by the user on a Likert-type scale (see chapter 5). Here, we examine the presence of similar patterns in user preferred designs, but at the local scale. The overall research objective is to understand how the *user rating* that is assigned by stakeholders interested in specific local sub-basins, but is used by the search algorithm to assess the desirability of the entire watershed plan (which includes decisions in other sub-basins beyond the stakeholders' region of interest), affects the IO algorithm's assignment of conservation practices in the entire decision space and design performance in the local sub-basin scale objective space. To achieve this research objective, two research questions were addressed:

1. For design alternatives preferred by participants, are there any patterns in the objective space of local sub-basins that are of interest to multiple stakeholders? When users have to consider multiple quantitative and subjective goals to evaluate their preference for decisions proposed in the local areas common to all of them, it

can be challenging to determine well-defined patterns in the objective function space that represents physical/environmental goals. However, similarities in the central tendencies can be used to assess the presence of potential patterns in the objective function space. Hence, such central tendencies were estimated in this study, and used to address this research question.

2. For design alternatives preferred by participants, are there any patterns in the decision space of local sub-basins that are of interest to multiple stakeholders? Multiple disagreements between stakeholders can arise due to the different subjective preferences (discussed earlier) for which practices should be implemented and where in their common area of interest. Hence, comparison across users at the sub-basin scale may help to determine similarities in decisions that the users may be more willing to negotiate for and/or adopt. Hence, in this study, we conducted a similarity assessment for all user-preferred designs to identify which sub-basin decisions could potentially yield less conflict or more conflict.

## **6.4 Methodology**

### **6.4.1. Spatially-explicit Subbasins of interest**

As mentioned in chapters 3, 4 and 5, the selection of specific conservation practices (i.e. decision variables) and goals (i.e. physical objective functions) for IO was prescribed by the researchers to all participating users, so that the user experiments could be conducted in a semi-controlled environment. Six groups of sub-basins of interest ( $SBint_q$ ) were identified to represent local regions in the watershed (See Figure 3.3 in chapter 3 and Table 6.1). The experiments with 20 participants in chapter 4 and chapter 5, were also employed for this study. Sixteen participants were assigned one of the six groups in Table 6.1 as their study area, while the other four users were asked to use the whole watershed as their study area.



**Table 6.1 ID numbers of SBs inside the SBint groups and participants focused on this areas**

Group	<i>IDs of SBint</i>	Participants
1	[10 11 14 15]	2, 9, 16
2	[12 19 20]	3, 24
3	[38 39 44 ]	4
4	[41 90 92 93]	7, 11, 18, 25
5	[58 59 61 53]	5, 13, 20
6	[103 105 106 121 122]	6, 14, 21

Note that the researchers selected both connected and unconnected *SBint* in specific areas of the watershed, in order to capture the effect of the hydrologic connectivity on the preference in design expressed by participants. In Table 6.1, the connected sub-basins are adjacent SBs where the upstream *SBint* (ID value = n) drains to the downstream *SBint* (ID value = n+1). When *SBint* are hydrologically unconnected, their *SBint* ID value does not have any upstream or downstream ID in the *SBint*.

#### 6.4.2. Objective Space Analysis

A histogram analysis was conducted at the End of the Search (EoS) to quantify the number of user-preferred designs (*user rating*  $R_3$ ) within particular range of multiple physical objective functions calculated at sub-basin scale (e.g., cost of implementing the prescribed practice in SB 10 in Table 6.1). The range varied depending on the bin size. The bin size was calculated using:

$$bin_{OFmk} = \frac{\max(OF_{mk}) - \min(OF_{mk})}{10} \quad (6.1)$$

Where,  $bin_{OFmk}$  is the bin size of the  $m^{th}$  physical objective function  $OF$  (for  $1 \leq m \leq 4$ ), in  $l^{th}$  SB in  $SBint_q$ , and for  $k^{th}$  participant. The  $OFm$  objective functions are found in table 6.2. The mathematical function can be found in Chapter 2, and Chapter 5, as well as in Piemonti et al. (2013) and Babbar-Sebens et al. (2015).

**Table 6. 2 Table with the objective function and the type of Objective Function**

$m$	$OF$	Type
1	Peakflow Reduction	Physical, related to flooding
2	Sediments Reduction	Physical, related to water quality
3	Nitrates Reduction	Physical, related to water quality
4	Cost	Physical, related to economics
5	User rating: 1 = "I do not like it", 2 = "Neutral", 3 = "I like it"	Subjective, related to user's unquantified goals and constraints in $SB_{int}$

Comparison of results across participants was done by relating the average of re-scaled OF values calculated using equation 6.2

$$Ratio_{OF_{mkj}} = Average \left( \frac{OF_{mlkj}}{Max(|OF_{mlkj}|)} \right) \quad (6.2)$$

Where  $Ratio_{OF_{mkj}}$  is the average value of the re-scaled physical objective function  $OF_{mlkj}$  in  $SB_{int_l}$  for the  $k^{th}$  user.

### 6.4.3. Decision space patterns

Similar to the objective space, the decision space was analyzed for all user-preferred design alternatives (*user rating*  $R_3$ ) at the End of the Search (EoS). Two variables were used to represent the different BMPs in the decision space: binary variables and real variables. In this study, the binary variables were used to represent the implementation (1) or non-implementation (0) of the cover crop BMP, whereas the real variables represented the filter strip width of the Filter Strips BMP. Piemonti et al., (2013), as well as Chapter 5, give detailed information about how the decision variables were incorporated in the optimization algorithm, and how they were used in the hydrological simulation model. In summary, design alternatives are transformed into a vector of decisions (also called chromosome in Genetic Algorithm terminology) of length  $CP*108$ , where CP is the number of conservation practices to be optimized in the watershed. Each gene ( $j$ ) inside the chromosome represents a specific SB, and the BMP (or, conservation practice) decision proposed for that SB.

#### 6.4.3.1. Binary variables

As mentioned above, cover crops are represented using binary variables. Hence based on frequency of implementation of cover crops (CC) in every sub-basin (SB), a probability of implementation of the CC ( $Prob_{jR3k}$ ) in all the design alternatives with user rating  $R_3$  was calculated using equation 6.3

$$Prob_{jR3k} = \frac{\sum_{l=1}^{L_{3k}} ccimp_{j3lk}}{L_{3k}} \quad (6.3)$$

where  $ccimp_{j3lk}$  is the binary variable value of the  $j^{th}$  gene that represents the CC decision in one of the SB in  $SB_{int}$ , in the  $l^{th}$  design alternative, and for the  $k^{th}$  participant.  $L_{3k}$  is the total number of designs rated  $R_3$  by the  $k^{th}$  participant. This metric allows determination of the percentage of design alternatives that include CC in the  $SB_{int}$ . The value is interpreted as the probability that the participant is likely to implement the conservation practice.

We also calculated the average  $Prob_{jR3k}$  for each of the groups of participants focused on the same  $SB_{int}$ . Such a group metric has the potential to provide an overall preference for cover crop in a sub-basin, across all users interested in the sub-basin. Equation 6.4 shows how this group metric was estimated:

$$AvgProb_{jR3} = \frac{\sum_{k=1}^N Prob_{jR3k}}{N} \quad (6.4)$$

Where,  $N$  is the total number of participants (in each group)

#### 6.4.3.2. Real variables

In this study, the real decision variables represent the filter strip width for the conservation practice filter strip (Chapter 3). The Mode of Filter Strip width calculated from values of the gene  $g$  representing the filter width for a specific SB in  $SB_{int}$  (see chapter 5 for details), and the Percentage Of Mode in solutions ( $POM_{gR3k}$ ) based on the percent of designs rated  $R_3$  that have the value of gene ( $g$ ) equal to the calculated mode  $g$  for those design alternatives. Equation 6.5 was used to calculate the  $POM_{gR3k}$

$$POM_{jR3k} = \frac{DEM_{jR3k}}{L_{3k}} * 100 \quad (6.5)$$

where  $DEM_{gR3}$  is the number of design alternatives with filter width design in the  $g^{th}$  gene equal to the calculated mode for each  $k^{th}$  participant,  $L_{3k}$  is the total number of designs rated  $R_3$ .

We also calculated the  $GPOM_{gR3}$  (Global Percentage Of Mode). This value represents the percentage of participants who are associated with a specific SB of interest and have the same Mode of Filter Strip. Equation 6.6 shows the parameters used for this metric. Before applied 6.6 we calculated the mode (most repeated value) across the participants, then:

$$GPOM_{jR3} = \frac{PDEM_{jR3}}{N} * 100 \quad (6.6)$$

Where  $PDEM_{gR3}$  is the number of participants with mode equal to the mode, and  $N$  is the total number of participants

## 6.5 Results and discussion

In this section, we have examined the potential effects of spatially-explicit user preferences for cover crops, filter strip widths, their locations, and their performance with respect to physical goals in  $SB_{int}$ , on the overall outcome of the search (i.e., End of Search or EoS). In Sub-section 6.5.1, we present results from the user experiments by examining the performance of the user-preferred ( $R_3$ ) designs in the Objective Space relevant to the sub-basins in  $SB_{int}$ . And, finally, in Sub-section 6.5.2, we examine the results from the same user experiments for possible patterns in the decision space of SB that indicate potential spatially-explicit preferences in design parameters.

### 6.5.1. Objective Space Analysis

Results in chapter 5 showed that there was not a very clear pattern in the objective space of preferred design alternatives at watershed scale, for those participants who were only interested in a specific set of SBs in  $SB_{int}$ . Note that during experimental setup process, researchers had assigned only specific set of SBs to these participants. Hence, this is also an indication that these users overlooked the performance of the objective function at the watershed scale during their personal decision-making process.

To assess for differences between participants in the sub-basin scale objective space (e.g., cost of implementing the prescribed practice in a specific SB), we generated

histograms that provide information on the number of designs within a given objective function bin size (Equation 6.1). The bin size varies based on the SB, objective function (OF), and participant. We noticed that, in general, participants interacting with Model B have relatively similar number of design alternatives in each bin of objectives functions  $OF_1$ ,  $OF_2$  and  $OF_3$  (regardless of whether the user is a Surrogate or a Stakeholder). However, it is possible to identify the range of performance (indicated by OF) that the participant may be more interested in, by identifying values of OFs associated with the majority of user-preferred designs. It is relevant to note that in majority of the participants'  $R_3$  designs, the values of  $OF_4$  (i.e. cost) were in the low range for majority of the sub-basins in  $SB_{int}$ . For example, for Participant 2 ( $SB_{int}$  shown in Figure 6.1) majority of the  $R_3$  designs in three (i.e. IDs 10, 11, and 14) out of four sub-basins have the values of performance  $OF_4$  in the low range. However, in SB 15 this is not true, indicating the possibility that this participant may not be concerned about the cost in SB 15. We also noticed that the other participants interested in this SB (Participant 9 and 16) also presented similar behaviors (even though they worked with a different simulation model: *Model B*). Therefore, there is a chance that this effect could be a consequence of the search algorithm and its exploration/exploitation process. Figure 6.1 also shows the correlations between the objective function values of each of the SBs. For 8 out of 23 SBs in  $SB_{int}$  groups (Table 6.1) a high correlation in their performance in the objective space was observed. We noticed that these 8 SBs were hydrologically connected SBs, indicating a relationship between hydrologic connectivity and sub-basin scale OF performance.

Figure 6.2 show a set of re-scaled plots (equation 6.2) that summarize the similarity in the performances of OFs between participants who were focused on the same set of sub-basins in  $SB_{int}$ . These plots show how close the absolute values of the average values of each objective function, (i.e.,  $Absolute(Average(OF_m))$ ) are to the absolute values of their maximum values (i.e.,  $Absolute(Max(OF_m))$ ), in every SB of  $SB_{int}$ . The results are consistent across all participants for  $OF_1$ ,  $OF_3$ , and  $OF_4$ . However, out of the 23 SBs in  $SB_{int}$ , two of the SB (SB15 and SB44, i.e., 9%) have preferred average values that are approximately half of the maximum value. Additionally, for some of participants the scaled average values in  $OF_4$  of their SB are contradictory to what other participants showed.

Specifically, in SB 38 (Participant 4), SB 63 (Participants 5 and 13), and SB 105 (Participants 6, 14, and 21), the average value of  $OF_4$  in the  $R_3$  designs have a positive cost value, contrary to negative values in other sub-basins and other participants. This indicates that, on an average, these participants liked design alternatives in these specific sub-basins of interest even if they do not get any revenue (i.e., cost) for their decisions in these sub-basins. An examination of the user behavior of these participants indicates that their confidence levels are either absent (Participants 4 and 5) or negative (Participants 13). However, the interaction with the system varies depending on the participant. This interaction shows that even when they are gathering information, their confidence levels in the design alternative's rating decrease.

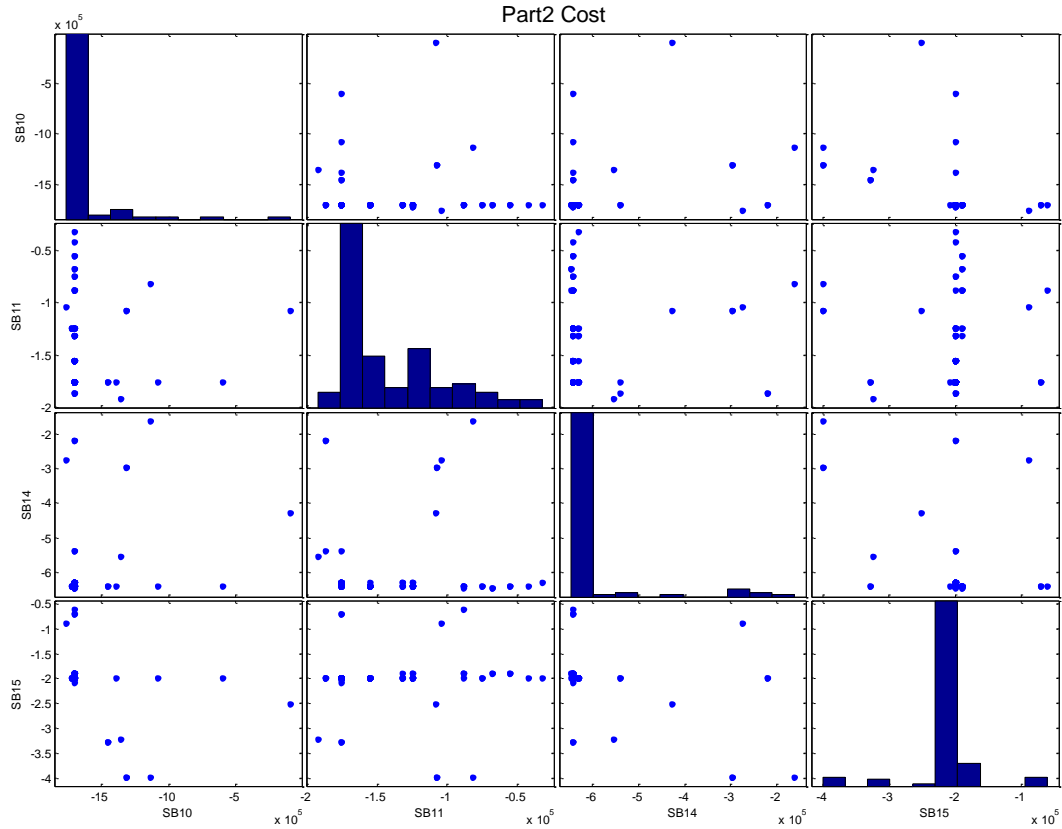


Figure 6.1 Histogram and correlations of SBint for Participant 2

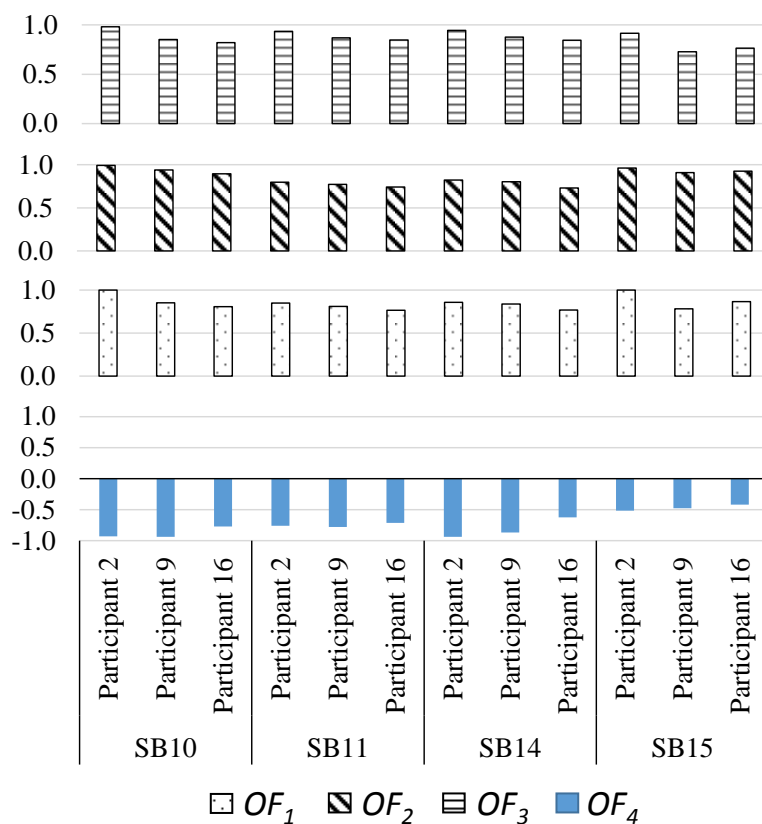


Figure 6.2 Group analysis of the Objective Space in for the participants in the  $SB_{int}$

### 6.5.2. Decision Space Analysis

For the analysis of the decision space of user-preferred design alternatives, the spatial location of SBs of interest ( $SB_{int}$ ) in the watershed was taken into consideration for each participant. Two  $SB_{int}$  groups (Groups 1 and 2) are located in the northern (or upstream) region; three groups (Groups 3, 4 and 5) were located in the mid region, and one (Group 6) was located in the southern region (or downstream) close to the reservoir. Each group has at least 2 hydrologically connected SBs. As mentioned earlier, any two SBs are considered hydrologically connected when their IDs are consecutively numbered and/or a stream flows through them. The unconnected SBs are those in the same region, but which stream does not drains to (or come from) an adjacent SB.

#### 6.5.2.1. Binary variables (Cover Crops)

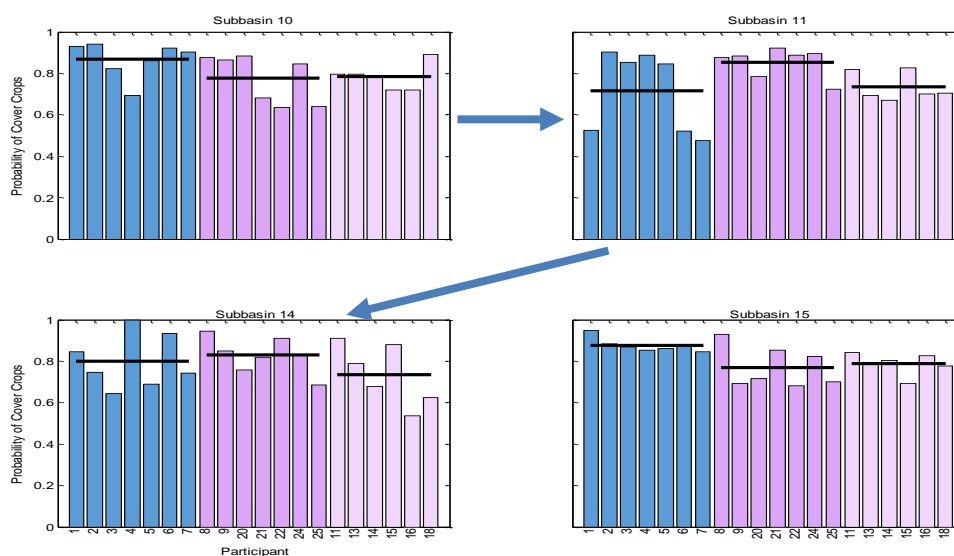
For each  $SB_{int}_q$  and each participant, we calculated and produced plots as the ones showed in Figure 6.3. Such a plot is useful to understand the  $Prob_{jR3k}$  for every participant (whether or not she/he was interested in specific sub-basin in  $SB_{int}$ ) and in every SB. The

blue arrows in Figure 6.3 represent the flow direction (and connection) of the SBs. Though Figure 6.3. shows results for only the  $SB_{int}$  with SB IDs 10, 11, 14, and 15, the majority of the participants (80%) have a  $Prob_{jR3k} > 0.5$  in all the SBs, regardless of which the group they belong, and all of the participants have a  $Prob_{jR3k} > 0.5$  in at least one SB. When  $Prob_{jR3k} < 0.5$ , it indicates that CC is implemented in the SB (associated with the  $j^{th}$  gene) for less than half of the  $R_3$  design alternatives ( $L_{3k}$ ).

All the participants have at least one SB where their preference for cover crops has  $Prob_{jR3k} < 0.5$ . For example, Participant 7 has a  $Prob_{jR37} < 0.5$  in 17% of the total SBs of interests. Further, cover crop was least desired in one of the SBs (SB103) since 61% of participants had their  $Prob_{jR3k} < 0.5$  in this SB. From the perspective of participants, Participant 25 is the participant with more SBs with a  $Prob_{jR325} < 0.5$  (with 7 out of 23 SBs), meaning that this participant did not desire the implementation of CC in the different SBs of interest. However, all his/her SBs that he/she was personally interested in (Group 4) the  $Prob_{jR325} > 0.5$ . Also, this participant has no statistically significant trend in confidence levels, and had very few interactions with the GUI towards information gathering. Therefore, his/her data could be considered as a noisy data set, giving less credibility to his/her selections.

Even when the majority of participants have  $Prob_{jR3k} > 0.5$ , there is a group of 5 participants (Participants 5, 6, 13, 20, 21) with  $Prob_{jR3k} < 0.5$  in at least one of their assigned SBs of interests. Among these participants, only two have a positive trend on their confidence levels and a fair amount of interaction with the interface in gathering information about the designs (Participants 6 and 21). These two participants were focused on  $SB_{int_6}$  and agree that SB106 should not have a cover crop implemented. We can conclude that due to the level of interaction, and increase their confidence levels; their preferences are clear regarding the distribution of the CC implementation.





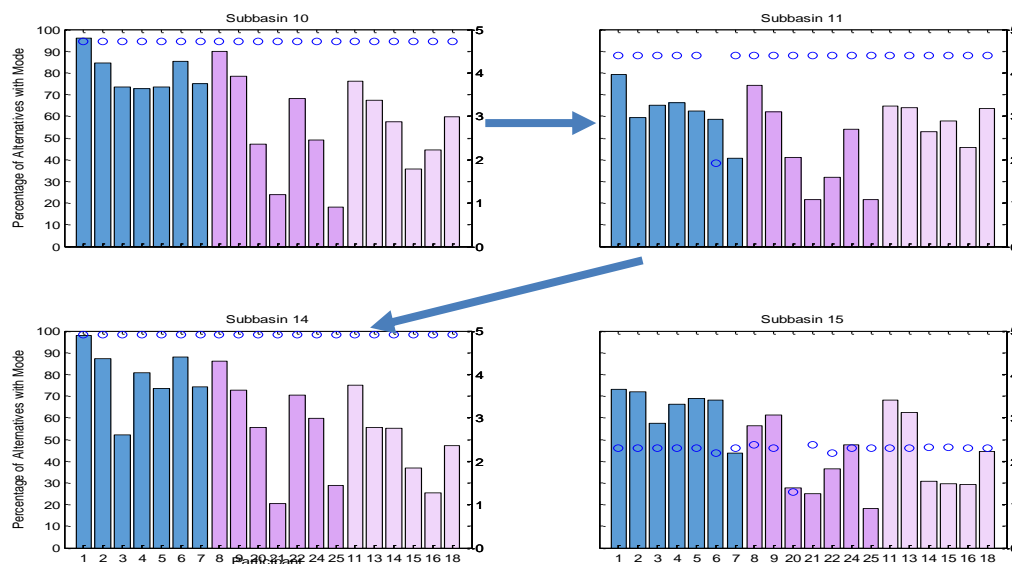
**Figure 6.3 Individual (bar plots) and average (black lines) values of probability of CoverCrops by SBint for Group 1**

In the evaluation of the averages by a group of participants, we observed that in 65% of the cases, Model A-Surrogates have a higher  $AvgProb_{jR3}$ , regardless if the SBs are located upstream or downstream. We also noticed that 6 out of 10 of the SBs in downstream showed probabilities of CC higher than for those SBs in their directly related upstream SB. Therefore, we can assume that the preferences and general performance of CC in the downstream region have a better effect in the reduction of the objective function a watershed scale, and are also preferred by the participants.

#### 6.5.2.2. Real variables (Filter Strip Widths)

The results for the real variable indicate that there is an agreement across most of the participants on the Mode values in the upstream region of the watershed. Figure 6.4 shows an example of these results. In this outcome  $GPOM_{jR3} = 100\%$  in SB10 and SB14, even if they were not concern about the design or performance in this SBs, the participant advertently (if they were focused on these SB IDs) or inadvertently (if they were not focused on these SB IDs) found design alternatives with similar filter width. Notice, for example, Participant 1 (a watershed-focused participant) has a  $POM_{jR3}$  very close to 100%, indicating that almost all the design alternatives that were rated  $R_3$  have the same value for filter strip width. In average, 19 out of the 23 SBs of interest present a  $GPOM_{jR3} > 50\%$ ,

meaning that there is a good agreement regarding the design of the filter strip width, even if the participant is not interested in the region.



**Figure 6.4 Example of the mode and percentage of the filter strip width for each participant, in SBint1**

However, we notice that there were ten SBs of interest, where the participants focused on the areas do not agree with the filter strip width selected by the majority of the participants. This is an evidence of how participants concentrated in a region may disagree with what is considered optimal for the rest of the community.

## 6.6 Conclusions and future work

The complexity of interactions in a watershed is not just limited to the relationship between the human and the natural system. Community interactions and agreements during decision making and collaboration are just as important, in order to identify acceptable watershed management plans. Understanding how participatory decision support systems (DSS) address the preferences of each participant give us a sense of how DSS environments can be improved. Such improvements can yield an adaptive design environment that is able to provide more opportunities for stakeholder-driven generation of scenarios for spatially explicit problems.

This work helped to understand different preferred design alternatives that meet the criteria of a group of individuals focused on different sub-regions in the watershed. This

understanding can support the negotiation and implementation of practices, as well as educate the stakeholders about the consequences of their selection and how it affects the other stakeholders. The safe and private environment for providing information about their preferences through a web-based outreach tool is a significant opportunity to support collaboration between stakeholders in participatory watershed planning processes. The WRESTORE platform is a web-based DSS tool that can be used at any time and place. Here, the initial explanation of the tool and how to use it was held in a classroom, but the actual experiment was performed at the participant's best convenience.

Knowing the developed and accepted patterns, as well as its uncertainty levels would help to improve the existing problem of iterative designs, not just because it can help designers to center in individual problems, but it also will allow the search algorithm to focus on the essential stakeholder-centric characteristics of the decisions, rather than to improve the designs based on only the watershed scale perspective. From this study, we can conclude that it is possible to identify the preferences at a local scale through histograms of the objective performances. This information (that is not currently used by the tool) could be added to the system to improve the interactive optimization algorithm's ability to find solutions that satisfy the user's preferences for certain types of decisions and/or criteria.

The average binary pattern allows us to compare the users even if their SBs of interest are different. It is a useful metric that can be used towards negotiation processes, because it determines the level of agreement in community to implement this practice. The real variable pattern also can offer insight into what values are more preferred in a design alternative. We notice that upstream designs are very consistent across participants. However, the designs of the downstream elements are in disagreement with the plans that the participant focused in the area may have for the land. This suggests that those participants who are concentrated on the local sub-regions, e.g.  $SB_{int}$  in this study, may be more affected by the location of their regions in their decision to agree or disagree with specific values for filter strip width.

## 6.7 ACKNOWLEDGMENT

We want to thank the funding agencies: National Science Foundation (Award ID # 1014693). Also we would like to especially thank to all of our collaborators from the different agencies and institutions: USDA, EPA, IUPUI and OSU, as well as all the participants that performed the test.

## 6.8 REFERENCES

- Ahmadi, M., Arabi, M., Fontane, D.G., and Engel, B.A., (2014), Application of multicriteria decision analysis with a priori knowledge to identify optimal nonpoint source pollution control plans, *J. Water Resour. Plann. Manage.*, 141(2), 1-9.
- Arabi, M., Govindaraju, R. S., and Hantush, M. (2006), Cost-effective allocation of watershed management practices using a genetic algorithm, *WRR*, 42, W10429.
- Artita, K., Kaini, P., and Nicklow, J. W. (2008), Generating alternative watershed-scale BMP designs with evolutionary algorithms, paper presented at the World Environmental and Water Resources Congress, Ahupua'a, HI.
- Babbar-Sebens, M., Barr, R.C., Tedesco, L. P., and Anderson, M. (2013), Spatial identification and optimization of upland wetlands in agricultural watersheds, *Ecol. Eng.*, 52, 130-142.
- Babbar-Sebens, M., Mukhopadhyay, S., Singh, V. B., and Piemonti, A. D., (2015), "WRESTORE- A web-based planning software for participatory optimization of conservation practices in watersheds," *Environ Mod and Soft.*, 69, 111-127.
- Cash, D.W., Clark, W.C., Alcock, F., Dickson, N., Eckley, N., and Jäger, J., (2002), *Salience, Credibility, Legitimacy and Boundaries: Linking Research, Assessment and Decision Making*, RWP02-046, J.F. Kennedy School of Government, Harvard University, Cambridge, MA.
- Cole, R.A., Feather, T.D., and Letting, P.K., (2002), Improving watershed planning and management through integration: a critical review of federal opportunities (USACE IWR Report 02-R-6). Alexandria, VA: US Army Institute for Water Resources.
- Evers, M., Jonoski, A., Maksimovič, C., Lange, L., Ochoa, S., Teklesadik, A., Cortes, J., Almoradie, A., Simões, N., Wang, L., and Makropoulos, C., (2012), Collaborative modelling for active involvement of stakeholders in urban flood risk management. *Natural Hazards and Earth System Science*, 12(9), 2821–2842.
- Gregersen, H.M., Ffolliott, P.F., and Brooks, K.N. (2007), *Integrated Watershed Management: Connecting people to their land*, Oxfordshire, UK: Cambridge University Press.
- Greiner, R., Patterson, L., and Miller, O., (2009), Motivations, risk perceptions and adoption of conservation practices by farmers, *Agricultural Systems*, 99, 86-104.
- Hoag, D., Luloff, A. E., and Osmond, D. L., (2012), *Lessons Learned from the NIFA-CEAP: How Farmers and Ranchers Make Decisions on Conservation Practices*, NC State University, Raleigh, NC.

- Johnson, N., Ravnborg, H.M., Westermann, O., and Probst, K., (2002), User participation in watershed management and research, *Water Policy*, 3, 507-520.
- Kaini, P., Artita, K., and Nicklow, J. W. (2012), Optimizing structural Best Management Practices using SWAT and Genetic Algorithm to Improve Water Quality Goals, *Water Resour. Mgmt.*, 26, 1827-1845.
- Kelly, M., Ferranto, S., Lei, S., Ueda, K., and Huntsinger, L., (2012). Expanding the table: the web as a tool for participatory adaptive management in California forests. *J. Environ. Mgmt.* 109, 1-11.
- Lethbridge, M.R., Westphal, M.I., Possingham, H.P., Harper, M.L., Souter, N.J., and Anderson, N. (2010), Optimal restoration of altered habitats, *Environ. Model. & Soft.*, 25(6), 737-746.
- Lubell, M.N., (2004), Collaborative watershed management: a view from the grassroots. *Policy Stud. J.*, 32,321–341.
- Mintzberg, H., and Westley, F., (2001), Decision making: It's not what you think, *MIT Sloan Management Review*, 42(3), 89–93.
- Perez-Pedini, C., Limbrunner, J.F., and Vogel, R.M. (2005), Optimal location of infiltration-based best management practices for storm water management, *J. of Water Resour. Plan. And Mgmt.*, 131(6), 441-448.
- Piemonti, A. D., Babbar-Sebens, M., and Luzar, E. J. (2013), Optimizing conservation practices in watersheds: Do community preferences matter? *WRR*, 49(10), 6425-6449.
- Piemonti, A.D., (2014), Correlating Human Preferences and Optimized Watershed Management Plans. Proceedings of the World Environmental and Water Resources Congress 2014: pp. 1961-1969
- Prokopy, L.S., Flores, K., Klotthor-Weinkauf, D., and Baumgart-Getz, A., (2008), Determinants of agricultural best management practice adoption: Evidence from the literature. *J. of Soil and Water Conservation*, 63(5), 300-311.
- Randhir, T.O., Lee, J.G., and Engel, B. (2000), Multiple criteria dynamic spatial optimization to manage water quality on a watershed scale, *Trans. of American Society of Agricultural Engineers*, 43 (2), 291–299.
- Seppelt, R., and Voinov, A., (2002), Optimization methodology for land use patterns using spatially explicit landscape models, *Ecological Model.*, 151, 125-142.

**CHAPTER 7. Human-centered design of spatial allocation of conservation practices in watersheds: Mining user responses to adapt search operators of an Interactive Genetic Algorithm**

**Adriana D. Piemonti<sup>1</sup>, Meghna Babbar-Sebens<sup>1</sup>, Snehasis Mukhopadhyay<sup>2</sup>, and Mahesh Yerram<sup>2</sup>**

<sup>1</sup> School of Civil and Construction Engineering, Oregon State University, Corvallis, Oregon, USA.

<sup>2</sup> Department of Computer and Information Sciences, Indiana University-Purdue University Indianapolis, Indianapolis, Indiana, USA.

To be submitted to the Journal of Water Resources Planning and Management.

## 7.1 Abstract

Over the last few years, incorporation of human stakeholder participation in modeling and multi-objective optimization of watershed planning and management problems has gained an increasing amount of interest. Interactive Genetic Algorithm (IGA) is one such participatory design technique that enables users to participate in the generation of desirable design alternatives by guiding the search process of the algorithm via a user feedback mechanism. However, IGA can suffer from poor convergence and performance in identifying “user-preferred” alternatives, especially if, for example, the decision space is enormous, users preferences are biased by the values of only a sub-set of decisions in addition to (or, in some cases, instead of) physical objective functions, and if the user’s feedback on the quality and acceptability of design alternatives is not always consistent. Here, we propose a modification of exploration-exploitation operators (e.g., selection, crossover, and mutation) used in an Interactive Genetic Algorithm to improve the convergence rates of IGA. These proposed operators, called Adaptive Human-guided Search (AHS) operators, use emerging patterns in the decision space and feedback from the participants to adapt the values of the search operators. The main purpose of this adaptation is to re-direct the search towards areas in the decision space that are preferred by the user. To evaluate this new approach, we created six simulated participants (SP) (three deterministic users and three stochastic users) based on a set of previous experiments with three real participants. The three human participants were short-listed from an original group of 20 participants after a careful selection process to identify users with most reliable feedback. The selection process was based on their confidence levels trends, and their interaction behavior on the graphical user interface supported by the Interactive Optimization software. Convergence was measured using the percent of design alternatives in the Genetic Algorithm population that had *user rating* (i.e., the user’s feedback) classified by the participants as “I like it”. Results show that for one of the deterministic SP, there is a convergence towards a 100% of designs after just 10 generations. However, the results with the stochastic SP does not show this convergence in any of the cases, This effect is attributed to the procedure selected to give one rate to the entire design, based on the combination of the *user rating* of the subbasins (SBs) of interest.

## 7.2 Introduction

Extensive research has been conducted previously to examine and improve optimization techniques used for water resources planning and management problems (Goldberg and Kuo, 1987; Seppelt and Voinov, 2002; Nicklow et al., 2010). However, when optimal or “near optimal” solutions are presented to the different users, it has been very challenging to persuade an individual (or community) to adopt and implement the proposed designs. Therefore, there is a growing interest on methods, in fields such as and Integrated Watershed Management, that support the inclusion of human stakeholders in the generation of design alternatives for land, water and soil management (Cole et al, 2002; Gregersen et al., 2007; Margerum and Robinson, 2015).

In Interactive Optimization methods (e.g., Interactive Genetic Algorithm (IGA)), users participate in the generation of *satisficing* (*satisfy* and *suffice*; Simon, 1956) design alternatives by guiding the search process of the underlying optimization algorithm via their feedback on the desirability of the emerging design alternatives. The feedback is collected via a parameter called *user rating*, and the user uses a graphical user interface (GUI) to provide values for *user rating* on a psychometric scale (e.g., Likert-type scale). However, little is known about how users determine values for *user rating* in a typical watershed optimization problem, and how that could influence the search performance of the optimization technique. This is because the user’s cognitive process to decide a value for *user rating* may be influenced by not only the performance of the design alternative, but also by the decisions proposed in the design alternative itself, and her/his biases to these values. Without the knowledge on a specific user’s rationale for *user rating*, it can be extremely challenging to implement Interactive Optimization methods in an efficient manner, especially when the algorithm has to navigate through large multi-dimensional decision spaces in order to identify user-preferred design alternatives, and with limited human interaction and minimum human fatigue.

In an IGA, the creation of and search for new design alternatives in decisions spaces are controlled by the selection, crossover, and mutation operators. A large variety of crossover and mutation operators has been developed in the literature (Gen and Cheng, 2000; Gong et al., 2005; Gong et al., 2008), and may be useful in an IGA. In this study, we focus on



the Interactive Genetic Algorithm called IGAMII (Babbar-Sebens and Minsker, 2011), which is embedded in a participatory watershed planning tool called WRESTORE (Babbar-Sebens et al., 2015). Currently, IGAMII uses constant probability values for its uniform crossover operator and mutation operator, respectively, to generate a new population in every iteration of the search. The crossover probability determines the chance with which a gene is allowed to exchange its value with a gene in a different chromosome of the population, whereas the mutation probability determines a low chance modification of a gene to a new random value. The uniform crossover assigns the constant crossover probability to every gene (i.e., individual decisions in binary or real number format) in population's chromosomes (i.e., a string of genes for multi-dimensional decisions). In the previous work (in chapters 5 and 6) on experiments with humans, who used Interactive Optimization to design plans for conservation plans in a watershed, it was illustrated that when users and the search algorithm collaborate to generate new design alternatives patterns in the decision space of design alternatives can emerge. These patterns indicate (a) what values of decision variables in her/his local area of interest (sub-basin scale) agree with a user's personal preference, and (b) how should values of decision variables in surrounding sub-basins be adapted by the search algorithm in order to maintain (or, improve) the overall performance of the entire decision space. Therefore, if the crossover and mutation operators could be adapted to preserve such emerging patterns in the design alternatives that are rated by a user as being desirable (i.e. "I like it" in this study), then the search process could be improved in the Interactive Genetic Algorithm to yield faster convergence rates; this hypothesis is examined and tested in this study.

### **7.3 Objectives**

This overall objective of this study is to develop and test human-guided search operators that adaptively learn for patterns in user-preferred alternatives generated by the Interactive Genetic Algorithm, and, as a result, improve the convergence rate of the search algorithm for generating design alternatives that conserve these learned patterns. The improved convergence rate is also expected to assist with the user fatigue issue due to the fewer number of generations that would be required to conduct the evolutionary search process, and at the same time find solutions that agree with the users' criteria.

To evaluate the performance of the Interactive Genetic Algorithm with the proposed modified operators, six (three deterministic and three stochastic) simulated users were developed by using the user feedback data gathered in a previous research experiment. The following questions were investigated:

- 1) What is the average rate of convergence? Using simulated user models, we examined the changes in percentage of preferred designs (“I like it”) when the search was conducted using the default search operators versus the adaptive human-guided search operators. This comparison serves as an initial test to prove the ability of the proposed operators to facilitate faster convergence to desired solutions.
- 2) At what rates to average probabilities of crossover and mutation change in each generation for the algorithm using the adaptive operators? The goal of the interactive genetic algorithm with the adaptive human-guided operators is to preserve features that are repeated in user-preferred design alternatives. When the operators change over time, it allows us to examine under what conditions and at what rate do the AHS operators begin to detect patterns in the decision space and conserve them.
- 3) How do the results of the adaptive human-guided operators vary, based on different types of users? Six types of simulated users (three deterministic and three stochastic) were developed and used to test the efficiency of proposed operators.

## **7.4 Methodology**

### **7.4.1. Interactive Optimization Algorithm**

WRESTORE’s interactive optimization method is an adaptation of the Interactive Genetic Algorithm with Mixed-Initiative Interaction (IGAMII) that was originally developed by Babbar-Sebens and Minsker (2011). IGAMII’s selection and search operators are based on the operators used in Non-dominated Sorting Genetic Algorithm (NSGA-II), proposed by Deb et al., (2002). In IGAMII, the objective function *user rating* is used as an additional objective function (in addition to physical objective functions), requiring that, in each generation, the user provides with a value for this rating for each of the design alternatives. Therefore, the total number of design alternatives (that represents the population of an IGA) presented to the user every interactive session, and the number of human-guided search sessions (that represents the micro-IGA number of generations)

must be carefully selected to avoid user's fatigue. User's fatigue (due to routine processes or long tasks) can affect the accuracy of the user's feedback data (Llora et al, 2005; Gong et al, 2008), introducing noise in the search, and potentially misleading the optimization process. Hence, because of limited population size and number of generations, the default search operators (i.e. selection, crossover, and mutation operators of NSGA-II) tend to be effective for only interactive optimization problems with few decision variables (or, short chromosomes).

#### **7.4.3. Proposed Adaptive Human-Guided Search (AHS) Operators**

In this research, we developed a modification of the search operators (crossover and mutation), and conducted test experiments to evaluate and compare the convergence rates with respect to the baseline (or, default) operators in IGAMII. Convergence rate is based on the percent of the algorithm's population that consists of design alternatives that meet the user's criteria. Figure 7.1 shows the flowchart of the Interactive Genetic Algorithm with Adaptive Human-Guided Search operators. The overall process for proposed algorithm is generally the same as followed by the NSGA-II. However, in each generation, the probability of crossover ( $P_{cross_{gi}}$ ) and the probability of mutation ( $P_{mut_{gi}}$ ) are recalculated and updated based on the *user rating* objective function values provided by the human.

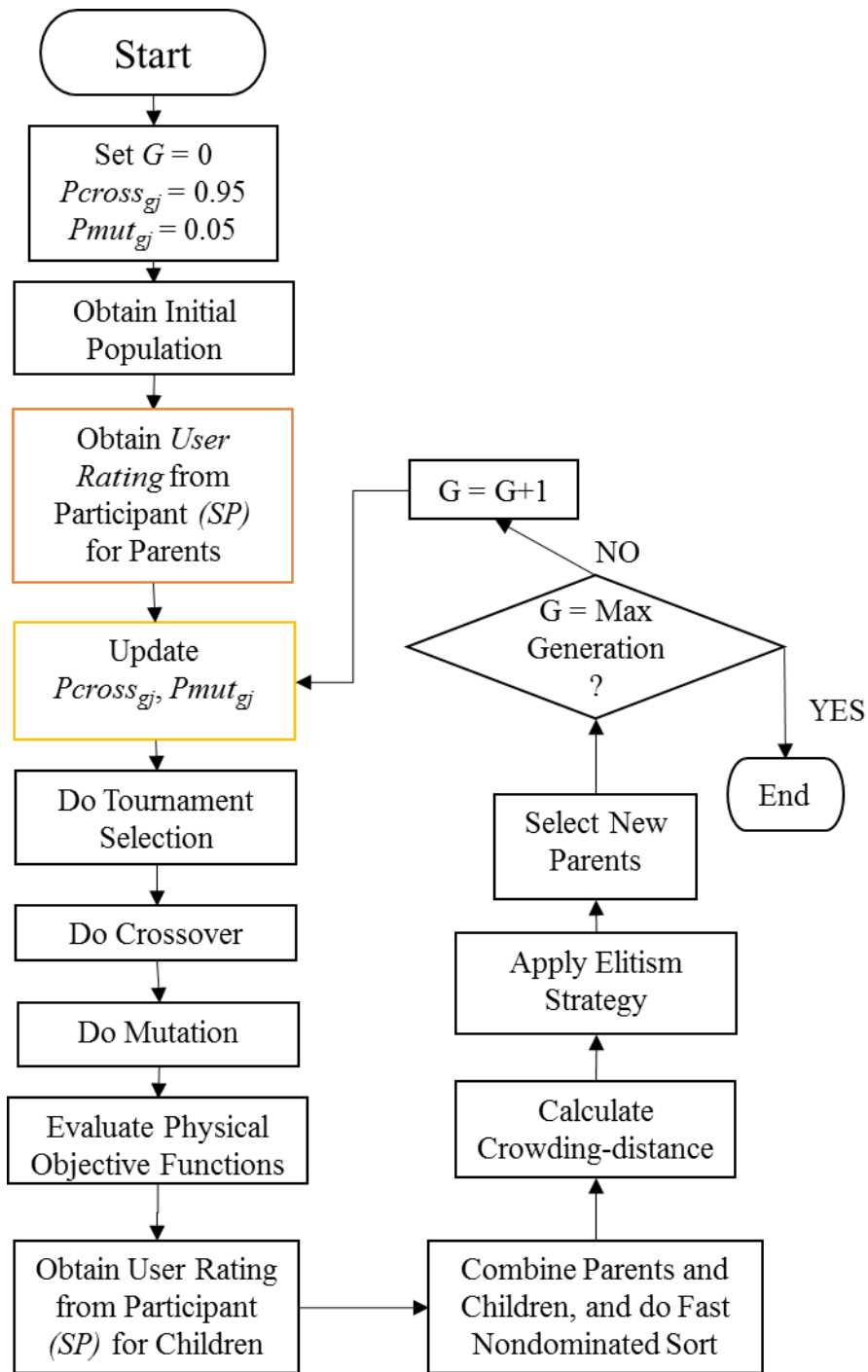


Figure 7.1 AHS flowchart

The following sections describe: 1) how the simulated participants (SP) were generated to estimate *User Rating* for testing the proposed operators, 2) how the crossover and mutation operators are updated in the proposed method, and 3) which metrics was used to

compare the results of proposed and baseline operators for each of the simulated participants.

#### 7.4.2. Simulated Participants (SP)

To test the performance of the modified operators, we designed six different simulated participants (SP). These SP were generated by using the information provided by real participants in previous experiments. In these previous experiments with real users, each participant had been assigned a group of three to five Subbasins (SBs) and had been asked to provide feedback via a Likert-type rating scale ( $R_1$  = “I do not like it” = 1,  $R_2$  = “Neutral” = 2,  $R_3$  = “I like it” = 3) and a confidence level for each of the design alternative proposed by the Interactive Genetic Algorithm (for more information regarding the feedback see Babbar-Sebens et al. 2015, and Chapters 3 and 5).

The task of the SP, in this study, was to predict the *user rating*, similar to the human participants, on the same Likert-type scale and for each of the new design alternatives generated by the interactive optimization technique. The model for simulating *user rating* was based exclusively on the decision space of assigned SBs in the design alternatives, and was built by using *user rating* data from actual experiments with three participants (Participant 3, 6 and 11). For each of these users, two different types of SP models were generated: deterministic SP and stochastic SP. The deterministic SP first estimated SB-specific values of *user rating* based on decisions in each SB that were in the human participant’s area of interest. The decisions in each of the SBs included combination of the conservation practices (or Best Management Practices, i.e., BMPs) prescribed by the genes of the Interactive Genetic Algorithm. The SB-specific *user rating* was estimated for every design alternative by using the following steps:

- 1) In  $SB_j$  (or SB of interest), count the number of total designs in the population with a particular design combination ( $DC_i$ ) where  $i = 1:4$ , and represent four different design combinations when cover crops and filter strips are chosen as candidate BMPs
  - $DC_1 = \text{YesCCYesFS}$  (Cover crop is implemented (value = 1) and FS width of equal to the mode of FSs),

- $DC_2 = \text{NoCCYesFS}$  (Cover crop not implemented and FS equal to the mode of FSs),
- $DC_3 = \text{YesCCNoFS}$  (Cover crop implemented and FS not equal to the mode of FSs), and
- $DC_4 = \text{NoCCNoFS}$  (Cover crop not implemented and FS not equal to the mode of FSs).

- 1) Notice that the value of  $i$  represents a combination vector of all the possible outcomes in the design. Hence, if we add an additional *BMP* the number of combinations that need to be considered will increase.
- 2) From the real users experiments, select all the designs rated  $R_3$ , and determine the design combination with the maximum number of the designs across the  $\max DC_i$ .
- 3) Use equation 7.1 to determine the scaled value of the count in each combination ( $ReScaledDC_i$ ). This scale value is a weight metric based on the number of designs that were classified in the design combination  $DC_i$

$$ReScaledDC_i = \frac{\text{num}(DC_i)}{\max(\text{num}(DC_i))} \quad (7.1)$$

- 4) Use Table 7.1 to determine the range of  $ReScaledDC_i$  value falls in and assign the corresponding rate.

**Table 7.1 Assignment of user rating based on scaled values**

Range	<i>user rating</i>
$0.66 < ReScaledDC_i \leq 1$	3
$0.33 < ReScaledDC_i \leq 0.66$	2
$0 < ReScaledDC_i \leq 0.33$	1

Because we are unaware of a user's internal process to select a final value for the user rating, we will assume that the participant used the average of *user ratings* of each of the SBs (in the participant's area of interest) to provide a single final *user rating* to the design. Hence, the final rating was determined to calculate the average user rating across SBs and rounded to the closest integer. Table 7.2 summarizes the *user rating* assigned for the four design combination ( $DC_i$ ) based on the implementation of the conservation practice.

**Table 7.2 Summary table of the selected rates in each SB based on the DCi**

SP	Subbasins	$DC_1$	$DC_2$	$DC_3$	$DC_4$
1	103	3	3	2	1
	105	1	3	3	2
	106	1	3	2	3
	121	3	3	1	2
	122	3	2	1	1
3	41	3	2	1	1
	90	3	2	1	1
	92	3	2	1	1
	93	3	2	1	1
5	12	3	1	3	2
	19	3	1	3	2
	20	3	1	3	2

The stochastic SP will provide the design *user rating* based on a probability ( $P_{jRi}$ ). This probability was calculated using equation 7.2

$$P_{jRi} = \frac{\sum_{l=1}^{L_i} DC_{li}}{L_i} \quad (7.2)$$

where  $P_{jRi}$  is the probability of design alternative ( $l$ ) to be rated  $R_i$  given a  $DC_{li}$  (design combination), and  $L_i$  is the total number of designs in  $R_i$ . Similar to the deterministic SP, the final *user rating* was determined by calculating the average *user rating* across all SBs in the real user's area of interest, and then rounded up to the closest integer.

### 7.4.3. Operators Update.

The operators are updated when patterns in design alternatives rated  $R_3$ , or preferred by the user, are present. The AHS is activated once the number of design alternatives rated  $R_3$  is at least half the number of designs presented to the participant. Figure 7.2 shows the first check and determination of activation of the AHS. This routine is called the activator, and will allow to access to process ①.

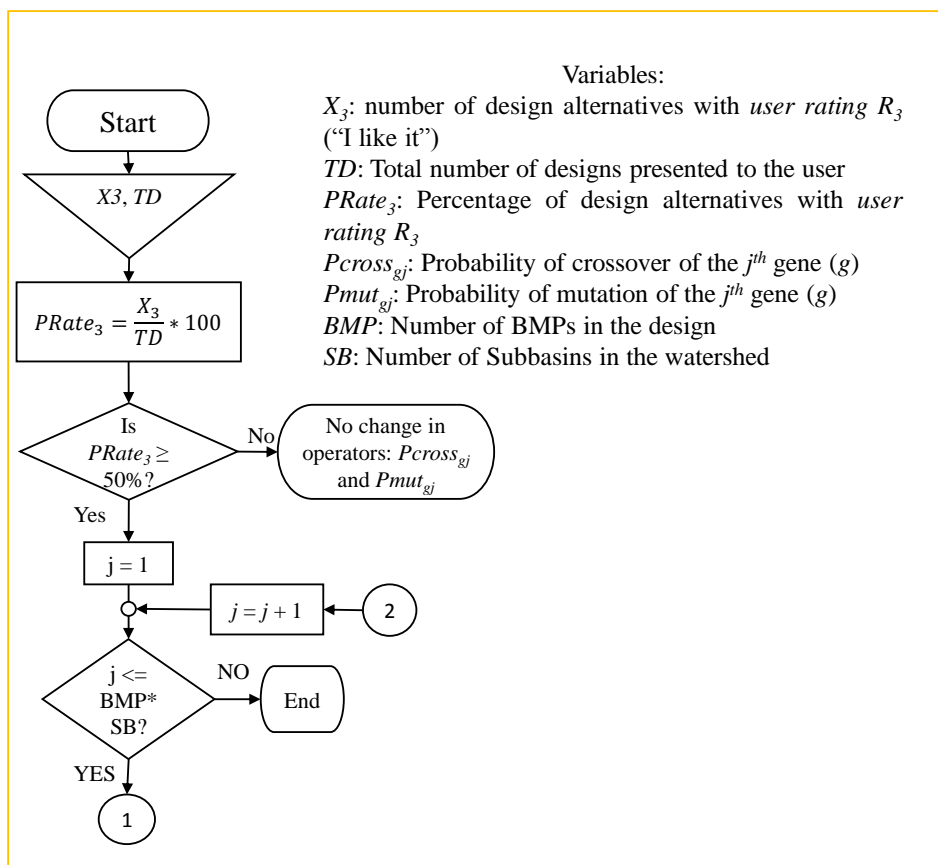


Figure 7.2 Flow chart for the activator of AHS

The update to crossover and mutation operators does not get activated unless at least 50% or more design alternatives are rated  $R_3$  in the population. Then, once it enters the update process (①) the following steps are executed for each of the genes in the entire chromosome:

- 1) Identify the type of variable in the gene,
- 2) identify patterns of the gene, using just those values that that were rated as preferred, and
- 3) modify the genes where a pattern is present in such a way that it will reduce the design.

These steps are represented in Figure 7.3 where the information provided shows those formulas used to calculate the pattern design. We expect that this update will generate a vector of  $Pcross$  and  $Pmut$  of the same size as the chromosome used in the decision space.



In this manner, just those genes presenting similitudes will be decrease their probability of changing in the following generation.

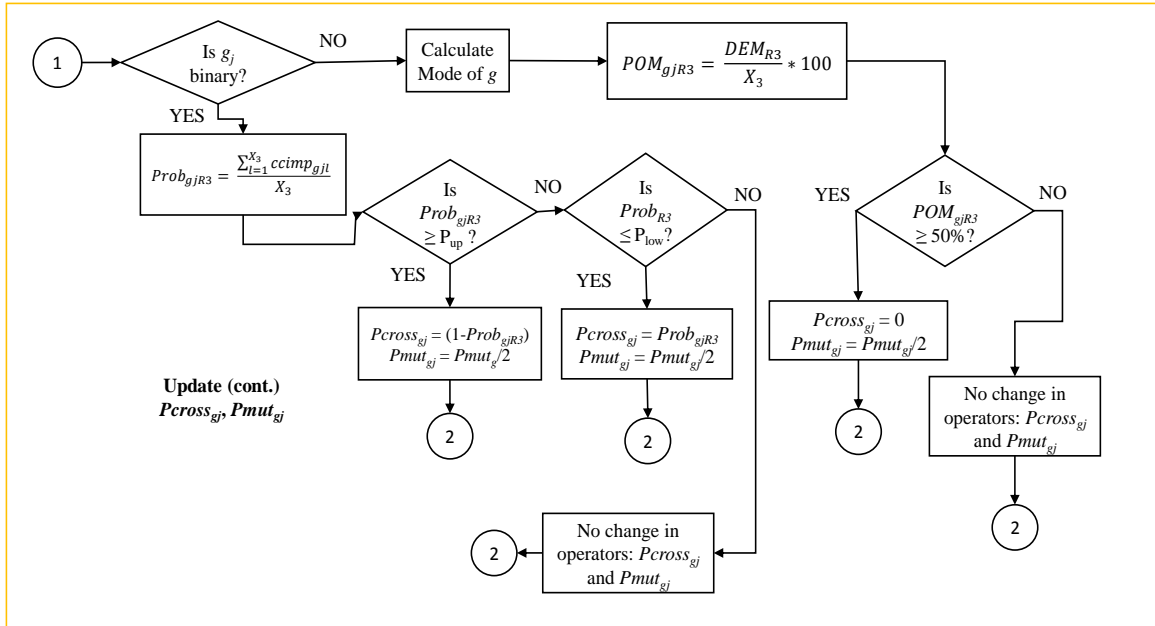


Figure 7.3 Internal process for the AHS after the activator have been set.

### 7.4.3. Metrics for Evaluating Convergence and Performance

This section describes the methods used for the evaluation and quantification of adaptive and default algorithms. We wanted to observe in which generation the design alternatives started to be rated in the  $R_3$  scale by the simulated user, and how it compares with the default parameters (i.e., without any change in the crossover or mutation rate). We also wanted to observe if there were changes in the objective space, and what were the final patterns in the set of SBs of interest ( $SB_{int}$ ) in each simulated user. The final observation is related to the generation where the operators start changing and what is the minimum value they reach.

#### 7.4.3.1. Percentage of user rating

In order to determine the percentage of designs at a rate, we used equation 7.3. This equation provides information of the changes in the rates in the different design alternatives. It is the indicator of the algorithm's converge to find solutions that satisfy the SP criteria.

$$PRa_{e_{kige}} = \frac{X_{kige}}{TD_e} * 100 \quad (7.3)$$

For this equation,  $PRate_{kige}$  is the percentage of solutions for the  $k^{th}$  SP for *user rating i* per generation ( $ge$ )  $X_{kige}$  is the number of design alternatives, and  $TD_{ge}$  is the total number of design alternatives in each generation. These values were calculated for the two types of SP (deterministic and stochastic) in order to observe the effect of a noisier dataset and how the algorithm reacts to the noise.

#### 7.4.3.2. Operators: Crossover and mutation

The operators' metric aims to understand the changes of the different generations in those SBs that are of interest to the user. In this work, we observed the changes through the generations for both operators. We calculate the average of all the  $P_{cross}$  and  $P_{mut}$  in the shown design alternatives. However, we give a close look to those  $SB_{int}$  used by the SP to select a rate for the design. If the algorithm is capturing patterns in the design, then there will be an exponential decrease of the  $P_{cross}$  and  $P_{mut}$ .

It is also important to understand at which generation is the change of the operators occurring, we could answer questions regarding how this changes may be associated with the percentage of user ratings, or they may establish the needed number of generations that allows to another process such as automated search, or a simple crowd of the desired alternatives.

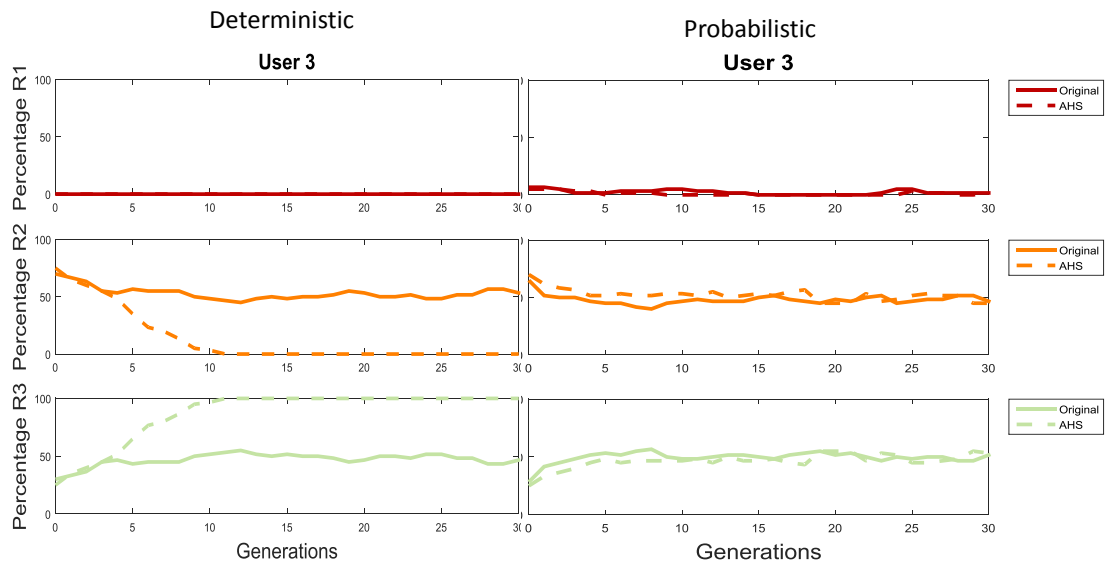
## 7.5 Results and discussion

In order to understand and provide insightful discussions and conclusions regarding the performance of the new modification, we decided to run two sets of the experiments, one using the default operators (referred as Default in the following subsections), and the other one using the new modification adaptive process (referred as Adaptive in the following subsections).

### 7.5.1. Percentage of user rating

Figure 7.4 shows the average result of the Simulated Participant (SP) User 3 for the percentage of user rating in each generation as an example of the behavior of the SPs. The left hand side represents the average for the deterministic SP, while the right hand side are

the results for the stochastic user. Two out of the three participants rated all the shown design alternatives  $R_3$  before the end of the generations (User 1 and User 3). However, User 5 does not seem to reach convergence of the designs rated  $R_3$ . Notice that the AHS has a rate of activation of 50% or more of the design alternatives rated  $R_3$ . This stage was never reached by the User 5.



**Figure 7.4 comparisons between the deterministic user and the Stochastic User for Participant 3**

Notice that the stochastic SP follows the default behavior pretty closely, indicating that there are not significant advantages of the AHS for this case. However, we could attribute this behavior to two different causes: 1) the noise related to the calculated probability of the selection of a rate in each SB, or 2) the fact that we assume the Users to be risk averse and calculate the average of each of the SBs in order to provide a unique rate for the design.

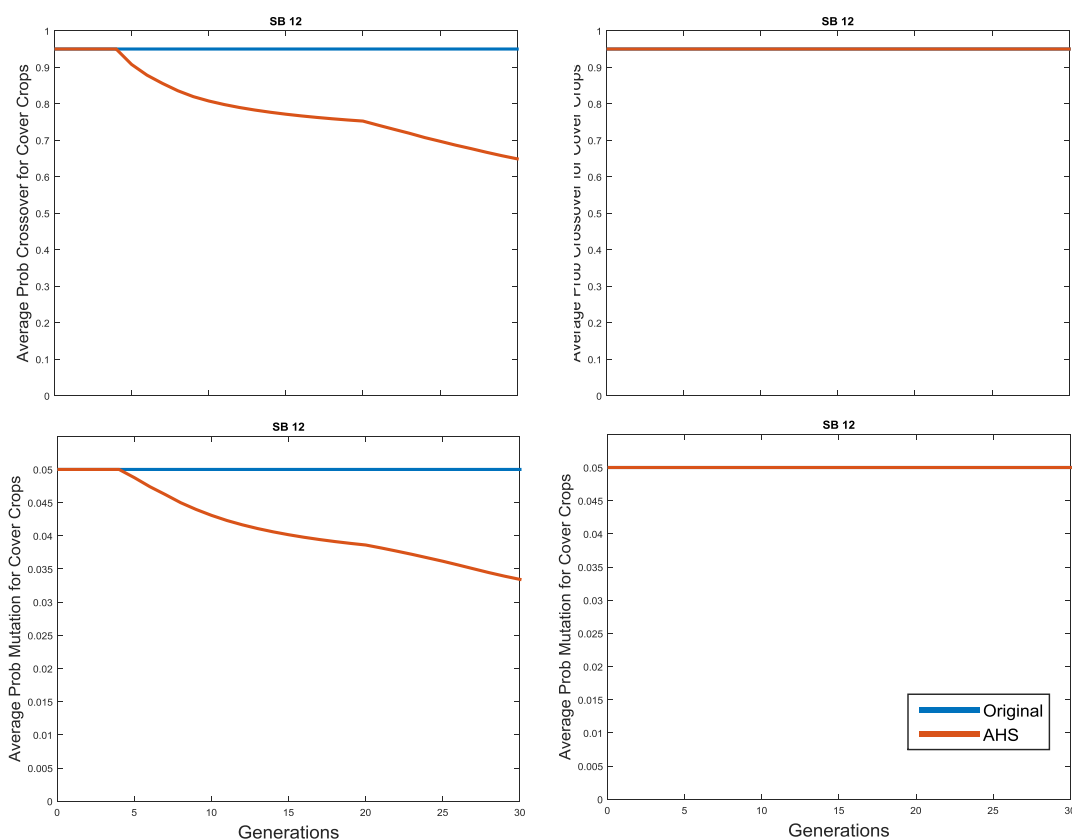
In either case, there is a need of a more exhaustive exploration of the selection of the final design alternative in order to determine the cases and exception where the technique should be use or avoid. Nevertheless, it can be say that for those users with a very predictable rating the implementation of AHS will allow a faster convergence to desirable designs.

### 7.5.2 Operators: Crossover and mutation changes

Figure 7.5 and 7.6 shows a typical behavior for the Probabilities of crossover and mutation for the SP. This example corresponds to SB12 as it is part of the SBs of interest for SP User 3. The percentage of solutions with user rating  $R_3$  are reflected in figure 7.4.

As we mention before, the deterministic SP reaches a convergence of 100%  $R_3$  design alternatives after the first 10 generations. All the SBs of interest for this SP User 3 (and for the other SP) seem to behave similarly, with some change in the rate of decay of the Probability of crossover, regardless the variable evaluated. This indicates that the AHS is working properly in trying to preserve those characteristics that the user desired

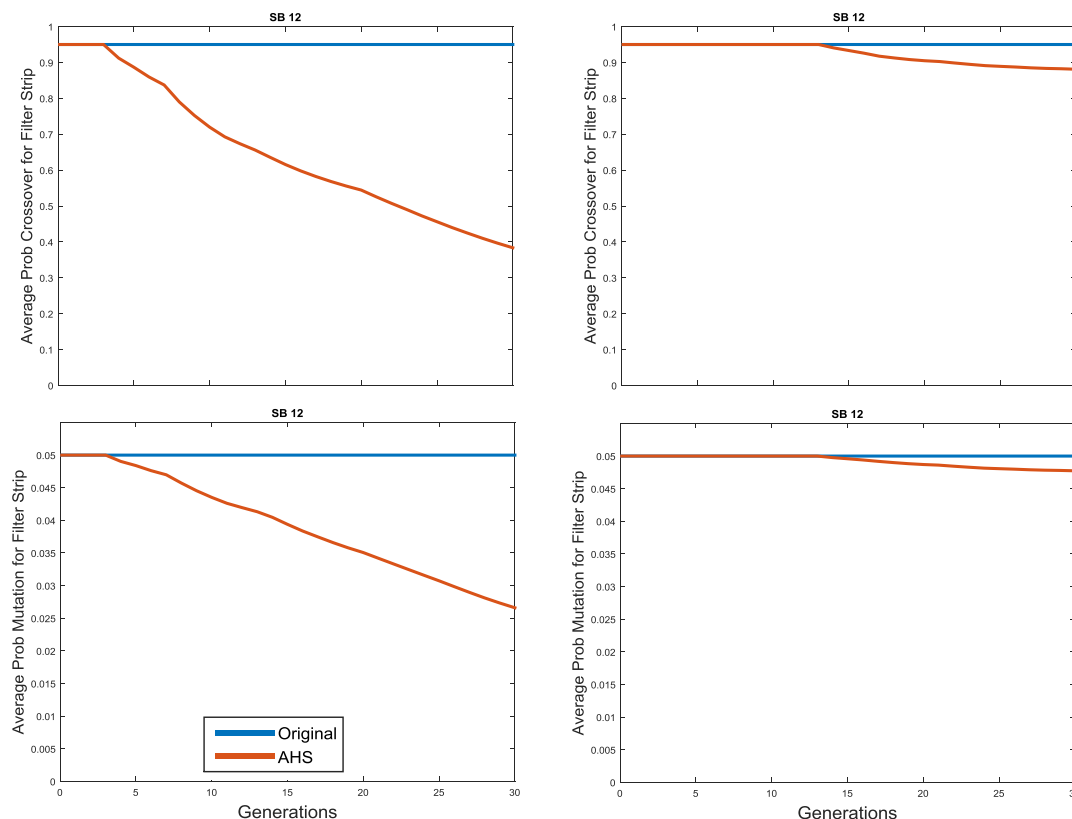
The operators for each of the SBs of interests are expected to decrease based on the selection of patterns. For the deterministic SP, in 100% of the cases (of the three performed runs) there is a decrease in both the binary and the real variable. However, for the stochastic SP, this trend is not observed in all the cases. One of the reasons is that the users did not reach the required 50% threshold of preferred designs in order to establish patterns as can be observed in figure 7.4.



**Figure 7.5** Probability of crossover for Cover Crops (up) and mutation (down) for SB 12 (User 3) for deterministic (left) and stochastic (right) SP through generations.

However, even when the pattern did not affect the binary variable for the stochastic case (as shown in Figure 7.5 right), there is an evidence of changes in all the probabilities of

crossover for filter strips. This changes occurs at a later stage and in a slower rate than the deterministic SP. But shows that there is still room to determine patterns and reach desirable solutions.



**Figure 7.6** Probability of crossover for Filter Strip (up) and mutation (down) for SB 12 (User 3) for deterministic (left) and stochastic (right) SP through generations.

## 7.6 Conclusions

Successful implementation of optimal plan designs required the participation of the stakeholders, not just in the first stages of the development plan, but also during the design, implementation, and maintenance of the system. Including Stakeholders at all stages in the process generates a sense of identification and responsibility about the causes and possible consequences of a management decision ().

The decision support systems (DSS) contribute to the generation of designs and plans that may consider not just physical and economical features, but that may include a qualitative component currently overrule by many developers. Understanding the complexity of how decision support systems and users' interactions affect not just the

objective space, but also the decision space, may help developers to improve the search of solutions, not just to satisfy the criteria, but also to deliver a better performance in the objective space.

This work explores the modification of a current interactive decision support system that uses a NonSorting Genetic Algorithm (NSGA-II) to search for optimal solutions to the spatial allocation of conservation practices in a watershed. This types of systems, although useful to explore large decision space areas, is exhausting for the users due to a large number of repletion needed for convergence. Therefore, knowing the generations where the adaptation process is initially activated can assist in providing a better limit for the necessary generations to reach the desired solutions.

Previous observation of the performance of this system suggested the presence of patterns in the designs that may be influencing the final classification of the design as a preferred (or not preferred) design.

The results suggested that there is an improvement of finding solutions if the user has a very clear idea of the desired solutions. Convergence may be reached after the first ten generations. However, we also notice that there is one user with not a clear convergence towards preferred alternatives. On the other hand, those users that present a noisy behavior are challenging to quantify, suggesting that a third element, besides objective and decision space, may be included in the search process in order to improve the satisfaction and percentage of user ratings. A proposal of modifying the 50% threshold have been suggested as part of the exploration of this novel addition to the interactive decision support system for spatial allocation of conservation practices.

Researchers have the plan to test two other simulated user, where the selection of the particular rate for the design is not related to the average rated of each SB, but rather to the maximum rate. Also, the idea of different combinations of SBs where the motivation is for specific designs may be explored to provide a wider set of possible user scenarios. Finally, tests with real users should be addressed in order to provide comparisons with simulated users and delivering their effect on the use of the modification.

## 7.7 Acknowledgements

We would like to acknowledge the funding agency - National Science Foundation (Award ID #1014693 and #1332385). We would also like to thank all of our collaborators from the different institutions: Indiana University-Purdue University Indianapolis, Oregon State University.

## 7.8 References

- Babbar-Sebens, M., and Minsker, B.S. (2011), Interactive Genetic Algorithm with Mixed Initiative Interaction for multi-criteria ground water monitoring design, *Applied Soft Comp.*, 12, 182-195.
- Babbar-Sebens, M., Mukhopadhyay, S., Singh, V. B., and Piemonti, A. D., (2015), “WRESTORE- A web-based planning software for participatory optimization of conservation practices in watersheds,” *Environ Mod and Soft.*, 69, 111-127.
- Cole, R.A., Feather, T.D., and Letting, P.K., (2002), Improving watershed planning and management through integration: a critical review of federal opportunities (USACE IWR Report 02-R-6). Alexandria, VA: US Army Institute for Water Resources.
- Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* 6 (2), 182-197.
- Gen, M., and Cheng, R., (1997), *Genetic Algorithms and Engineering Design*, New York: John Wiley and Sons
- Goldberg, D., and Kuo, C., (1987), Genetic algorithms in pipeline optimization, *J. of Computing in Civil Eng.*, 1(2), 128-141.
- Gong, D., Zhou, Y., and Li, T., (2005), Cooperative Interactive Genetic Algorithm based on user’s preference, *Int. J. of Info. Techn.*, 11(10), 1-10.
- Gong, D., Guo, G., Lu, L., and Ma, H., (2008), Adaptive interactive genetic algorithms with individual interval fitness, *Progress in Natural Scs.*, 18, 359-365.
- Llorca, X., Sastry, K., Goldberg, D. E., Gupta, A., and Lakshmi, L., (2005), Combating user fatigue in iGAs: partial ordering, support vector machines, and synthetic fitness. In Genetic and Evolutionary Computation Conference, GECCO 2005, Proceedings, Washington DC, USA, pp.1363–1370.
- Margerum, R.D., and Robinson, C.J., (2015), Collaborative partnerships and the challenges for sustainable water management, *Curr Opin Environ Sustain*, 12, 53-58.
- Nicklow, J., Reed, P, Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B., Ostfeld, A., Singh, A., and Zechman, E., (2010), State of the Art for genetic algorithms and Beyond in Water Resources Planning and Management, *J. of Water res. Plann. And Mgmt.*, 136(4), 412-432.
- Seppelt, R., and Voinov, A., (2002), Optimization methodology for land use patterns using spatially explicit landscape models, *Ecological Model.*, 151, 125-142.
- Simon, H. A., (1956), Rational choice and the structure of the environment. *Psychological Review*, 63, 129–138.

## **CHAPTER 8. Final discussion**

### **8.1 Conclusions**

With the growing use of Information and Communication Technologies (ICT) in decision support systems (DSS), the need for improving our understanding of the user-DSS interactions has never been more critical. Interactive DSS provide a framework for communicating and educating stakeholders and decision makers on critical resource management problems, and for supporting collaborative and participatory watershed planning activities. Among the multiple technologies available for communication, internet-based computing technologies have had a rapid influx in our lives over the last few decades, and have now become some of the most commonly used in our society. One of the biggest strengths of using web platforms for user-DSS interactions is that they create a remotely-accessible, secure, and private communication environment, which people can use to examine and express their own personal preferences, test hypotheses and ideas, simulate scenarios of decisions, and interact with others in their social network.

This research has investigated how individuals and optimization-based, human-centered design methods can be coupled in a web-based, watershed decision support system (DSS) called WRESTORE (Watershed REStoration using Spatial-Temporal Optimization of REsources- <http://wrestore.iupui.edu/>). The main goal of WRESTORE is to assist stakeholders in a watershed community with the generation of user-preferred design alternatives of spatial allocations for conservation practices in an agricultural watershed. WRESTORE was originally developed for the Eagle Creek Watershed, in Indianapolis, IN, with the goal of providing a more democratic venue for stakeholder-driven design of watershed plans.

This dissertation contributes to the assessment of the user-DSS interactions in WRESTORE, and to the development of a novel human-centered search/optimization algorithm that incorporates stakeholder participation in the development of design alternatives for watershed restoration. The examination of how real users interact with WRESTORE is done via Usability techniques adopted from the field of Human-



Computer Interaction. This assessment of the quality of a user's interactions is an important contribution of this study because it provides mechanisms via which the WRESTORE DSS can ascertain the reliability of the user-provided evaluations, which are later used by WRESTORE's underlying optimization algorithm to identify user-preferred regions of the objective space and decision space. At the same time, this work also improves our understanding of how human-centered interactive genetic algorithms can be improved upon for faster convergence to satisficing solutions (based on a user's evaluation of alternatives on the GUI), and, thereby, have lesser impact on human fatigue.

The following sections describe the conclusions in each of the sections of this dissertation work. Section 8.1.1 discuss the findings from Chapter 4 on usability metrics that have been proposed for assessing quality of user data used in Interactive Genetic Algorithms (IGA). Sections 8.1.2 and 8.1.3, describe the most relevant findings from chapters 5 and 6 on the plans generated at the watershed scale and at the local scale respectively, while section 8.1.4 summarize the general conclusions for the new adaptive human-guided search operators proposed for IGA applications.

### **8.1.1. Usability Metrics for IGA**

Usability metrics were found to be effective methods for tracking user behavior in a DSS, when real-time data from users is collected via the DSS to guide the search process of the Interactive Optimization algorithm within the DSS. Tracking user behavior can also help assess whether a user's evaluation of the design alternatives generated via the algorithm are reliable or not. In Chapters 4, the main goal of this work was to conduct an observational study of the interactions of multiple users with the Graphical User Interface (GUI) of WRESTORE during interactive optimization experiments. During this observational study, we were able to record and analyze multiple usability metrics (such as response times, clicking events and confidence levels), and evaluate the differences and similarities in user behaviors and interactions between two different types of participants, surrogates (volunteer non-stakeholder group who assisted with testing of the tool) and stakeholders (end users who would be expected to use the tool). Overall, the results in this section provided three significant contributions:

- 1) Usability metrics for participatory design tools based on information technologies were proposed and validated as measures for determining reliability of user-generated data.
- 2) Differences between how surrogates and stakeholders use such interactive DSS for designing alternatives were evaluated, and
- 3) Suggestions were proposed for possible improvements in GUI of similar web-based watershed DSS that support continual user interactions. These recommended improvements are expected to also address the current gap in the evaluation of how effective Environmental Decision Support Systems can be in supporting needs of end-users; investigations of these gaps have also been advocated by many other researchers (Jakeman et al., 2008; McIntosh et al., 2011).

Results indicate that overall time taken by participants in both (surrogate and stakeholder) groups decreased over time, as participants became more efficient in navigating the GUI and using the GUI's features. Therefore, surrogates can potentially be used as proxies for stakeholders for analyses related to overall task times. For the assessment of the time spent and mouse clicks in information gathering areas of the GUI, the results concluded that the stakeholders were more engaged with features of GUI that yielded useful decision-aiding information for the users. These results suggest that knowing the consequences of the design alternatives before evaluating them for being acceptable and/or satisfactory from a user's perspective was of greater importance to the stakeholders than to the surrogates.

It was also noticed that the majority of the stakeholders showed an increase in their mean confidence levels over time. A comparison of trends in mean confidence levels showed that Positive and No trends were directly related with whether participants spent more or less time and effort in gathering information, respectively. Those participants with negative trends have the same average in time spend in the information gathering areas, but there is a 30% less activity for the clicks events. Additionally, it was also observed that some participants may learn slower, and for them a change in their

confidence level trends may be observed if they are given the opportunity to continue engaging with the tool for gathering information over repeated sessions.

### 8.1.2. Watershed Scale Plans Generated from IGA

The differences between the design alternatives, investigated in Chapter 5, demonstrate the ability of the interactive genetic algorithm to incorporate individual preferences in the design of the entire watershed plan. Individual preferences related to two different spatial scales were examined in this chapter – participants who were interested in decisions, goals, or both for the entire watershed, and participants who were interested in decisions, goals, or both in a local area of the watershed. Chapter 5 also examined relationships among users (stakeholders and surrogates), usability metrics, and patterns in the watershed-scale design alternatives generated during the experiment with WRESTORE.

In general, results show that the majority (55 %) of the participants presented a higher percentage of design alternatives with user rating  $R_3$  (i.e., “I like it”) than design alternatives with user rating  $R_1$  (i.e., “I don’t like it”). This further suggests that these majority of participants found the interactive genetic algorithm (IGA) to be beneficial in finding more scenarios of watershed plans that also agreed with their individual preferences. In this chapter, the design alternatives generated via the IGA were also compared with those generated via a typical non-interactive GA that did not include the user-evaluations (i.e. user ratings) as an additional objective function. The comparison was conducted by calculating the distances between the set of non-dominated design alternatives found via the non-interactive optimization, and each of the sets of IGA-generated design alternatives classified based on their user ratings (i.e.,  $R_1$ ,  $R_2$ , and  $R_3$ ). The IGA-generated design alternatives classified as “I do not like it” have a greater distance and are more spread out in objective space than those design alternatives classified as “I like it”. In total, 80 % of the participants presents a  $DR_1(A,B) > DR_3(A,B)$ . The separation between sets of  $R_1$ ,  $R_2$ , and  $R_3$  in the objective space is more evident for those participants who evaluated the performance of design alternatives at the entire watershed-scale (Piemonti, 2014).

The analysis of IGA-generated design alternatives in their decision space helped identify spatial patterns in decisions across participants and across space in the entire watershed. For binary decision variables (e.g. for cover crops practice where a decision variable value of 1 indicated that the cover crops were implemented in that subbasin (SB), where as a value of 0 indicated otherwise) probabilities of implementation were calculated for  $R_3$  design alternatives to identify SB where the practice was most likely to be implemented. The standard deviation in these probabilities across participants demonstrated SBs where participants most likely agreed or disagreed with each other. The results showed that the stakeholder's group have the lowest variance in the majority of the SBs (70 out of 108 with standard deviation less than 0.13). While the surrogates group for model A have a similar behavior, with just 64 out 108 SBs; the surrogates model B have just 49 out of 108 SBs with standard deviation less than 0.13. The SBs with highest standard deviation are located between the Northwest and South area of the watershed in all the groups

Similarly, for real decision variables (e.g., for filter strips practice where the decision variable value indicated the width of the filter strip in a SB), the mode of filter strips was founded for all the  $R_3$  design alternatives, to identify the preferences in a design parameter. The variability was calculated using the second moment around the mode and it shows that there is a higher agreement across participants in the selection of design alternatives with a particular filter strip width. Again, *Model A-Surrogates* and *Model B-Stakeholders* showed a higher agreement with 90 and 98 SBs respectively with a second moment around the mode that is lower than 3 m<sup>2</sup> and where the highest differences are located close to the central area of the watershed. On the other hand, *Model B-Surrogates* have just 82 SBs were the second moment around the mode is lower than 3 m<sup>2</sup>. However, the distribution of the highest disagreements are also located towards those SBs in the central area.

### **8.1.3. Local Scale Plans Generated from IGA**

In Chapter 6, the focus of the analysis was narrowed to specific local SBs that were of interest to individual participants, to assess whether there were any demonstrated user-preferences in practices or goals in these SBs. The chapter also evaluated potential

relationships between the type of users, the usability metrics, and patterns in the IGA-generated user-preferred design alternatives for the spatial allocation of conservation practices in SBs of interest.

Patterns in the values of the binary decision variables in local SBs of interest allowed us to identify probabilities of implementation of the practice in a particular SB, even if the participants were or were not directly concerned with the SBs of interest. The majority of participants have  $Prob_{jR3k} > 0.5$  in all of their SBs of interest, except for a group of 5 participants (Participants 5, 6, 13, 20, 21) who had  $Prob_{jR3k} < 0.5$  in at least one of their assigned SBs of interests. Among these five participants, only two (Participants 6 and 21) had a positive trend in their mean self-confidence levels, and had a fair (in average, 57% of click events and 51% of time spent in information gathering ) amount of interaction with the interface in gathering information about the designs. In summary, we observed that all the participants had at least one SB of interest ( $SB_{int}$ ) where the probability of implementation of cover crop is greater than 0.5. An important observation is that 61% of the participants (regardless of the group or model) had a probability of cover crop lower than 0.5 for  $SB_{int}$  103. This demonstrates that even agreements on where the practices should not be implemented can be determined through this processes.

The results in the real variable showed more agreement in the selected filter strip width (actual design value), regardless of the group or model. However, the percentages of solutions with filter strip width equal to the mode varied from participant to participant. Participants in Model A seemed to select the majority of the design alternatives with the preferred filter strip width equal to the mode (i.e., 96 % of SB have an  $GPOMR_3 > 50\%$  - where  $GPOMR_3$  is the Group percentage of design alternatives with filter strip width equal to the mode). Assessment of the performance of  $R_3$  (i.e. "I like it") design alternatives in objective space indicated that for 17% of the SBs of interests participants seemed to like the designs even if the performance in the SBs was not the best. This suggests that the performance of physical objective functions (i.e. cost, peak

flow reductions, nitrate reductions, and sediment reductions) in those particular SBs was not a significant factor in the participant's reasoning process.

#### **8.1.4. New Adaptive Operators for IGA**

Finally, in Chapter 7, we examined whether the observed patterns in decision space identified in the user experiments could be used towards improvement of the IGA-based search algorithm. A novel human-guided search operators for IGA were proposed and developed, which identify and learn from patterns in the design space of user-preferred alternatives and, then, use the patterns to adapt the mechanisms of selection, crossover, and mutation. The goal of these operators is to generate new design alternatives that preserve the learned patterns. The proposed modification was done on default selection, crossover, and mutation operators in the multi-objective Nondominated Sorting Genetic Algorithm – II (NSGA-II; Deb et al., 2002), which are used by the underlying IGA approach in WRESTORE. The goal of these modifications was to assist the multi-objective IGA algorithm to rapidly converge to design alternatives, especially when the number of decision variables are significantly large, and the design space is immensely complex. Rapid convergence is also expected to facilitate a lower level of human fatigue in participating humans.

Simulation experiments were conducted with different types of user models (deterministic and stochastic user models), which were generated from the data of the real humans who had participated in user-experiments reported in Chapter 4, 5, and 6. Experiment results suggest that the new adaptive operators certainly improve (two out of three simulated participants) the convergence rates for deterministic users (i.e., when the user has a very clear idea of what she/he desires in a solution and always gives consistent user ratings). For example, for two of the three deterministic users, a convergence of 100 percent (i.e. 100% of IGA population consisted of  $R_3$  alternatives) was reached after the first ten generations. On the other hand, those simulated users who presented a noisy user evaluation (i.e. stochastic users) demonstrated poorer convergence that was similar to the convergence of the IGA with default NSGA-II operators. It is recommended that additional user types should be tested in the future to examine the conditions under which the adaptive operators work better than or similar to the default search operators.

## 8.2 Future Research

Even though the concept of participation and human-centered design has been widely studied in different fields, such as web page development and software development, and ergonomics, its application in watershed planning and management problems is relatively new, providing a vast opportunity for future research topics and long-term advancements in the Systems Analysis and Hydroinformatics fields. Some of the areas that require additional research investigations include improvement of optimization-based interactive search algorithms, development and evaluation of new usability metrics, detection and evaluation of users' conceptual learning processes when they are immersed in interactive DSS for participatory planning and design, and monitoring of the actual rates of adoption of the user-preferred design alternatives in the communities.

While the research described in this dissertation investigates the usability of GUI features in one decision support system (i.e. WRESTORE), it is recommended that similar usability evaluations of other DSS should be also performed and investigated. Additionally, future studies should also examine how Usability metrics can provide insights into user types, especially when participants engage in synchronous experiments that allow them to interact and collaborate on problems where they agree or disagree with each other.

The kind of optimization-based human-centered design platform investigated in this research is expected to create new opportunities for including capabilities of web-based, social networking tools within DSS. Such networking environments are expected to help connect diverse sets of individuals in collaborating during the design process, educate them based on their needs, and create a more user-friendly and secure environment that can serve as a platform for negotiation processes in a community. Finally, development and implementation of similar systems will generate more venues for bottom-up decision making efforts that communities can initiate themselves and/or in collaboration with NGOs, state and federal agencies, and other institutions.

On the subject of interactive optimization algorithms, additional investigations are recommended so that better methods can be developed for using individual participant's

and multiple participants' personal feedback in the generation of solutions that satisfy subjective criteria, while at the same time informing and educating the stakeholders about the benefits of simulated scenarios. In this research, we incorporated the observed patterns in the decision space to guide the search operators in algorithm and tested the new operators with experiments involving simulated participants. Therefore, there remains a need to test these new operators with real human participants, as well as test different selected thresholds and mechanisms to guide the adaptation of search operators. Finally, since multiple stakeholders are generally interested in a few SBs, it is recommended that distributed and coordinated approaches for interactive optimization should also be investigated so that individual participants can conduct interactive search for decision relevant to their local areas, but still be able to influence a compromise search for the entire watershed.



## REFERENCES

- Abbott, M.B., (1991), *Hydroinformatics: Information technology and the aquatic environment*, Great Britain: Billing & Sons Ltd.
- Abbott, M.B., (1999), Introducing hydroinformatics, *J. of Hydroinformatics*, 1, 3-19.
- Adeyemo, J.A., (2011), Reservoir operation using multi-objective evolutionary algorithms-A review, *Asian J. of Scientific Research*, DOI: 10.3923/ajsr.2011
- Ahmadi, M., Arabi, M., Fontane, D.G., and Engel, B.A., (2014), Application of multicriteria decision analysis with a priori knowledge to identify optimal nonpoint source pollution control plans, *J. Water Resour. Plann. Manage.*, 141(2), 1-9.
- Arabi, M., Govindaraju, R. S., and Hantush, M. (2006), Cost-effective allocation of watershed management practices using a genetic algorithm, *WRR*, 42, W10429.
- Arabi, M., Frankenberger, J. R., Engel, B. A., and Arnold, J.G. (2007), Representation of agricultural conservation practices with SWAT, *Hydrol. Processes*, 22(16), 3042–3055.
- Arnold, J.G., Srinivasan, R., Muttiah, R.S., and Williams, J.R., (1998), Large Hydrologic Modeling and Assessment: Part I. Model Development, *JAWRA*, 34(1), 73-89.
- Artita, K., Kaini, P., and Nicklow, J. W. (2008), Generating alternative watershed-scale BMP designs with evolutionary algorithms, paper presented at the World Environmental and Water Resources Congress, Ahupua’A, HI.
- Avigad, G., and Moshaiov, A., Interactive evolutionary multiobjective search and optimization of set-based concepts, *IEEE Trans. on Sys. Man and Ciber. Part B*, 39(4), 1013-1027.
- Babbar-Sebens, M., and Minsker, B.S. (2011), Interactive Genetic Algorithm with Mixed Initiative Interaction for multi-criteria ground water monitoring design, *Applied Soft Comp.*, 12, 182-195.
- Babbar-Sebens, M., Barr, R.C., Tedesco, L. P., and Anderson, M. (2013), Spatial identification and optimization of upland wetlands in agricultural watersheds, *Ecol. Eng.*, 52, 130-142.
- Babbar-Sebens, M., Mukhopadhyay, S., Singh, V. B., and Piemonti, A. D., (2015), “WRESTORE- A web-based planning software for participatory optimization of conservation practices in watersheds,” *Environ Mod and Soft.*, 69, 111-127.
- Baron, J., (2000), *Thinking and deciding* (3rd ed.). New York: Cambridge University Press.
- Barreteau, O., and Abrami, G. (2007), Variable time scales, agent-based models, and role-playing games: The PIEPLUE river basin management game, *Simulation Gaming*, 38(3), 364-381.
- Beer, D., (2009), Power through the algorithm? Participatory web cultures and the technological unconscious, *New media & Society*, 11(6), 985-1002.

- Bekele, E.G., and Nicklow, J.W., (2005), Multi-objective Management of Ecosystem Services by Integrative Watershed Modeling and Evolutionary Algorithms, *Water Res. Research*, 41(10), W10406.
- Bonnel, J., and Koontz, T., (2007), Stumbling forward: The organizational challenges of building and sustaining collaborative watershed management. *Society and Natural Resources*, 20(2),153-167.
- Brauman, K.A., Daily, G.C., Duarte, T.K., and Mooney, H.A., (2007), The nature and value of ecosystem services: an overview highlighting hydrologic services, *Annu. Rev. Environ. Resour.*, 32, 67–98.
- Burkey, J. (2006), *A non-parametric monotonic trend test computing Mann-Kendall Tau, Tau-b, and Sens Slope written in Mathworks-MATLAB implemented using matrix rotations. King County, Department of Natural Resources and Parks, Science and Technical Services section. Seattle, Washington. USA.* [Online]. Available:  
<http://www.mathworks.com/matlabcentral/fileexchange/authors/23983>
- Castelletti, A. and Socini-Sessa, R. (2007), Bayesian Networks and participatory modelling in water resource management, *Environ. Model. & Soft.*, 22(8), 1075-1088.
- Cash, D.W., Clark, W.C., Alcock, F., Dickson, N., Eckley, N., and Jäger, J., (2002), *Saliency, Credibility, Legitimacy and Boundaries: Linking Research, Assessment and Decision Making*, RWP02-046, J.F. Kennedy School of Government, Harvard University, Cambridge, MA.
- Clark, G.D., (1980), The Indiana Water Resource; availability, uses, and needs, Governor's Water Resource Study Commission, State of Indiana, Indiana Department of Natural Resources, Indianapolis, Indiana.
- Clark, J., and Aufderheide, P., (2009), Public Media 2.0: Dynamic, engaged publics, WRLC, <http://hdl.handle.net/1961/5394>
- Clearfield, F., and Osgood, B.T., (1986), Sociological aspects of the adoption of conservation practices, Washington D.C, USA: Soil Conservation Services.
- Cole, R.A., Feather, T.D., and Letting, P.K., (2002), Improving watershed planning and management through integration: a critical review of federal opportunities (USACE IWR Report 02-R-6). Alexandria, VA: US Army Institute for Water Resources.
- Cohen, A. and Davidson, S., (2011). The watershed approach: Challenges, antecedents, and the transition from technical tool to governance unit. *Water Alternatives*, 4(1), 1-14.
- Clausen, J. C., and Meals, D. W. (1989), Water quality achievable with agricultural best management practices, *J. of Soil and Water Cons.*, 44(6), 593-596.
- Cox, P.G. (1996), Some issues in the design of agricultural decision support systems, *Ag. Systems*, 52(2/3), 355-381.
- Cunha, M. and Sousa, J., (1999), Water Distribution Network Design Optimization: Simulated Annealing Approach, *J. Water Resour. Plann. Manage.*, 125(4), 215-221.

- Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* 6 (2), 182-197.
- de Groot, R.S., Wilson, M.A., and Boumans, R.M.J., (2002), A typology for the classification, description and valuation of ecosystem functions, goods and services, *Ecological Economics*, 41, 393-408.
- de Kok, J.L., Titus, M, and Wind, H.G., (2000), Application of fuzzy sets and cognitive maps to incorporate social science scenarios in integrated assessment models A case study of urbanization in Ujung Pandang, Indonesia, *Integrated Assessment*, 1(3), 177-188.
- Delipetrev, B., Jonoski, A., and Solomatine, D.P., (2013), Development of a web application for water resources based on open source software, *Computers & Geosciences*, 62, 35-42.
- Diez, E., and McIntosh, B.S., (2009), A review of the factors which influence the use and usefulness of information systems, *Environ. Model. & Soft.*, 24(5), 588-602.
- Elmahdi, A., and McFarlane, D., (2012), Integrated multi-agency framework: sustainable water management, *Proceedings of the ICE-Water management*, 165(6), 313-326.
- Eppstein, David, (1996), ICS 161: Design and Analysis of Algorithms. Lecture notes for March 12, 1996. <https://www.ics.uci.edu/~eppstein/161/960312.html>
- Estes, W.K. (1994), *Classification and Cognition*, New York, NY: Oxford University Press-Clarendon Press.
- Eusuff, M., and Lansey, K., (2003), Optimization of Water Distribution Network Design Using the Shuffled Frog Leaping Algorithm, *J. Water Resour. Plann. Manage.*, 129(3), 210-225.
- Evers, M., Jonoski, A., Maksimovič, C., Lange, L., Ochoa, S., Teklesadik, A., Cortes, J., Almoradie, A., Simões, N., Wang, L., and Makropoulos, C., (2012), Collaborative modelling for active involvement of stakeholders in urban flood risk management. *Natural Hazards and Earth System Science*, 12(9), 2821–2842.
- Evers, M., Jonoski, A., Almoradie, A., and Lange, L., (2015), Collaborative decision making in sustainable flood risk management: A socio-technical approach and tools for participatory governance, *Environmental Sciences & Policy*, 55, 335-344.
- Feather, P.M., and Amacher, G.S., (1994), Role of information in the adoption of best management practices for water quality improvement. *Agricultural Economics*, 11(2-3), 159-170.
- Fonseca, C.M., and Fleming, P.J., (1995), An overview of evolutionary algorithms in Multiobjective Optimization., *Evolutionary Computation*, 3(1), 1-16.
- Fraternali, P., Castelletti, A., Soncini-Sessa, R., Ruiz, C.V., and Rizzoli, A.E., (2012), Putting humans in the loop: Social computing for Water Resources Management, *Environ. Modelling and Soft.*, 37, 68-77.

- Fry, D.L., Battles, J.J., Collins, B.M., and Stephens, (2015), Sierra Nevada Adaptive Management Project. Fire and Forest Ecosystem Health report.
- Geem Z.W., Kim, J.H., and Loganathan, G.V., (2001), A new heuristic optimization algorithm: harmony search. *Simulation*, 76(2), 60–68.
- Gen, M., and Cheng, R., (1997), *Genetic Algorithms and Engineering Design*, New York: John Wiley and Sons
- Goldberg, D., and Kuo, C., (1987), Genetic algorithms in pipeline optimization, *J. of Computing in Civil Eng.*, 1(2), 128-141.
- Gong, D., Zhou, Y., and Li, T., (2005), Cooperative Interactive Genetic Algorithm based on user's preference, *Int. J. of Info. Techn.*, 11(10), 1-10.
- Gong, D., Guo, G., Lu, L., and Ma, H., (2008), Adaptive interactive genetic algorithms with individual interval fitness, *Progress in Natural Scs.*, 18, 359-365.
- Gregersen, H.M., Ffolliott, P.F., and Brooks, K.N. (2007), *Integrated Watershed Management: Connecting people to their land*, Oxfordshire, UK: Cambridge University Press.
- Gregory, R. (2000), Using stakeholder values to make smarter environmental decisions, *Environ.*, 42(5), 34-44.
- Greiner, R., Patterson, L., and Miller, O., (2009), Motivations, risk perceptions and adoption of conservation practices by farmers, *Agricultural Systems*, 99, 86-104.
- Ha, W. (2010), Extended evaluation for quality of service in the community of web service, paper presented at International Symposium on Computational Intelligence and Design.
- Haklay, M., and Tobon, C., (2003), Usability evaluation and PPGIS: towards a user-centred design approach, *Int. J. Geo. Info. Sci.*, 17(6), 577-592.
- Haddad, O.B., Afshar, A., and Marino, M.A., (2006), Honey-Bees Mating Optimization (HBMO) Algorithm: A new Heuristic approach for water resources optimization, *Water Res. Mgmt.*, 20, 661-680.
- Hamalainen, R.P., Kettunen, E., Ehtamo, H., (2001). Evaluating a framework for multistakeholder decision support in water resources management, *Group Decision and Negotiation*, 10, 331–353.
- Hare, M., Gilbert, N., Medugno, D., Asakawa, T., Heeb, J. and Pahl-Wostl, C. (2001). The Development of an Internet Forum for Long-term Participatory Group Learning about Problems and Solutions to Sustainable Urban Water Supply Management In L. M. Hilty and P. W. Gilgen (Eds.), *Sustainability in the Information Society*, pp. 743-750. Marburg: Metropolis Verlag.
- Hearst, M.A., (1999), Trends and controversies: mixed-initiative interaction, *IEEE Intel. Syst.*, 14 (5), 14-23.
- Heger, A.S., Duran, F.A., Frysinger, S., and Cox, R.G. (1992), Treatment of Human-Computer Interface in a Decision Support System, paper presented at IEEE International Conference on Systems, Man, and Cybernetics (SMC).

- Heidegger, M., (1977), *The question concerning technology and other essays*, New York: Harper Torchbooks.
- Helsel, D.R., and Hirsch, R.M., (2002), *Statistical Methods in Water Resources Techniques of Water Resources Investigations*, Book 4, chapter A3. U.S. Geological Survey. 522 pages.
- Hey, D.L., Kostel, J., and Montgomery, D., (2009), An ecological solution to the flood damage problem, paper presented at Finding the Balance between floods, flood protection and river navigation Conference, St. Louis, Mo.
- Hoag, D., Luloff, A. E., and Osmond, D. L., (2012), *Lessons Learned from the NIFA-CEAP: How Farmers and Ranchers Make Decisions on Conservation Practices*, NC State University, Raleigh, NC.
- Hornbæk, K. (2006), Current practice in measuring usability: challenges to usability studies and research, *Int. J. Human-Computer Studies*, 64(2), 79-102.
- Ice, G. (2004), History of innovative Best Management Practice development and its role in addressing water quality limited waterbodies, *J. Environ. Eng.*, 130(6), 684-689.
- ISO DIS 9241-210 (2008). Ergonomics of human-system interaction – Part 210: Humancentred design process for interactive systems. ISO.
- Jaber, M.Y. (2011), *Learning curves: theory, models, and applications*, Boca Raton, FL: CRC Press.
- Jakeman, A.J., Voinov, A.A., Rizzoli, A.E., and Chen, A.H. (2008), *Environmental Modelling, Software and Decision Support*. The Netherlands. Elsevier B.V.
- Jankowski, P., Robischon, S., Tuthill, D., Nyerges, T., and Ramsey, K. (2006), Design considerations and evaluation of a collaborative spatio-temporal decision support system, *Transac. in GIS*, 10(3), 335-354.
- Johnson, L. E. (1986), Water resource management decision support system, *J. Water Resour. Plann. Mgmt.*, 112(3), 308-325.
- Johnson, N., Ravnborg, H.M., Westermann, O., and Probst, K., (2002), User participation in watershed management and research, *Water Policy*, 3, 507-520.
- Kaini, P., Artita, K., and Nicklow, J. W. (2012), Optimizing structural Best Management Practices using SWAT and Genetic Algorithm to Improve Water Quality Goals, *Water Resour. Mgmt.*, 26, 1827-1845.
- Kahneman, D., and Tversky, A., (1979), Prospect theory: An analysis of decisions under risk. *Econometrika*, 47, 263–291.
- Kahneman, D., and Tversky, A., (1984). Choices, values, and frames. *American Psychologist*, 39, 341–350.
- Kaunda-Bukenya, N., Tadesse, W., Fu, Y., Tsegaye, T., and Wagaw, M., (2012), Spatial Decision Support System (SDSS) for Stormwater Management and Water Quality Assessment, in Voudouris, K., and Voutsas, D., Published: April 5<sup>th</sup>, 2012, ISBN: 978-953-51-0486-5. InTech
- Kelly, M., Ferranto, S., Lei, S., Ueda, K., and Huntsinger, L., (2012). Expanding the table: the web as a tool for participatory adaptive management in California forests. *J. Environ. Mgmt.* 109, 1-11.

- Kelly, R. A., and Merritt, W. S. (2010), The Role of Decision Support Systems (DSS) in planning for improved water quality in coastal lakes, in *Decision Support Systems in Agriculture, Food and the Environment: Trends, Applications and Advances*, edited by B. Manos, N. Matsatsinis, K., Paparrizos and J., Papathanasiou, pp. 48-73, IGI Global, Hersey, PA.
- Kirchoff, C.J., Lemos, M.C., and Dessai, S. (2013), Actionable knowledge for environmental decision making: broadening the usability of climate sciences, *Annu. Rev. Environ. Resour.*, 38, 393-414.
- Krettek, J., Braun, J., Hoffmann, F., Bertram, T., Ewald, T., Schubert, H.G., and Lausch, H., (2009), Interactive evolutionary multiobjective optimization for hydraulic valve controller parameters, in: Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics, 816–821.
- Kronaveter, L., and Shamir, U., (2009), Negotiation support for cooperative allocation of a shared water resource: Methodology, *J. of Water Res. Plan and Mgmt.*, 135 (2), 60-69.
- Kushniruk, A., Patel, V., Cimino, J.J., and Barrows, R.A., (1996), Cognitive evaluation of the user interface and vocabulary of an outpatient information system, Proceedings of the AMIA-Annu Fall Symp, 22-26.
- Lansey, K. and Mays, L., (1989), Optimization Model for Water Distribution System Design, *J. Hydraul. Eng.*, 115(10), 1401-1418.
- Legates, D.R., and McCabe, G.J., (1999), Evaluating the Use of “Goodness-of-fit” measures in Hydrologic and Hydroclimatic Model Validation, *Water Res. Research*, 35(1), 233-241.
- Lethbridge, M.R., Westphal, M.I., Possingham, H.P., Harper, M.L., Souter, N.J., and Anderson, N. (2010), Optimal restoration of altered habitats, *Environ. Model. & Soft.*, 25(6), 737-746.
- Little, J.D.C. (1979), Decision support system for marketing managers, *J. Marketing*, 43(3), 9-26.
- Liu, Y., Gupta, H., Springer, E., and Wagener, T. (2008), Linking sciences with environmental decision making: experiences from an integrated modeling approach to supporting sustainable water resources management, *Environ. Model. & Soft.*, 23, 846-858.
- Llorca, X., Sastry, K., Goldberg, D. E., Gupta, A., and Lakshmi, L., (2005), Combating user fatigue in iGAs: partial ordering, support vector machines, and synthetic fitness. In Genetic and Evolutionary Computation Conference, GECCO 2005, Proceedings, Washington DC, USA, pp.1363–1370.
- Lodhi, A. (2010), Usability heuristics as an assessment parameter: for performing usability testing, paper presented at International Conference on Software Technology and Engineering.
- Lubell, M.N., (2004), Collaborative watershed management: a view from the grassroots. *Policy Stud. J.*, 32,321–341.
- Maier, H., Simpson, A., Zecchin, A., Foong, W., Phang, K., Seah, H., and Tan, C., (2003), Ant Colony Optimization for Design of Water Distribution Systems, *J. Water Resour. Plann. Manage.*, 129(3), 200-209.

- Margerum, R.D., and Robinson, C.J., (2015), Collaborative partnerships and the challenges for sustainable water management, *Curr Opin Environ Sustain*, 12, 53-58.
- Maringanti, C., Chaubey, I., and Popp, J., (2009), Development of a Multi-objective Optimization Tool for the Selection and Placement of best Management Practices for Nonpoint Source Pollution Control, *Water Res. Research*, 45(6), W06406/1.
- Mathur, Y.P., and Nikam, S.J., (2009), Optimal reservoir operation policies using genetic algorithms, *Int. J. of Eng. And Tech.*, 1(2), 184-187
- Matsuura, K., Willmott, C., Legates, D., 2009. WebWIMP: the web-based water-budget interactive modeling program. <<http://climate.geog.udel.edu/~wimp/>> (last accessed 28.12.15).
- Maxted, J.T., Diebel, M.W., and Vander Zanden, M.J., (2009), Landscape planning for agricultural non-point source pollution reduction. II. Balancing watershed size, number of watersheds and implementation effort, *Environ. Mgmt.*, 43(1), 60-68.
- Mazerolle, M.J., (2004), Mouvements et reproduction des amphibiens en tourbières perturbées, Ph.D. dissertation, Dept. Forest. And Geomat., Univ. du Quebec, Quebec, Montreal, Canada.
- McCown, R.L., (2002), Changing systems for supporting farmers' decisions: problems, paradigms, and prospects. *Agric. Syst.*, 74, 179-220.
- McFadden, J.E., Hiller, T.L., and Tyre, A.J. (2011), Evaluating the efficacy of adaptive management approaches: is there a formula for success? *J. of Environ. Mgmt.*, 92,1354–1359.
- McIntosh, B.S., Ascough II, J.C., Twery, M., Chew, J., Elmahdi, A., Haase, D., Harou, J.J., Hepting, D., Cuddy, S., Jakeman, A.J., Chen, S., Kassahun, A., Lautenbach, S., Matthews, K., Merritt, W., Quinn, N.W.T., Rodriguez-Roda, I., Sieber, S., Stavenga, M., Sulis, A., Ticehurst, J., Volk, M., Wrobel, M., van Delden, H., El-Sawah, S., Rizzoli, A., and Voinov, A. (2011), Environmental decision support systems (EDSS) development-Challenges and best practices, *Environ. Model. & Soft.*, 26(12), 1389-1402.
- McKinney, D.C., and Lin, M., (1994), Genetic algorithm solution of groundwater management models. *Water Res. Research.*, 30(6),1897–1906
- Mendoza, G.A., and Martins, H., (2006), Multi-criteria decision analysis in natural resource management: A critical review of methods and new modelling paradigms, *Forest Ecology and Mgmt.*, 230, 1-22.
- MEA (2005), *Ecosystems and Human Well-being: Current State and Trends*. Island Press, Washington, D.C.
- Metcalf, S.S., Wheeler, E., BenDor, T., Lubinski, K.S., and Hannon, B.M., (2010), Sharing the floodplain: mediated modelling for environmental management, *Environ. Model. & Soft.*, 25(11), 1282-1290.
- Mintzberg, H., Raisinghani, D., and Théorêt, A., (1976), The Structure of "Unstructured" Decision Processes. *Administrative Science Quarterly*, 21 (2), 246-275

- Mintzberg, H., and Westley, F., (2001), Decision making: It's not what you think, *MIT Sloan Management Review*, 42(3), 89–93.
- Moote, M.A., McClaran, M.P., and Chickering, D.K., (1997), Theory in Practice: applying participatory democracy theory to Public land planning, *Environ. Management*, 21(6), 877-889
- Mostert, E., Craps, M., and Pahl-Wostl, C., (2008), Social learning: the key to integrated water resources management?, *Water International*, 33(3), 293-304.
- Mysiak, J., Giupponi, C., and Rosato, P. (2005), Towards the development of a decision support system for water resource management, *Environ. Model. & Soft.*, 20(2), 203-214.
- Nash, J.E., and Sutcliffe, J.V., (1970), River Flow Forecasting Through Conceptual Models: Part I. A Discussion of Principles, *Journal of Hydrology*, 10(3), 282-290.
- Neitsch, S.L., Arnold, J.G., Kiniry, J.R., and Williams, J.R., (2005), Soil and Water Assessment Tool Theoretical Documentation, Version 2005, Tempe, TX: USDA-ARS Grassland, Soil and Water Research Laboratory. Web. July, 2012. <<http://swatmodel.tamu.edu/documentation/>>
- 
- Nemeth, C. (2004), *Human Factors Methods for Design*, Boca Raton, FL: CRC press, pp.149
- Newell, A., and Rosenbloom, P.S. (1981), Mechanisms of skill acquisition and the law of practice, in *Cognitive skills and their acquisition*, Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., pp. 1-51.
- Newman, J.E., (1997), Our Changing Climate, In: *The Natural Heritage of Indiana*, edited M.T. Jackson, Indiana University Press, p.85-99.
- Nicklow, J., Reed, P, Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B., Ostfeld, A., Singh, A., and Zechman, E., (2010), State of the Art for genetic algorithms and Beyond in Water Resources Planning and Management, *J. of Water res. Plann. And Mgmt.*, 136(4), 412-432.
- Nielsen, J. (1994), Usability inspection methods, paper presented at Conference Companion on Human Factors in Computing.
- Nyerges, T., Jankowski, P., Tuthill, D., and Ramsey, K. (2006), Collaborative water resource decision support: results of a field experiment, *Ann. Assoc. Amer. Geograph.*, 96(4), 699-725.
- Osman, M., (2010), Controlling uncertainty: A review of human behavior in complex dynamic environments, *Psychological Bulletin*, 136(1), 65-86.
- Özesmi, U., and Özesmi, S.L. (2004), Ecological models based on people's knowledge: a multi-step fuzzy cognitive mapping approach, *Ecol. Model.*, 176(1-2), 43–64.
- Pahl-Wostl, C., (2005). Information, public empowerment, and the management of urban watersheds. *Environ. Model. and Soft.*, 20 (4), 457–467.
- Palmer, R.N., (1998), A history of shared vision modeling in the ACT-ACF comprehensive study: a modeler's perspective. *Coord. Water Resour. Environ.* 221-226.



- Peng, Y., Kou, G., Wang, G., Wu, W., and Shi, Y., (2011), Ensemble of software defect predictors: an AHP-based evaluation method, *Int. J. of Information Tech. & Decision Making*, 10(1), 187-206.
- Patel, V.L., Kushniruk, A.W., Yang, S., and Yale, J.F., (2000), Impact of a computer-based patient record system on data collection, knowledge organization, and reasoning, *JAMA*, 7(6), 569-585.
- Perez-Pedini, C., Limbrunner, J.F., and Vogel, R.M. (2005), Optimal location of infiltration-based best management practices for storm water management, *J. of Water Resour. Plan. And Mgmt.*, 131(6), 441-448.
- Piemonti, A. D., Babbar-Sebens, M., and Luzar, E. J. (2013), Optimizing conservation practices in watersheds: Do community preferences matter? *WRR*, 49(10), 6425-6449.
- Piemonti, A.D., (2014), Correlating Human Preferences and Optimized Watershed Management Plans. Proceedings of the World Environmental and Water Resources Congress 2014: pp. 1961-1969
- Poch, M, Comas, J., Rodriguez-Roda, I., Sanchez-Marre, M., and U. Cortes, (2004) Designing and building real environmental decision support systems, *Environ. Model. and Soft.*, 19(9), 857–873.
- Porzecanskik, I., Saunders, L.V., and Brown, M.T., (2012), Adaptive management fitness of watersheds. *Ecology and Society*, 17, 29.
- Power, D.J., and Sharda, R. (2007), Model-driven decision support systems: Concepts and research directions, *Decision Support Systems*, 43(3), 1044-1061.
- Priscoli, J.D., (2004), What is public participation in water resources management and why is it important,? *Water International*, 29 (2), 221-227.
- Prokopy, L.S., Flores, K., Klotthor-Weinkauff, D., and Baumgart-Getz, A., (2008), Determinants of agricultural best management practice adoption: Evidence from the literature. *J. of Soil and Water Conservation*, 63(5), 300-311.
- Rabotyagov, S.S., Campbell, T.D., White, M., Arnold, J.G., Atwood, J., Norfleet, M.L., Kling, C.L., Gassman, P.W., Valcu, A., Richardson, J., Turner, R.E., and Rabalais, N.N., (2014), Cost-effective targeting of conservation investments to reduce the northern Gulf of Mexico hypoxic zone, *PNAS*, 111(52), 18530-18535.
- Rahelizatovo, N.C., and Gillespie, J.M., (2004), The Adoption of Best-Management Practices by Louisiana Dairy Producers. *J. of Agricultural and Applied Economics*, 36, 229-240.
- Randhir, T.O., Lee, J.G., and Engel, B. (2000), Multiple criteria dynamic spatial optimization to manage water quality on a watershed scale, *Trans. of American Society of Agricultural Engineers*, 43 (2), 291–299.
- Rhoades, R.E., 1999, Participatory Watershed Research and management: where the shadow falls, IIED., Gatekeeper (81).
- Runkel, R., Crawford, C., and Cohn, T., (2004), Load Estimator (LOADEST): A FORTRAN program for estimating constituent loads in streams and rivers, Techniques and Methods. Book 4, Chapter A5.

- Savic, D. A., and Walters, G. A., (1997), Genetic Algorithms for Least Cost Design of Water Distribution Networks, *J. of Water Res. Plann. and Mgmt.*, 123 (2), 67-77.
- Seppelt, R., and Voinov, A., (2002), Optimization methodology for land use patterns using spatially explicit landscape models, *Ecological Model.*, 151, 125-142.
- Sidiropoulos, E., and Tolikas, P., (2008), Genetic algorithms and cellular automata in aquifer management. *Applied Mathematical Modelling*, 32(4), 617-640.
- Siegfried, T., Bleuler, S., Laumanns, M., Zitzler, E., and Kinzelbach, W., (2009), Multiobjective groundwater management using evolutionary algorithms, *IEEE Trans Evol Comput.*, 13(2), 229-242.
- Simon, H. A., (1955), A behavioral model of rational choice. *Quart. J. of Economics*, 59, 99–118.
- Simon, H. A., (1956), Rational choice and the structure of the environment. *Psychological Review*, 63, 129–138.
- Simon, H. A., (1957), Models of man, social and rational: Mathematical essays on rational human behavior. New York: Wiley
- Simon, H.A., (1977), The New Science of Management Decision (3rd revised edition; first edition 1960) Prentice-Hall, Englewood Cliffs, NJ.
- Simpson, A.R. and Goldberg, D.E. (1994). “Pipeline optimisation via genetic algorithms: from theory to practice”, paper in Water Pipeline Systems, by Miller, D.S. (Editor), Mechanical Engineering Publications, London, 309–320.
- Slocum, T.A., Cliburn, D.C., Feddeman, J.J., and Miller, J.R. (2003), Evaluation the usability of a tool for visualizing the uncertainty of the future global water balance, *Cart. And Geo. Info. Sci.*, 30(4), 299-317.
- Smit, S., Tutusaus, M., Curry, E., Messervey, T., and D’andrea, Z., (2015) Business drivers for adopting SMART water technology, E-Proceedings of the 36<sup>th</sup> IAHR World Congress.
- Smith, M., de Groot, D., Perrot-Maître, D., and Bergkamp, G., (2006), *Pay – Establishing payments for watershed services*. Gland, Switzerland: IUCN. Reprint, Gland, Switzerland: IUCN, 2008.
- Strange, E.M., Fausch, K.D., Covich, A.P., (1999), Sustaining ecosystem services in human-dominated watersheds: biohydrology and ecosystem processes in the South Platte River Basin, *Environ. Mgmt.*, 24(1), 39-54.
- Tedesco, L.P., Pascual, D.L., Shrake, L.K., Hall, R.E., Casey, L.R., Vidon, P.G.F., Hernly, F.V., Salazar, K.A., Barr, R.C., Ulmer, J., Pershing, D., 2005. Eagle Creek Watershed Management Plan: An Integrated Approach to Improved Water Quality. Eagle Creek Watershed Alliance, CEES Publication 2005-07, IUPUI, Indianapolis, 182p.
- Tilak, O., Babbar-Sebens, M., and Mukhopadhyay, S. (2011), Decentralized and partially decentralized reinforcement learning for designing a distributed wetland system in watersheds, paper presented at IEEE International Conference on Systems, Man, and Cybernetics (SMC).

- Tsitsiklis, J.N., (1984), Problems in decentralized decision making and computation (Doctoral dissertation). Massachusetts Inst of Tech Cambridge Lab for Information and Decision Systems.
- Tullis, T., and Albert, B., *Measuring the User Experience: collecting, analyzing and presenting usability metrics*. Waltham, MA: Elsevier Inc, 2013.
- Tversky, A., (1969), Intransitivity of preferences. *Psychological Review*, 76, 31–48.
- Tversky, A., and Kahneman, D., (1981), The framing of decisions and the psychology of choice. *Science*, 211, 453–458.
- Vervoort, J.M., Kok, K., van Lammeren, R., and Veldkamp, T., (2010), Stepping into futures: Exploring the potential of interactive media for participatory scenarios on social-ecological systems, *Futures*, 42, 604-616.
- Voinov, A., and Bousquet, F. (2010), Modelling with stakeholders, *Environ. Model. and Soft.*, 25(11), 1268-1281.
- von Neumann, J., and Morgenstern, O., (1944), Theory of games and economic behavior. Princeton, NJ: Princeton University Press.
- Wang, X.M., Qin, Z.L., and Hu, Y.D., (2001), An interactive algorithm for multicriteria decision making: the attainable reference point method, *IEEE Trans. on Sys. Man and Ciber. Part A*, 31(3), 194-198.
- White, K.L., and Chaubey, I., (2005), Sensitivity Analysis, Calibration, and Validations for a Multisite and Multivariable SWAT Model, *JAWRA*, 41(5), 1077-1089.
- Williams, J.R., Smith, C.M., Roe, J.D, Leatherman, J.C., and Wilson, R.M., (2012), Engaging watershed stakeholders for cost-effective environmental management planning with “Watershed Manager”, *J. of Natural Res. And Life Scs. Edu.*, 41, 44-53.
- Wood, F.B., Siegel, E.R., LaCroix, E.M., Lyon, B.J., Benson, D.A., Cid, V., and Fariss, S. (2003), A practical approach to E-Government web evaluation, *IT Professional*, 5(3), 22-28.
- Wurbs, R., (1993), Reservoir-System Simulation and Optimization Models, *J. Water Resour. Plann. Mgmt.*, 119(4), 455-472.
- Yelle, L.E. (1979), The learning curve: historical review and comprehensive survey, *Decision Sciences*, 10(2), 302-328.
- Yeomans, J.S., Huang, G.H., and Yoogalingam, R. (2003), Combining simulation with evolutionary algorithms for optimal planning under uncertainty: An application to municipal solid waste management planning in the regional municipality of Hamilton-Wentworth, *J. of Environ. Info.*, 2(1), 11-30.
- Yigitcanlar, T., and Okabe, A., (2002), Building online participatory systems: towards community based interactive environmental decision support systems, Research Monograph. United Nations University, Institute of Advance Studies, May 2002, Tokyo, Japan
- Zorrilla, P., Carmona, G., De la Hera, A., Varela-Ortega, C., Martinez-Santos, P., Bromley, J., and Henriksen, H.J., (2010), Evaluation of Bayesian Networks in

participatory water resources management, Upper Guadiana Basin, Spain,  
*Ecology and Society*, 15(3), 12.

# APPENDIX

**Chapter 4.** Results per participant for: a) Introspection Session vs. Time spent, b) HS session vs Time spent, c) Total percentage of time in each Area of Interest (AOI, Info = information gathering, DM = evaluation, Other = other), d) Total percentage of clicks in each AOI, and e) Confidence level trend. Notice: Period = Epoch

## Model A-Surrogates

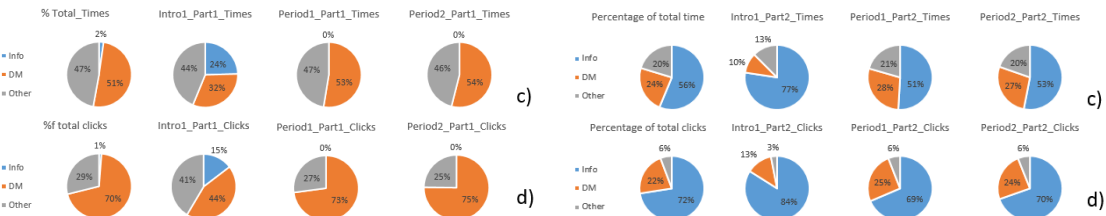
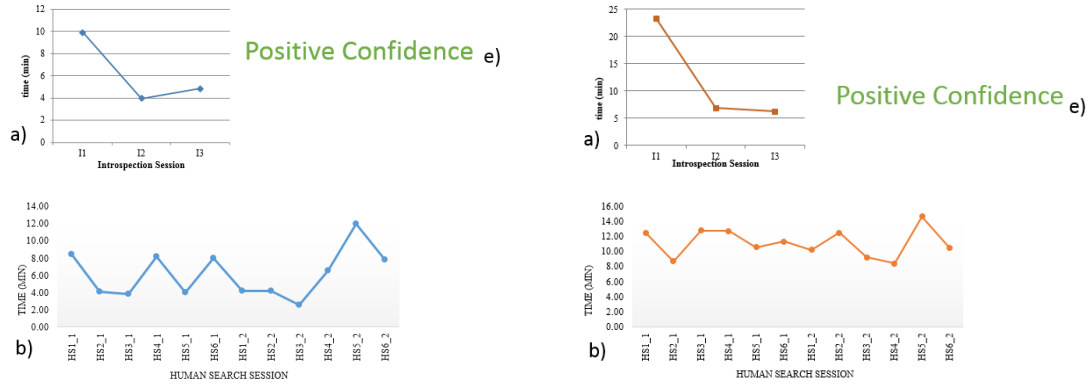
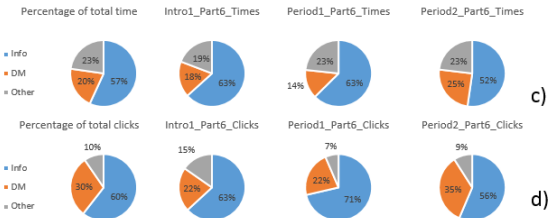
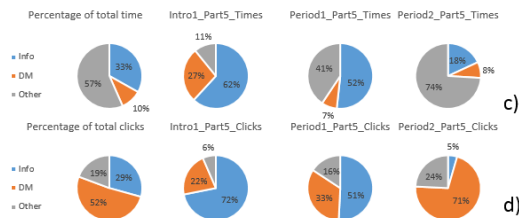
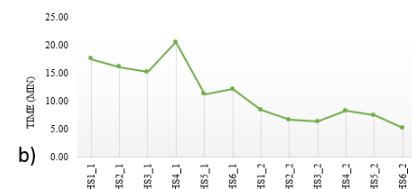
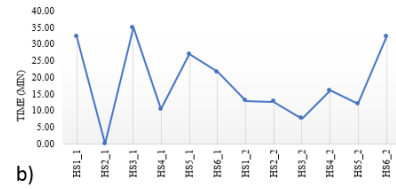
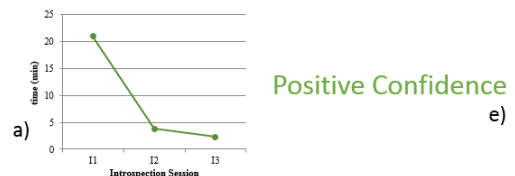
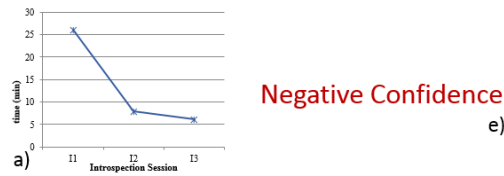


Figure A.1 Usability metrics Participant 1

Figure A.2 Usability metrics Participant 2

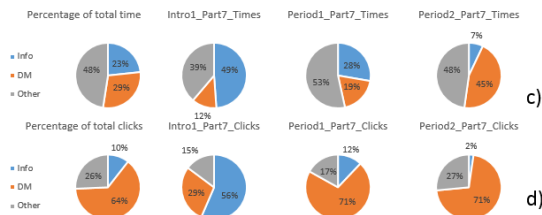
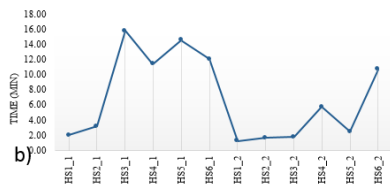
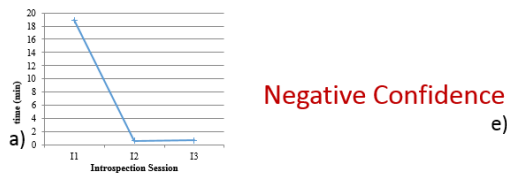
Figure A.3 Usability metrics Participant 3

Figure A.4 Usability metrics Participant 4



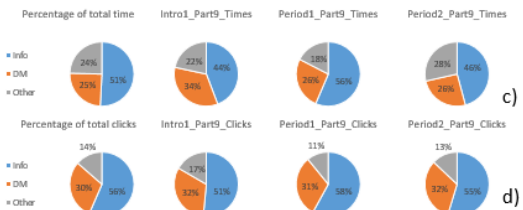
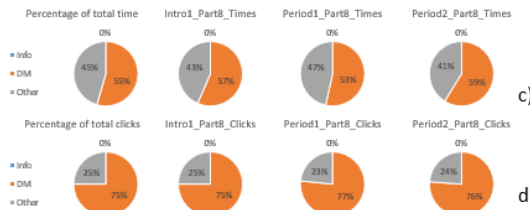
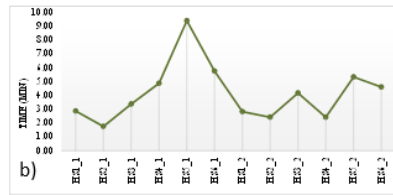
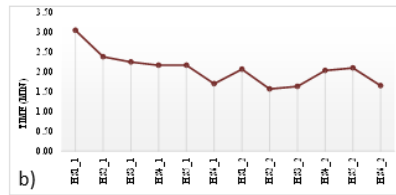
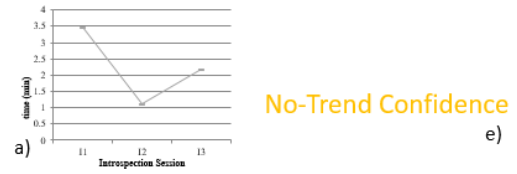
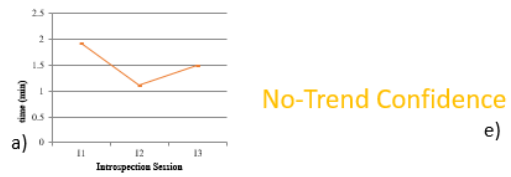
**Figure A.5 Usability metrics Participant 5**

**Figure A.6 Usability metrics Participant 6**



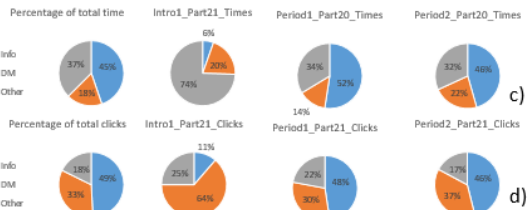
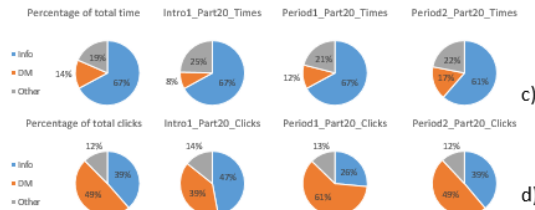
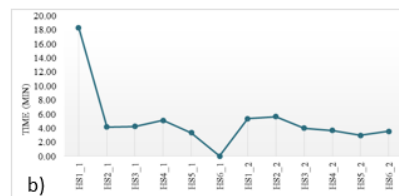
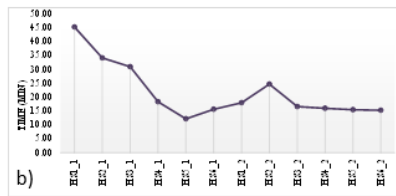
**Figure A.7 Usability metrics Participant 7**

### Model B-Surrogates



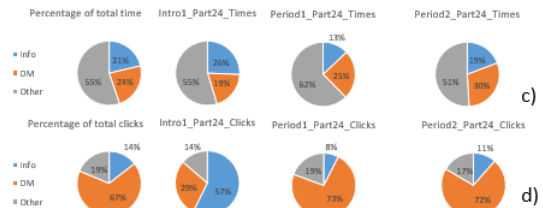
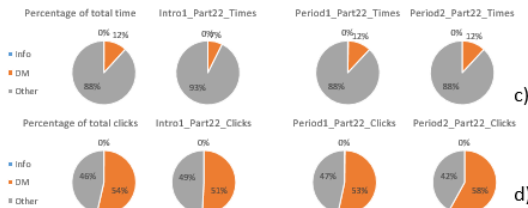
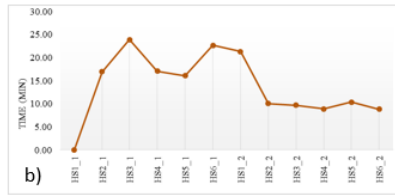
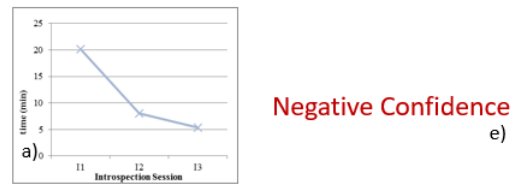
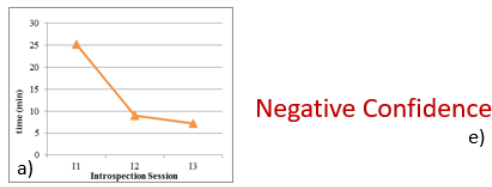
**Figure A.8 Usability metrics Participant 8**

**Figure A.9 Usability metrics Participant 9**



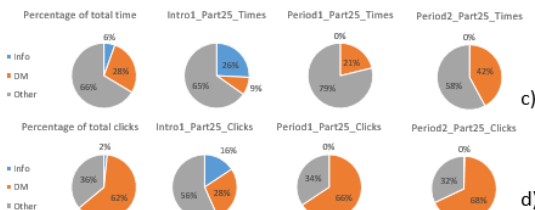
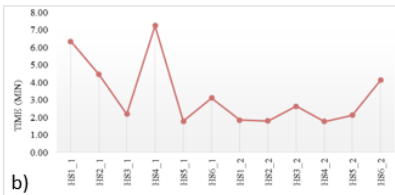
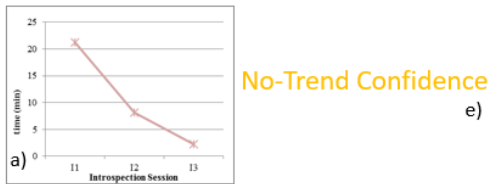
**Figure A.10 Usability metrics Participant 20**

**Figure A.11 Usability metrics Participant 21**



**Figure A.12 Usability metrics Participant 22**

**Figure A.13 Usability metrics Participant 24**



**Figure A.14 Usability metrics Participant 25**



### Model B-Stakeholders

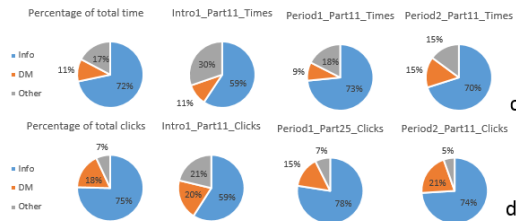
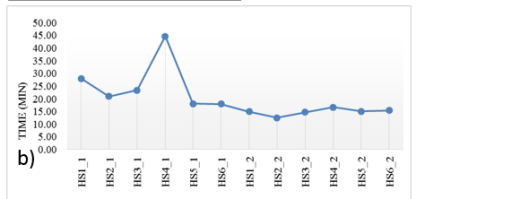
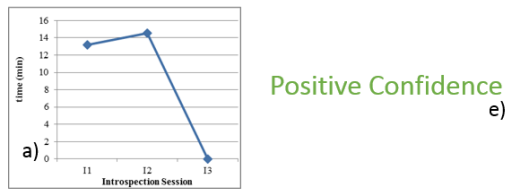


Figure A.15 Usability metrics Participant 11

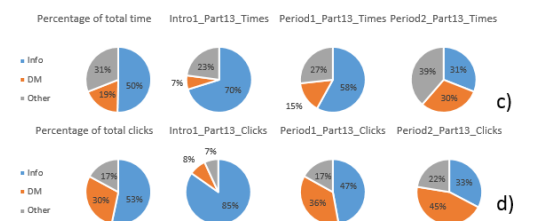
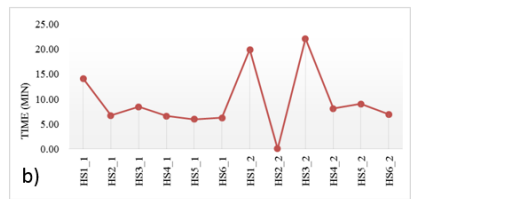
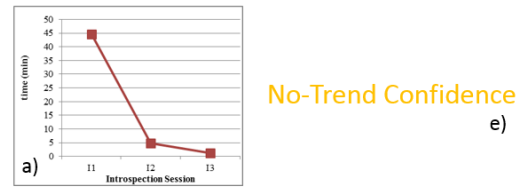


Figure A.16 Usability metrics Participant 13

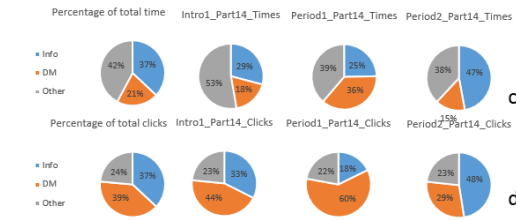
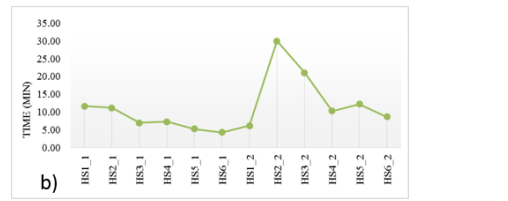
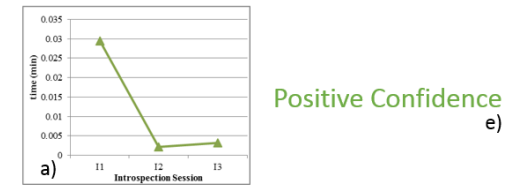


Figure A.17 Usability metrics Participant 14

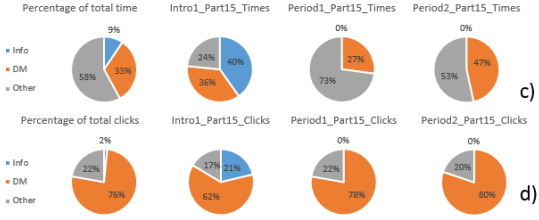
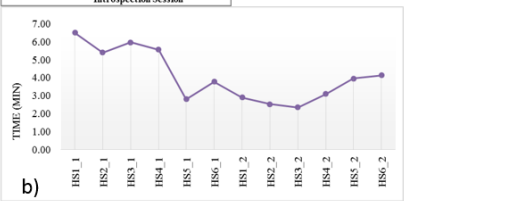
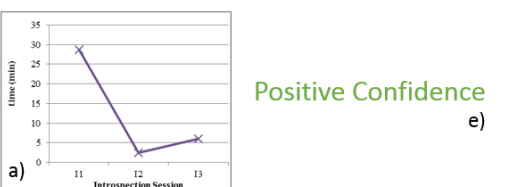


Figure A.18 Usability metrics Participant 15

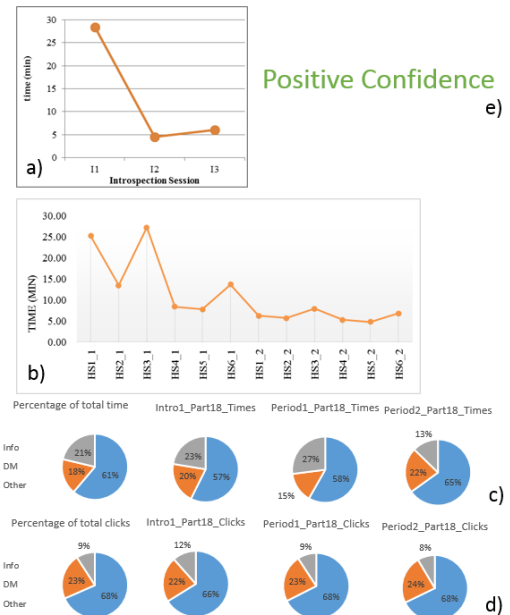
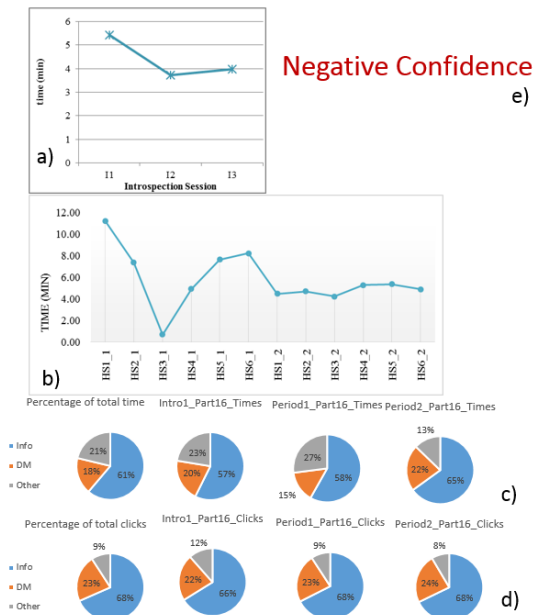


Figure A.19 Usability metrics Participant 16

Figure A.20 Usability metrics Participant 18

## Chapter 5

### *Model A-Surrogates*

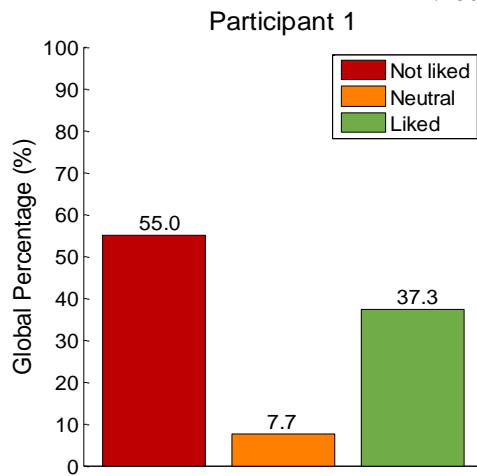


Figure A.21 EoS Percentage of design alternatives per user rating Participant 1

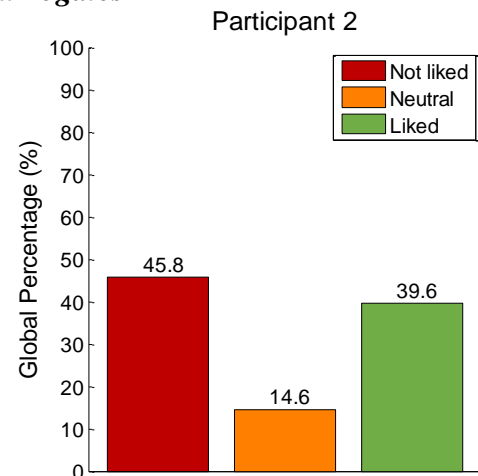


Figure A.22 EoS Percentage of design alternatives per user rating Participant 2

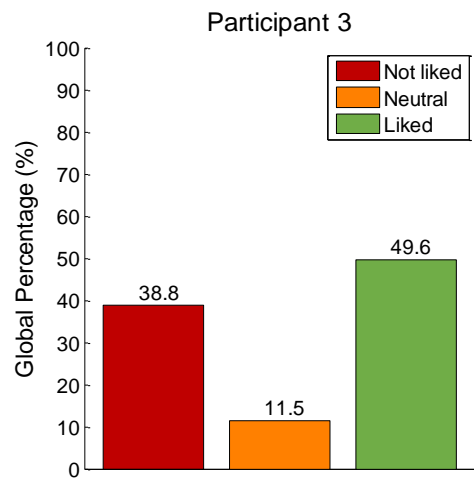


Figure A.23 EoS Percentage of design alternatives per user rating Participant 3

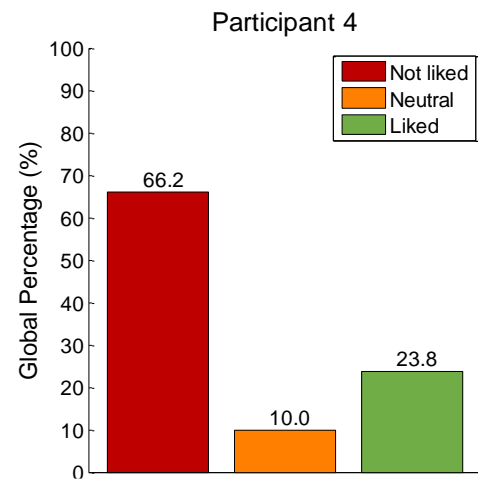


Figure A.24 EoS Percentage of design alternatives per user rating Participant 4

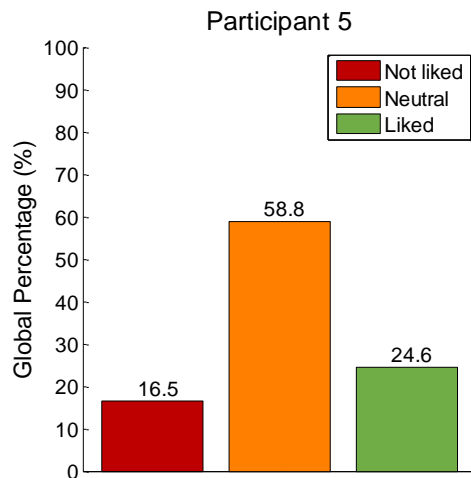


Figure A.25 EoS Percentage of design alternatives per user rating Participant 5

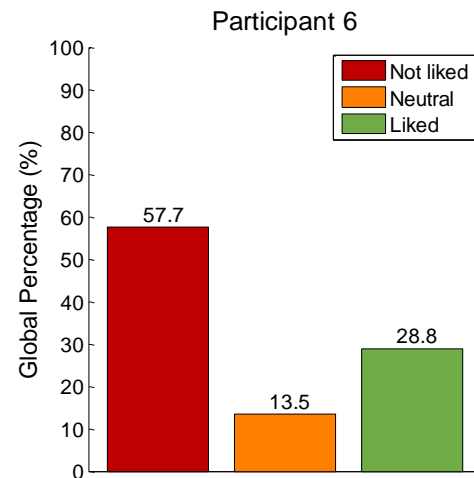


Figure A.26 EoS Percentage of design alternatives per user rating Participant 6

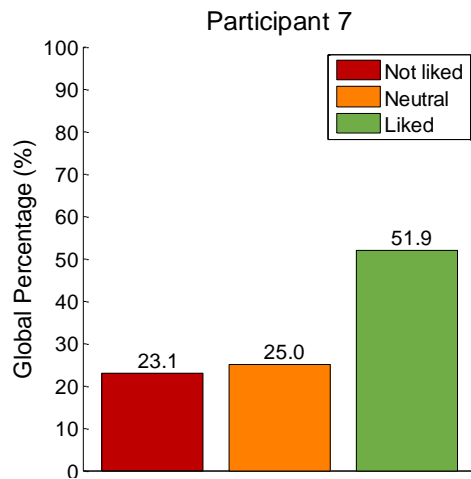


Figure A.27 EoS Percentage of design alternatives per user rating Participant 7

*Model B-Surrogates*

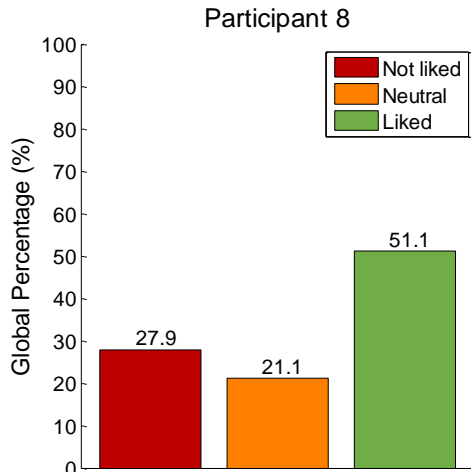


Figure A.28 EoS Percentage of design alternatives per user rating Participant 8

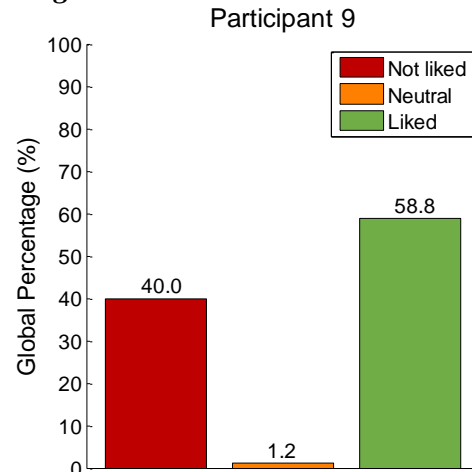


Figure A.29 EoS Percentage of design alternatives per user rating Participant 9

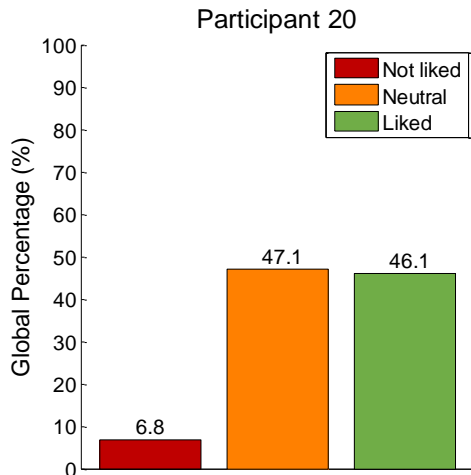


Figure A.30 EoS Percentage of design alternatives per user rating Participant 20

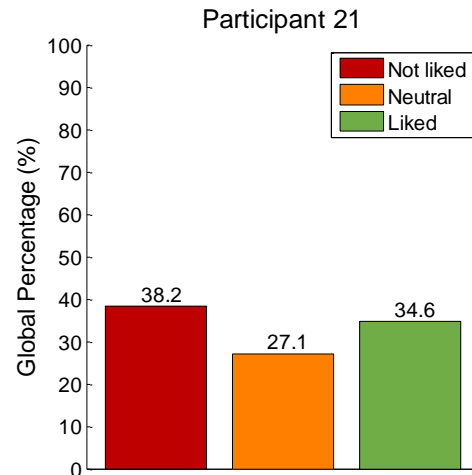
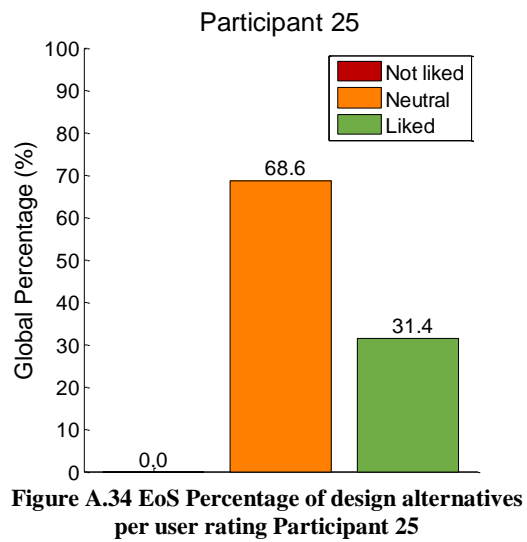
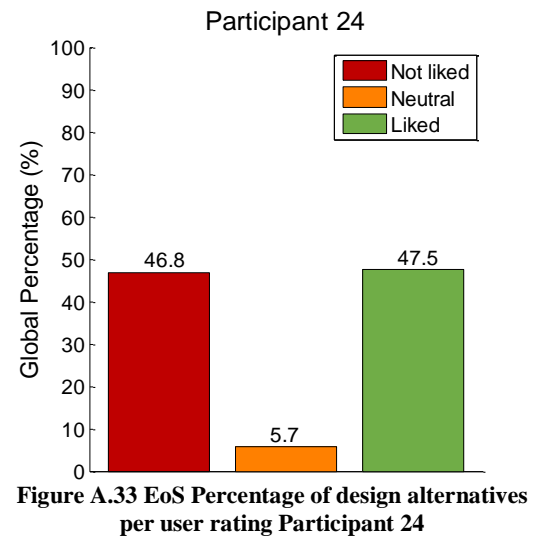
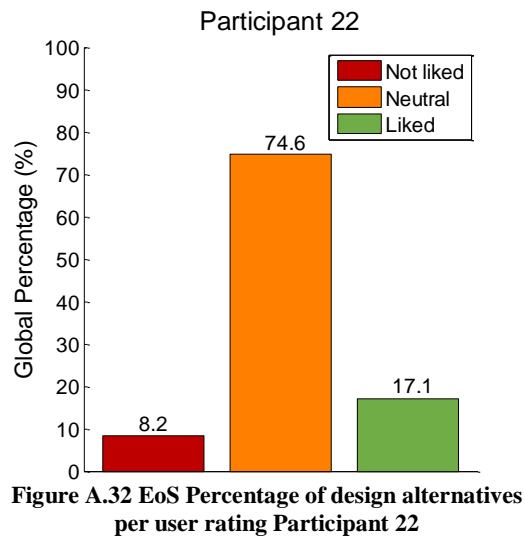


Figure A.31 EoS Percentage of design alternatives per user rating Participant 21



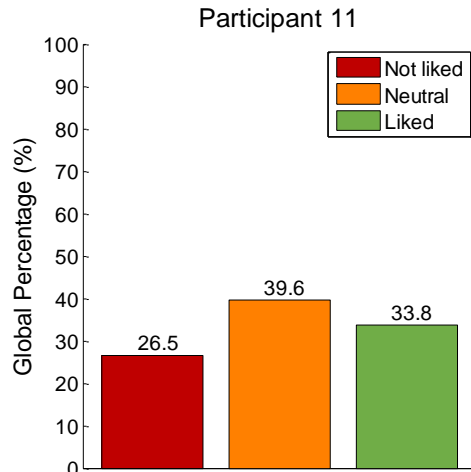
*Model B-Stakeholders*

Figure A.35 EoS Percentage of design alternatives per user rating Participant 11

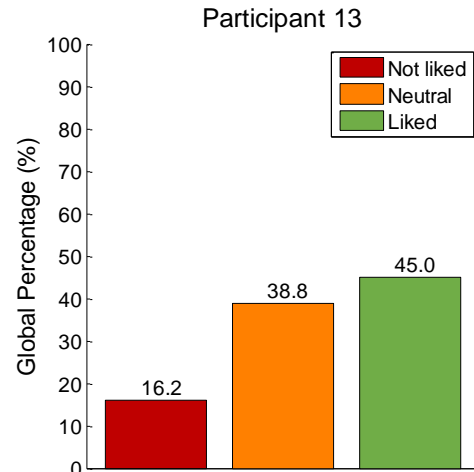


Figure A.36 EoS Percentage of design alternatives per user rating Participant 13

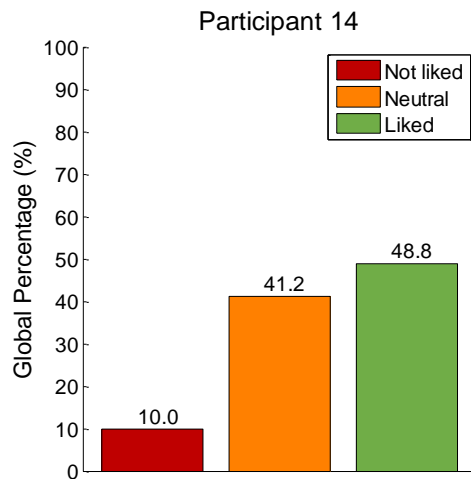


Figure A.37 EoS Percentage of design alternatives per user rating Participant 14

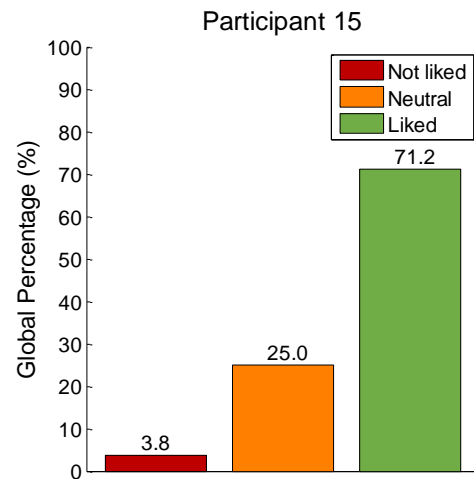


Figure A.38 EoS Percentage of design alternatives per user rating Participant 15

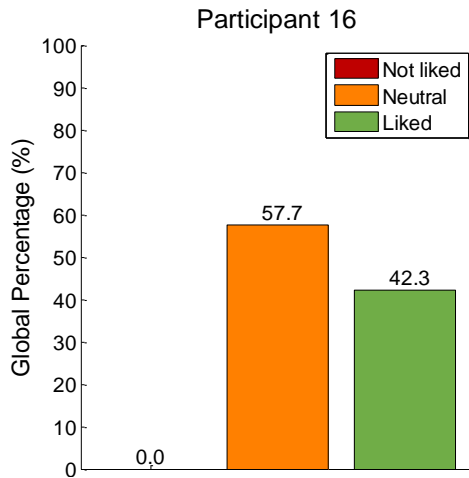


Figure A.39 EoS Percentage of design alternatives per user rating Participant 16

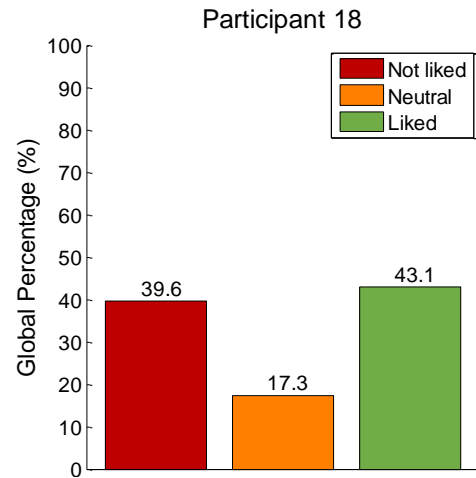


Figure A.40 EoS Percentage of design alternatives per user rating Participant 18

**DS (During Search) Percentage of design alternatives per user rating**

*Model A-Surrogates*

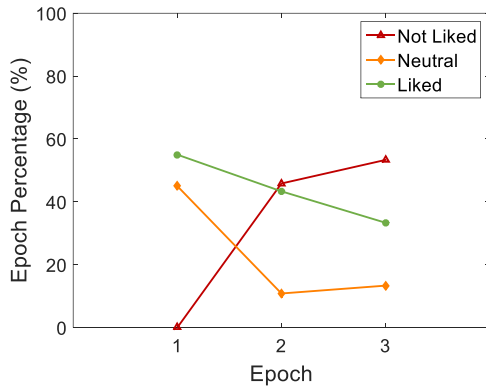


Figure A.41 DS Percentage of design alternatives per user rating Participant 2

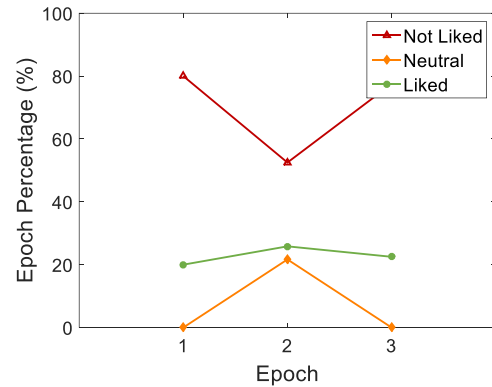


Figure A.42 DS Percentage of design alternatives per user rating Participant 4

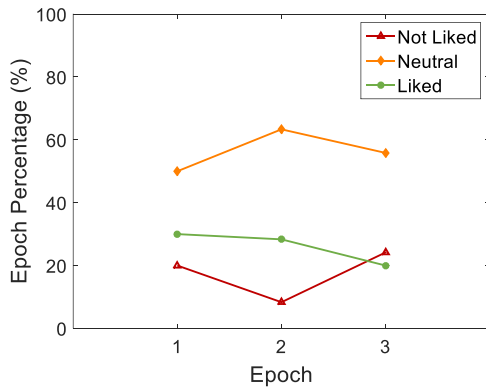


Figure A.43 DS Percentage of design alternatives per user rating Participant 5

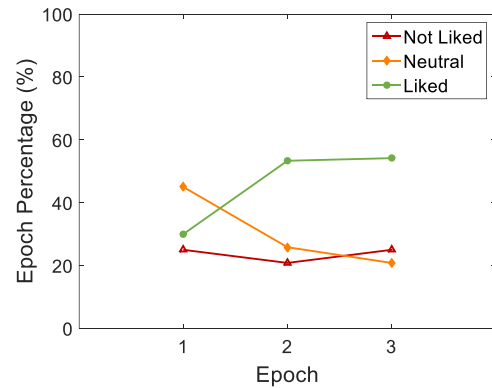
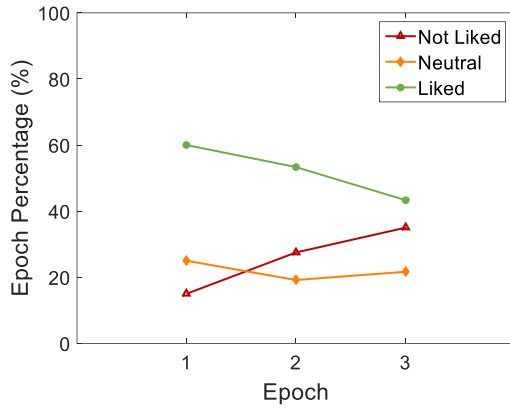


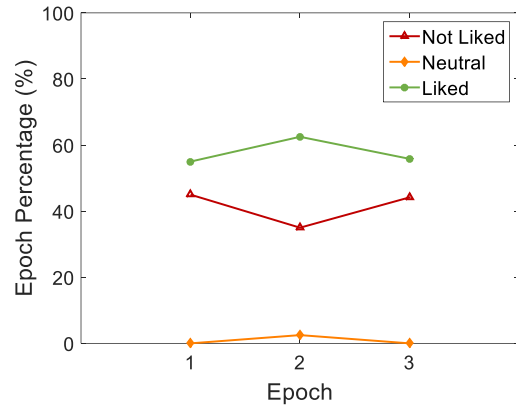
Figure A.44 DS Percentage of design alternatives per user rating Participant 7



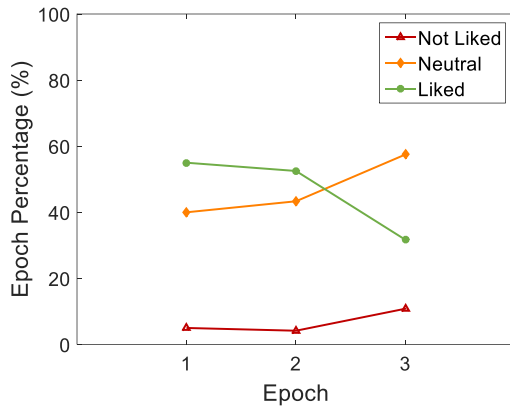
*Model B-Surrogates*



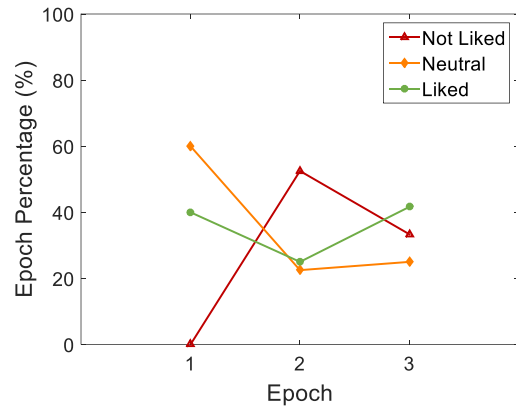
**Figure A.45** DS Percentage of design alternatives per user rating Participant 8



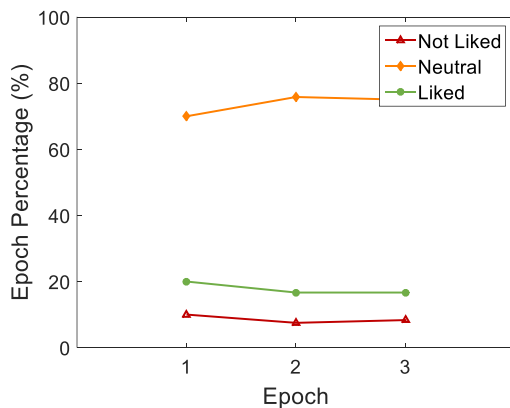
**Figure A.46** DS Percentage of design alternatives per user rating Participant 9



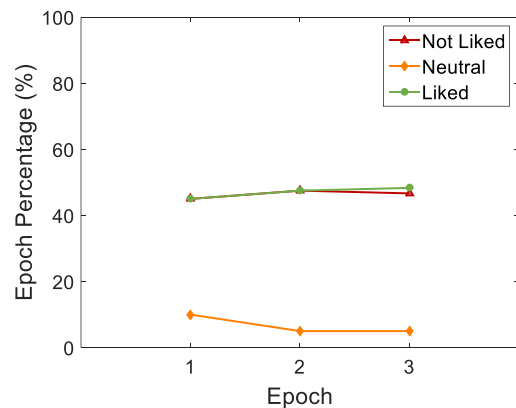
**Figure A.47** DS Percentage of design alternatives per user rating Participant 20



**Figure A.48** DS Percentage of design alternatives per user rating Participant 21



**Figure A.49** DS Percentage of design alternatives per user rating Participant 22



**Figure A.50** DS Percentage of design alternatives per user rating Participant 24

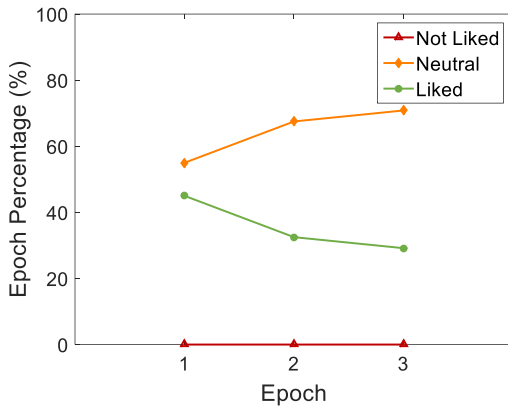


Figure A.51 DS Percentage of design alternatives per user rating Participant 25

*Model B-Stakeholders*

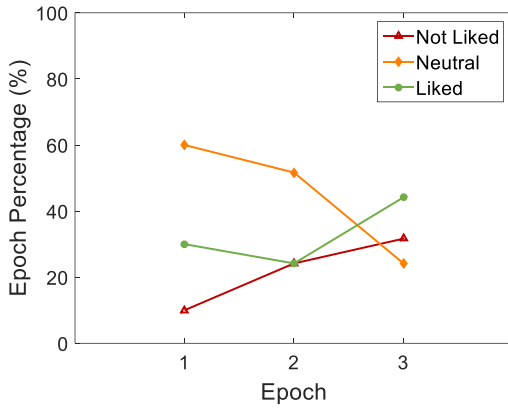


Figure A.52 DS Percentage of design alternatives per user rating Participant 11

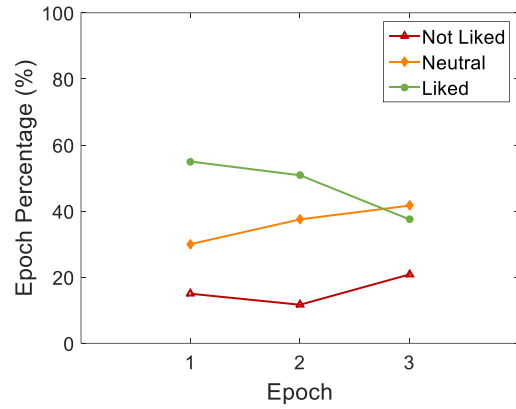


Figure A.53 DS Percentage of design alternatives per user rating Participant 13

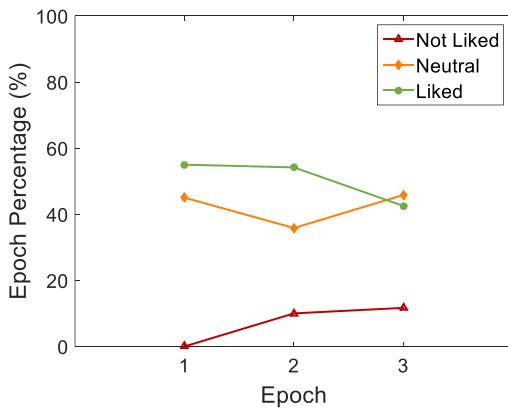


Figure A.54 DS Percentage of design alternatives per user rating Participant 14

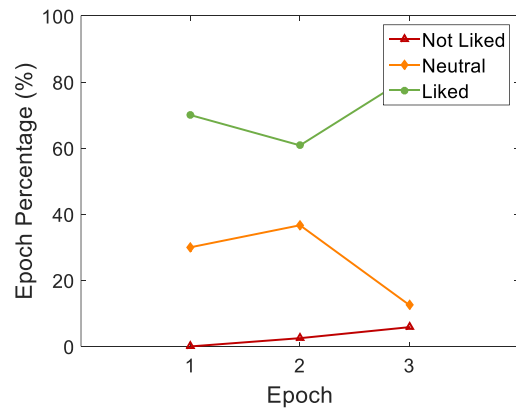


Figure A. 55 DS Percentage of design alternatives per user rating Participant 15

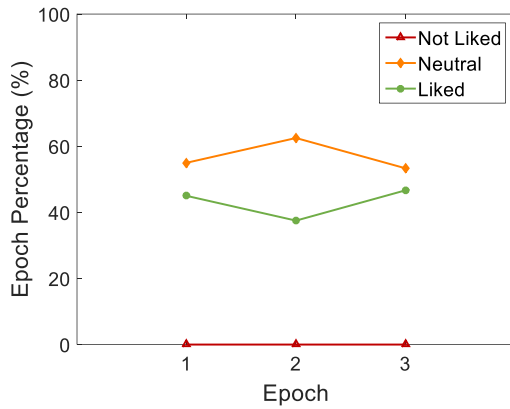


Figure A.56 DS Percentage of design alternatives per user rating Participant 16

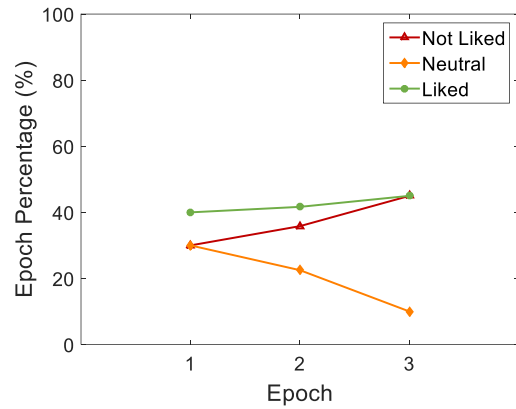


Figure A.57 DS Percentage of design alternatives per user rating Participant 18

**EoS (End of Search) Similarities in Objective Space per user per rating**

*Model A-Surrogates*

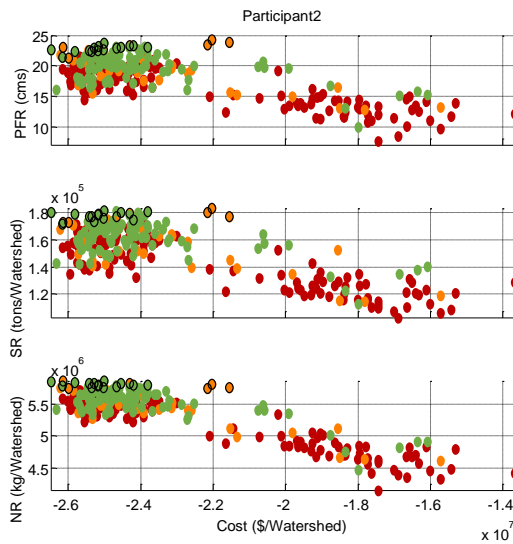


Figure A. 58 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 2

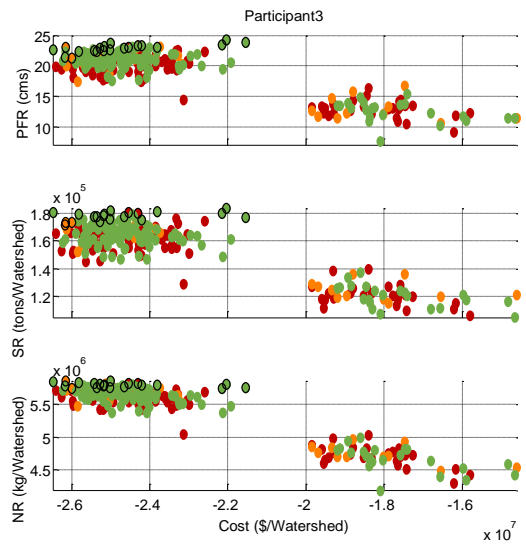
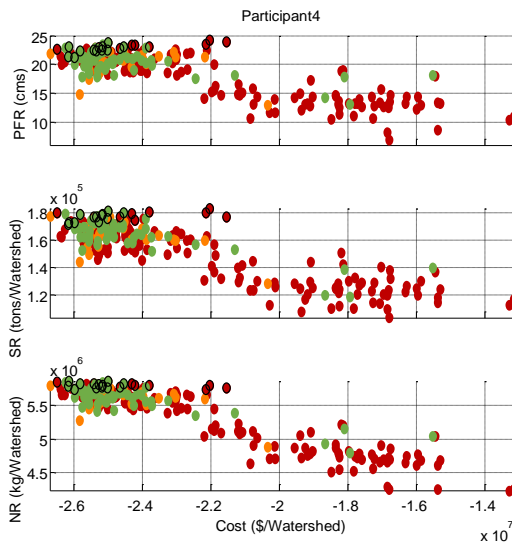
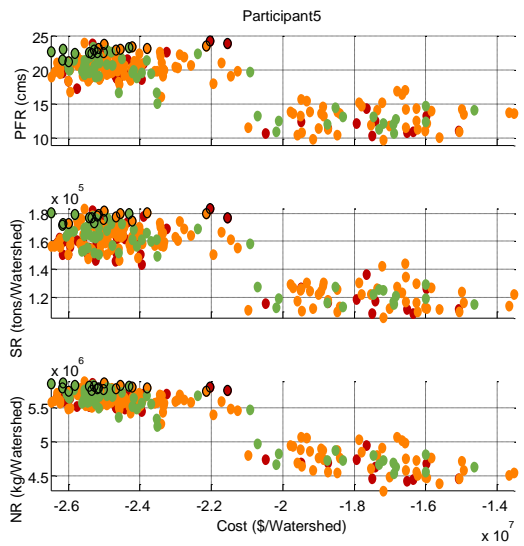


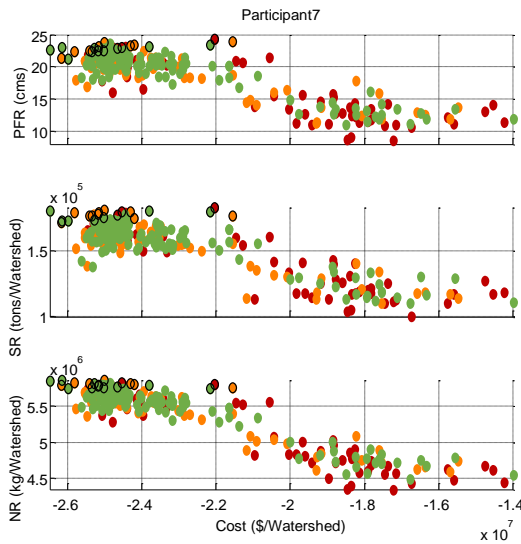
Figure A. 59 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 3



**Figure A.60 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 4**

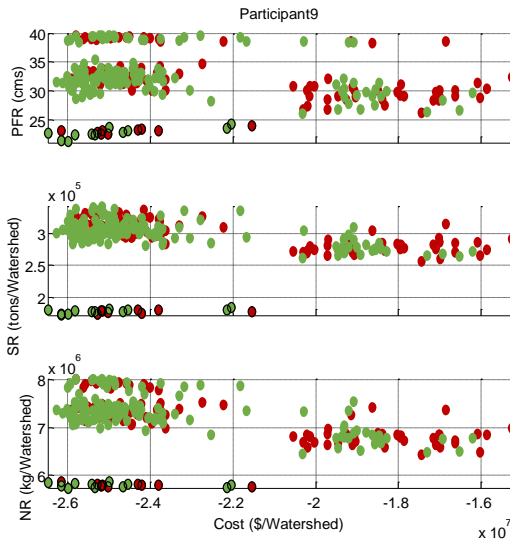


**Figure A.61 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 5**

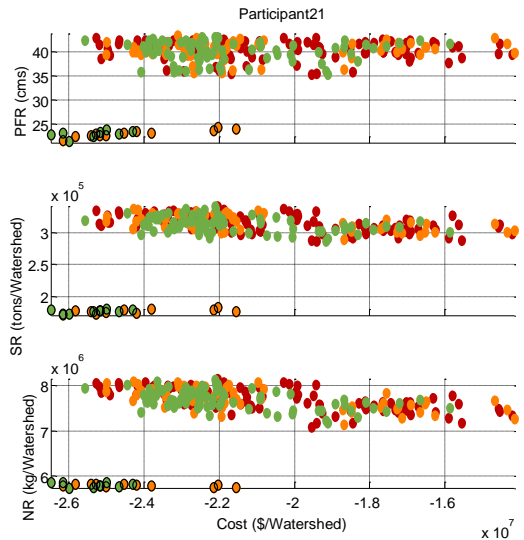


**Figure A. 62 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 7**

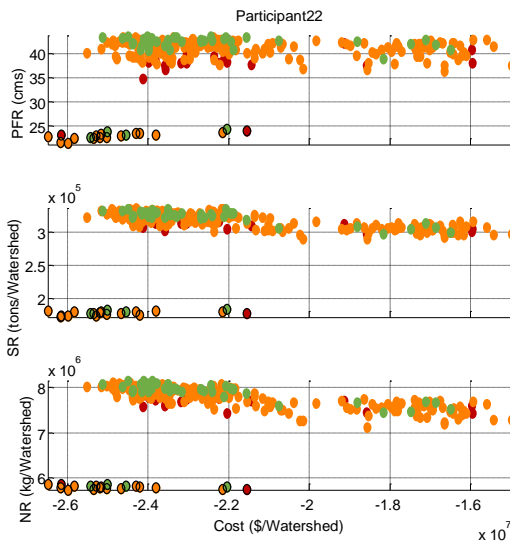
*Model B-Surrogates*



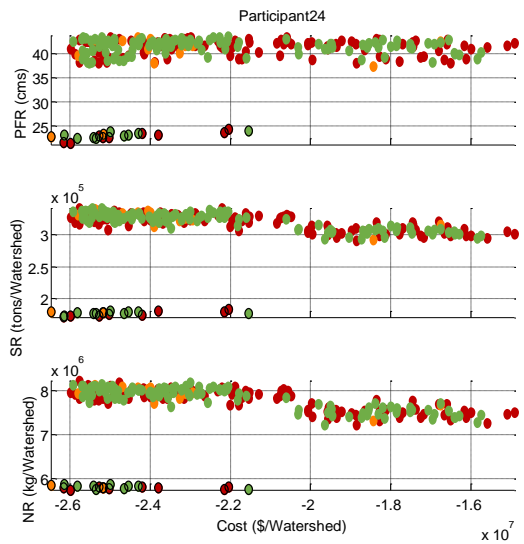
**Figure A.63** Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 9



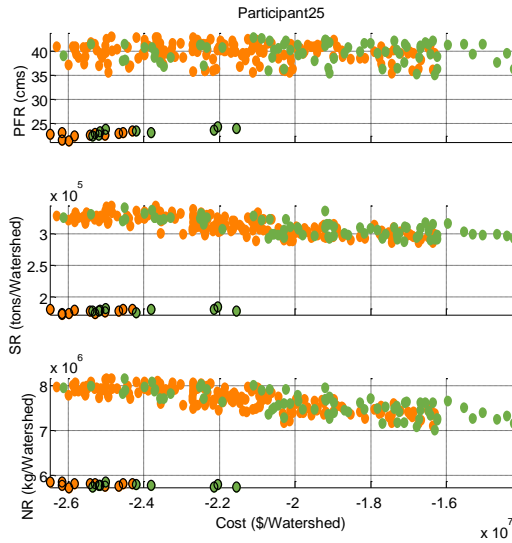
**Figure A.64** Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 21



**Figure A.65** Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 22

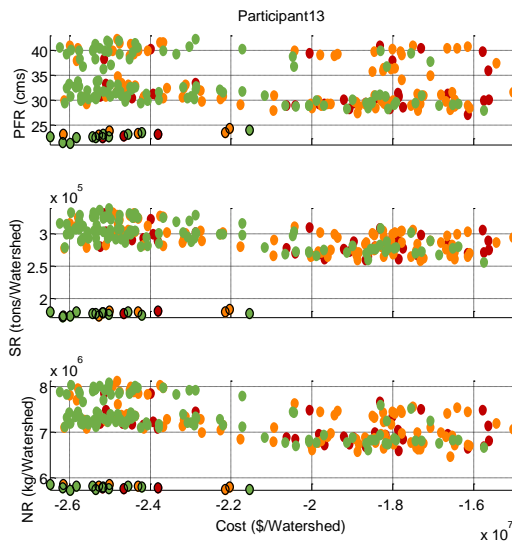


**Figure A.66** Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 24

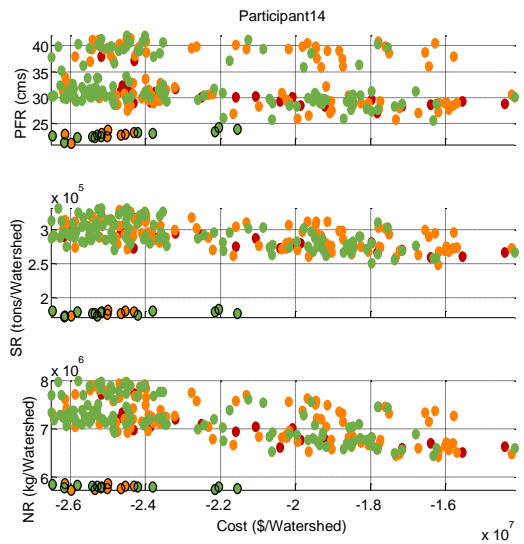


**Figure A. 67 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 25**

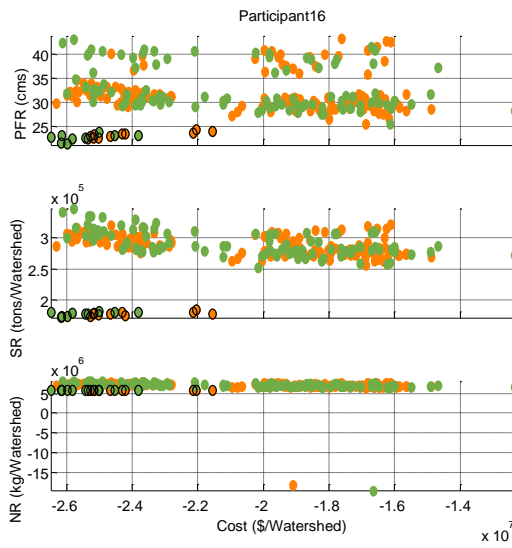
*Model B-Stakeholders*



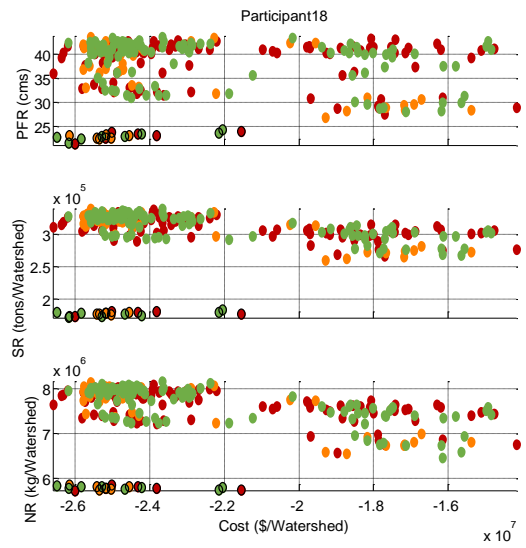
**Figure A.68 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 13**



**Figure A.69 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 14**

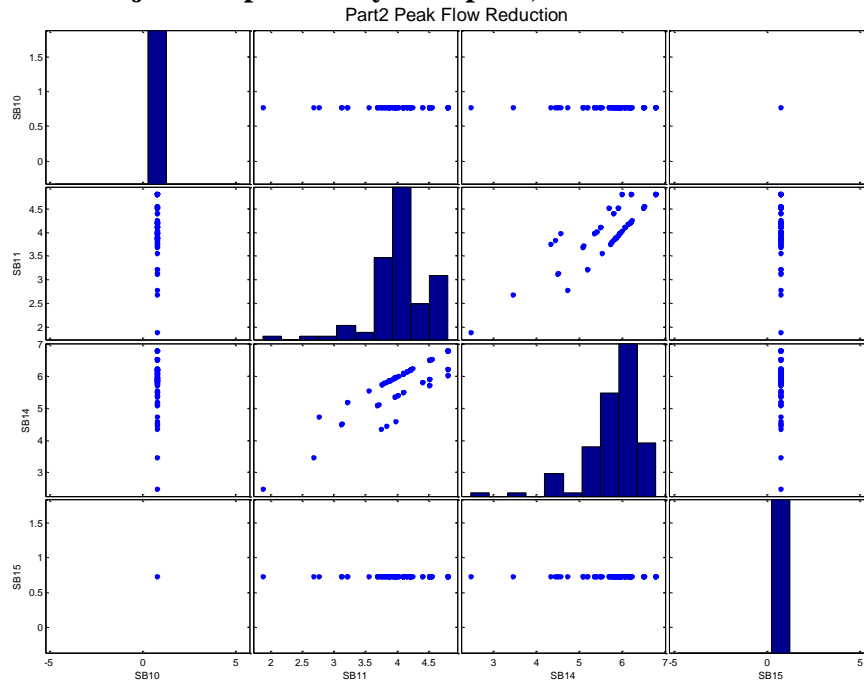


**Figure A.70 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 16**

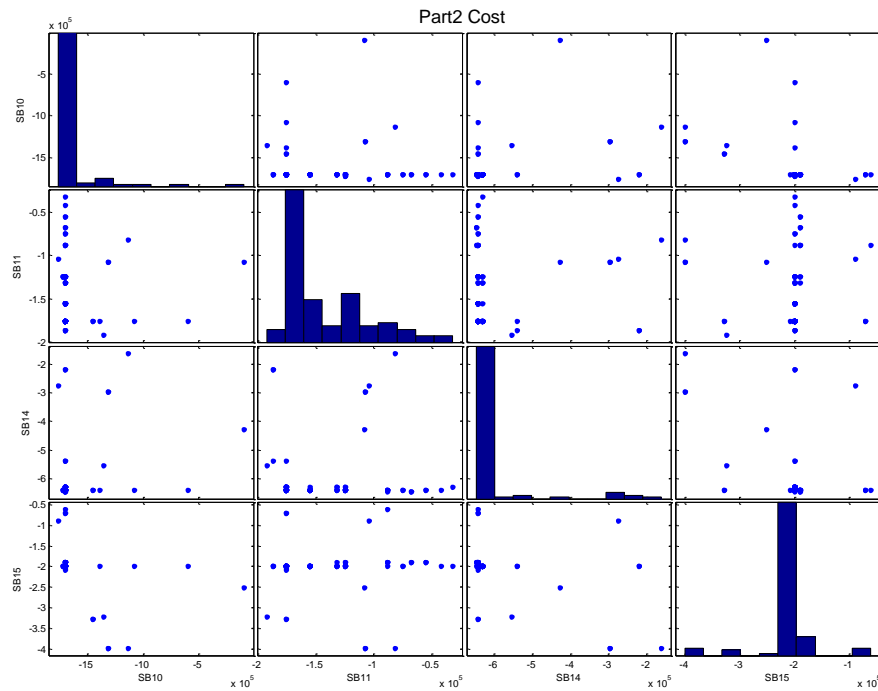


**Figure A.71 Objective Space performance for PFR (Peak Flow Reduction), SR (Sediment Reduction) and NR (Nitrate Reduction) vs. Cost per user rating for Participant 18**

**Chapter 6**  
**Subbasins of interest Objective space analysis: Gplots, for SBint<sub>1</sub>**



**Figure A.72** Matrix plot for performance of Peak flow reduction (PFR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 2



**Figure A.73** Matrix plot for performance of Cost of the design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 2



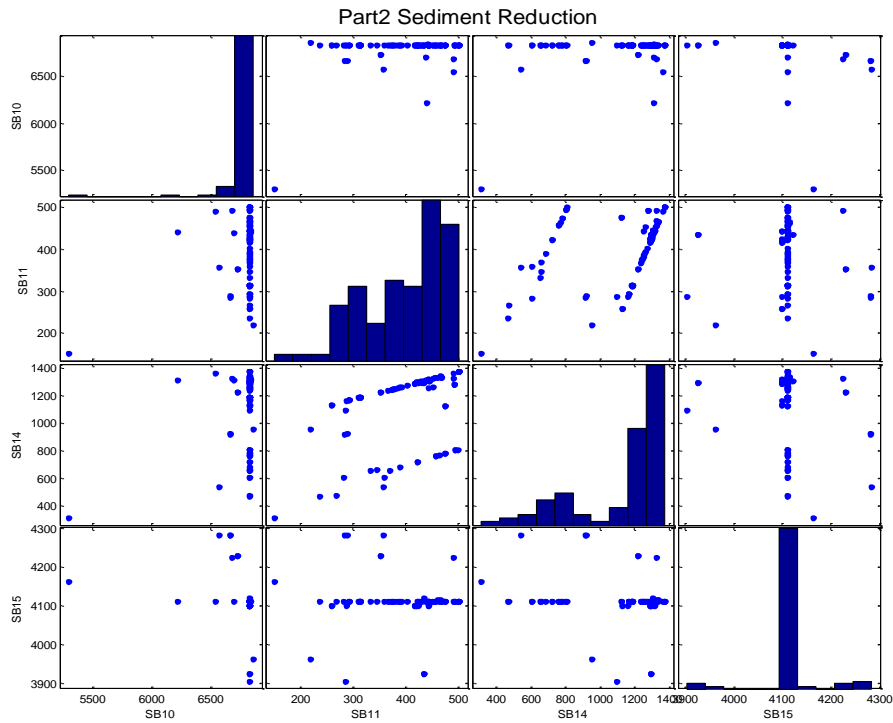


Figure A.74 Matrix plot for performance of Sediment reduction (SR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 2

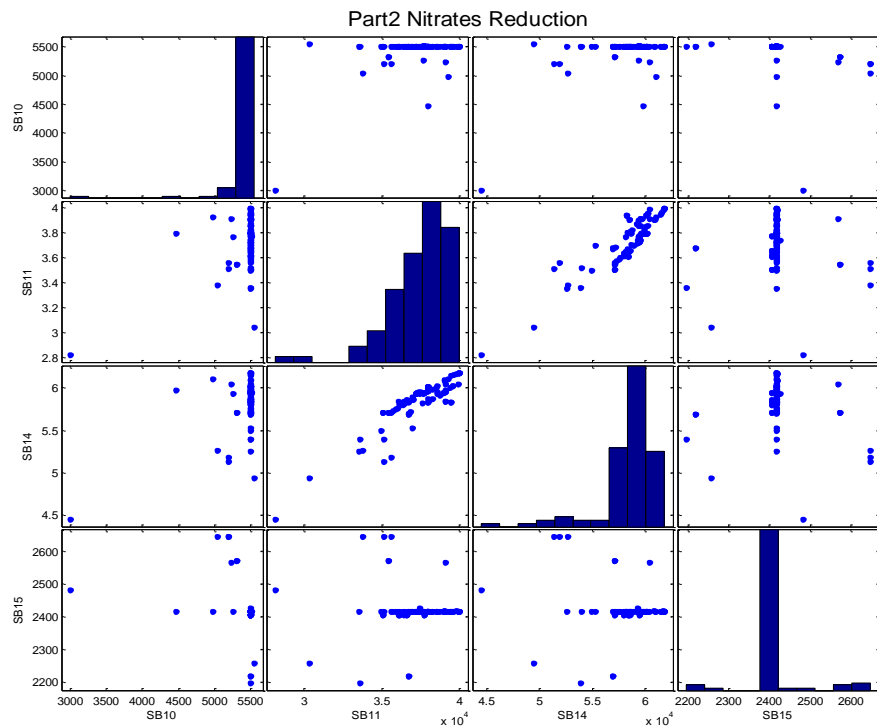


Figure A.75 Matrix plot for performance of Nitrates reduction (NR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 2

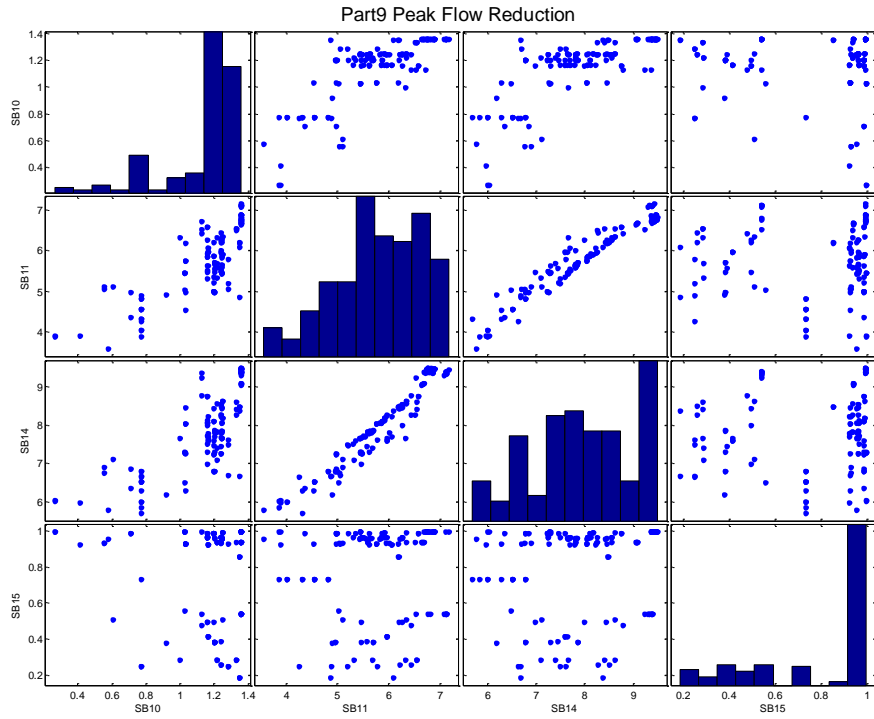


Figure A.76 Matrix plot for performance of Peak flow reduction (PFR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 9

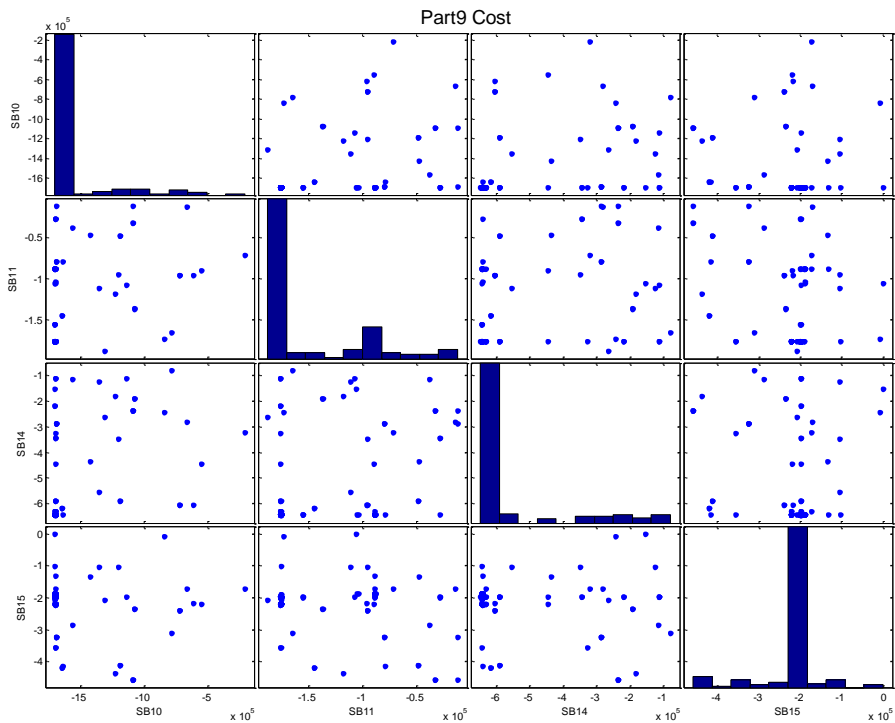


Figure A. 77 Matrix plot for performance of Cost of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 9

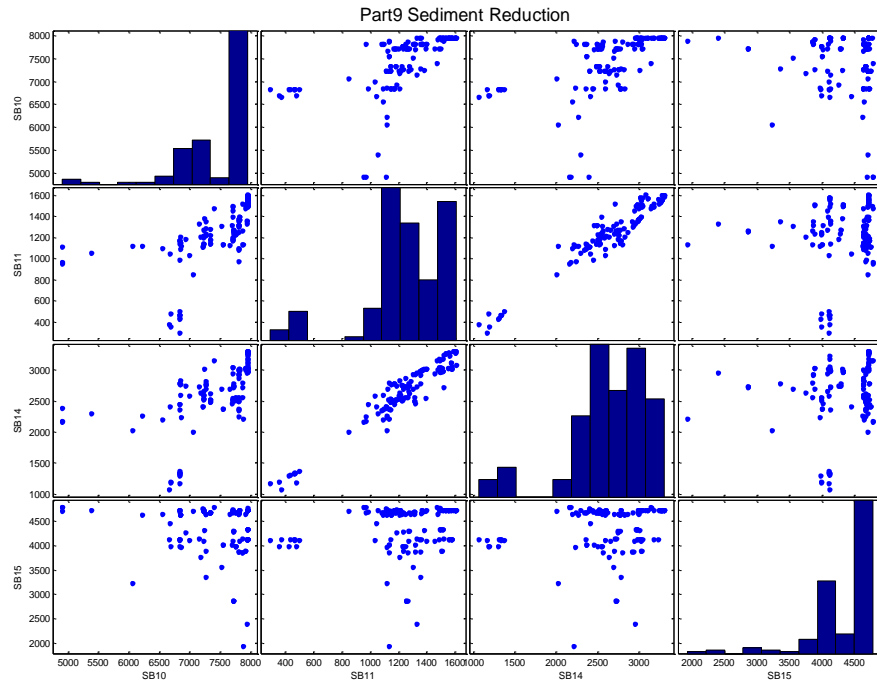


Figure A.78 Matrix plot for performance of Sediment reduction (SR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 9

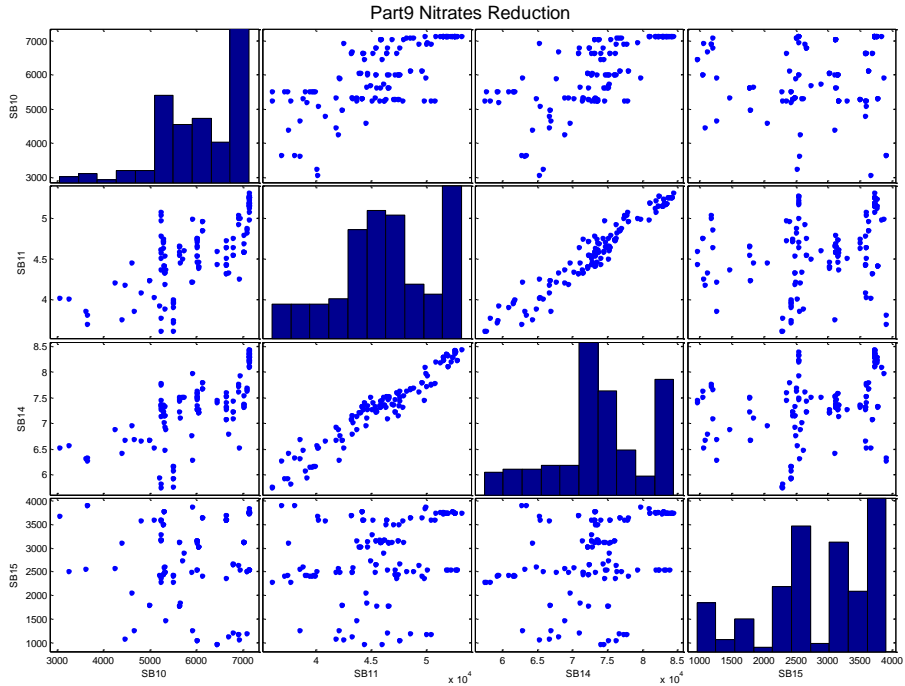


Figure A.79 Matrix plot for performance of Nitrate reduction (NR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 9

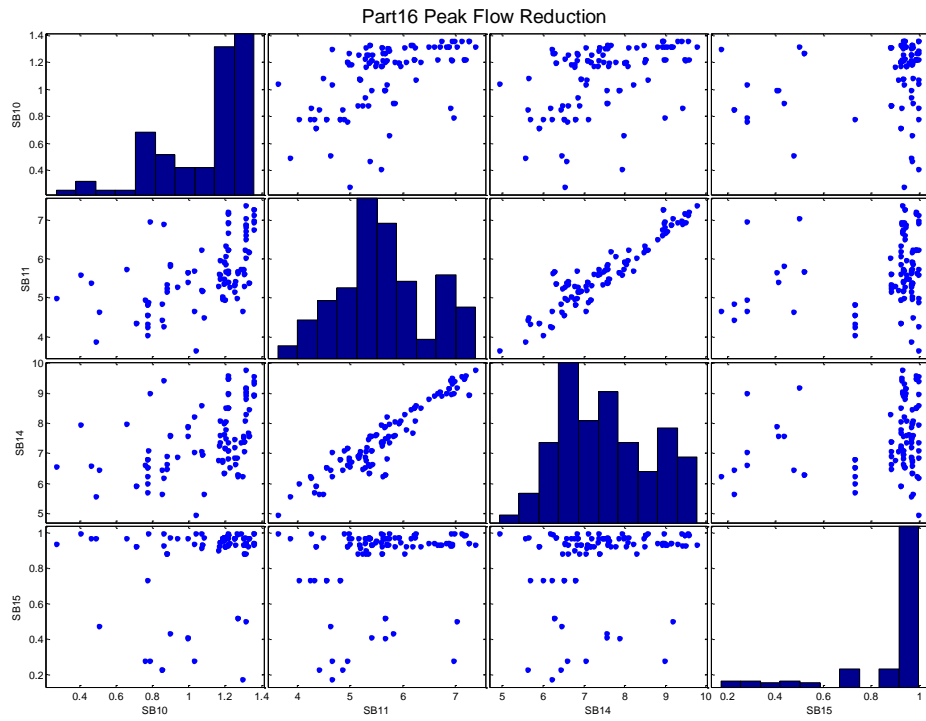


Figure A. 80 Matrix plot for performance of Peak flow reduction (PFR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 16

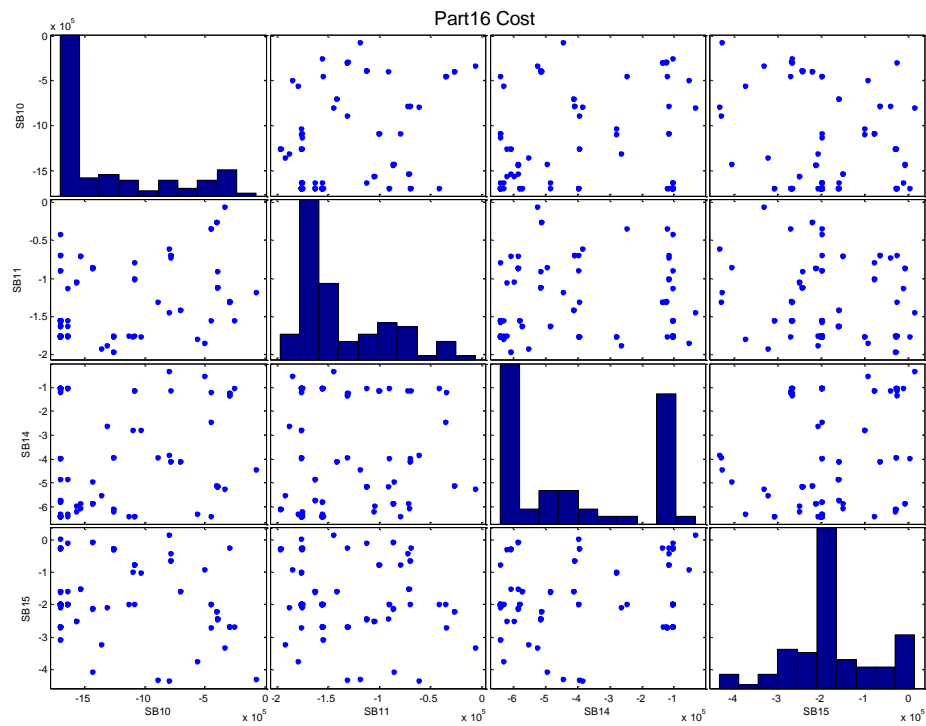


Figure A.81 Matrix plot for performance of Cost of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 16

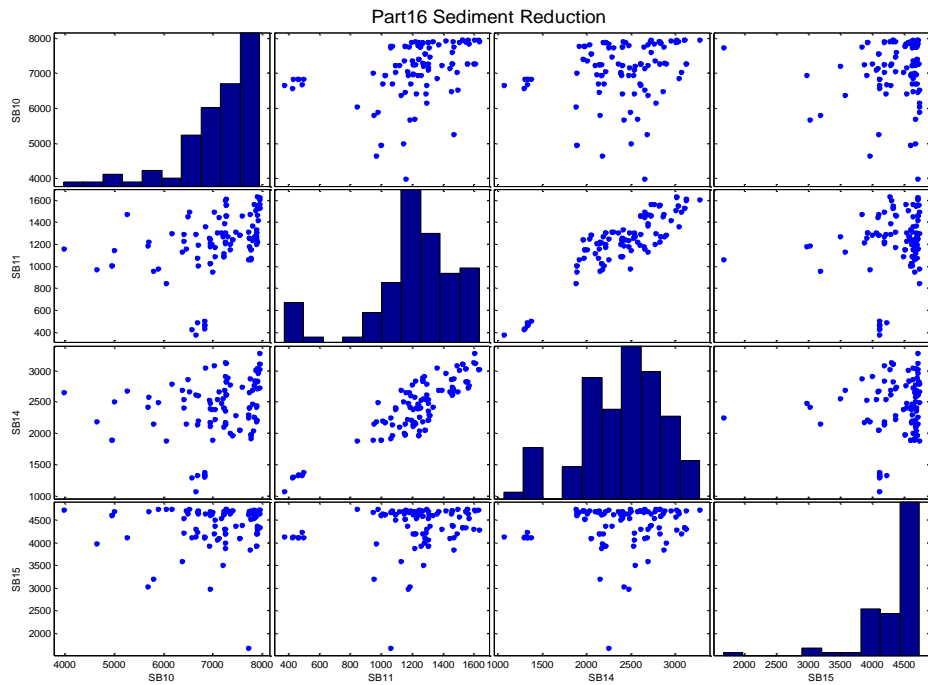


Figure A.82 Matrix plot for performance of Sediment Reduction (SR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 16

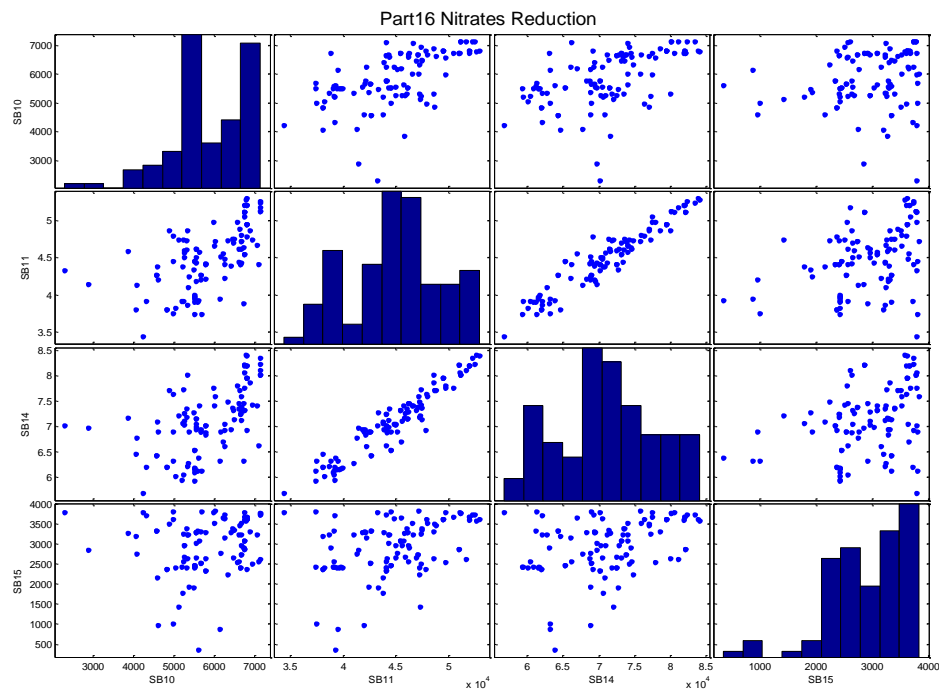


Figure A.83 Matrix plot for performance of Sediment Reduction (SR) of design alternatives rated “I like it” in Subbasins of interest for group 1/Participant 16

**Group analysis for SBint 2 to 6**

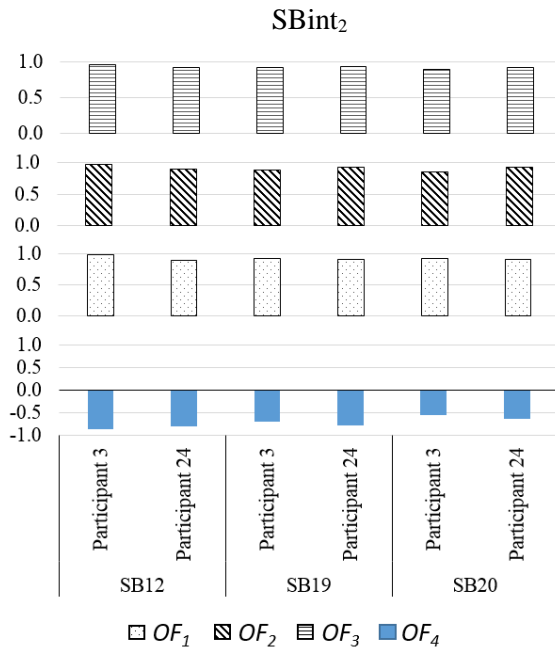


Figure A.84 Ratio group analysis for SB of interest in group 2

**SBint<sub>3</sub>**

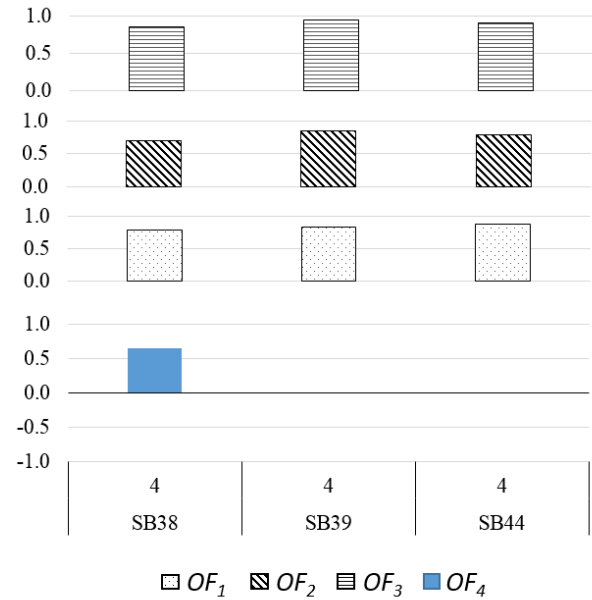


Figure A.85 Ratio group analysis for SB of interest in group 3

**SBint<sub>4</sub>**

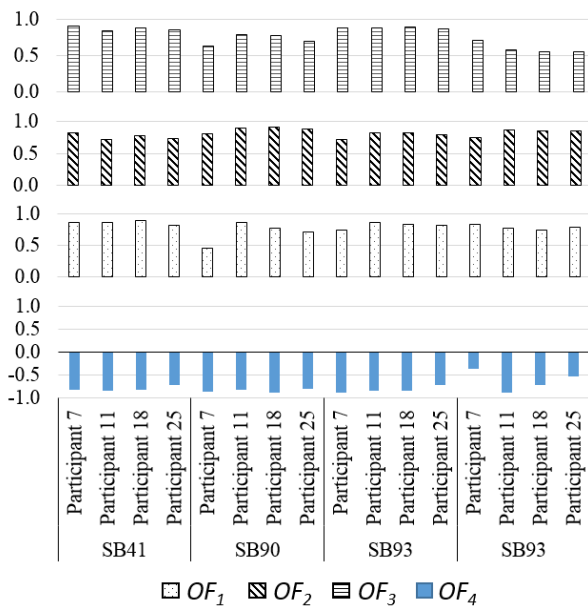


Figure A.86 Ratio group analysis for SB of interest in group 4

**SBint<sub>5</sub>**

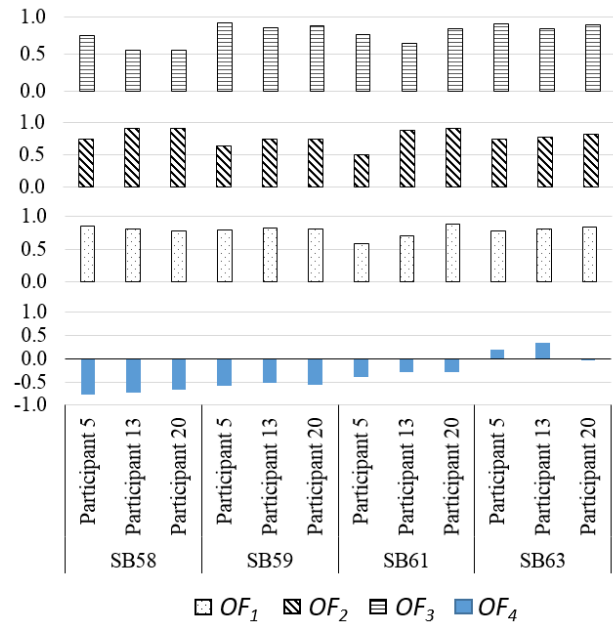


Figure A.87 Ratio group analysis for SB of interest in group 5

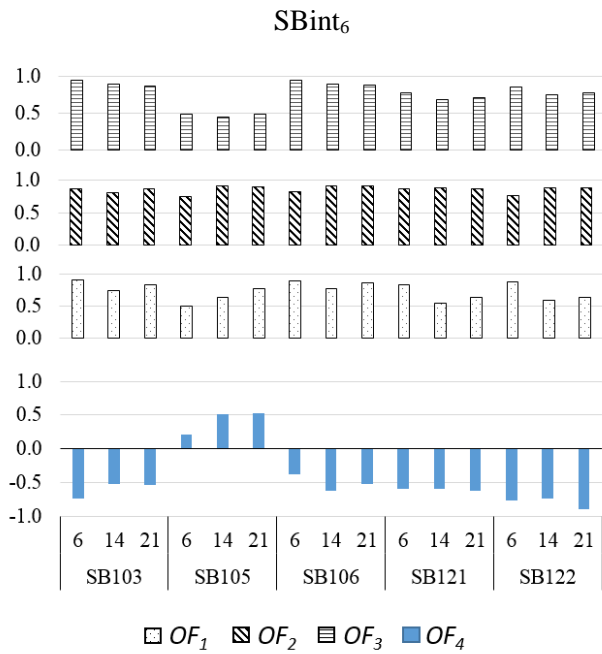


Figure A.88 Ratio group analysis for SB of interest in group 6

### Decision Space Analysis

Individual bar plots for Probability of CoverCrops by Participants, and Average by group

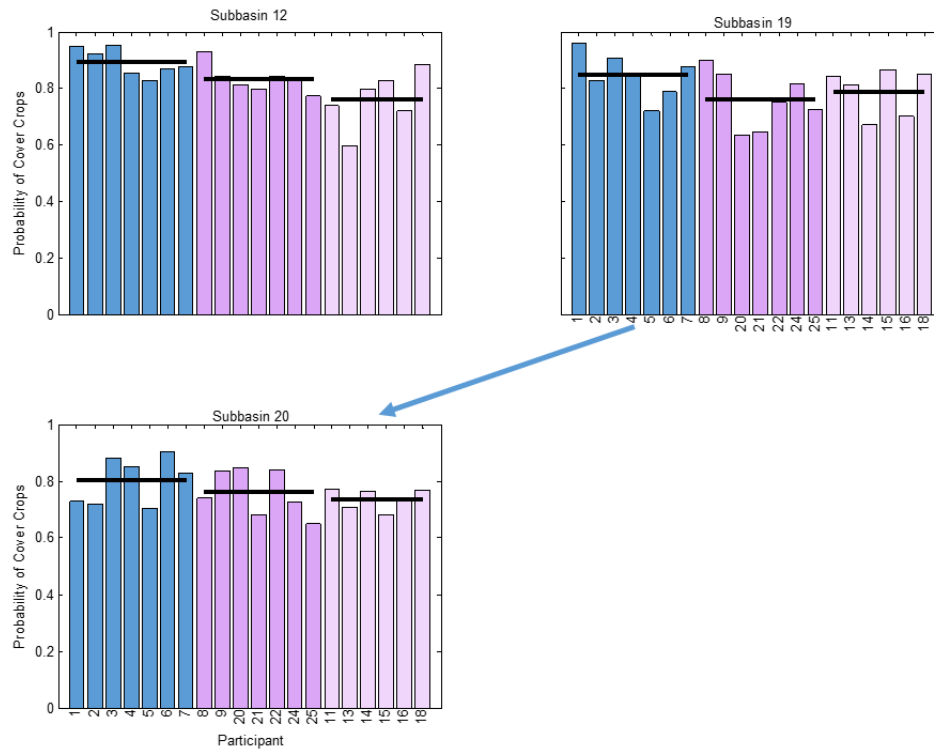
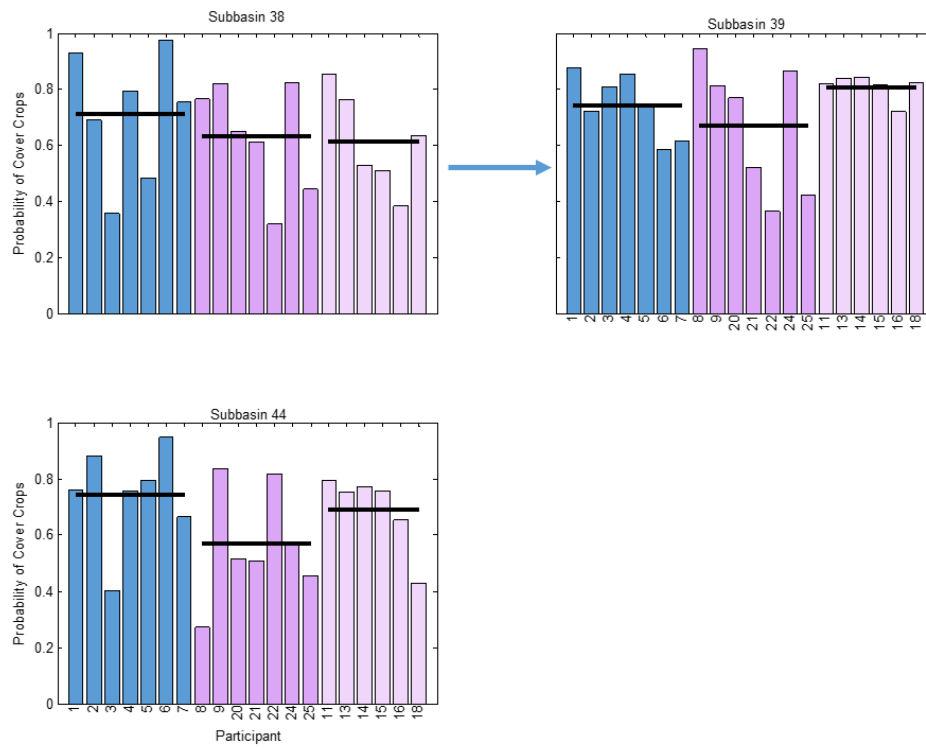
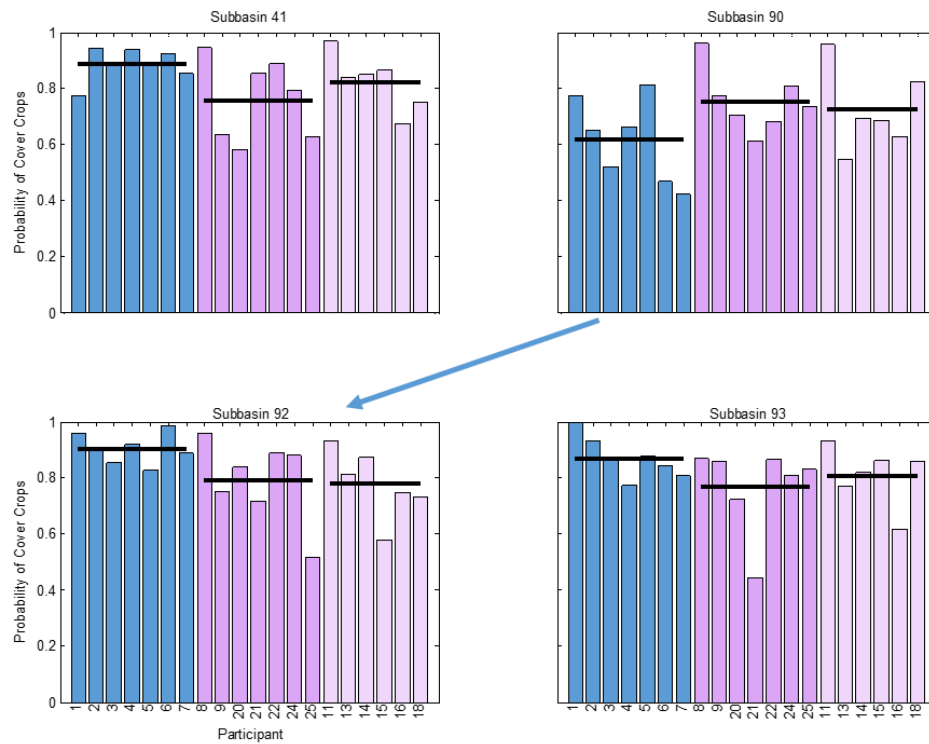


Figure A.89 Bar plots for Probabilities of Cover Crop in SB of interest in group 2

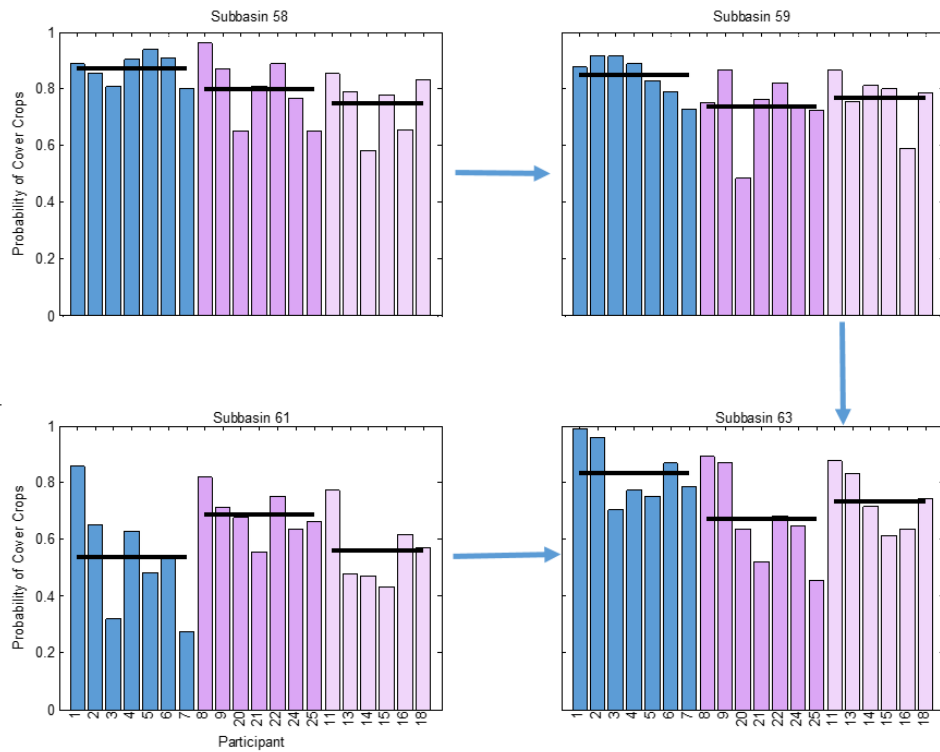


**Figure A.90** Bar plots for Probabilities of Cover Crop in SB of interest in group 3

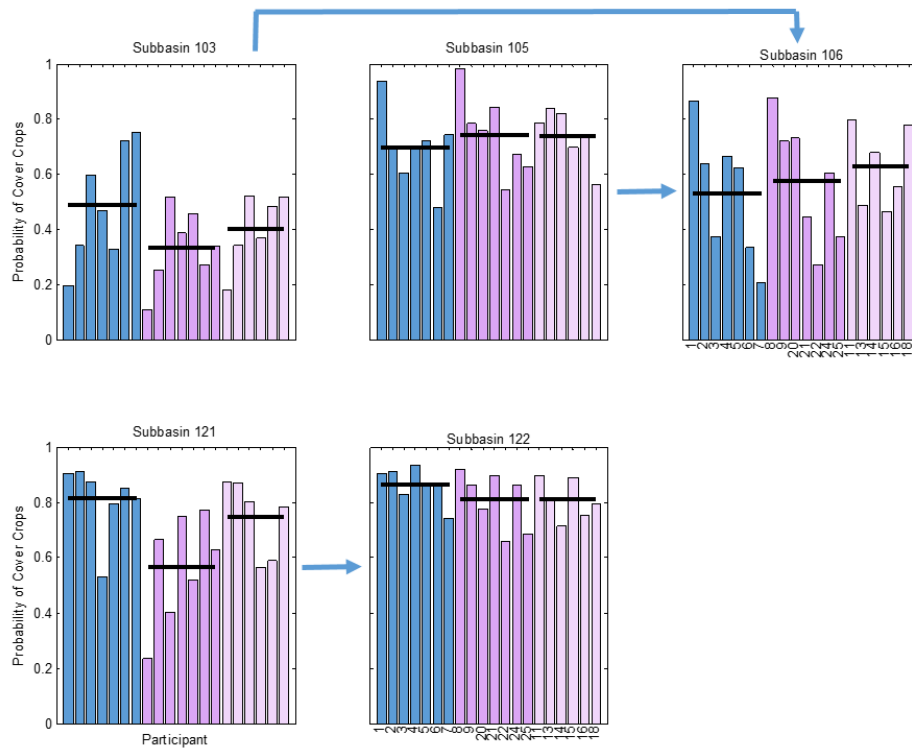


**Figure A.91** Bar plots for Probabilities of Cover Crop in SB of interest in group 4





**Figure A.92 Bar plots for Probabilities of Cover Crop in SB of interest in group 5**



**Figure A.93 Bar plots for Probabilities of Cover Crop in SB of interest in group 6**

Individual bar plots for Mode of Filter Strip Width and percentages of designs with Filter Strip width equal to the mode for SBInt 2 to 6

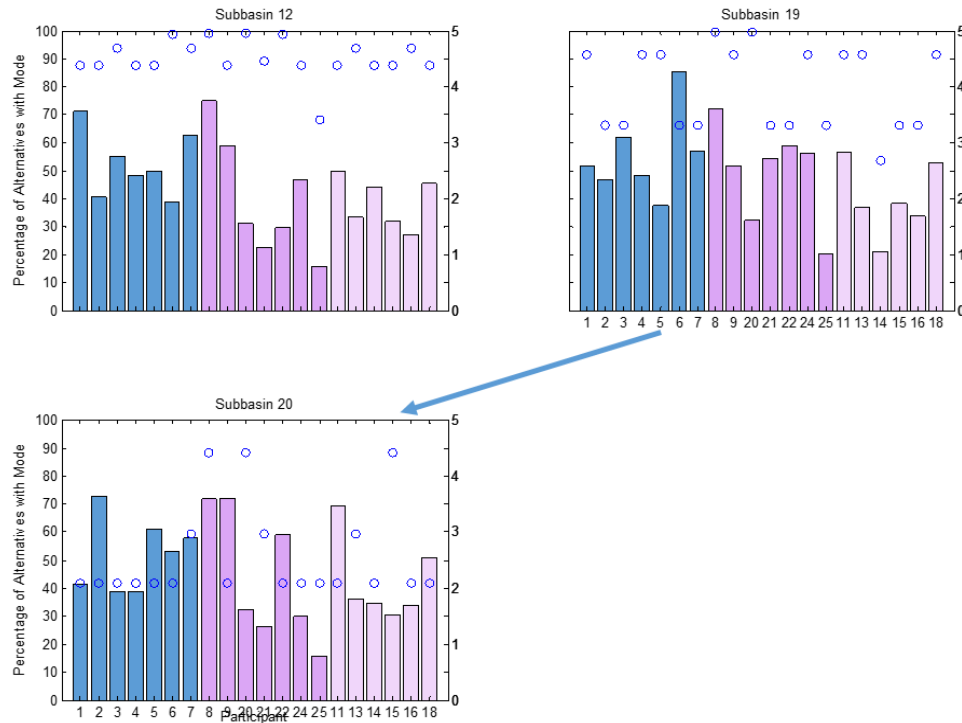


Figure A.94 Bar plots for Filter Strip mode and Percentage of design alternatives equal to the mode in SB of interest for group 2

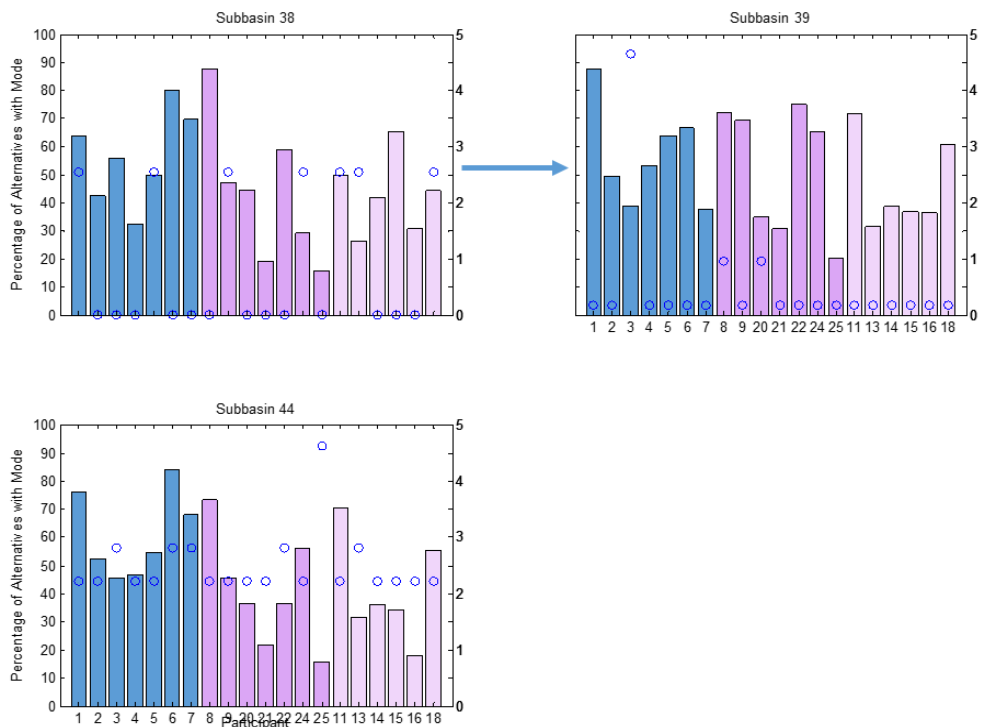


Figure A.95 Bar plots for Filter Strip mode and Percentage of design alternatives equal to the mode in SB of interest for group 3

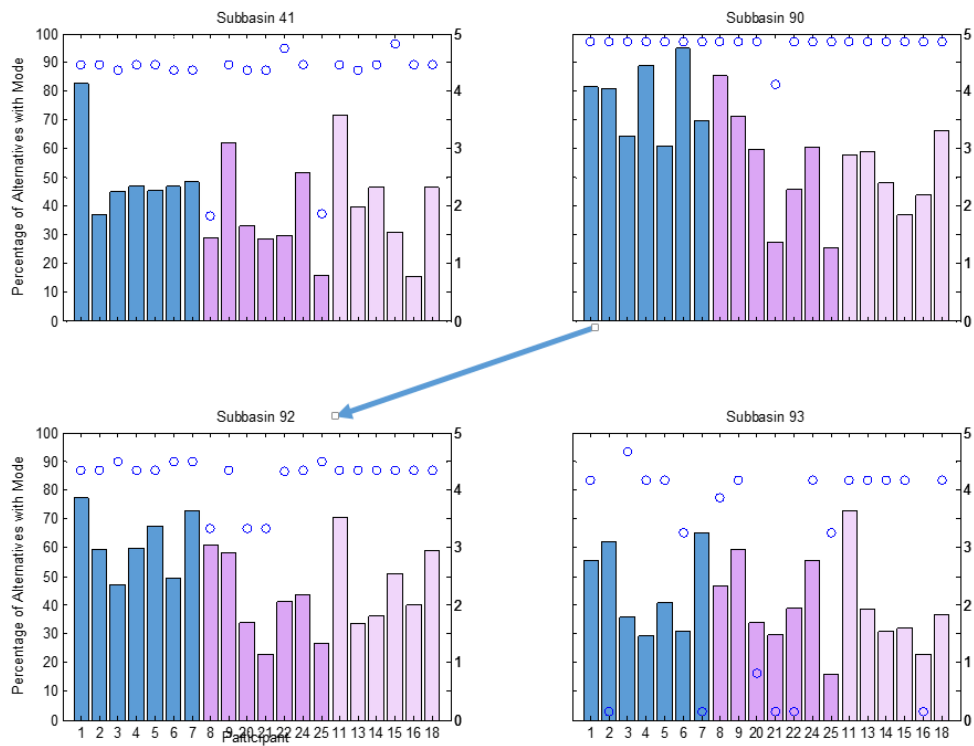
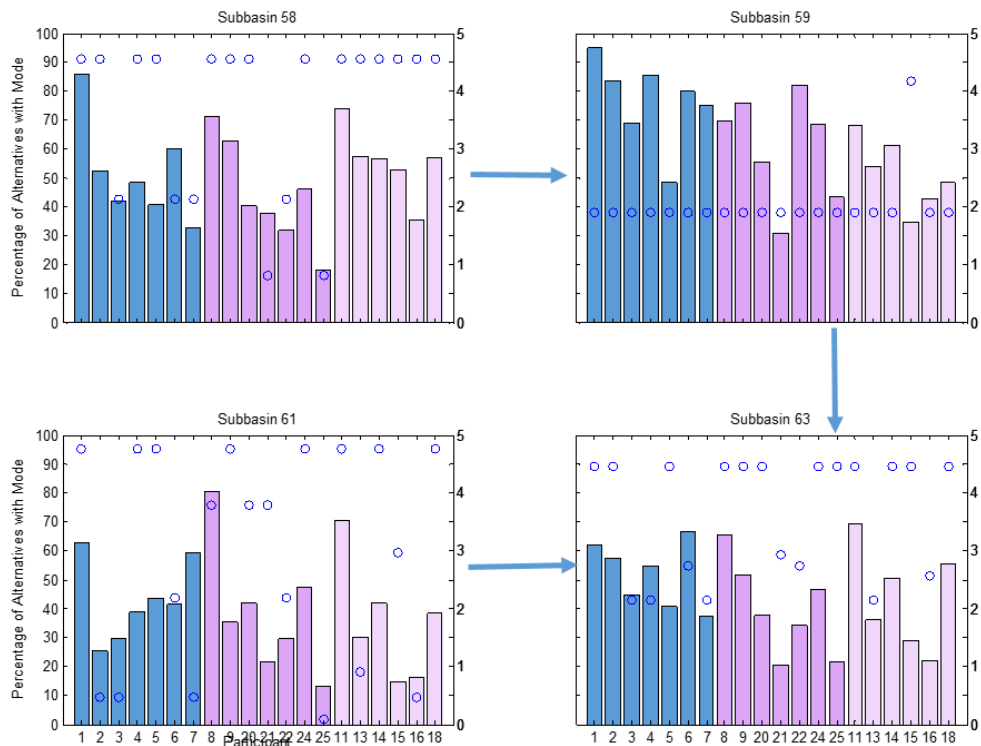
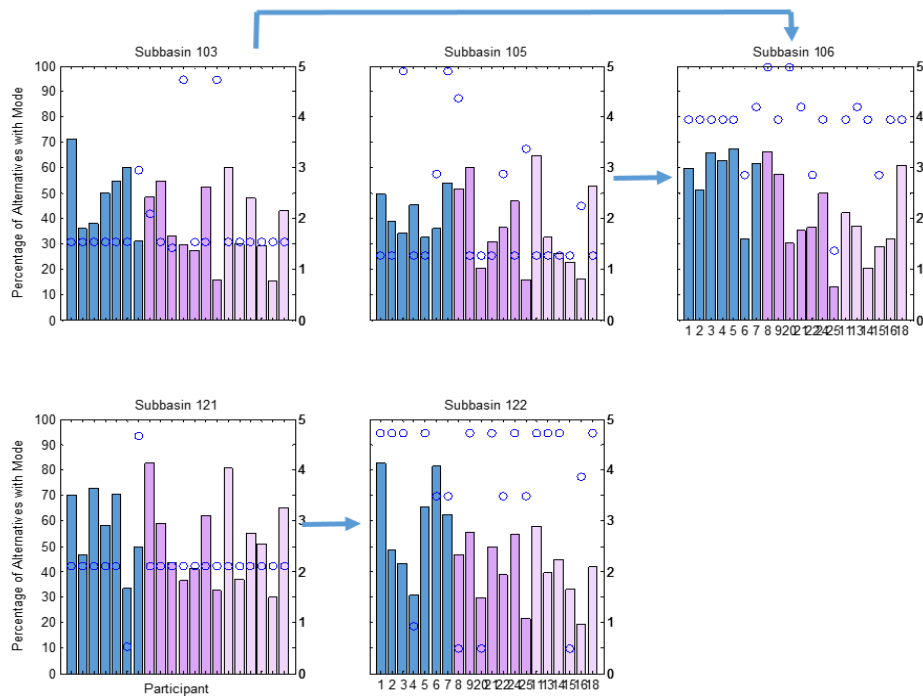


Figure A.96 Bar plots for Filter Strip mode and Percentage of design alternatives equal to the mode in SB of interest for group 4



**Figure A.97** Bar plots for Filter Strip mode and Percentage of design alternatives equal to the mode in SB of interest for group 5



**Figure A.98** Bar plots for Filter Strip mode and Percentage of design alternatives equal to the mode in SB of interest for group 6

Chapter 7

Simulated User 1 and 5 results

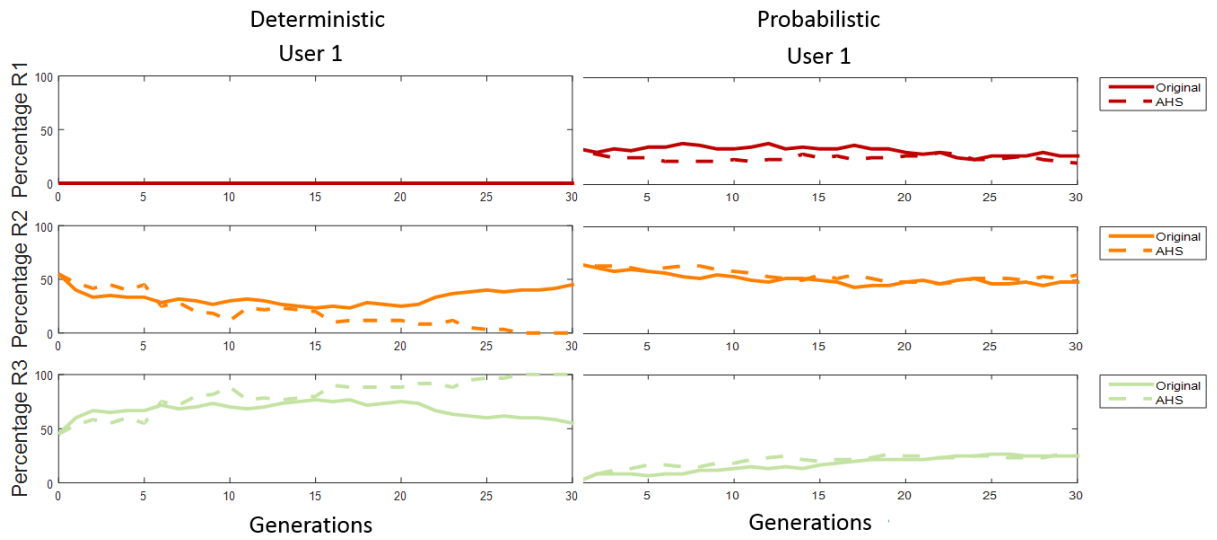


Figure A.99 Percentage of design alternatives in different generations for: Deterministic Simulated user 1 (left) and Probabilistic Simulated user 1 (right)

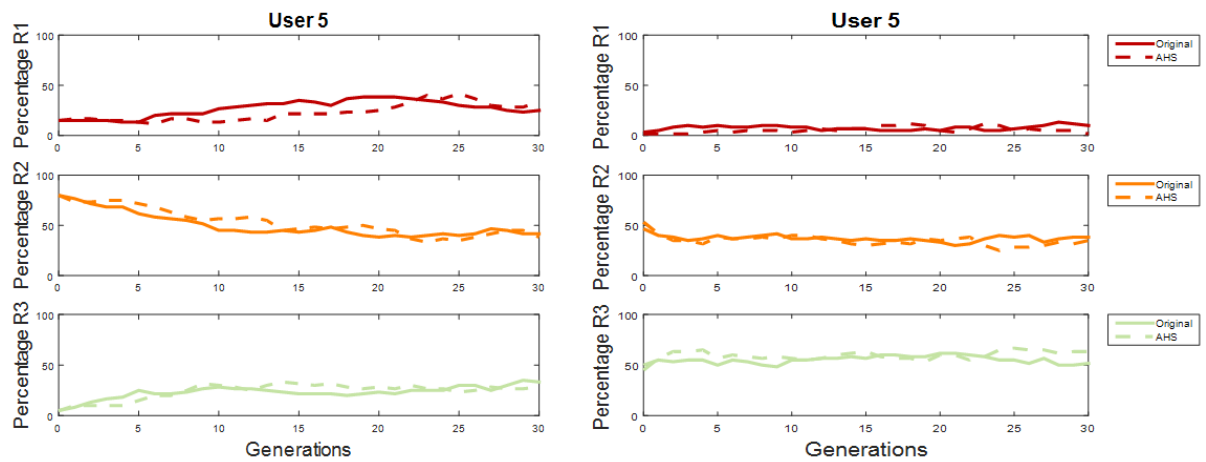


Figure A. 100 Percentage of design alternatives in different generations for: Deterministic Simulated user 5 (left) and Probabilistic Simulated user 5 (right)